

CONFERENCE PROCEEDINGS

2025 IMEKO TC6 INTERNATIONAL CONFERENCE ON

METROLOGY AND DIGITAL TRANSFORMATION

BENEVENTO, ITALY SEPTEMBER 3-5, 2025

M4D CONF

BENEVENTO2025

Welcome Message from the General Chairs

On behalf of the Organizing Committee, we warmly welcome you to the **2nd International Conference on Metrology and Digital Transformation** (***M4DConf 2025***), organized by IMEKO TC-6. We are proud to mark the second edition of this rapidly growing conference, which has quickly established itself as a premier global event in the field of metrology for digital transformation. This milestone is a testament to the dedication of our organizers, the contributions of our participants, and the increasing global interest in this dynamic field. Since its inception, M4DConf has served as an international meeting place for researchers, professional engineers, and academics to discuss cutting-edge advancements in metrology, measurement, and digital transformation. The conference fosters collaboration between experts in measurement, instrumentation, and industrial testing and those driving innovation in digital metrology. The diverse participation from fields such as data science, artificial intelligence, and cybersecurity has enriched the conference, creating a vibrant platform for interdisciplinary exchange. This 2nd edition of M4DConf will be hosted at Complesso Sant'Agostino, an Augustinian convent dating back to the XIII century, from University of Sannio, Benevento, Italy, a prestigious location that reflects the conference's commitment to excellence. After the challenges of the global pandemic, which required virtual formats for many events, we are thrilled to welcome you in person to this inspiring venue. Located in the heart of Benevento, the conference center offers a perfect backdrop for intellectual exchange and networking.

M4DConf 2025 continues its mission to advance the state-of-the-art in metrology for digital transformation. The program will focus on key topics such as metrological traceability, sensor network metrology, digital calibration, legal metrology, and digital infrastructure for measurement. We aim to explore how these advancements drive innovation across industries, from smart manufacturing to sustainable technologies. The M4DConf 2025 Technical Program is robust and diverse, featuring three keynote speeches, 10 technical oral sessions over three days, one tutorial session, and two panel discussion events. We received 85 extended abstracts from around the world, each rigorously peer-reviewed by our dedicated program committee and additional reviewers. Based on quality, originality, and relevance, 72 papers were selected for presentation. We extend our gratitude to all reviewers for their meticulous efforts in ensuring a high-quality program. Accepted and presented papers will be submitted for inclusion in the IMEKO Proceedings website, and for inclusion in *SCOPUS* Database (as 'Conference paper').

This year's keynote speeches will feature distinguished experts in metrology and digital transformation:

- **Frank Härtig**, *Physikalisch-Technische Bundesanstalt, Germany,* will address "Status of Developments in Digital Quality Infrastructure and Traceable AI".
- **Sundeep Bhandari**, *National Physical Laboratory (NPL), United Kingdom*, will discuss "Trust and Confidence in AI - The Role of National Metrology Institutes".
- **Luca Mari**, *Università Cattaneo - LIUC, Castellanza, Italy*, will present on "Foundational Concepts of Measurement in the Context of a Formal Ontology".

We are honored to have these thought leaders share their insights and expertise with our community. This edition also includes exciting initiatives:

- A Tutorial on "***Advancing Automotive Precision: The Evolution of GNSS Technology and Sensor Fusion***", organized by Domenico Di Grazia, Salvatore Crasta, and Domenico Rega, from STMicroelectronics, aimed at exploring emerging trends into the latest developments in GNSS technology.
- A Panel Discussion on "***Driving Change: Digital Transformation in Metrology***", organized by Blair Hall from Measurement Standards Laboratory, New Zealand, to promote diversity and mentorship in the field.
- A roundtable "***Metrology & Marine Power Equipment***" (hosted by Shanghai Marine Equipment Research Institute) organized by National Metrology Data Construction and Application Base for Marine Power Equipment, fostering dialogue on the digital transformation of metrology, which offers enhanced reliability, safety, and performance through digital twin technology and wireless sensor networks, while addressing challenges such as adapting measurement uncertainty methods, ensuring hardware security, and establishing traceability in complex, real-time data streams. The discussion will prioritize developing international standards, overcoming hardware bottlenecks like low power consumption and anti-interference, and fostering collaboration between national metrology institutions and industry to support robust software tools and data interoperability.

Several awards will recognize outstanding contributions, including the "**Best Conference Paper Award**," the "**Best Paper Presented by a Young Researcher**," and the "**Best Paper Presented by a Woman**." These accolades celebrate excellence and inspire the next generation of researchers and professionals.

Organizing M4DConf 2025 has been a monumental effort, requiring coordination across technical programs, logistics, and social events. We extend our heartfelt thanks to our organizing team, volunteers, and supporting organizations, including IMEKO TC-6, for their invaluable contributions. Special thanks go to sponsors of the M4DConf 2025 conference for their unwavering support in managing the complex details of the conference. We invite you to immerse yourself in the technical sessions, engage with colleagues, and enjoy the cultural and natural beauty of Benevento. Your feedback is invaluable to us as we strive to enhance the quality of M4DConf and achieve lasting excellence. It is with great honor and pride that we welcome you to Benevento, Italy and to M4DConf 2025. We hope you find the conference both professionally enriching and personally enjoyable.

Warm regards,

Ioan Tudosa*, University of Sannio, Italy*
Sascha Eichstädt*, Physikalisch-Technische Bundesanstalt (PTB), Germany*
M4DConf 2025 General Chairs

IMEKO M4DConf 2025 Committee

General Chairs

Ioan Tudosa, *University of Sannio, Italy*
Sascha Eichstädt, *Physikalisch-Technische Bundesanstalt (PTB), Germany*

Technical Program Chair

Francesco Picariello, *Universitas Mercatorum, Italy*

International Program Committee

Khaled M. Ahmed, *SASO, Saudi Arabia*
Lerina Aversano, *University of Foggia, Italy*
Jon Bartholomew, *EMI, United Arab Emirates*
Diego Nahuel Coppa, *National Institute for Industrial Technology, Argentina*
Dolores del Campo, *EURAMET*
Sascha Eichstädt, *Physikalisch-Technische Bundesanstalt (PTB), Germany*
Hugo Gasca, *National Metrology Center, Mexico*
Damir Ilic, *University of Zagreb, Croatia*
Blair Hall, *Measurement Standards Laboratory, New Zealand*
Kruno Milicevic, *FERIT and Random Red Ltd. Osijek, Croatia*
Vukan Ogrizović, *University of Novi Sad, Serbia*
Francesco Picariello, *Universitas Mercatorum, Italy*
Cui Shan, *National Metrology Centre, Singapore*
Anjali Sharma, *CSIR-National Physical Laboratory, India*
Ioan Tudosa, *University of Sannio, Italy*
Zoltan Zelenka, *Bundesamt für Eich- und Vermessungswesen (BEV), Austria*
Cristian Zet, *Technical University of Iasi, Romania*

PROGRAM

Wednesday, September 3

Session 1.1 - General Session - PART I
Room: Auditorium Sant'Agostino

1 **From the past to the Future: <Calibration Certificates for Radioactivity @PTB>**
Stefan Röttger and Annette Röttger (Physikalisch-Technische Bundesanstalt, Germany)

6 **Digital Transformation: Interoperable Processes and Services Based on a Cloud Native Architecture**
Alexander Reissert, Samuel Eickelberg, Petra Tsesmetzi and Abdul Rehman (Physikalisch-Technische Bundesanstalt, Germany)

11 **Advanced Metrological Knowledge Representation in the D-SI Metadata Model**
Daniel Hutzschenreuter, Matthias Bernien, Frauke Gellersen, Moritz Jordan and Benedikt Seeger (Physikalisch-Technische Bundesanstalt, Germany)

19 **Approximative Bayesian Approach for Uncertainty Evaluation in Machine Learning-Based Hardness Measurement**
Junnosuke Takai, Yukimi Tanaka, Masahiro Yoshioka and Katsuhiro Shirono (National Institute of Advanced Industrial Science and Technology, Japan)

Session 1.2 - General Session - PART II
Room: Room SA3

24 **A Torque Sensor with a Unified Coupling Interface for SWPDT System**
Yuehua Chen and Jiahui Zheng (Ningbo University, China); Xianglong Ma (China); Gang Zhang (Ningbo University, China)

28 **Innovative Approach in Industry 5.0: Use of Multiple Sensors in a Real Production Process to Enhance Workers' Well-Being**
Matteo Zendri, Francesco Pilati, Francesca Calabrese and Qingwei Cai (University of Trento, Italy)

32 **Calibration of Storage Tanks in the Oil and Gas Industry Based on the Concept of Digital Metrology and Big Data Analytics - PART I**
Ankica Milinkovic (Calibration Laboratory Vekom Geo, Serbia & University of Jaen, Spain); Sanja Tucikešić (University of Banja Luka, Bosnia and Herzegovina)

38 **Calibration of Storage Tanks in the Oil and Gas Industry Based on the Concept of Digital Metrology and Big Data Analytics - PART II**
Ankica Milinkovic (Calibration Laboratory Vekom Geo, Serbia & University of Jaen, Spain); Sanja Tucikešić (University of Banja Luka, Bosnia and Herzegovina)

44 **Structural Connection Parameter Measurement for Precision Motion Systems Using Accelerometers: a Preliminary Study**
Li Li, Yang Zheng and Ziqing Jia (Harbin Institute of Technology, China)

Session 2.1 - Calibration and Comparison in a Digital Era - PART I
Room: Auditorium Sant'Agostino

49 **Interoperability of DCCs for Weights and Weighing Instruments Through Standardization**
Julian Haller (Sartorius Lab Instruments GmbH & Co. KG, Germany); Gisa Foyer (Physikalisch-Technische Bundesanstalt, Germany)

55 **Towards Seamless Digitalisation in TUBITAK UME Mass Laboratory**
Emrecan Keskin and Beste Korutlu (TUBITAK National Metrology Institute, Turkey)

60 **Automated Calibration System with Digital Calibration Certificates Support**
Serhii Kursin, Oleh Velychko, Ihor Pototskyi and Andrii Bachynskyi (SE Ukrmetrteststandard, Ukraine)

65 **Transition to DCC-Based Calibration Management: Findings from Proof-of-Concept Investigations**
Tuukka Mustapää (Beamex Oy Ab, Finland)

70 **Dcclib - Consolidation of Digital Calibration Certificate (DCC) Tools into a Unified Python Library**
Jan Henry Loewe, Justin Sebastian Jagieniak and Shanna Schönhals (Physikalisch-Technische Bundesanstalt, Germany)

75 **Development and Application Study of a Calibration Certificate Anomaly Detection System**
Dong-Hun Ryu, Chaewook Lim (Korea Testing Laboratory, South Korea)

Session 2.2 - General Session - PART III

Room: Room SA3

80 **Checking Traceability by Analyzing DCC Blockchain Transactions**
Cristian Zet (Technical University of Iasi, Romania); Gabriel Constantin Dumitriu (Individual Enterprise GCD, Romania); Cristian Fosalau (Technical University of Iasi, Romania)

85 **Automatic Method of Assessment and Grading of Students in a Measurement Laboratory**
Cristian Zet, Anamaria Hariton, Cristian Fosalau (Technical University of Iasi, Romania)

89 **Quantum Communications for Distributed Measurement Systems: Current Situation and Research Trends**
Arman Neyestani, Ioan Tudosa, Luca De Vito, Fatemeh Khalesi and Sergio Rapuano (University of Sannio, Italy)

95 **Quantum Channel Characterization in QKD: A Metrological Perspective**
Fatemeh Khalesi and Ioan Tudosa (University of Sannio, Italy); Francesco Picariello (Universitas Mercatorum, Italy); Arman Neyestani and Sergio Rapuano (University of Sannio, Italy)

101 **Machine Learning-Based Classification and Identification of Arabic Dialects Through Acoustic Feature of Guttural Sounds**
Seyed Ebrahim Hosseini, Shahbaz Pervez and Hamed Al-Tairi (Whitecliffe, New Zealand)

106 **Measuring Privacy: Critical Reflections and Directions for a Metrology-Based Approach**
Eulalia Balestrieri (University of Sannio, Italy); Ilaria Amelia Caggiano (Università Degli Studi Suor Orsola Benincasa, Italy); Francesco Picariello (Universitas Mercatorum, Italy); Ioan Tudosa (University of Sannio, Italy)

Session 3.1 - Calibration and Comparison in a Digital Era - PART II

Room: Auditorium Sant'Agostino

112 **Digital Metrological Expert - Design of a Software for Automated Key Comparison Data Analysis in a Digital World**
Daniel Hutzschenreuter, Clifford Brown, Moritz Gafert, David Urban and Wafa El Jaoua (Physikalisch-Technische Bundesanstalt, Germany)

119 **Digital Transformation Applications in Mechanical Quantities - Hardness Measurements**
Moritz Loewit (Bundesamt für Eich- und Vermessungswesen, Austria); Cihan Kuzu (Turkiye Bilimsel ve Teknolojik Arastirma Kurumu, Turkey); Alessandro F.L. Germak and Andrea Prato (INRiM - National Institute of Metrological Research, Italy); Febo Menelao (Physikalisch-Technische Bundesanstalt, Germany); Miha Hiti (ZAG, Slovenia); Tatiana Apostol (I.P. Institutul Naţional de Metrologie, Moldova)

125 **Automating Photometric Measurements Using LabVIEW-Python-Based AI: Enhancing Precision in Luminous Intensity, Responsivity, Illuminance, and Flux Analysis at Saudi Standards, Metrology & Quality Org**
Mohammed Dhafer M Almelfi, Ahmed El-Matarawey and Saad Bin qoud (National Metrology and Calibration Center & Saudi Standards, Metrology & Quality Org-SASO, Saudi Arabia); Rawan Almutairi (Calibration Specialist, Saudi Arabia)

131 **Thermometry Machine Learning Model for Digitized Metrological Calibration of Platinum Resistance Thermometer**
Ahmed El-Matarawey (NIS, Saudi Arabia & SASO, Saudi Arabia); Saad Bin qoud (SASO-NMCC, Saudi Arabia)

Session 3.2 - Advancing Sensor Network Metrology for Industry 4.0 and IoT Applications

Room: Room SA3

137 **Sensor Fault Diagnosis Using Spectral Principal Component Analysis and CNN Deep Learning**
Jianqiang Mou and Shan Cui (National Metrology Centre, Singapore)

143 **Setup of a Distributed Sensor Network for Acquiring Environmental Data**
Alicja Wiora and Jozef Herbert Wiora (Silesian University of Technology, Poland)

148 **Metrology-Driven Standardization of Sensor Networks in Mining: a RAMI 4.0 Approach to Sustainable and Efficient Ventilation Systems**
Armin Shirbazo, Binghao Li, Seher Ata, Hamed Lamei Ramandi and Serkan Saydam (UNSW, Australia)

154 **Infrastructure Requirements for Metrological Distributed Sensor Networks**
Martin Koval (Czech Metrology Institute, Czech Republic); Gertjan Kok (VSL Dutch Metrology Institute, The Netherlands); Maximilian Gruber (Physikalisch-Technische Bundesanstalt, Germany); Shahin Tabandeh (VTT-MIKES, Finland); Martin Staněk (Czech Metrology Institute, Czech Republic)

Session 4.1 - Digitalisation of Measurements for Industrial Processes: Enabling Smart and Sustainable Manufacturing

Room: Auditorium Sant'Agostino

160 **Single-Device Integration of Legal Metrology and Third-Party Software via Virtualization**
Nicola Zingirian (University of Padova, Italy); Marco Profeti (Sampi S.p.A, Italy); Federico Botti (Click & Find s.r.l., Italy)

165 **Supporting Medicines Manufacturing Through Semantic Technologies**
Nina Peric, Moulham Alsuleman, Joao Gregorio, Paul Duncan and Michael Chrubasik, National Physical Laboratory, United Kingdom)

171 **Designing a Fatigue Monitoring Sensor System with Industry 5.0 Principles**
Qingwei Cai, Francesco Pilati, Francesca Calabrese and Matteo Zendri (University of Trento, Italy)

176 **Preliminary Experimental Assessment of an IoT-Based Fatigue Monitoring System for Industrial Operators**
Enrico Picariello (University of Sannio, Italy); Francesco Picariello (Universitas Mercatorum, Italy); Ioan Tudosa (University of Sannio, Italy)

Session 4.2 - General Session - PART IV

Room: Room SA3

182 **Soil Digitalization Using Micro-Sensors**
Federico Fina (Roma Tre University, Italy); Massimo Piotto (University of Pisa, Italy); Simone Contardi (University of Pisa, Italy & Sensichips Srl, Italy); Fabio Leccese (Roma Tre University, Italy)

187 **Characterization of the Lighting System of Hospital Rooms**
Marian-Andrei Vieru, Cristian-Gyozo Haba (Technical University of Iasi, Romania)

193 **Machine-Readable Data for Measurements of Total Luminous Flux Using Goniophotometer: A Step Toward Digitalized Metrology**
Khaled Abdelftah (SASO, Saudi Arabia)

198 **A Python-based graphical uncertainty calculator with optimal propagation of uncertainty and Monte-Carlo evaluation possibility**
Khaled Abdelftah (SASO, Saudi Arabia)

203 **Metrological Data Application Solutions in the Field of Marine Power**
Hao Wu (Shanghai Marine Equipment Research Institute, China)

Thursday, September 4

Session 5.2 - Metrologically Validated Digital Twins for Medical Diagnostics and Industrial Applications
Room: Room SA3

206 **Barriers to Implementing Digital Twin Technologies in Industrial Settings**
Jozef Herbert Wiora, Alicja Wiora and Faisal Saleem (Silesian University of Technology, Poland)

211 **Metrological Validation of a Digital Model for a CMM Including Digital Bias Correction**
Marcel van Dijk, Walter Knulst, Devrim Nalbantoglu and Gertjan Kok (VSL Dutch Metrology Institute, The Netherlands)

216 **Towards Safety and Accuracy of Hydrogen Refuelling Stations Through Digital Twins**
Christian Klaus, Ricardo Soruco Aloisio, Claudia Koch and Matthias Prellwitz (Bundesanstalt für Materialforschung und -prüfung, Germany)

Session 6.1 - Just Ask the Metrologist: Eliciting Domain Knowledge for Digitalisation in Metrology
Room: Auditorium Sant'Agostino

222 **Metrology and Architectural Description: a Case Study of the Hybrid Comparison System**
Blair Hall (Measurement Standards Laboratory of New Zealand, New Zealand); Diego Coppa (National Institute of Industrial Technology, Argentina); Anjali Sharma (CSIR-National Physical Laboratory, India)

227 **Bridging Knowledge Gaps in Metrology: a Requirement Elicitation Use Case for DCC**
Anjali Sharma (CSIR-National Physical Laboratory, India); Niharika Bhatia (Vellore Institute of Technology, India)

232 **Communities of Practice in Metrology**
Blair Hall (Measurement Standards Laboratory of New Zealand, New Zealand)

237 **Provenance Information in Metrological Traceability: Application and Modeling**
Ryan M White (National Research Council Canada, Canada & NRC, Canada); Julia Neumann (Physikalisch-Technische Bundesanstalt, Germany); Jean-Laurent Hippolyte (National Physical Laboratory, United Kingdom); Blair Hall (Measurement Standards Laboratory of New Zealand, New Zealand); Thiago Menegotto (Instituto Nacional de Metrologia, Qualidade e Tecnologia, Canada)

Session 6.2 - Enabling smart grid technologies for the digitalization of electrical power systems
Room: Room SA3

243 **On-Line Frequency Forecasting Using Convolutional Neural Networks**
Théo Chacou Bertoldi (Polytech Clermont, Campus des Cézeaux, Aubière, France); Viktoriya Mostova, Silvia Iuliano and Alfredo Vaccaro (University of Sannio, Italy)

249 **Metrological Characterization of a Vehicle's Charging Profile for Smart Charging Applications**
Chiara Franzoni, Antony Vasile, Davide Astolfi, Dmitrii Vasenin, Alessandro Musatti, Marco Pasetti and Stefano Rinaldi (University of Brescia, Italy)

255 **A Comprehensive Methodology Based on SCADA Data Analysis for Diagnosing Static Errors Affecting Wind Turbine Performance**
Davide Astolfi (Università di Brescia, Italy); Silvia Iuliano (University of Sannio, Italy); Antony Vasile (University of Brescia, Italy); Alessandro Canali (University of Perugia, Italy); Marco Pasetti (University of Brescia, Italy); Francesco Castellani (University of Perugia, Italy); Alfredo Vaccaro (University of Sannio, Italy)

261 **Partitioning Algorithm for Integrating Electric Vehicles into Residential Renewable Energy Communities**
Manuela Minetti, Andrea Bonfiglio, Maria Martino and Renato Procopio (University of Genova, Italy)

Session 7.1 - Pathways to digital transformation - PART I
Room: Auditorium Sant'Agostino

267 **Pathways to Digitalisation at the Measurement Standards Laboratory of New Zealand**
Blair Hall and Annette Koo (Measurement Standards Laboratory of New Zealand, New Zealand)

272 **"SASO Uncertainty Machine" - Advanced Pythonic ML Algorithm for Estimating Uncertainty in General Calibration Services at Saudi Standards, Metrology, and Quality Organization-SASO-KSA**
Saad Bin qoud (SASO-NMCC, Saudi Arabia); Ahmed Elmataraway (A. Matarawey, Saudi Arabia); Ahmed El-Matarawey (NIS, Saudi Arabia & SASO, Saudi Arabia)

278 **The DX Schema as a Modular Concept for Metrological Certificates and Reports**
Justin Sebastian Jagieniak, Shan Lin, Moritz Jordan, Muhammed-Ali Demir and Lutz Doering (Physikalisch-Technische Bundesanstalt, Germany); Thomas Engel (Siemens AG, USA); Wiebke Heeren, Jan Henry Loewe, Shanna Schönhals and Siegfried Hackel (Physikalisch-Technische Bundesanstalt, Germany)

284 **Metrological AI Reliability Verification**
Volker Zeuner and Gulian Couvreur (Federal Institute of Metrology METAS, Switzerland)

Friday, September 5

Session 8.1 - Digital Transformation of the Quality Infrastructure - PART I

Room: Auditorium Sant'Agostino

289 **Knowledge Management Foundations in the Design of the IPEM-SP Quality Infrastructure Portal: Strategies and Challenges**
Robson Santos da Silva (Technological Research Institute of São Paulo, Brazil); Roberto M Araujo, Filho (Universidade do Estado do Rio Grande do Norte, Brazil); Marcos Helenono Guerson de Oliveira Junior; Eduardo Mario Dias and Maria Lídia Rebello Pinho Dias Scoton (Universidade de São Paulo, Brazil)

295 **Digital Transformation for the United States Air Force Metrology and Calibration Program**
Jeremy Latsko, Michael S Brockway, Salvatore Capra and Evan Elliott (United States Air Force Metrology and Calibration, USA)

301 **Measurement Data and Information of Non-Automatic Weighing Instruments as Structured Data**
Tatyana Sheveleva and Gisa Foyer (Physikalisch-Technische Bundesanstalt, Germany)

307 **Digitalized Third-Party Validation for Calibration Service: a System Design Example**
Hiroshi Watanabe, Yoshitaka Shimizu, Katsuhiro Shirono and Toshiyuki Fujimoto (National Institute of Advanced Industrial Science and Technology, Japan)

313 **Digitalization of Calibration Workflow at National Institute of Metrology Thailand (NIMT)**
Sunantiya Parana, Nititorn Kenyota, Naruesorn Nanna, Praiya Thongluang, Narin Chanthawong and Jariya Buajarern (National Institute of Metrology, Thailand)

317 **"SASO Proficiency Test Machine" - Advanced Pythonic AI Algorithms for Automating and Validating ISO 13528 & ISO 5725-2 at Saudi Standards, Metrology, and Quality Organization- SASO-KSA**
Ahmed El-Matarawey, Saad Bin qoud, Fahad A. Almuhlaki and Rayan A AlYousefi (SASO-NMCC, Saudi Arabia)

Session 8.2 - Pathways to digital transformation - PART II

Room: Room SA3

323 **A Timing Accuracy Assessment System Prototype for Multiple NTP Servers**
Deepak Sharma, Divya Singh Yadav, Preeti Kandpal, Bharath Vattikonda, Ashish Agarwal (CSIR - National Physical Laboratory & Academy of Scientific and Innovative Research, India)

328 **The Regional Metrology Organisations Coordination Working Group of the CIPM FORUM on Metrology and Digitalization. Helping Emerging NMIs on Their Path to Digital Transformation**
Nikita D Zviagin (D.I. Mendeleyev Institute for Metrology, Russia)

332 **Data Quality Characteristics for Improved Metrology in Sensor Networks**
Mads Johansen (FORCE Technology, Denmark); Anupam Prasad Vedurmudi (Physikalisch-Technische Bundesanstalt, Germany); Martha Arbayani Zaidan (University of Helsinki, Finland); Milos Davidovic (Vinca Institute of Nuclear Sciences, Serbia); Gertjan Kok (VSL Dutch Metrology Institute, The Netherlands); Maitane Iturrate-Garcia (Federal Institute of Metrology METAS, Switzerland); Shahin Tabandeh (VTT-MIKES, Finland)

337 **Research on Digital Methods for Testing and Calibration Schemes**
Xinmeng Liu, Yulin Qi and Weiqun Cui (National Institute of Metrology, China)

343 **A Modular Windows-Based Intelligent API for Traceable Drone Positioning Using UWB-OptiTrack Fusion and AI-Based Residual Learning**
Ihtisham Ul Haq, Luigi D'Alfonso, Giuseppe Fedele and Francesco Lamonaca (University of Calabria, Italy)

Session 9.1 - Digital Transformation of the Quality Infrastructure - PART II
Room: Auditorium Sant'Agostino

349 **Swiss Quality Infrastructure in Transition**
Peter Blattner, Oscar De Feo and Fabiano Assi (Federal Institute of Metrology METAS, Switzerland)

353 **Advancing Digital Quality Infrastructure: Transforming Laboratory Processes for Enhanced Efficiency and Reliability**
Anna Maria Elert, Lena Meyer, Nanine Brunner and Michael Melzer (Bundesanstalt für Materialforschung und -prüfung, Germany); Claudia Koch (BAM Federal Institute for Materials Research and Testing & Berlin Institute of Technology, Germany)

359 **Towards an Inclusive and Agile Implementation Roadmap for a Digital Quality Infrastructure**
Jens Niederhausen (Physikalisch-Technische Bundesanstalt, Germany)

364 **National Metrology Laboratory of the Philippines Digital Calibration Certificate: DigiCert**
Ahdrian Camilo Gernale (Department of Science and Technology, Philippines & Asian Institute of Management, Philippines); Roj Gian D. Gorospe, Mark Josep C Nicasio and Nathaniel Ken A. Aquino (University of Santo Tomas, Philippines)

370 **Development of Frontend Interface for Digital Calibration Certificate for AC High Current Source Parameters**
Paramita Guha (CSIR-National Physical Laboratory Delhi India, India & AcSIR, India); Arun Ram RT (Scientist, Nigeria); Manish Kumar Tamrakar and Shri Krishan (Technical Officer, India); Priyanka Jain (Scientist, Nigeria)

376 **The Digital Reference Material Document: from Paper Certificates to Interoperable Data Objects in Digital Quality Infrastructure**
Matthias Prellwitz, Claudia Koch, Silke Richter, Johannes van de Kreeke, Michael Melzer and Mehran Monavari (Bundesanstalt für Materialforschung und -prüfung, Germany)

Session 9.2 - General Session - PART V
Room: Room SA3

382 **Using FAIR Principles and Data Quality Assessment to Enhance the Reliability of Digital Twins for Subsea Asset's Integrity Monitoring**
Miguel Burg Demay, Luiz Eduardo De Farias, Gustavo Donatelli and Andre Luiz Meira De Oliveira (Certi, Brazil)

387 **An Overview of Metrology Knowledge Storage: Taxonomies, Ontologies and Controlled Vocabularies**
Clifford Brown, Julia Neumann and Tomasz Sołtysiński (Physikalisch-Technische Bundesanstalt, Germany)

391 **Lightweight Passive Monitoring for Soft Anomaly Classification in Wired Networks on Resource Constrained Microcontrollers**
Prabin Dhakal (Institute of Engineering, Nepal); Francesco Picariello (Universitas Mercatorum, Italy); Basanta Joshi and Nanda Bikram Adhikari (Institute of Engineering, Nepal)

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

From the past to the future: <Calibration certificates for radioactivity @PTB>

Stefan Röttger[1], Annette Röttger[1]

[1] *Physikalisch-Technische Bundesanstalt (PTB), Stefan.Roettger@PTB.de, Annette.Roettger@PTB.de*

Abstract **– The XML-DCC (Digital Calibration Certificate) was developed at PTB for use in a variety of applications. This contribution refers to the field of radioactivity. It involves representing the calibration item, the measurement method, the detected nuclides and the measured activity digitally, and referring them to a specific reference time.**
While this specific schema is in the tradition of 100 years of calibration certificates as defined by Hans Geiger, it also points towards the future by employing a machine-interpretable digital representation of the SI in the field of ionising radiation.

I. INTRODUCTION

A. Measuring radioactivity: A new field evolves

Accurate measurement has had an institutional home in Germany since 1887. The approval of the first budget for the Physikalisch-Technische Reichsanstalt (PTR) on 28 March 1887 marked the birth of the first large-scale state research institution and the beginning of a success story. The industrialist Werner von Siemens and the scientist Hermann von Helmholtz are regarded as the PTR's founding fathers. Thanks largely to their vision and persistent commitment, the German Reichstag finally approved the PTR's first annual budget on 28 March 1887. This laid the foundations for the first state-funded, non-university, large-scale research institution committed to basic research, free of material interests, and to supporting industry with current problems [1].
At the beginning of the 20th century, under the presidency of Emil Warburg, the PTR increasingly focused on the so-called 'New Physics'. This included X-rays, new ideas about atomic structure, Einstein's special theory of relativity, and quantum physics based on the black body and electron properties. The renewal of the PTR's research content was achieved above all by outstanding researchers such as Hans Geiger, who set up the PTR's first radioactivity laboratory and quickly turned it into a world-class scientific facility.
By the early 1900s, the effects of radioactivity were already well understood, forming the basis for the development of detectors that are still in use today. For example, radiation can blacken a photographic plate, generate ions in gases, and trigger flashes of light, known as scintillations, when it hits certain materials.
Geiger's counting methods have a long history. Not only did they lead to the invention of various detectors, such as the counting chamber with longitudinally entering beams in the proportional range (created by Rutherford and Geiger in 1908), the sphere counter (created by Rutherford and Geiger in 1912), the tip counter (created by Geiger in 1913), the Geiger-Müller counting tube (created in 1928), and the self-extinguishing counting tube (created by Adolf Trost in 1936), but they also led to the development of recording and amplifying devices [2].
The invention of the Geiger–Müller counter tube in 1928 represented a new development in the field of ionising radiation detectors. The construction of electronic amplifiers for this tube, however, was driven forward by Geiger's students and the working groups of Walther Bothe and his colleagues in Germany, as well as various institutes in Italy, France, the USA and England.
In 1912, Hans Geiger was appointed head of the newly founded Laboratory for Radioactivity at the Physikalisch-Technische Reichsanstalt in Berlin. He had previously spent five years in Manchester as a research assistant to Ernest Rutherford, gaining experience in the new scientific field of radioactivity research. In Berlin, his role was to measure the activity of radioactive materials, maintain radioactive measurement standards, and develop new measurement methods in accordance with the PTR's mandate.
Geiger expanded the laboratory into a central testing centre for radioactive standards. By the time he left the PTR in 1925 to take up a professorship in experimental physics at Kiel University, his team had determined the activity of eighty-five radium standards, six ore samples, five radium compresses, two radioactive ointments, a radioactive bread additive and a radioactive cat fur.
Radioactivity was one of the new discoveries that attracted the interest of scientists and aroused the enthusiasm of the press and public alike. In 1896, Henri Becquerel became the first person to detect radiation emitted by uranium in Paris. By the start of the First World War, its basic properties were well understood, although they were still the subject of much controversy and debate in scientific circles. Other radioactive elements had been discovered,

and a distinction had been made between alpha, beta and gamma rays. The first radioactive decay series had been established, and the radioactive decay law had been formulated. Radioactive decay was also recognised as a random process. Important scientific discoveries were based on investigations involving radioactive rays, including the discovery of the atomic nucleus and isotopes, and then neutrons and nuclear fission in the 1930s.

In the age of digitalisation and Industry 4.0, it is clear that paper-based calibration certificates are no longer sustainable. With an increasing number of sensors and devices requiring calibration, printing, mailing, storing and processing paper calibration certificates not only creates a bottleneck, but also increases the chance of errors and mistakes at each step in the chain. Electronic (digital)

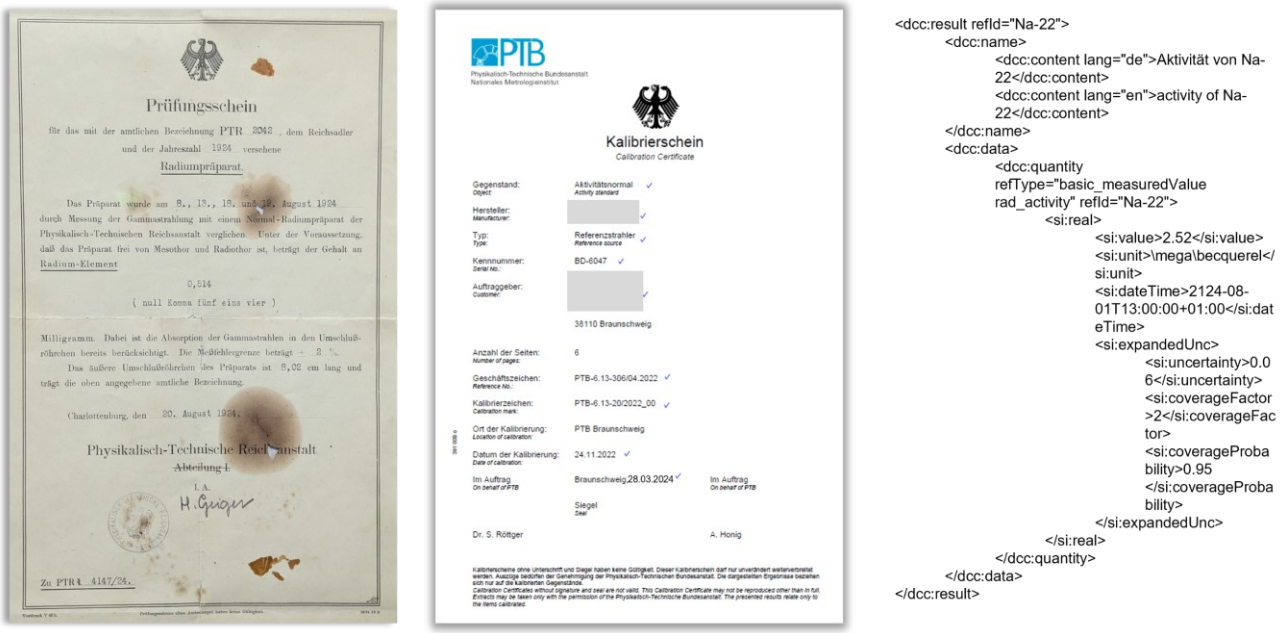


Figure 1: Calibration certificate (left) from 1924 (Hans Geiger), Typical calibration certificate (middle) from 2024 (Anja Honig) and a possible structure for a machine interpretable activity statement (right) maybe before 2124.

Radioactivity research in the early 20th century gave rise to many new fields of study, including atomic, nuclear and neutron physics, radiochemistry, radiobiology and radiology. From the time of Geiger to the present day, the history of the Laboratory for Radioactivity at the Physikalisch-Technische Bundesanstalt has been continued. For over 100 years, the results have been passed on using calibration certificates. Even if these have changed over the course of time, the field of radioactivity has its own requirements, as can be seen in the historical, current and future calibration certificates, a possible digital calibration certificate (DCC) – see Figure 1.

The information included in the radioactivity calibration certificate became more detailed and included even fundamental and nuclear data used and needed for further usage of the calibration data. Since this data might even change with new developments in physics or better experiment with smaller uncertainties, newer fundamental data might even influence the calibration results afterwards.

II. DIGITAL CALIBRATION CERTIFICATES

calibration certificates mitigate these problems. We need to distinguish between different levels of "digitalness" [3]. Here we focus on calibration certificates of level three or higher according to the extended utility model of [3] - i.e. certificates with machine-readable (level 3), machine-interpretable (level 4) and/or machine-executable (level 5) content.

Several formats for digital calibration certificates and similar documents have been proposed, including some that append machine-interpretable data to human-readable PDF documents (see e.g. [4, 5]). Other approaches focus solely on having a machine-interpretable document such as XML (Extensible Markup Language) [6]. The advantages of using pure XML over PDF-with-XML (or PDF-with-JSON [Portable Document Format; JavaScript Object Notation]) are that no additional steps are required to extract the machine-readable data from the PDF, and there is no potential for a mismatch between human-readable and machine-readable documents. On the other hand, XML certificates require additional information to enable customers to extract the information in a meaningful way, for example in the form of Extensible

Stylesheet Language Transformations (XSLT) style files. These would need to be distributed with the calibration certificate or hosted online.

We decided to implement calibration certificates in the XML format developed and hosted by PTB (Physikalisch-Technische Bundeanstalt), which is described in the next section.

III. PTB DCC

In this document, the XML-based format for digital calibration certificates developed at PTB is referred to by the abbreviation "DCC" [6, 3]. This DCC is based on an XML schema in XSD format. It is maintained and hosted by PTB [1]. The current release is version 3.3.0[2]. The schema is complemented by a vocabulary database[3], good practice examples[4], and usage guides (e.g., [7, 8]). The vocabulary database and usage guides are developed and maintained by technical committees (TCs) within DKD[5], the German Calibration Service at PTB.

XSLT style sheets may be used to render XML DCCs for human readability [9]. External files, including a human-readable representation of the document, may also be included in the DCC (where permitted by the schema) as base 64 binary strings.

The DCC schema depends on the D-SI schema [10] for the representation of physical quantities such as measurement results, and on a modified version of the W3C xmldsig schema for digital signatures, and otherwise stands alone. Further harmonisation and modularisation of the digital quality infrastructure has been proposed, which would involve the introduction of a generic schema “DX” on which the schemas for the DCC and other documents, such as Digital Reference Materials or Digital Test Certificates, could be based. [11].

Each DCC document contains different sections (and an optional signature(s)) which are: administrative data, measurement results, comments and included documents.

Many elements of the DCC can have the optional “refType” attribute. RefTypes allow information to be stored and extracted in a harmonised way, making DCCs truly machine-interpretable. To that end, each calibration community maintains a list of community-specific refTypes with definitions and usage notes. Currently, PTB hosts a database of refTypes[6]. RefTypes consist of a prefix and base, separated by an underscore. The prefix indicates the domain or metrological area, while the prefix “basic” is reserved for refTypes used by several communities. Only the combination of prefix and base uniquely identifies a refType.

Future versions of the DCC schema[7] will support definitions of measurands in external vocabulary databases, reducing the need for measurand-specific refTypes.

IV. DCC FOR RADIOACTIVITY

The steps to implement a DCCs for a new calibration community are: identifying community-specific information that needs to be represented in a DCC. This must then be prepared as: A list of refTypes (with definitions and usage notes) to encode the information in a machine interpretable way. Ideally, all this would be documented in a DKD expert report or similar publication.

We are currently in the process of implementing DCCs for radioactivity, that is, the calibration of radionuclide standards with calibrated activity (or activity concentration), and the calibration of detectors that measure activities in Becquerel and related quantities. Since this is very much work in progress - input and feedback from the community is very welcome and necessary at this stage.

B. Prefixes and refTypes

Ionising radiation metrology, which includes the BIPM service categories of dosimetry, radioactivity, and neutron measurements, is part of DKD TC 2, “high frequency and optics”. The structure of refTypes and refType prefixes for the fields of metrology included in TC 2 are currently under discussion. In this text, we use the symbol “Y_” as a placeholder for the future refType prefix that will cover the new refTypes for radioactivity.

C. Calibration Item

Calibration items in the field of radioactivity comprise devices and artefacts. Artefacts are available in different shapes including point-like sources, area sources, liquid solutions, volume samples in complex matrices, and gaseous samples. The simplest case: A single-nuclide radioactivity standard provided in the form of a liquid solution, characterised in terms of its activity. Other cases, as well as DCCs for devices measuring radioactivity, will have to be considered in future works.

[1] See https://www.ptb.de/dcc/
[2] https://www.ptb.de/dcc/dcc.xsd
[3] https://digilab.ptb.de/dkd/refType/vocab/
[4] https://wiki.dcc.ptb.de/en/gp_home
[5] https://www.ptb.de/cms/en/metrological-services/dkd.html
[6] https://digilab.ptb.de/dkd/refType/vocab/
[7] https://www.ptb.de/dcc/v3.4.0-rc.1/

The definition of the calibration item(s) is part of the administrative data section of the DCC. In general, a DCC may have multiple calibration items (e.g., a set of weights). In the field of radioactivity, each calibration certificate commonly only covers a single calibration item. Still, we propose to assign an `id` attribute to the calibration item, to keep this structure as general as possible.

We propose at least the following information to be included:

- An `id` attribute to uniquely identify the calibration item within the DCC.
- A `refType` attribute specifying the type of sample.
- Child elements `name`, `description`, and `manufacturer` filled according to the DCC schema.
- The `identifications` element should contain at least one identification with refType `basic_serialNumber`, with a value corresponding to the serial number that the physical source is labelled with. Further identifiers may be listed.
- Physical properties of the calibration item of interest to the customer are listed under the `itemQuantities` element. These depend on the source type. Simple physical quantities, such as the mass of the sample solution, can be listed immediately, using the embedded D-SI schema. We propose to specify complex properties, such as the composition of the solution, within the `statements` element, and use the `refId` formalism to refer to them from the calibration item description.

D. Further definitions

The `statements` element within the administrative data section may contain general statements about the calibration laboratory or calibration certificate. This is also the element where further detailed information shall be placed that can then be referenced from elsewhere in the DCC. We propose to place definitions of complex properties of the calibration item here.

All radionuclides referenced in the results section should be listed, even if they were not detected in the sample. We propose to identify nuclides unambiguously by their InChI string `InChI` and isomeric state as given by the J (nuclear spin), π (parity), and n (energy level) quantum numbers. Additional information of interest to the customer may be listed here, such as the half-life to be used when converting the activity to a different reference time. Another option currently under consideration is to reference an external database of radionuclides.

E. Measurement results

For a single-nuclide radioactivity standard, the quantity of interest is typically the activity at a given reference time. This quantity will be listed in the measurement results section in the form of a `si:real` quantity, whose child elements contain value, unit, and measurement uncertainty. It is important that the activity value be unambiguously associated to the reference time for which it is valid. We propose to list the reference time in the `si:dateTime` child element of `si:real`. The element `si:dateTime` contains a time stamp in ISO 8061 format [12], with the timezone defined by its offset to UTC.

An example for a single-nuclide measurement result contains the refTypes `basic_measuredValue` and `Y_activity`, in combination with the reference back to the previously defined `nuclide_Am241`, identify this result as the measured activity of Am-241, with value e.g. 2.52 MBq at 13:00 Central European Time (UTC+1) on August 1, 2022. For mixture samples comprising multiple radionuclides, we propose to add one `result` element per nuclide within the same `measurementResult` element.

The DCC schema specifies a general structure for each of the metadata elements. However, the details of how to implement information in a manner that is machine readable in the field of radionuclide metrology is still under developement.
We propose to define refTypes for all measurement quantities needed for radionuclide metrology calibration certificates, such as activity, activity per mass, activity per volume, and activity ratios. A full list of proposed refTypes is in preparation.
We also propose to define refTypes for the decision threshold and limit of detection, to quantify the absence of impurities. These would be used in conjunction with a refType defining a quantity, such as `Y_activity`.

V. SUMMARY AND CONCLUSION

Digital calibration certificates (DCCs) are becoming increasingly relevant in metrology, quality assurance, and compliance-driven industries due to their efficiency, traceability, and interoperability.

Main advantages are:

1. Enhanced Traceability and Authenticity:
   Digital certificates can be cryptographically signed, ensuring they are tamper-proof and traceable to the issuing authority. This boosts trust and simplifies audits.
2. Automation and Integration:
   DCCs can be directly integrated into laboratory information management systems (LIMS), and

quality control platforms. This enables automated updates, flagging of overdue calibrations, and better lifecycle management of instruments.

3. Reduced Errors and Administrative Load:
   Manual transcription from paper certificates is error-prone. DCCs eliminate this by using standardized formats (e.g., XML, JSON, or formats aligned with the Digital Calibration Certificate standard by PTB), reducing human error and time spent on data entry.
4. Interoperability and Standardization:
   Efforts like the SI Digital Framework, D-SI (Digital System of Units), and PTB's Digital Calibration Certificate initiative aim to harmonize DCCs globally. This facilitates cross-border calibration recognition and cooperation between labs and regulatory bodies.
5. Support for Industry 4.0:
   DCCs are foundational for smart manufacturing and the Industrial Internet of Things (IoT), where equipment needs to autonomously verify calibration status to ensure process accuracy in real time.
6. Regulatory and Compliance Pressure:
   Regulatory agencies and quality standards (e.g., ISO/IEC 17025) increasingly expect digital traceability. DCCs support easier compliance with such requirements.
7. Sustainability and Cost-Effectiveness:
   Eliminating paper and reducing manual processes align with sustainability goals and reduce long-term operational costs.

In summary, digital calibration certificates are becoming essential for modern, automated, and data-driven operations for calibration laboratories, research institutes and industry. They're not just a convenience but a strategic enabler for digital transformation in metrology and beyond. It is therefore essential to adopt them in every field of metrology.

References

[1] *PTB-Infoblatt,* PTB, 2018.

[2] *Sonderdruck der PTB-Mitteilungen,* PTB, 2012.

[3] D. A. Czarny, J. Diemer und M. Schacht, „Szenarios für die Digitalisierung der Normen und Normung," DIN/DKE, https://www.dke.de/resource/blob/2272804/0a674443fb9a40f87ae5387e5b2fd2ba/idis-whitepaper-1-de---download-data.pdf, 2021.

[4] R. Coulon, F. G. Toro and C. Michotte, "Machine-readable data and metadata of international key comparisons in radionuclide metrology," *Measurement Science and Technology,* vol. 34, no. 7, p. 074009, 2023.

[5] G. B. Toro, M. Wollensack, M. Zeier, C. Blaser, C. Hof, M. Stathis, P. Blattner, F. Stuker, N. Basic and F. Grasso, "PDF/A-3 solution for digital calibration certificates," *Measurement: Sensors,* vol. 18, no. 2665-9174, p. 100282, 2012.

[6] S. Hackel, S. Schönhals, L. Doering, T. Engel and R. Baumfalk, "The Digital Calibration Certificate (DCC) for an End-to-End Digital Quality Infrastructure for Industry 4.0," *Sci,* vol. 5, no. 1, p. 11, 2023.

[7] G. Foyer, M. Häfner, J. Haller, C. Müller-Schöll and S. Osang, "Instructions on how to use the DCC schema to create a digital calibration certificate for weights and weight sets," DKD, DKD Expert Report, 2024.

[8] G. Bartl, D. Heißelmann, K. Janzen, R. Liske, C. Sander, G. Bartl and C. Sander, "Anleitung zur Nutzung des DCC-Schemas für die Erstellung eines digitalen Kalibrierscheins für Parallelendmaße," DKD, DKD Expert Report, 2024.

[9] M.-A. D. Söylev-Öktem, M. Jordan, T. Krah, S. Schönhals, S. Hackel, F. Härtig, T. Schrader, J. Loewe, L. Doering, B. Gloger, J. Jagieniak and G. Söylev-Öktem, A Human Readable Form of the DCC, IMEKO TC6 International Conference on Metrology and Digital Transformation: IMEKO, 2022.

[10] D. Hutzschenreuter, F. Härtig, H. Wiebke and others, "SmartCom Digital System of Units (D-SI) Guide for the use of the metadata-format used in metrology for the easy-to-use, safe, harmonised and unambiguous digital transfer of metrological data - Second Edition," Zenodo, 6 2020. [Online]. Available: https://doi.org/10.5281/zenodo.3816686. [Accessed 2 5 2025].

[11] G. S. Ö. Jagieniak, S. Hackel, F. Härtig, J. Loewe, B. Gloger and J. Jagieniak, "Digital Schema X and the Future of the Digital Calibration Certificate," IMEKO, IMEKO TC6 International Conference on Metrology and Digital Transformation, 2022.

[12] I. C. Secretary, "Date and time --- Representations for information interchange Part 1: Basic rules," International Organization for Standardization, ISO 8601-1:2019, 2019.

[13] S. Hackel, F. Härtig, T. Schrader, A. Scheibner, J. Loewe, L. Doering, B. Gloger, J. Jagieniak, D. Hutzschenreuter and G. Söylev-Öktem, "The fundamental architecture of the DCC," *Measurement: Sensors,* vol. 18, no. 2665-9174, p. 100354, 2021.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Digital Transformation: Interoperable processes and services based on a cloud native architecture

Alexander Reissert [1], Samuel Eickelberg [1], Petra Tsesmetzi[1], Abdul Rehman[1]

[1]*Physikalisch-Technische Bundesanstalt, Abbestr. 2-12, 10587 Berlin, alexander.reissert@ptb.de*

***Abstract* – In accordance with the Unix philosophy, the implemented software architecture guidelines are designed to prioritise simplicity, modularity and composability. The objective is to develop services that excel in a specific area, with the ability to integrate seamlessly with other services to manage complex tasks.**

In this paper, we address the issues of interoperability and cross-institutional data sharing by proposing a cloud-native software architecture, the Operation Layer. This innovative solution integrates standalone legacy IT systems and paves the way for a comprehensive digital transformation. Furthermore, the revised and extended digital calibration workflow illustrates the potential of interoperable processes.

I. INTRODUCTION

It is an unavoidable fact that all organisations have legacy software and information architectures in place, and that these will continue to grow over time. New information systems will be introduced to solve a further set of problems. However, the exchange of data between these systems is negligible. It is not automated, and nor has it ever been intended. Data sharing within organisations can present certain challenges. Across institutions and other fields, it is not yet a common practice.

It is evident that digital transformation encompasses more than merely the digital support of processes or the digitisation of documents. Rather, it signifies a comprehensive transformation of the organisational structure and working environment, thereby facilitating streamlined and more efficient workflows.

In essence, the implementation of digital transformed and interoperable processes will facilitate the generation of new data and information regarding procedures and systems in real time. This, in turn, will generate new insights and information, which will result in improvements and innovation. This data tier provides a solid foundation for emerging technologies, such as machine learning and artificial intelligence [1].

However, the quality of the data is a key factor in determining the usefulness and outcome of the digitally transformed processes. This is a well-known input-output problem that has become more prominent due to the increased availability of large data sets. The creation of a digital ecosystem is intended to facilitate seamless communication and collaboration among all active participants.

Physikalisch-Technische Bundesanstalt (PTB) is undergoing a progressive digital transformation, encompassing its working groups, laboratories and workflows. This initiative is being implemented to enhance operational efficiency, facilitate seamless data exchange, and ensure the integrity and quality of data. The staff have expressed concerns about the change management process and the individual workflows of the highly specialised working groups, suggesting that these may present an insurmountable challenge.

The Operation-Layer (OP-Layer) is a proposed solution to address the integration problem by bridging the gap between information systems. It achieves this through a highly scalable and distributed architecture consisting of loosely coupled services. The OP-Layer provides highly adaptable services for the horizontal communication flow of working groups, integrating and enabling vertical communication via APIs to other information systems and departments. The OP-Layer is an integration layer that comprises services and APIs. It can also be described as a data tier that lays the foundation for a coherent and consistent digital transformation.

II. INTEROPERABLE PROCESSES

According to Cambridge Dictionary [2], the term "interoperability" is defined as "the degree to which two products, programs, etc. can be used together, or the quality of being able to be used together".

Interoperability can be viewed from four distinct perspectives: technical, syntactical, semantic and organisational. The following descriptions are taken from the ETSI white paper [3] and are intended to facilitate understanding of the complexity.

> **Technical Interoperability** is usually associated with hardware/software components, systems and platforms that enable machine-to-machine communication to take place. This kind of interoperability is often centred on (communication) protocols and the infrastructure needed for those protocols to operate.
>
> **Syntactical Interoperability** is usually associated with data formats. Certainly, the messages transferred by communication protocols need to

have a well-defined syntax and encoding, even if it is only in the form of bit-tables.

Semantic Interoperability is usually associated with the meaning of content and concerns the human rather than machine interpretation of the content. Thus, interoperability on this level means that there is a common understanding between people of the meaning of the content (information) being exchanged.

Organisational Interoperability, as the name implies, is the ability of organisations to effectively communicate and transfer (meaningful) data (information) even though they may be using a variety of different information systems over widely different infrastructures, possibly across different geographic regions and cultures. Organisational interoperability depends on successful technical, syntactical and semantic interoperability.

In this paper, the concept of interoperability is constrained to the realms of technical and syntactical interoperability. The primary focus is on harmonised interfaces and data exchange formats. In order to implement interoperable processes successfully, it is essential that the organisational dimension is resolved at a management level. In the absence of interdepartmental interoperability, it is implausible that any technical solution will be capable of overcoming this issue, let alone one that is applicable to different institutions.

Moreover, the term 'interoperability' already suggests the existence of a complex process with multiple subprocesses that must interlock seamlessly. The data are of paramount importance and are invariably intended for a specific purpose. Depending on the initial request, the items are frequently transported and converted during the evaluation process. Nonetheless, it is imperative to ensure the integrity of the data throughout the entire of the process chain. In the context of complex process chains, data is generated from a multitude of sources, employing a variety of methodologies and created by different individuals. These data are subsequently compiled and evaluated in a unified manner. In order to facilitate a successful and harmonised digital transformation of complex processes, it is imperative to undertake a comprehensive analysis of the entire lifecycle of the required and processed data. It is essential that this data is standardised, universally machine-readable, and internationally recognisable as far as possible with regard to future extensions [4].

In consideration of the aforementioned factors, the selection of an appropriate IT infrastructure is imperative that facilitates the management of the data in question, in conjunction with the implementation of suitable interfaces for the purpose of modularisation. It is evident that, with the establishment of a suitably flexible system structure, a multitude of individual processes will be able to be combined in a harmonious manner in the future. The aforementioned interoperability presents a series of opportunities, including the potential for the development of existing infrastructures, the adaptation to novel developments, and the provision of support for the future digital metrology quality infrastructure. [5]

III. CLOUD NATIVE ARCHITECTURE

The Operation Layer (OP-Layer) employs the concept of a central business process middleware or integration layer in a research environment characterised by highly specialised, tailored applications. Rather than using off-the-shelf software, open-source components are employed to address the specific requirements of the research institute and the metrology domain. The OP-Layer constitutes a platform of interconnected, yet loosely coupled services, the purpose of which is to facilitate interaction for a variety of interoperable processes across multiple departments and laboratories within PTB. The following elements are to be considered in order to ensure the effective management of data:

- The harmonisation of data structures and flows through standardised interfaces (REST, JSON),
- Security and traceability are to be ensured by using state-of-the-art single sign-on identification (OpenID [6],
- Direct communication with the electronic file system ("E-Akte") is to be facilitated, and
- High availability and scalability are to be guaranteed through deployment in a managed cluster environment.

All services constituting OP-Layer's interconnected service landscape adhere to the same coding guidelines, module structure, and communication interface design. It is evident that all services are developed in two artefacts: a service and a service-API artefact. The former comprises all the business logic that is required, in addition to a data layer and REST controllers, if these are deemed necessary. The latter consists of the API features that are required in order to communicate with the service. The aforementioned elements comprise data transfer objects (DTOs), client classes that serve as the one-to-one counterparts to the service's REST controllers, controller routes, and other configuration items.

JSON Web Tokens (JWTs) are used to prevent unauthorised access to and among the services. These tokens include the rights and roles of the user initiating the request. Each request header contains such a token. The

service receiving the request will interpret it. PTB operates its own identity provider. This is used to grant federated users specific rights and roles. These are tailored to the services. It also allows PTB to trace each operation back to a user. Use the OpenID-based identity and access management solution Keycloak, which is part of the commonly used resources within OP-Layer. This achieves the required result.

The Kubernetes design principles[1] fulfil the majority of the aforementioned requirements for a modern distributed infrastructure approach [7]. The subsequent bullet points delineate the requisite criteria:

- **Scalability** - is defined as the ability to expand or grow in size or capacity. In the context of applications, scalability refers to the capability of a system to handle increasing demand or workload. Horizontal scaling, a method of enhancing scalability, involves the execution of the same application multiple times. This approach ensures that the application's request load is distributed across multiple instances, thereby mitigating the impact of high demand. It is evident that Kubernetes is equipped with the inherent functionality of load balancing and horizontal scaling capabilities, which are available without the necessity for additional configuration or customisation.
- **Portability** - Deployment operations have undergone significant changes. Containers and pods enable applications to run on any platform and cluster. Kubernetes provides a range of container runtimes, facilitating application portability and straightforward deployment for an agile development approach.
- **High Availability** - It should be noted that backups and replicas address the high availability requirement. Whilst backups concentrate on data restoration at a specific point in time, replicas support business continuity. Kubernetes is designed to ensure high availability at both the application and infrastructure level.
- **Security** - It is essential that security be addressed at the cluster, application and network levels. Furthermore, it is imperative to implement stringent rights and role management for each user, thereby ensuring that only authorised operations are executed by applications in the cluster. In a distributed architecture with loosely-coupled services, a session-less authentication and authorisation approach using tokens, which are placed in each HTTP REST request, has proven itself to be state-of-the-art. This approach ensures that each operation can be traced back to the user initiating any defined process in any use case. Moreover, the API endpoints are secured via Transport Layer Security (TLS).

[1] https://kubernetes.io/docs/concepts/overview/

IV. SERVICE INFRASTRUCTURE

This section outlines the most significant services that have been implemented to date and those that are scheduled for completion within the current year. During implementation, great care was taken to ensure that these were implemented as generically as possible so that the code base can be reused in other use cases.

A. *E-File*

The E-File is an integrated service for electronic file management. It is based on a structured filing system for secure, revision-safe file management. In addition, the E-File provides functionalities for collaboration and support for case processing and special procedures.[2]

The interoperability of processes is best shown when using E-File, as it is the center piece of numerous procedures within PTB. So whenever documents and attached meta data are part of a procedure, there will be interoperability between E-File service and other services within OP-Layer. Every described use case described in this paper interoperates with E-File.

B. *DigiSeal*

Establishing a digital workflow by interconnecting different information systems with a central electronic sealing capability is the key design aim for a formal process. Currently, two file types are supported: PDF files allow embedding electronic seals from the PDF/A-1 standard onward. Digital Calibration Certificate (see section v.) XML files can also take an embedded electronic seal according to the W3C XML Signature[3] recommendation.

The Digital Seal Service within OP-Layer [8] interacts with E-File Service in the following way. Digital Seal Service is job-controlled, meaning that it checks periodically via E-File Service for circulars containing documents to seal. The interoperability between E-File and DigiSeal through OP-Layer occurs as follows:

1. The service receives circulars directed at the organizational unit (OU) "Siegelstelle" (sealing authority), and extracts documents in these circulars.
2. The service opens the documents, extracting any records, as actual physical files, such as PDF or XML, from them.
3. Digital Seal Service sends theses files to *digiSeal server* through its proprietary SOAP interface. The *digiSeal server* is a third-party product developed by secrypt GmbH[4]. PTB runs a server instance, as part

[2] https://www.bva.bund.de/DE/Services/Behoerden/Verwaltungsdienstleistungen/E-Akte/e-akte_node.html

[3] https://www.w3.org/TR/xmldsig-core/

[4] See https://www.secrypt.de/en/homepage/

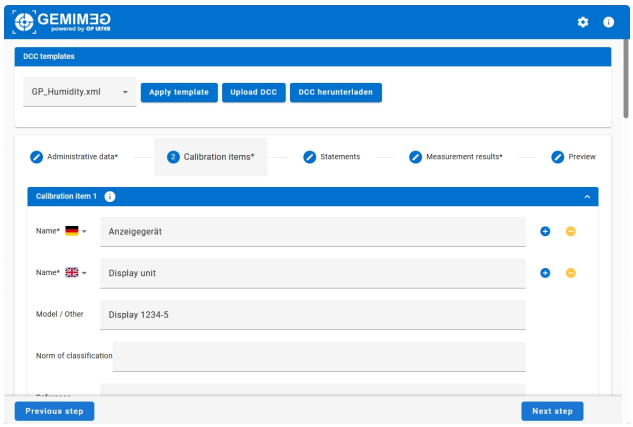


Fig. 1. GEMIMEG 2.0 Tool Editing Mode

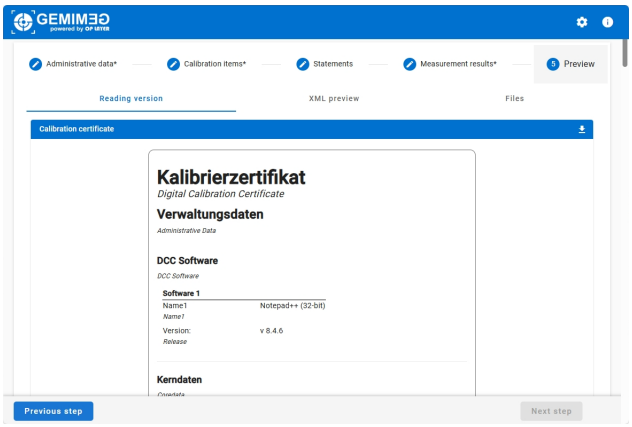


Fig. 2. GEMIMEG 2.0 Tool Preview Mode

of the central OU "Siegelstelle".

4. The sealed files are placed in the documents, which were received though steps one and two. The E-File Service then redirects the circular(s) to the respective initiator(s) of the original procedure(s).

C. Certificates and ID generation

Every report, certificate, conformity declaration and other such documents have unique identifiers, which follow specific patterns, depending on the type of document, and the issuing department. The identifier generation patterns have been implemented in a backend service within OP-Layer, which not only generates new document identifiers, but also stores already generated ones in order to keep them unique. This service, called *Document ID generator*, can interoperate with E-File, when a document, which is to be a result of e. g. a calibration procedure. The resulting calibration certificate will then receive a unique document identifier from Document ID generator, and, if required, an electronic seal from Digital Seal Service (see section iv.-B.).

For measurement instruments, calibration and conformity certificates are issued by PTB. They are stored in a central database called the *MICert* database. The OP-Layer service to manage certificates in this database is currently under development. It will automatically, and periodically pull any new certificates from E-File and store them in the MICert database. The database has a publicly available web interface for customers and other authorities, so keeping it up-to-date is another aspect of how interoperability advances digital transformation.

V. DIGITAL CALIBRATION CERTIFICATE WORKFLOW

As part of the digital transformation strategy and propagation, the Digital Calibration Certificate (DCC [9]) serves as the cornerstone to provide a digital document format for official calibration certification. As it is XML-based, DCC allows for machine-readable, interoperable data exchange. The first showcase implementation of DCC within OP-Layer was concerned with calibration applications submitted via the E-Service Portal, forwarding its data via E-File to the calibration laboratory and providing a valid DCC [7], [10].

At the current stage of development, the use of DCCs is integrated into several processes already in place within OP-Layer. The following sections describe the usage and integration of DCCs in different use cases.

A. Embedded electronic seal

For processing DCC XML files, a document in E-File must contain a DCC XML file which is to be sealed. The Digital Seal Service will check though the DCC Service, if the provided XML is a valid DCC according to the DCC schema. If this is the case, Digital Seal Service will process it further according to steps three and four, as described in section iv.-B..

B. Gemimeg Tool 2.0

The GEMIMEG Tool 2.0[5] is a stand-alone version of two components: A frontend and a backend service. The frontend contains everything required to run the DCC display and editing components as an Angular[6] service. The backend contains the DCC Service and in addition the Cache Service as well as several common components, integrated within one backend instance. OP-Layer's security components are not part of this version, because its main goal is to be able to run independently even on client computers, which don't need to be connected to the internet. Figure 1 shows the editing of a DCC conveniently in a browser window. It currently supports five languages: English, German, French, Spanish and Portuguese, in order

[5] https://gemimeg-tool.ptb.de
[6] https://angular.dev/

to address different NMIs and related institutions around the globe.

The GEMIMEG Tool also has a preview mode, so that the currently edited DCC can be previewed in a human-readable form. This preview can also be downloaded as PDF. The preview mode is white-labeled, so that an institution wishing to use the tool to generate DCCs can place their own logo in the resulting documents. Figure 2 shows the white-labeled preview mode in a browser window.

Carving out the backend DCC and Cache Services from OP-Layer's backend service ecosystem is comparably easy, as they are already loosely coupled and communicate via open and harmonized RESTful interfaces. The only integration that was made is integrating both Cache and DCC services into one GEMIMEG Backend service. As mentioned above, the security features comprising of JWT-based authentication and IAM (Identity Access Management) integration (e.g. Keycloak) have been entirely removed from the code base.

The frontend separation required to first create a library of re-usable components for Angular, and make it available for NPM[7] so that it can be integrated within OP-Layer's frontend as well as GEMIMEG frontend from the same code base. This way, and feature development and maintenance done for one of them automatically finds their way into the other.

C. SI Digital Framework

The DCC schema uses the D-SI schema to include dimensions. By doing this, the DCC service interoperates with another OP-Layer backend service, the D-SI service. It validates SI units, assures correct conversions and verifies number formats [11].

VI. CONCLUSIONS AND FUTURE WORK

In this paper, the concept of interoperability is introduced and defined in terms of technical and syntactical interoperability. To ensure the profitable implementation of interoperable processes, the primary focus should be on harmonised interfaces and data exchange formats. Furthermore, to implement interoperable processes successfully, it is essential that the organisational dimension is resolved at a management level. In the absence of interdepartmental interoperability, it is implausible that any technical solution will be capable of overcoming this issue, let alone one that is applicable to different institutions.

If a process management system is in place that ensures linear and concise processes with clearly defined responsibilities, then the proposed business process middleware Operation Layer can unfold its full potential in a cross-institutional data sharing infrastructure. This paves the way for a holistic digital transformation of a highly automated infrastructure, with data sharing at its core. Enabling future projects such as predictive and automated calibration of measuring instruments.

[7]Node Package Manager is the recommended package manager for Node.js, which Angular is built on.

REFERENCES

[1] Christoph Jacob. Designing innovative ecosystems and introducing digital smart services using examples of the value chain from building investor to facility management. *Creating Innovation Spaces: Impulses for Start-ups and Established Companies in Global Competition*, pages 99–117, 2021.

[2] Cambridge English Dictionary. Meanings & definitions, 2025. https://dictionary.cambridge.org/dictionary/english/interoperability?q=Interoperability [Accessed: (2025-04-28)].

[3] Hans Van Der Veer and Anthony Wiles. Achieving technical interoperability. *European telecommunications standards institute*, 2008.

[4] C Kulka-Peschke, S Eickelberg, A Keidel, M Meiborg, and A Oppermann. A1. 3-digital transformation of processing metrological services. *Lectures*, pages 33–34, 2023.

[5] Anke Keidel and Sascha Eichstädt. Interoperable processes and infrastructure for the digital transformation of the quality infrastructure. In *2021 IEEE International Workshop on Metrology for Industry 4.0 & IoT (MetroInd4.0 & IoT)*, pages 347–351, 2021.

[6] Alexander Oppermann and Samuel Eickelberg. Digitale Transformation in der Metrologie: Harmonisierung digitaler Identitäten mittels OpenID. *Werkwandel - Zeitschrift für angewandte Arbeitswissenschaft - Ausgabe #1 - Februar,2022*, pages 26–30, 2022.

[7] Alexander Oppermann, Samuel Eickelberg, and Manuel Meiborg. Digital transformation: Towards process automation in a cloud native architecture. *Acta IMEKO*, 12(1):1–6, 2023.

[8] Alexander Reissert, Anke Keidel, and Samuel Eickelberg. Digital transformation: Building an event driven electronic seal service within a cloud native metrological ecosystem. *Measurement: Sensors*, page 101489, 2024.

[9] Siegfried Gustav Hackel, Frank Härtig, Julia Hornig, and Thomas Wiedenhöfer. The digital calibration certifcate. *PTB-Mitteilungen*, 127(4):75–81, dec 2017.

[10] Samuel Eickelberg, Thomas Bock, Matthias Bernien, and Alexander Oppermann. Integrating a calibration labaratory worflow into a metrological digital ecosystem: A case study. *Acta IMEKO*, 12(1):1–6, 2022.

[11] Richard JC Brown, Jan-Theodoor Janssen, and Louise Wright. Why a digital framework for the si? *Measurement*, 187:110309, 2022.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Advanced metrological knowledge representation in the D-SI metadata model

Daniel Hutzschenreuter[1], Matthias Bernien[2], Frauke Gellersen[1], Moritz Jordan[1], Benedikt Seeger[1]

[1] *Physikalisch-Technische Bundesanstalt, Bundesallee 100, 38112 Braunschweig, Germany,*
daniel.hutzschenreuter@ptb,de
[2] *Physikalisch-Technische Bundesanstalt, Abbestraße 2-12, 10587 Berlin, Germany*

***Abstract* – Universal, unambiguous, safe, and easy to use metadata models are a critical prerequisite for the efficient digitalisation of services and tools in metrology. The D-SI is providing users in metrology and their stakeholders with a simple, yet powerful framework of metadata models for an efficient implementation of data based on the International System of Units. Recent extensions of the D-SI address needs of advanced application such as digital certificates of calibration and findings in respect to the ongoing digitalization under the International Committee of Weights and Measures. Updates are presented comprising improvements of semantics, kinds of quantities, statistical distributions of measurement uncertainties and representations of dependencies of uncertainties.**

I. INTRODUCTION

Universal, unambiguous, safe, and easy to use metadata models are a critical prerequisite for the efficient digitalisation of services and tools in metrology. The General Conference on Weights and Measures has declared in its 26th meeting in 2022 [1] that it is a core mission for the metrology organisations and their partner organisations in standardisation, accreditation, and science to work towards commonly accepted digital representations of metrological data. These representations shall be based on the International System of Units (SI) [2] and implemented by the use of the FAIR (findable, accessible, interoperable, reusable) data principles [3]. A SI Digital Framework [4] and digital representations for the units of measurement of the SI-brochure [5] are under development in respect to this mission by the International Bureau of Weights and Measures and the Forum on Metrology and Digitalization of the International Committee of Weights and Measures [6]. These developments build an authoritative and trusted international digital reference for all knowledge related to the realization of and metrological traceability of measurements to the SI.

Metadata models such as the D-SI [7] are important tools to implement data exchange formats for humans and machines, that are linking the international references with a wide range of practical applications. These are, for example, measurement data in Digital Calibration Certificates [8], assets of research data, data ready for the training and verification of Machine-Learning, or data being exchanged through programming interfaces of software. The D-SI is in particular providing a consistent framework of metadata models combining the representation of measured quantity values (numbers, units) with semantics of the international Vocabulary in Metrology [9] and the Guide on the Expression of Uncertainties in Measurement [10] based on needs from both, the international perspective of metrology and users in industry.

Code Example 1 shows a quantity value for real and scalar data from the D-SI in Extensive Markup Language (XML) format [11]. The root element is si:real which is encapsulating data and metadata that are common to a majority of applications. Mandatory information are the numerical value of the quantity (si:value) and the associated unit of measurement in the SI (si:unit). Additional data are a text label (si:label), a time-stamp for the date and time attributed to the quantity (si:dateTime), and metadata on the uncertainty of the measurement (si:univariateMeasurementUncertainty). The uncertainty is provided by a standard measurement uncertainty (si:standardMU).

Example 1. Real quantity value

```
<si:real>
  <si:label>effective area of displacer
      </si:label>
  <si:value>0.01</si:value>
  <si:unit>\metre\tothe{2}</si:unit>
  <si:dateTime>2025-09-03 T13:42:00Z
      </si:dateTime>
  <si:univariateMeasurementUncertainty>
    <si:standardMU>
      <si:valueStandardMU>0.002
          </si:valueStandardMU>
      <si:distribution>normal
          </si:distribution>
    </si:standardMU>
  </si:univariateMeasurementUncertainty>
</si:real>
```

It defines the value of the standard uncertainty (si:valueStandardMU) and has the optional field (si:distribution) to define a Probability Density Function (PDF) of the uncertainty. Simple strings such as 'normal' and 'uniform' were used to annotate basic kinds of PDFs. Other elements in the D-SI are constant, complex and multivariate quantity values.

This paper presents recent advancements of the D-SI considering needs of priority that we have identified in discussions with partners from the international committees of metrology (CIPM FORUM-MD) and stakeholders of the application in DCCs. Section II will give an outline of general concepts that have been added to the D-SI metadata model in 2023. Section III gives an introduction to the future representation of Probability Density Functions (PDFs) for a better description of statistical properties of uncertainties of measurement. Furthermore, Section IV will outline the inclusion of uncertainty models of dependencies and resulting benefits for the propagation of measurement uncertainties across chain-links of calibration. This work will be concluded by a discussion of the updates in Section V and an outlook to future developments is given in Section VI.

II. NEW GENERAL CONCEPTS IN THE D-SI

New general concepts comprise extensions to the existing data models that have been requested by user communities of the D-SI. These extensions have been identified to be of high priority to improve the interoperability and useability of metrological data for a wider number of users and are available in D-SI version 2.2.0 since 2023.

A. Kinds of quantities

Quantity values using SI-base units as reference for the magnitude can simplify calculations such as the conversion and interoperation between data. To enable higher levels of an unambiguous interoperation it is however necessary to extend the information on the unit by metadata on the underlying kinds of quantities and scales of measurement [12]. One example are data for radioactive decay with the unit Becquerel and frequency with the unit Hertz. Both units have the SI-base representation 1/s and adding a mix of both quantity values only based on seeing, it is the same units of measurement would be an incorrect choice. Another example is the interoperation of quantities representing absolute temperature and differences of temperature. Depending on the kind of quantity (aspects) and units such as Degree Celsius and Kelvin, conversion of data needs to be able to distinguish between different measurement scale types and apply the correct calculations for these scales. The need to be able to assess information on the quantity for interoperation is in general relevant, if quantities of dimension one are present such as ratios of quantities of the same kind and counting quantities.

Example 2. Quantity and significant digit

```
<si:real>
  <si:label>effective area of displacer
      </si:label>
  <si:quantityTypeQUDT>Area
      </si:quantityTypeQUDT>
  <si:value>0.01143</si:value>
  <si:unit>\metre\tothe{2}</si:unit>
  <si:signifcantDigit>-3
      </si:signifcantDigit>
  <si:dateTime>2025-09-03 T13:42:00Z
      </si:dateTime>
</si:real>
```

The D-SI is addressing this need by introducing an abstract element for the quantity kind [9, def. 1.2] and the list of identifiers for kinds of quantities provided by the QUDT ontology [13] was adopted to the D-SI. Example 2 shows how the QUDT identifier can be written as element si:quantityTypeQUDT. Users can access metadata on the definition of the kind of quantity from QUDT by calling the resolvable persistent identifiers <prefix>/<ID>.ttl with the prefix qudt.org/vocab/quantitykind and ID ‘Area’ from the example.

B. Significant digits

Knowledge on the significant digits of the magnitude of a quantity value is a requirement in practical applications where inference on the accuracy of data is deduced from printed (significant) digits in addition to values of uncertainty. The D-SI recommends to directly include contributions of rounding of values related to significant digits within the measurement uncertainty value as a clear way to enable a consistent and easy propagation of measurement uncertainty by machines. For cases, where users are required to provide additional information, an option was added to identify the significant digits. Using this option shall not replace the requirement to properly handle all uncertainty contributions within a budget leading to the reported uncertainty value.

A review of existing definitions for significant digits for an adoption in the D-SI turned out to lead to no useful reference. IEC [14] on basic terms in IT defines “least significant digit”. However, it is not disclosed how to identify this digit in data. DKD Guideline L-13 [15] defines significant digits from the metrological perspective, where the leading two left digits of the uncertainty value are used to identify the significant digits of the value. However, it considers to provide the number of significant digits which is ambiguous, as these are counted starting from non-zero digits which will not work for digital numbers close to zero. ISO 80000 [16] addresses the need to identify significant digits by introducing the property of a rounding range. This rounding range is a power of ten defining the digit where rounding is applied which is more robust regarding the

digital representation of numbers than counting digits.

It was decided to introduce the element si:significantDigit in the D-SI which is defined to provide the integer number of the exponent of the power of ten of the rounding range as it is invariant considering different representations of numerical numbers and as users can quickly access the properties of significant digits considering the before mentioned specifications of IEC, ISO and DKD.

Example 2 shows the significant digit number -3. It implies that the value 0.01142 has the significant digits 0.011 and the rounding range is 10^{-3}. Even if the value -3 is an exponent and not an actual identification of the significant digit in the sense of a specific place in the mantissa of a number, it was decided to still call it the significant digit to increase the recognition by users.

C. Semantics of uncertainty of measurement

As the development of the SI Digital Framework is progressing and as new implementations in the form of ontologies become increasingly important, the role of semantics was carefully reviewed with the aim to close potential gaps that would limit a future interoperability. Semantics do on the one hand consider the terms used in data, such as identifiers for the units and kinds of quantities. Using ontologies such as the SI Reference Point and QUDT are important resources to enhance the semantic interoperability. On the other hand, element names and their underlying data types in the data model carry additional information on concepts that need to be interoperable across applications and domains. To be able to adhere to this requirement, it is important to take care when implementing the D-SI metadata model in specific formats such as XML. The XML implementation is based on an XML Schema Definition (XSD) file which sets up the element names and all underlying data types, whereby XML is coming with a set of universal primitive types such as integers, decimal (floating point numbers), strings, and Booleans. Using these primitive types directly to represent data creates only a weak semantic representation. Take for example the property of the label and the PDF of the measurement uncertainty. If both values would only be of the primitive type string, a machine would not be able to identify the difference in the meaning. Also, it is more difficult in this case to inform humans and machines in a machine-actionable way about links between these types and elements in XML and semantics form other sources such as controlled vocabularies for semantics of the VIM and GUM. To circumvent such kinds of limitations, the D-SI follows a strict strategy of typing where each element and property is provided by a unique and dedicated data type [11]. The real quantity value element has for example the type si:realQuantityType in the XSD which represents a subclass of a quantity value. All primitive data types are encapsulated by metrology specific types such as a type for the unit, the value of the magnitude and the PDF.

Example 3. Alignment of semantics of D-SI implementations for real quantity value.

```
// XML
<element name=”real” type=”realQuatityType”>

// TURTL (ontology)
https://ptb.de/si#Real rdf:type owl:class,
        rdfs:subClassf dsio:QuantityValue;

// JSON-LD context
"real": {
 "@id": "https://ptb.de/si#Real",
 "@type": "https://ptb.de/si/realQuantityType"
}
```

Furthermore, uncertainty statements were semantically aligned regarding their dimensions as univariate measurements uncertainties (standard, expanded, coverage interval uncertainties) and bi- and multivariate uncertainties (coverage regions). To establish type definitions of all uncertainty models in the D-SI that are ready for future semantical interoperation, the new elements si:standardMU, si:expandedMU, and si:coverageIntervalMU were introduced. At the same time, we deprecated the old definitions si:expandedUncertainty and si:coverageInterval as their use of primitive types was not in line with the needs for future semantics.

As one side-effect, these updates also facilitated the alignment of various implementations of the D-SI. The XML implementation defines basic element names and semantic data types supporting the machine-readability of data. The implementation of the D-SI ontology instantiates a comprehensive knowledge graph adding meaning to the elements (classes of entities) and their properties in relation to other knowledge graphs such as the SI Digital Framework for reasoning by machines [17]. A JSON-LD (Java Script Object Notation – Linked Data) context definition of the D-SI [18] requires to associated element names with unique IDs of entities and data types which aligns with both, the ontological and XML representations. Example 3 outlines the links between the elements.

III. PROBABILITY DENSITY FUNCTIONS

Using PDFs to model measurement uncertainties is a fundamental practice in metrology. To increase the digital maturity of interoperability, it is necessary to introduce parameterized entities for PDFs within data models. Semantics of PDFs require to identify the kind of a distribution such as the normal or student-t distribution and to align unambiguous types of parameters such as expectation values, variance, or degrees of freedom.

A. Requirements of D-SI applications

The D-SI metadata model aims for an easy-to-use and universal implementation of PDFs that is not dependent on

any specific formatting of data such as XML or JSON. A representation should be compatible to the current practice in the D-SI where information can be included in a single line of code. PDFs defined by the GUM and its supplements are of relevance and the implementation should also support distributions from a wider range of applications. Furthermore, the representations need to adhere to FAIR data principles and allow semantic interoperability. This requirement includes the availability of persistent identifiers (PIDs) for all PDFs. The PIDs should allow to link to metadata on the distributions such as knowledge graphs and machine-actionable formulas that support a use and evaluation within software. Finally, existing and established standards should be reused where possible.

B. Approach for representing PDFs

A review of existing standards for representing PDFs in a machine-usable format comprised openly available knowledge bases, ontologies, and software libraries. The number of existing approaches is surprisingly huge, as many specific solutions have been established. The reason for this development is probably the lack of commonly adopted international standards which poses a domain-overarching interoperability challenge. Thus, our investigation focused on comprehensive systems such as PDFs in METAS's UncLib [19], Wolfram software [20], UncertML [21], or ProbOnto [22]. The later implementation, ProbOnto, is an ontology for more than 150 different types of univariate and multivariate distributions. It gives a structured listing of codes for distributions and their parameters which are considered to be used within the D-SI. Examples of the codes are shown in Table 1.

Table 1. ProbOnto PDF code examples.

PDF type code	**Parameter codes**
Uniform1	minimum, maximum
Normal1	mean, stdev
StudentT2	mean, scale, degreesOfFreedom

The proposal for the representation of the PDFs in the D-SI will use the codes of ProbOnto to define PIDs for distributions. The formalism to write the PID is the following:

<T>?<P>=<V>;…;<P>=<V>

where <T> is a placeholder to enter the PDF type code of the distribution, "?" is a separator indicating that the list of parameters begins, the entries <P> are codes of distribution parameters, "=" is a value separator, <V> are the numerical values of the parameters (double precision floating point numbers allowing scientific exponent), and ";" is used for separating two or more parameters. This notation is a standard for parameterized URLs.

Example 4. PID of normal distribution in Example 1.

```
// PID
Normal1?mean=0.01;stdev=0.002
//D-SI element notation
<si:distribution>
    Normal1?mean=0.01;stdev=0.002
        </si:distribution>
```

Example 4 outlines a PID for a normal distribution. The tag 'Normal1' indicates, that the PDF is a normal distribution. Parameter 'mean=0.01' is the mean value of this distribution and 'stdev=0.002' is the value of its standard deviation (square root of variance value).

C. FAIR services to resolve PDFs

The PID is only a simple semantic notation of the distribution of a measurement. To increase interoperability, it is required to establish additional FAIR services that link substantial metadata to the PIDs. Such services could be web services for example, that are looking up properties and metadata from ProbOnto and provide it to users in a semantically meaningful way by using ontologies or similar encodings. As the PID is implemented on the formalism of parametrized URLs, resolvers for the metadata can be implemented as simple RESTful Get requests [23] in the form of

<web service root URL><PID>.

IV. DEPENDENCIES OF UNCERTAINTIES

So far, the D-SI models for measurement uncertainty are based on the reporting of simple uncertainty numbers, which are outcomes of the evaluation of uncertainty budgets. The budget-specific mathematical combination of different components of uncertainty is known and defined in the GUM under the principle of the linear propagation of uncertainties [10, section 7.1.4]. A better identification and ability to reuse information on individual components of uncertainty can help to improve the accuracy of data processing along chains of subsequent measurements and measurement models. It also helps to uncover hidden correlation between data. By introducing dependencies of uncertainty to the D-SI, the use of such information on components of uncertainties will be enabled.

A. Comparison between conventional approach and measurement data considering dependencies

The hands-on example of calibrating leaks in gas flow measurement will be used to explain the concept of dependencies, how it is implemented in the D-SI, and how

it could improve the propagation of measurement uncertainties across DCC in calibration chains. In the calibration of test standards for low gas flow leaks, a model based measurand is established by the interpolation of flows coming from two reference leaks. Let Q_1 and Q_2 be the gas flows of the reference leaks R_1 and R_2. Then the interpolated gas flow of the test leak is

$$Q = \alpha(Q_2 - Q_1) + Q_1 \tag{1}$$

The parameter α depends on the readings of the reference leaks using a mass spectrometer, but for simplicity of the example, we consider it to be a constant value without uncertainty. The values of the gas flows of the references are calibrated individually. The calibrations are made using the same primary standard for flow. The flows are calculated according to Equation 2 where p_1 and p_2 are pressure measurements that are multiplied with the volume difference of the primary standard over time. The differences of time are Δt_1 and Δt_2. The volume difference is calculated by multiplying the effective area $A = A_1 = A_2$ of a calibrated displacer by its change in length Δl_1 and Δl_2.

$$Q_1 = p_1 A_1 \frac{\Delta l_1}{\Delta t_1}, \qquad Q_2 = p_2 A_2 \frac{\Delta l_2}{\Delta t_2} \tag{2}$$

The combined standard measurement uncertainties of Q_1 and Q_2 are calculated according to Equation 3 following the law of the propagation of uncertainties. The coefficients c_* are the partial derivatives of Q_i with respect to the variables p_i, A, Δt_i and Δl_i. The standard measurement uncertainties of the inputs are the values $u(*)$.

$$\begin{aligned} u^2(Q_i) = {} & c_{p_i}^2 u^2(p_i) + c_{A_i}^2 u^2(A) \\ & + c_{\Delta l_i}^2 u^2(\Delta l_i) + c_{\Delta t_i}^2 u^2(\Delta t_i), \\ & i = 1{,}2 \end{aligned} \tag{3}$$

It is obvious, that Q_1 and Q_2 are correlated through the component A, but also through the reference instruments that are used to measure the pressure, change of length and time. So far, this correlation has not been taken into account explicitly in calibrations of Q. We are assessing the influence of A in the example in this paper. The other contributions of correlation are not considered for simplicity.

If the calibration model for Q does not consider correlation, then its combined standard measurement uncertainty is

$$u_{uncor}^2(Q) = (1-\alpha)^2 u^2(Q_1) + \alpha^2 u^2(Q_2). \tag{4}$$

The calculation of the correlated uncertainty is done by

$$\begin{aligned} u_{cor}^2(Q) = {} & (1-\alpha)^2 u^2(Q_1) + \alpha^2 u^2(Q_2) \\ & + 2\alpha(1-\alpha)\, u(Q_1, Q_2) \end{aligned} \tag{5}$$

using the covariance value

$$u(Q_1, Q_2) = c_{A_1} c_{A_2} u^2(A). \tag{6}$$

Table 2 shows data of a calculation close to a real-world application. There is a clear difference in the uncertainty value considering correlation.

Table 2. Example of values of uncertainty with and without correlation.

Input quantity values	
$p_1 = 3.250000$ Pa, $\Delta l_1 = 4.76000 \cdot 10^{-3}$ m, $\Delta t_1 = 234.000$ s, $p_2 = 35.300$ Pa, $\Delta l_2 = 4.08000 \cdot 10^{-3}$, $\Delta t_2 = 200.600$ s, $A = 9.0256 \cdot 10^{-5}$ m²	$u(p_1) = 3.3 \cdot 10^{-5}$ Pa, $u(\Delta l_1) = 2.9 \cdot 10^{-7}$ m, $u(\Delta t_1) = 5.0 \cdot 10^{-2}$ s, $u(p_2) = 3.5 \cdot 10^{-2}$ Pa, $u(\Delta l_2) = 2.9 \cdot 10^{-7}$ m, $u(\Delta t_2) = 5.0 \cdot 10^{-2}$ s, $u(A) = 6.9 \cdot 10^{-8}$ m²
Model-based output quantity values	
$Q_1 =$ $5.9669 \cdot 10^{-9}$ N·m/s $Q_2 =$ $6.48 \cdot 10^{-8}$ N·m/s	$u(Q_1) =$ $7.7 \cdot 10^{-12}$ N·m/s $u(Q_2) =$ $8.4 \cdot 10^{-11}$ N·m/s
$Q = 2.0799 \cdot 10^{-8}$ N·m/s, ($\alpha = 0.2521008$)	$u_{uncor}(Q) =$ $2.2 \cdot 10^{-11}$ N·m/s $u_{cor}(Q) =$ $2.4 \cdot 10^{-11}$ N·m/s

B. Existing approaches

Two important tools explicitly applying dependencies are METAS's UncLib which is developed by Wollensack [19] and the GUM Tree Calculator (GTC) which is developed by Hall [24]. Common to both tools is the ability to propagate measurement uncertainty and correlation along measurement models that relate the measurement uncertainty of calculated quantities to input quantities. Uncertainties of input quantities in the models are standard measurement uncertainties; each uncertainty has a unique identifier and it is possible to quantify correlation between inputs. These inputs are in our example the properties p_i, A, Δt_i and Δl_i. The uncertainties of output quantities in the measurement models are combined standard measurement uncertainties which are provided as lists of the components of the uncertainty budget. In our example, these components are $c_{p_i} u(p_i)$, $c_{A_i} u(A)$, etc. Each of the listed items has as additional information the identifier of the uncertainty of the input quantity it is depending on.

Despite the conceptual similarities of both tools, there are also differences. The latest version of UncLib considers in addition the usage of non-linear models to

propagate uncertainty and the application of Monte-Carlo-type methods for a simulation of properties of uncertainty. Thus, the uncertainty contributions are implemented by the sensitivity coefficients (Jacobian matrix values c_*) and providing the PDFs related to the input quantities. Correlation between input quantities is implemented by decomposing the correlation matrix into a set of virtual input quantities whose dependencies to the actual input quantities are used to reconstruct correlation data. A specific Eigenvalue decomposition method is required to calculate the virtual inputs. GTC is focusing on the linear propagation of uncertainties. It provides the dependencies by the values $c_* u(*)$. In addition, correlation between inputs can be expressed through a separate list of correlation values for each input quantity. GTC also allows explicit identification of interim quantities and their dependencies.

C. Modelling of Dependencies in the D-SI

The requirements to ensure an easy-to-use D-SI and the integration with existing models for uncertainty are guiding the modelling of dependencies. User will need calculation software (middleware) to create and process data. The software must be able to handle correlation between input quantities, underlying units of measurements, and types of quantities in a straightforward way and, where appropriate, preventing redundancy of information.

In this respect, the existing D-SI models for univariate uncertainties are extended by an optional identifier field (si:uniqueID) to provide unique identifiers of the uncertainty components of input quantity values. Furthermore, a list of correlation values (si:correlations) can be added to univariate uncertainty statements. Example 5 outlines the input quantity A_1 in respect to the before mentioned properties. The quantity value is providing a standard measurement uncertainty element with the unique identifier 'ID:A1'. As A_1 and input quantity A_2 ('ID:A2') are identic, the list of correlations is providing correlation value one. This information is written in the D-SI as a list item (si:correlation) that has the numerical correlation value (si:valueCorrelation) and the unique identifier of the input quantity it is related to (si:correlatedQuantityID). Correlation elements are only written in the D-SI if their values are non-zero and for identifier pairs where si:uniqueID is not equal to si:correlatedQuantityID. Correlation between inputs would typically be specified within one DCC.

For writing the uncertainty of an output quantity that is derived from dependencies, a new univariate measurement uncertainty model is introduced (si:dependenciesMU). This model provides a list of the components (si:dependency) of the uncertainty. Each dependency consists of a numerical value (si:valueComponentMU) which is the product $c \cdot u$ (uncertainty of an input quantity times the sensitivity coefficient from the uncertainty budget) and a reference to the unique identifier of the input quantity (si:inputQuantityID). The unit of measurement and type of quantity of the parent D-SI element (si:real) apply to the numerical value which is in line with other uncertainty models. Example 6 outlines this new model to represent the value and uncertainty of gas flow Q including correlation. Within this approach unique identifiers of the input quantities and their uncertainty component provide a way of identifying and quantifying correlations between uncertainties stemming from different DCCs.

Example 5. Dependencies in D-SI – input quantity

```
<si:real>
  <si:label>effective area of displacer A1
      </si:label>
  <si:value>0.01</si:value>
  <si:unit>\metre\tothe{2}</si:unit>
  <si:univariateMeasurementUncertainty>
    <si:standardMU>
      <si:valueStandardMU>0.002
          </si:valueStandardMU>
      <si:distribution>normal
          </si:distribution>
      <si:uniqueID>ID:A1</si:uniqueID>
    </si:standardMU>
    <si:correlations>
      <si:correlation>
        <si:valueCorrelation>1
            </si:valueCorrelation>
        <si:correlatedQuantityID>ID:A2
            </si:correlatedQuantityID>
      </si:correlation>
    </si:correlations>
  </si:univariateMeasurementUncertainty>
</si:real>
```

Example 6. Dependencies in D-SI – output quantity

```
<si:real>
  <si:label>gas flow Q</si:label>
  <si:value>2.50</si:value>
  <si:unit>\metre\tothe{2}</si:unit>
  <si:univariateMeasurementUncertainty>
    <si:dependenciesMU>
      <si:dependency>
        <si:inputQuantityID>ID:1
            </si:inputQuantityID>
        <si:valueComponentMU>0.025
            </si:valueComponentMU>
      </si:dependency>
      <si:dependency>
        <si:inputQuantityID>ID:2
            </si:inputQuantityID>
        <si:valueComponentMU>0.025
            </si:valueComponentMU>
      </si:dependency>
      […]
    </si:dependenciesMU>
  </si:univariateMeasurementUncertainty>
</si:real>
```

A complete example of input and output quantity data in D-SI XML format and example calculations based on the R application are available [26].

V. DISCUSSION

The recent updates of the D-SI lay a foundation for applications that need higher levels of machine-interoperability in the future. These comprise the ability to interoperate units of measurement and kinds of quantities in a common framework but also to integrate with digital approaches for representing semantics in metrology and chains of traceability.

Metrological communities that are fostering the harmonization of the use of kinds of quantities are still at the beginning of their work. There is a clear need for a machine-actionable representation of kinds of quantities defined by the International System of Quantities [16], however, no digital representation exists today that is fully covering the whole system. For this reason, the D-SI may adopt further representation systems in addition to QUDT in the future. Important and emerging developments are the kinds of quantities from BIPM's SI Digital Framework [4], IEC's Common Data Dictionary codes [25], and developments such as the M-Layer registry of units, quantities and measurement scales [12].

It was already mentioned that the provision of PDFs is posing a new interoperability challenge. In addition to the suggested inline codes in the D-SI, proper services are required to provide machine-actionable metadata. The final technical implementation of the representation of the PDF's parameters is still to be determined and consultations with stakeholders are ongoing. While the URL-based string representation has advantageous regarding querying of metadata, an alternative hierarchical representation of key-value-pairs of parameters is considered to ease parsing of data in software.

The D-SI is aiming to establish a basic web service covering distributions from the GUM [10]. Initial ideas on this include the creation of PIDs for each distribution that can be resolved to machine-readable metadata derived from a common ontology. Additional APIs could be added to allow to draw random samples of the distribution, to create plots, and to interoperate information across different formats of representation. Moreover, some statistical analysis also requires the provision of products of PDFs (multiplication). The digital representations and web services will need to allow this in future.

Modelling of uncertainties based on dependencies is posing a great opportunity but also a great challenge to the whole user community. The technical implementation is accompanied by many additional questions that need to be solved. A thoughtful definition of the unique identifiers is required for example. One requirement is to limit the length of the identifiers to a fixed number of characters to enable efficient traversing of data by software. At the same time, the identifiers can act as persistent and resolvable links in the sense of FAIR data. Furthermore, in the case of UncLib, the identifiers can also serve as random seeds for Monte Carlo simulation. Another challenge is the need of additional training capacities to enable users from a wider number of communities to properly apply the dependencies. Semantic representations of knowledge in the GUM can help to make the concepts of the linear propagation of uncertainties more transparent to humans and machines. However, a lot more easy-to-understand and hands-on training courses will be needed.

Software enabling an easy use of the D-SI and the dependency-based treatment of measurement uncertainties based on GTC and UncLib is another important component supporting users. Seeger et al. are developing for example the 'dccQuantities' open-source software library for Python, which will allow the automated application of relevant mathematical operations [27].

Finally, it is left open, if an additional model for the representation of uncertainties based on Monte-Carlo simulation data will be introduced. Storing lists of full samples is already possible in applications such as DCCs which are based on the D-SI. However, users are hesitating to store the potentially large amounts of data. Investigation and exchange with users will continue to discuss, if alternative approaches such as the approximation of PDFs and summary models (e.g., histograms) of Monte-Carlo data may be adopted in the future.

VI. SUMMARY AND CONCLUSION

The presented updates of general concepts such as types of quantities and semantics, the introduction of PDFs and modelling of dependencies lead to a substantial advancement of the capabilities of the D-SI metadata model for exchanging values of quantities in digital applications. By these updates, the development of core models in the D-SI are completed. Future developments will provide maintenance of the core-models and extensions such as introducing additional representation systems for kinds of quantities or advancing the services helping to interoperate units, quantities, PDFs, and semantics, as well as implementations in various formats such as XML, JSON-LD, and ontologies.

VII. ACKNOWLEDGEMENTS

We like to sincerely thank Michael Wollensack, Blair Hall, Hans Koch, and Giacomo Lanza for their indispensable and inspiring ideas and work which is pushing the D-SI development. Furthermore, we like to acknowledge the efforts in digitalization and harmonization undertaken by calibration laboratories worldwide, the BIPM, the CIPM, and within EURAMET which provide valuable guidance for the D-SI.

REFERENCES

[1] General Conference on Weights and Measures, 'Resolution 2 of the 27th CGPM (2022), On the

global digital transformation and the International System of Units'. BIPM. [Online]. Available: https://doi.org/10.59161/CGPM2022RES2E
[2] BIPM, 'The International System of Units (SI), 9th edition'. 2019. [Online]. Available: https://www.bipm.org/documents/20126/41483022/SI-Brochure-9.pdf
[3] M. D. Wilkinson et al., 'The FAIR Guiding Principles for scientific data management and stewardship', Sci Data, vol. 3, p. 160018, 2016, doi: 10.1038/sdata.2016.18.
[4] BIPM, 'SI Digital Framework, Digital references for FAIR measurement data'. [Online]. Available: https://si-digital-framework.org/.
[5] BIPM, 'SI Reference Point'. [Online]. Available: https://si-digital-framework.org/SI.
[6] CIPM, 'Forum on Metrology and Digitalization (FORUM-MD)'. [Online]. Available: https://www.bipm.org/en/committees/fo/forum-md
[7] [1] Daniel Hutzschenreuter et al., 'SmartCom Digital System of Units (D-SI) Guide for the use of the metadata-format used in metrology for the easy-to-use, safe, harmonised and unambiguous digital transfer of metrological data - Second Edition', Zenodo, EMPIR, 2020, doi: 10.5281/zenodo.3816686.
[8] [1] S. G. Hackel et al., 'The Digital Calibration Certificate (DCC) for an End-to-End Digital Quality Infrastructure for Industry 4.0', Sci., vol. 5(1), p. 11, 2023, doi: 10.3390/sci5010011.
[9] JCGM, 'JCGM 200:2012 - International vocabulary of metrology – Basic and general concepts and associated terms (VIM) - 3rd edition'. BIPM, 2008. [Online]. Available: https://www.bipm.org/documents/20126/2071204/JCGM_200_2012.pdf.
[10] JCGM, 'JCGM 100:2008(E), Evaluation of measurement data - Guide to the expression of uncertainty in measurement (2008)'. BIPM, 2008. [Online]. Available: https://doi.org/10.59161/JCGM100-2008E.
[11] PTB, 'D-SI XML GitLab repository'. [Online]. Available: https://gitlab1.ptb.de/d-ptb/d-si/xsd-d-si.
[12] M. J. Kuster, 'Toward a metrology-information layer for digital systems', ACTA IMEKO, vol. 12, no. 1(18), pp. 1–4, 2023.
[13] QUDT.org, 'Website of Quantities Units Dimensions and Types (QUDT) Ontology'. [Online]. Available: https://qudt.org/
[14] ISO/IEC JTC 1, ISO/IEC 2382:2015 Information technology -Vocabulary, 2015.
[15] DKD, DKD-L 13-1 Rounding of Results and Measurement uncertainties in Calibration Certificates, Nov. 2021. doi: 10.7795/550.20220114.
[16] ISO/TC 12, ISO 80000-1:2022 Quantities and units - Part1: General, 2022.
[17] M. Jordan, 'D-SI Ontology GitLab repository'. [Online]. Available: https://gitlab1.ptb.de/d-ptb/d-si/ontology-d-si.
[18] D. Hutzschenreuter, 'D-SI JSON-LD context'. [Online]. Available: https://www.ptb.de/si/SI_Format.jsonld.
[19] M. Wollensack, METAS UncLib. (2025). [Online]. Available: https://www.metas.ch/metas/en/home/fabe/hochfrequenz/unclib.html.
[20] Wolfram, 'The Ultimate Univariate Probability Distribution Explorer'. [Online]. Available: https://blog.wolfram.com/2013/02/01/the-ultimate-univariate-probability-distribution-explorer
[21] 'Uncertainty Markup Language: UncertML'. [Online]. Available: https://web.archive.org/web/20161029215725/http://www.uncertml.org/.
[22] 'ProbOnto - the Ontology and Knowledge Base of Probability Distributions'. [Online]. Available: https://sites.google.com/site/probonto.
[23] [1] Wikipedia, 'REST (Representational State Transfer)'. [Online]. Available: https://en.wikipedia.org/wiki/REST.
[24] B. Hall and J. Borbely, GUM Tree Calculator. (2024). MSL. [Online]. Available: https://gtc.readthedocs.io/.
[25] IEC/SC 3D, 'IEC 61360-4 - IEC/SC 3D - Common Data Dictionary'. [Online]. Available: https://cdd.iec.ch/.
[26] D. Hutzschenreuter, 'D-SI Metadata Model – example data and software scripts for handling uncertainty dependencies'. Zenodo, 2025. doi: 10.5281/zenodo.15276391.
[27] B. Seeger, E. V. Stehr, and T. Bruns, 'Python-based development libraries and web applications to process and store data using a JSON representation XML-DCC', Measurement: Sensors, p. 101480, 2025, doi: 10.1016/j.measen.2024.101480.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Approximative Bayesian approach for uncertainty evaluation in machine learning-based hardness measurement

Junnosuke Takai[1], Yukimi Tanaka[1], Masahiro Yoshioka[1], Katsuhiro Shirono[1]

[1] Research Institute for Engineering Measurement, National Metrology Institute of Japan, National Institute of Advanced Industrial Science and Technology (AIST), AIST Tsukuba Central 3, Umezono 1-1-1, Tsukuba, Ibaraki 305-8563, Japan, Email:takai.jun@aist.go.jp

***Abstract* – In recent years, the application of machine learning to the field of metrology has been increasingly explored. In the evaluation of measurement uncertainty using machine learning, it is generally considered necessary to evaluate it through combining two types of uncertainties: aleatoric uncertainty, which arises from randomness, and epistemic uncertainty, which arises from systematic factors. However, the methods for evaluating these uncertainties have not yet been established. In this study, we evaluate measurement uncertainty in the Vickers hardness measurement using a Convolutional Neural Network. For this evaluation, we employ Monte Carlo Batch Normalization as an approximation of a Bayesian Neural Network to evaluate epistemic uncertainty. As a result, it was found that a reasonable evaluation is possible for materials similar to those in the training data.**

I. INTRODUCTION

Machine learning (ML) is a technology that enables the construction of models capable of representing functions based on data, even when the analytical form of the function is unknown or highly complex. In recent years, it has been applied across a wide range of fields. In particular, deep learning, a class of machine learning methods that allows for the construction of highly flexible and complex models, has seen widespread adoption. In the field of metrology as well, a lot of studies employing deep learning techniques have been reported.

For example, one of the authors, Tanaka, together with co-authors, reported a study on the Vickers hardness measurement that employed a Convolutional Neural Network (CNN), a type of deep learning model [1]. The Vickers hardness is measured using the diagonal length of an indentation created when a pyramidal indenter is pressed into a material. When this diagonal length is measured manually from an image of the indentation, there are some challenges including time consuming tasks and human depending differences in the results. Furthermore, in cases where the diagonal length is determined automatically from image features such as gradients, accurate measurement may be difficult due to factors such as rough surfaces or cracks. Tanaka et al. demonstrated that the use of a CNN for diagonal length measurement enables high efficiency and high accuracy evaluation across a wide variety of materials.

However, the evaluation of uncertainty when using machine learning has not yet been fully established. Thompson et al. (2021) stated that "Traditional uncertainty evaluation of the kind systematized by the GUM framework makes the assumption that the measurement model is known. Uncertainty evaluation in ML deviates from this paradigm in one key respect". One of the essential requirements for ensuring metrological traceability of measurements is the evaluation of measurement uncertainty. Establishing metrological traceability is significant in conformity assessment and quality control. In the context of the hardness measurements, the evaluation of measurement uncertainty is also required to ensure traceability.

Regarding measurement uncertainty in the use of machine learning, Hüllermeier et al. (2021) proposed that the uncertainty in machine learning outputs can be decomposed into two types: aleatoric uncertainty and epistemic uncertainty. They suggest that by evaluating these two components, it becomes possible to evaluate measurement uncertainty even when machine learning is employed [3]. Aleatoric uncertainty reflects the randomness in the data, whereas epistemic uncertainty represents the uncertainty arising from the imperfection of the model. Several methods have been proposed to estimate these two types of uncertainties [2–4].

In this study, we report on the evaluation of uncertainty in the measurement of the Vickers hardness using a CNN. For the estimation of epistemic uncertainty, we apply Monte Carlo Batch Normalization (MCBN), one of the methods proposed in previous research [4]. The structure of this paper is as follows. In Section 2, we provide an overview of the two types of uncertainties in machine learning and describe the MCBN. Section 3 presents our experiments on hardness measurement and uncertainty

evaluation using the MCBN, along with the corresponding results. Section 4 discusses and interprets the results presented in Section 3. Finally, Section 5 provides a summary of this study.

II. UNCERTAINTY EVALUATION THROUGH MONTE CARLO BATCH NORMALIZATION

A. Aleatoric uncertainty and epistemic uncertainty

In this study, we consider the problem of supervised learning, where a measurement model is represented (trained) using known measurement data (training data), and predictions are made for an unknown estimate via the regression. It is assumed that the training data are independently and identically distributed (i.i.d.), and generated from an unknown probability density function P. Associating the measurement data with the probability density function P implies that, even if complete knowledge of P were available, some variation in the output estimate would remain. This irreducible uncertainty is referred to as aleatoric uncertainty. On the other hand, in practice, it is not possible to determine the probability density function P perfectly. The uncertainty arising from this incomplete knowledge is referred to as the epistemic uncertainty. According to Hüllermeier et al. [3], the epistemic uncertainty can be further decomposed into two components: model uncertainty, which reflects whether the machine learning model is capable of accurately representing the measurement model, and approximation uncertainty, which results from the model's limited approximation capability due to the finiteness of the data.

However, in the case of flexible models such as ones developed by deep learning, it is generally assumed that the measurement model could be represented with high accuracy, provided that enough data were available. Therefore, in this study, where a deep learning model is employed, only approximation uncertainty is treated as epistemic uncertainty.

For uncertainty evaluation in measurement using machine learning, it is necessary to evaluate both aleatoric and epistemic uncertainties.

B. Monte Carlo Batch Normalization

Prior to discussing the evaluation of uncertainty, we provide an overview of Batch Normalization (BN), a technique widely used in deep learning to improve training stability and accelerate convergence. In deep neural networks, as illustrated in Fig. 1, input values are propagated through multiple layers, where each layer applies nonlinear transformations at each node, eventually producing a predicted output. While deeper networks tend to have greater expressive capacity, they are also more prone to instability and inefficiency during training.

One possible approach to addressing these challenges is BN. In typical deep learning frameworks, training data are divided into small subsets called mini-batches, and model parameters are updated based on each

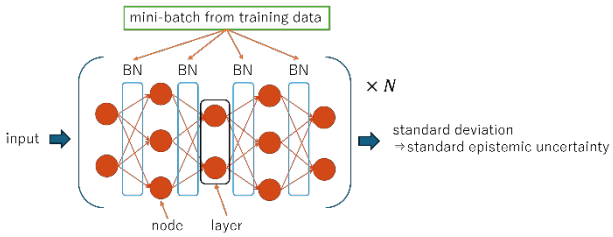


Fig. 1 Schematic diagram of the MCBN. N is the number of Monte Carlo sampling (sampling number of the mini-batches).

mini-batch. BN normalizes the output of each node in a layer to have zero mean and unity variance based on the data of each mini-batch. During the prediction, the normalization is performed using the parameters optimized during training (see Fig. 1).

We then introduce Monte Carlo Batch Normalization (MCBN), proposed by Teye et al. [4], as a method for evaluating epistemic uncertainty. Several fundamentally different approaches have been proposed for evaluating epistemic uncertainty in the deep learning [2]. One such approach is the Bayesian Neural Network (BNN), which evaluates uncertainty by performing the Bayesian inference on the parameters of the underlying probability distribution P. In BNN, a predictive posterior distribution of the output estimates represents the epistemic uncertainty. However, due to the extremely large number of parameters in deep learning models, obtaining the predictive posterior distribution via direct Bayesian inference, such as that using Markov Chain Monte Carlo (MCMC) methods, is often computationally impractical.

To address this issue, Teye et al. proposed the MCBN as an approximate method for deriving the predictive posterior distribution. As mentioned previously, standard BN uses fixed normalization parameters during the prediction. In contrast, MCBN introduces stochasticity by randomly sampling mini-batches from the training data even during the prediction phase and applying the BN based on the sampled mini-batch. By repeating this sampling, multiple prediction estimates can be obtained for the same input data. According to Teye et al., the resulting distribution of the prediction estimates can be interpreted as an approximation of the predictive posterior in the BNN. In this study, we evaluate epistemic uncertainty by analyzing the distribution of prediction estimates generated through MCBN.

III. EXPERIMENTAL AND ANALYTICAL METHODS

This section provides an overview of the uncertainty analysis in the Vickers hardness measurement using CNNs. For details regarding the Vickers hardness measurement procedure, see [1].

The Vickers hardness is determined by measuring the lengths of the diagonals of the indentation formed when an indenter is pressed into a material. In [1], two types of CNNs are used to measure the diagonal lengths from images of the indentations. Each indentation image is captured at one of the following magnification levels:

$m = 10,\ 20,\ 40,\ 50,\ 100$. First, a coarse estimation of the four corner positions of the indentation is obtained using the first CNN. Then, around each estimated corner, a magnified image is extracted, so that each magnified image contains only one visible corner.

For each magnified image, a second CNN is used to predict the corner position more precisely, from which the diagonal length is computed. As the CNN outputs are in pixel units, the results are converted into physical lengths based on the corresponding magnification factor m.

Since the output of the first CNN does not significantly influence the final measurement, only the uncertainty associated with the second CNN is evaluated. Following [1], this CNN is referred to as the CNN-CP (CNN for the Corner Positions). The CNN-CP takes a 100×100-pixel image as input and consists of a 10-layer neural network. The activation function used is ReLU, and the BN is applied after each layer. The network is implemented using the PyTorch library, trained and optimized using the Adam optimizer with a smooth L1 loss function [5]. A mini-batch of 64 is used. Training is performed for the first 100 epochs with a learning rate of 10^{-4}, followed by an additional 200 epochs with a learning rate of 10^{-6}.

The training dataset consists of indentation images obtained from reference blocks with nominal Vickers hardness values of 200 HV, 300 HV, 400 HV, 500 HV, 600 HV, 700 HV, 800 HV, and 900 HV, as well as manually labeled indentation images of titanium dioxide (TiO_2). Further details are available in [1].

The standard epistemic uncertainty, u_{eps}, is estimated using the MCBN. As the measured value is defined as the average of the two diagonals, and each diagonal requires two corner predictions, a total of four CNN-CP outputs is needed for a prediction estimate. To account for possible correlations among the four outputs, the average diagonal length is calculated in each Monte Carlo sample. The standard deviation of these averages is then used to determine u_{eps} as follows:

$$u_{\mathrm{eps}}(x) = \sqrt{\frac{\sum_i \left(f_{\mathrm{MC},i}(x) - \bar{f}_{\mathrm{MC}}(x)\right)^2}{N-1}}, \qquad (1)$$

where x is the input image, $f_{\mathrm{MC},i}(x)$ is the average diagonal length predicted in the i-th Monte Carlo sample, and $\bar{f}_{\mathrm{MC}}(x)$ is the mean of all $N = 5000$ samples. Note that the standard epistemic uncertainty depends on the input.

The standard aleatoric uncertainty, u_{ale}, is derived from the training dataset. Since outliers may be present in the training data, a robust estimation method is employed. Specifically, the median absolute deviation (MAD) between the predicted estimates and the measurement value is calculated. Based on the properties of the normal distribution, the MAD is scaled by 1.4826 to obtain the standard uncertainty:

$$u_{\mathrm{ale}}(x) = c_m \times 1.4826 \times \mathrm{Med}\left(\left\{\left|g\left(x_{\mathrm{train},i}\right) - t_{\mathrm{train},i}\right|\right\}_{i \in T}\right) \qquad (2)$$

where $g\left(x_{\mathrm{train},i}\right)$ is the CNN output for the i-th training input, $t_{\mathrm{train},i}$ is the manually labeled pixel coordinate of the corner position, T is the set of indices for the training data, $\{a_i\}_{i \in A}$ is the set elements of which are variables a_i associated with i in the set A, $\mathrm{Med}(A)$ is the median of elements in the set A, and c_m is a sensitivity coefficient with respect to diagonal length determined solely by the magnification factor m. Importantly, c_m is independent of the content of the input image. Consequently, in this study, the aleatoric uncertainty is treated as a function of magnification only and does not vary with other input features.

The standard uncertainty in the diagonal length prediction, denoted by $u_{\mathrm{CNN}}(x)$, is calculated as the square root of the sum of the squares of the standard epistemic and aleatoric uncertainties:

$$u_{\mathrm{CNN}}(x) = \sqrt{u_{\mathrm{eps}}^2(x) + u_{\mathrm{ale}}^2(x)} \qquad (3)$$

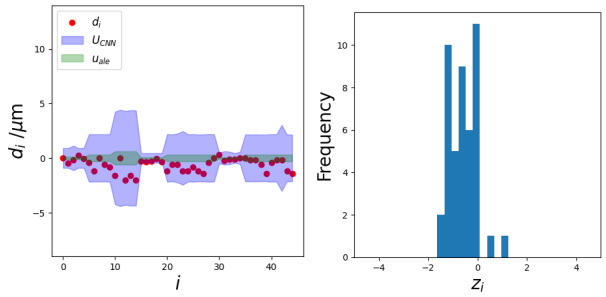
For simplicity, we do not handle other uncertainty sources associated with hardness measurement, as they are beyond the scope of this study. The expanded uncertainty is computed using a coverage factor $k = 1.96$, giving: $U_{\mathrm{CNN}}(x) = k u_{\mathrm{CNN}}(x)$

IV. RESULTS AND DISCUSSION

Two types of test data are evaluated: one measured on materials that are included in the training dataset, and another measured on materials not included in the training data. The results measured on the same materials as those in the training data are shown in Fig. 2. The data were obtained using images of reference blocks with nominal values of 200 HV, 600 HV, and 900 HV. Fig. 2(a) shows the relationship between the difference d_i between the measured diagonal lengths $y_{\mathrm{test},i}$ and predicted diagonal lengths $f\left(x_{\mathrm{test},i}\right)$ and the evaluated expanded uncertainty $U_{\mathrm{CNN}}(x_{\mathrm{test},i})$. The points represent the differences $d_i = f\left(x_{\mathrm{test},i}\right) - y_{\mathrm{test},i}$, where the blue shaded area represents the expanded uncertainty $U_{\mathrm{CNN}}(x_{\mathrm{test},i})$, and the green shaded area represents the standard aleatoric uncertainty $u_{\mathrm{ale}}(x)$. In comparison with the sensitivity coefficient c_m depending on magnification, the smallest standard epistemic uncertainty was $u_{\mathrm{eps}}\left(x_{\mathrm{test},11}\right)/c_m = 2.66$, and the largest was $u_{\mathrm{eps}}\left(x_{\mathrm{test},32}\right)/c_m = 3.88$.

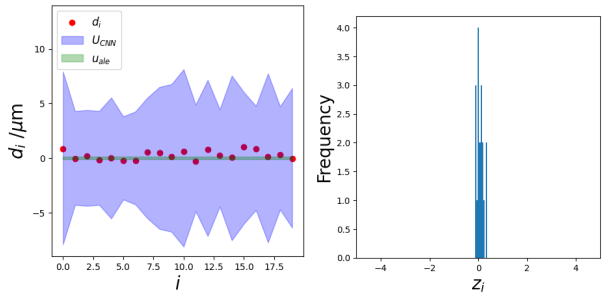
Since the standard aleatoric uncertainty was $u_{\mathrm{ale}}\left(x_{\mathrm{test},11}\right)/c_m = u_{\mathrm{ale}}\left(x_{\mathrm{test},32}\right)/c_m = 0.53$, the difference in standard epistemic uncertainty is significantly larger than the standard aleatoric uncertainty. This implies that, if a common epistemic uncertainty were quantified from the test data, this difference would not be reflected in the quantified uncertainty, and the uncertainty would be over- or under-evaluated. Therefore, it is necessary to evaluate the uncertainty according to the characteristics of each image, as done in this study.

To verify the evaluated uncertainty, Fig. 2(b) shows a histogram of the ratio between the difference d_i and the evaluated standard uncertainty $u_{\mathrm{CNN}}(x_{test,i})$, defined as $z_i = d_i / u_{\mathrm{CNN}}(x_{\mathrm{test},i})$.



(a) (b)

Fig. 2 Measurement of 200 HV, 600 HV and 900 HV reference blocks (a) relationship between d_i *and* $U_{CNN}(x_{test,i})$ *(b) histogram of* $z_i = d_i/u_{CNN}(x_{test,i})$



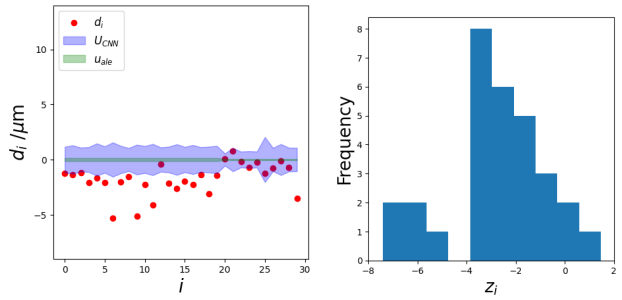
(a) (b)

Fig. 3 Measurement of CaF_2 *(a) relationship between* d_i *and* $U_{CNN}(x_{test,i})$ *(b) histogram of* $z_i = d_i/u_{CNN}(x_{test,i})$

For the case measured on the same material as the training data, all differences d_i fall within the expanded uncertainty $U_{\mathrm{CNN}}(x_{\mathrm{test},i})$. Furthermore, as seen in the histogram, the uncertainty is not excessively large. Therefore, it can be said that the estimated uncertainty is reasonable. It is considered that MCBN was applicable in this case.

Next, Figs. 3 and 4 show the results for two materials different from those in the training data: CaF_2 and Si_3N_4, respectively. The structure of Figs. 3 and 4 is the same as that of Fig. 2. As shown in Fig. 3(b), the maximum absolute value of z_i is less than 0.4. Therefore, in the case of CaF_2, it is presumed that either the aleatoric uncertainty, the epistemic uncertainty, or both were over-evaluated. On the other hand, in the case of Si_3N_4, the results indicate that the uncertainty was under-evaluated. There is a tendency for the predicted average diagonal length $f(x_{\mathrm{test},i})$ to show smaller values. Such bias should be evaluated as epistemic uncertainty.

Thus, the evaluation method in this study may not provide valid uncertainty evaluation for materials different from those used in training. Two possible reasons are considered. One is that the mini-batches used in the MCBN for evaluating epistemic uncertainty consist of training data (standard reference blocks and TiO_2). Since the input for materials different from the training data comes from a data region different from that of the training data, it is considered that the BNN using the MCBN did not perform properly. The second reason is that the aleatoric



(a) (b)

Fig. 4 Measurement of Si3N4 (a) relationship between d_i *and* $U_{CNN}(x_{test,i})$ *(b) histogram of* $z_i = d_i/u_{CNN}(x_{test,i})$

uncertainty is evaluated using the training data. Since the variation in measurement values differs depending on the material, aleatoric uncertainty should be evaluated for each material.

V. CONCLUSION

In this study, uncertainty evaluations in the Vickers hardness measurement using a convolutional neural network were conducted. To evaluate the uncertainty of the CNN outputs, it is necessary to evaluate two types of uncertainties: aleatoric uncertainty and epistemic uncertainty. In the present study, epistemic uncertainty was evaluated using Monte Carlo batch normalization, and aleatoric uncertainty was estimated from the difference between the training data and the predicted values.

For data obtained from the same materials as those used in the training dataset, it was confirmed that the proposed evaluation method allows for a valid assessment of uncertainty. The quantified standard uncertainty varied significantly for each image, indicating the necessity of evaluating measurement uncertainty not comprehensively, but individually for each image. On the other hand, for data obtained from materials different from those in the training dataset, a valid uncertainty assessment remains challenging, but it is an important future task.

REFERENCES

[1] Tanaka, Yukimi, Yutaka Seino, and Koichiro Hattori. "Automated Vickers hardness measurement using convolutional neural networks." The International Journal of Advanced Manufacturing Technology 109.5 (2020): 1345-1355.

[2] Thompson, Andrew, et al. "Uncertainty evaluation for machine learning." (2021).

[3] Hüllermeier, Eyke, and Willem Waegeman. "Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods." Machine learning 110.3 (2021): 457-506.

[4] Teye, Mattias, Hossein Azizpour, and Kevin Smith. "Bayesian uncertainty estimation for batch normalized deep networks." International conference on machine learning. PMLR, 2018.

[5] The Linux Foundation, PyTorch, https://pytorch.org/ (accessed on 13th Mar. 2025).

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

A Torque Sensor with a Unified Coupling Interface for Wireless Power and Data Transmission

Yuehua Chen [1], Jiahui Zheng [2], Xianglong Ma [3], Gang Zhang [4]

[1] *Ningbo University, China, chenyuehua@nbu.edu.cn*
[2] *Ningbo University, China, 2111087079@nbu.edu.cn*
[3] *None, China, theseveneira@outlook.com*
[4] *Ningbo University, China, zhanggang@nbu.edu.cn*

***Abstract* – This paper presents a single-coil technical solution for a strain-type torque sensor based on magnetic coupling resonance, enabling simultaneous wireless power and data transmission. The power and data sharing a common set of coupling coils, the system leverages frequency-division multiplexing (FDM) and carrier modulation technologies to independently control distinct operating frequencies, directly modulate data onto the power carrier via amplitude-shift keying (ASK), and perform data pickup and demodulation. A prototype torque sensor was developed, achieving an output power of 5 W and a data transmission rate of 20 Kb/s.**

***Keywords:* wireless power and data transfer, torque sensor, frequency-division multiplexing**

I. INTRODUCTION

In high-precision measurements, strain torque sensors are the most mature and reliable. The torque sensor is installed on the rotating shaft system, and torque measurement is relatively difficult. The torque sensor needs to be divided into two parts, the rotor and the stator, to solve the matching problem between rotating and stationary components. Due to its two independent parts, there are problems with power supply and communication between the two parts. Currently, the mainstream technical solutions for strain sensors are: conductive slip ring type for contact, and wireless power supply plus wireless communication or battery power supply plus wireless communication for non-contact. However, the above technical solutions cannot meet the requirements of high reliability and easy maintenance. Therefore, how to design a strain torque sensor that is easy to install, easy to use, and has high reliability is a technical problem that urgently needs to be solved.

In recent years, the proposal of the magnetic coupling resonance concept [1-2] has achieved medium-distance wireless power transfer (WPT). Subsequently, after ten years of development based on this, magnetic-coupled wireless power transfer (MC - WPT) has gradually matured, and a wireless power and signal synchronous transmission system has also been developed based on this technical foundation.

The basic principle of wireless power and signal synchronous transmission is to use wireless power transfer technology and, on this basis, add a signal transmission topological structure to achieve the synchronous transmission of wireless power and signals. Compared with traditional wired power transmission, wireless power transfer technology has advantages such as non-contact, flexibility, safety, and reliability, and has been widely applied in various fields, such as electric vehicle charging, medical equipment, lighting, and electronic products.

Currently, wireless power transfer (WPT) technology can be transmitted through various media, including electric and magnetic fields. According to different transmission mechanisms, it can be divided into three methods: inductive, resonant, and radiative. At present, the mainstream wireless power transfer (WPT) technologies mainly include electromagnetic wave radiative, electromagnetic inductive, and magnetic-coupled resonant types.

A. Electromagnetic Wave Radiative Wireless Power Transfer Technology

The electromagnetic wave radiative wireless power transfer technology adopts the principle of microwave power transmission, that is, it converts electrical energy into microwaves through a highly directive transmitting antenna, and then uses a receiving antenna to collect the microwaves and convert them into direct current for subsequent circuits. This technology has the ability of long-distance transmission. For example, in the 20th century, the United States proposed the Space Solar Power Station (SSPS) plan [3]. By deploying satellites in outer space, the solar energy in outer space is converted into electrical energy and transmitted to the ground through microwaves.

B. Electromagnetic Inductive Wireless Power Transfer Technology

The electromagnetic inductive wireless power transfer technology is an electric energy transmission method based on the electromagnetic induction phenomenon. This technology is similar to the principle of a transformer and has high requirements for the coupling coefficient between coils. When there is a position offset, it will have a great impact on the power and efficiency of wireless power transfer, and the transmission distance is relatively short, only a few centimeters, so its application is limited.

C. Magnetic-Coupled Resonant Wireless Power Transfer Technology

The magnetic-coupled resonant wireless power transfer technology uses a high-frequency power supply to form a resonant circuit through the coil inductance and capacitance, generating a strong high-frequency alternating current, thus generating a high-frequency electromagnetic field through the coil. The receiving coil is also in a resonant state, receiving the electromagnetic signal from a distance and converting it into a high-frequency alternating current. Through the mutual coupling of the two coils, wireless power transfer is achieved. This technology has advantages such as high transmission power, high transmission efficiency, and strong anti-interference ability.

This technology was first proposed in a paper published in the journal "Science" by the Massachusetts Institute of Technology in the United States [1-2]. In 2007, it demonstrated its results to the outside world: at a frequency of 9.9 MHz, a 60W light bulb was successfully lit with an efficiency of 40% at a distance of 2m. The release of this research result has elevated the practical application of wireless power transfer to a new level. Subsequently, scholars from various countries have invested in research in this field.

Zhang [4] first applied this technology to medical equipment to charge implantable devices, verifying its feasibility in the medical field. Takehiro [5] proposed applying this technology to the field of wireless charging for electric vehicles and explored the precise frequency and efficiency required to achieve electromagnetic coupling based on antenna theory, circuit theory, experimental observations, and electromagnetic calculations [6]. Subsequently, Cannon B.L [7] explored the possibility of supplying power to multiple receivers, expanding the application scenarios of magnetic-coupled wireless power, enabling power supply to multiple receivers, as shown in Figure 1.2.

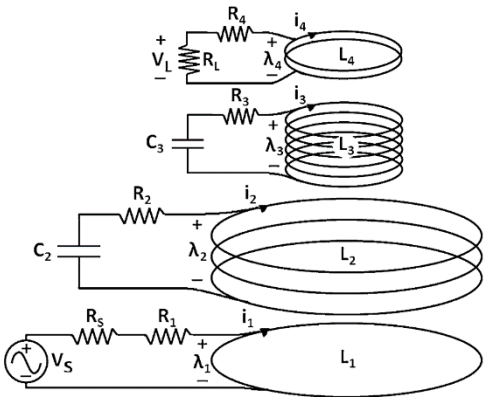


Fig. 1. Schematic diagram of one to many magnetic coupled radio energy transmission

The Korea Advanced Institute of Science and Technology (KAIST) has carried out a large amount of research in the field of electric vehicles, mainly focusing on detailed research in aspects such as dynamic power supply [8], electromagnetic relay [9], and coil structure [10].

Based on the above research foundation, this paper proposes a technical solution for a single - coil strain torque sensor based on magnetic - coupled resonant wireless power and data synchronous transmission. This solution solves the two functions of power transmission and signal transmission of the torque sensor simultaneously through a single system. The advantage of this technical solution is that it greatly improves the convenience of use and maintenance of the torque sensor, thus effectively enhancing the competitiveness of related products.

II. CONTRIBUTION

This paper designs a wireless power and data synchronous transmission system according to the measurement dimensions of the torque sensor. The system block diagram is shown in Fig.2. Based on this, a torque sensor experimental prototype is developed, The influence of rotational speed and metal media on the system was explored, and the reliability of the system is verified through experiments.

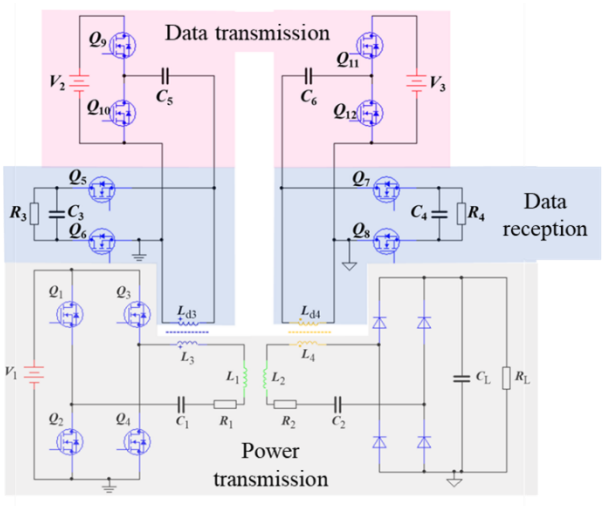


Fig. 2. wireless power and data synchronous transmission circuit block diagram

III. RESULTS

Based on the above mentioned system, this paper subsequently manufactures the 2000 Nm torque sensor, including the mechanical structure processing of the rotor and stator, the pasting of strain gauges, the production and debugging of the circuit board, as shown in Fig.3. Subsequently, the wireless power and data synchronous transmission system is tested. The waveforms are shown in Fig.4. Point 1 is the waveform at the data loading end; Point 2 is the waveform at the data receiving end; Point 3 is the waveform after envelope filtering of the data; Point 4 is the data output waveform. It can be seen from Figure 4 that the demodulation effect of the data transmission waveform signal is good, and the bit-error rate is low.



Fig. 4. Physical diagram of the 2000 Nm torque sensor

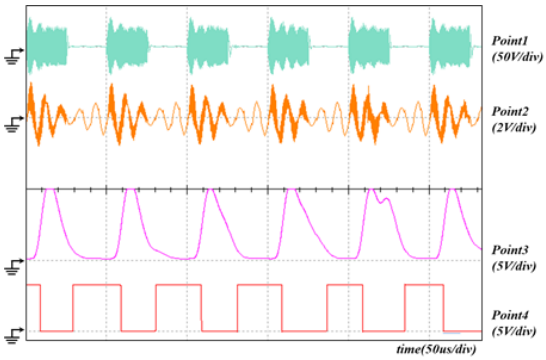


Fig. 4. the wireless power and data synchronous transmission system waveform diagram

IV. CONCLUSIONS

This paper proposes a technical solution for a single - coil strain - type torque sensor based on magnetic - coupled resonant wireless power and data synchronous transmission. Based on a 2000 Nm torque sensor, a system experimental prototype is designed. The influence of rotational speed and metal media on the system is explored, and the reliability of the system is verified through experiments.

REFERENCES

[1] Karalis A, Joannopoulos J D, Soljacic M, “Efficient wireless non-radiative mid-range energy transfer”, Ann. Phys., vol.323, No.1, January 2008, pp.34-48.

[2] Andre Kurs, Aristeidis Karalis, Robert Moffatt, et al, “Wireless power transfer via strongly coupled magnetic resonances”, Science, vol.317, No.6, July 2007, pp.83-86.

[3] Glaser P E, “Power from the Sun: Its Future”, Science, vol.162, No.3856, November 1968, pp.857-861.

[4] Zhang F, Liu X, Hackworth S A, et al, “In vitro and in vivo studies on wireless powering of medical sensors and implantable devices”, Life Science Systems & Applications Workshop, 2009, vol.1, pp.84-87.

[5] Imura T, Okabe H, Hori Y, “Basic experimental study on helical antennas of wireless power transfer for Electric Vehicles by using magnetic resonant couplings”, IEEE Vehicle Power and Propulsion Conference, 2009, vol.1, pp.936-940.

[6] Imura T, Okabe H, Uchida T, et al, “Study of Magnetic and Electric Coupling for Contactless Power Transfer Using Equivalent Circuits”, IEEJ Trans. Ind. Appl., vol.130, No.1, 2010, pp.84-92.

[7] Cannon B L, Hoburg J F, Stancil D D, et al, “Magnetic Resonant Coupling As a Potential Means for Wireless Power Transfer to Multiple Small

Receivers", IEEE Trans. Power Electron., vol.24, No.7, July 2009, pp.1819-1825.
[8] Suh I S, Kim J, "Electric vehicle on-road dynamic charging system with wireless power transfer technology", International Electric Machines & Drives Conference, 2013, vol.1, pp.234-240.
[9] Ahn D, Hong S, "A Study on Magnetic Field Repeater in Wireless Power Transfer", IEEE Trans. Ind. Electron., vol.60, No.1, January 2013, pp.360-371.
[10] Lee W S, Son W I, Oh K S, et al, "Contactless Energy Transfer Systems Using Antiparallel Resonant Loops", IEEE Trans. Ind. Electron., vol.60, No.1, January 2013, pp.350-359.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - M4DConf 2025
Benevento, Italy, September 3-5, 2025

Innovative Approach in Industry 5.0: Use of Multiple Sensors in a Real Production Process to Enhance Workers' Well-Being

Matteo Zendri[1], Francesco Pilati[2], Francesca Calabrese[3], Qingwei Cai[4]

[1]*Research Fellowship, Department of Industrial Engineering, University of Trento, matteo.zendri@unitn.it*
[2]*Associate Professor, Department of Industrial Engineering, University of Trento, francesco.pilati@unitn.it*
[3]*Assistant Professor, Department of Industrial Engineering, University of Trento, francesca.calabrese@unitn.it*
[4]*PhD Student, Department of Industrial Engineering, University of Trento, qingwei.cai@unitn.it*

Abstract – This paper presents a novel application for Industry 5.0: the first global implementation of multiple wearable sensors in a real industrial environment.

With minimal investment in IoT infrastructure, the proposed system enables the collection of useful data on operator activities, including self-reported fatigue after specific tasks.

These data are fundamental for analyzing the relationship between assembly tasks and self-reported fatigue by operators, allowing optimization of product design and assembly processes.

By adopting this approach, industries can enhance worker well-being and decrease the incidence of occupational diseases. The final goal of the project is to develop and validate an algorithm capable of predicting, in real-time, the perceived fatigue experienced by operators upon completing specific tasks.

I. INTRODUCTION

Industry 5.0 is a new concept that follows the evolution of Industry 4.0. This previous concept is centered on efficiency optimization and improvement of the production processes using IoT, data analysis and Artificial Intelligence. With Industry 4.0 there are innovative concepts like collaboration between man and machine, adaptability and continuous training of the operators. With Industry 5.0 the focus is moved from the manufacturing processes to human centricity and well-being. This is necessary due to the particular European situation and especially in Italy where the population is becoming older without an increase in birth in new generations.

A. Birth rate

In the following figure, the official Eurostat report explaining the birth trend in European countries is shown [1].

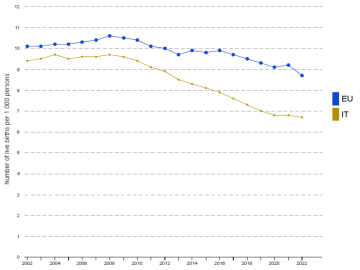


Figure 1: *Birth rate trends (2002-2022).*

B. Population age

In the following figures, the official Eurostat report shows the ageing of the European population [1].

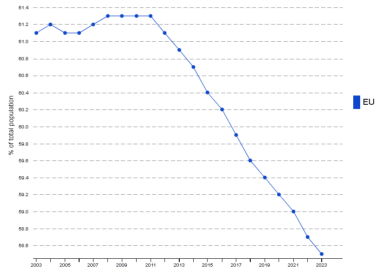


Figure 2: *European population between 20-64 years (2002-2022).*

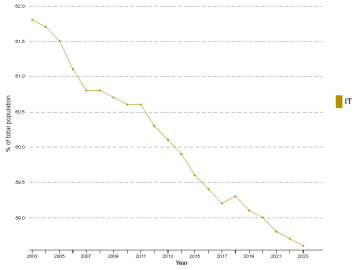


Figure 3: *Italian population between 20-64 years (2002-2022).*

C. Occupational diseases

In the following graph it is possible to see a continuous increase of occupational diseases with the exception of the period 2020-2023, when the Covid-19 pandemic led to a temporary decrease in cases. Each disease impacts production and represents a high cost and profit loss for companies, as the operator is unable to continue the activity, thereby reducing productivity. Multiple factors may contribute to these conditions, such as repetitive movement at high velocity, handling of heavy components, or assembly tasks performed with poor posture. These causes are not always easy to detect, and assessing the severity of the impact on the operator can be difficult [2].

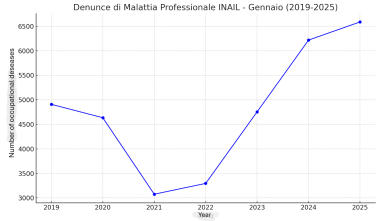


Figure 4: *Occupational diseases over the last 6 years.*

For these reasons, Industry 5.0 introduces a new human-centric approach, aiming to preserve operator health over time.

II. LITERATURE REVIEW

In comparison, the study by Bonakdar et al. (2025) [3] analyzed fatigue during manual handling tasks (MHT) using a methodology based on inertial measurement units (IMUs) to monitor joint angles and coordinative variability. The key findings revealed an increase in both mean and peak joint angles as fatigue intensified, except for the knee joint, where both decreased. Additionally, a decline in coordination between adjacent joints was observed, as indicated by reduced information transmission measured through mutual information theory. This study reports tests conducted in a controlled laboratory environment, tracking three distinct activities performed repeatedly over time: lifting, carrying, and moving. To collect data for these activities, eight participants volunteered for the experiment and were asked to wear six IMUs each. The accuracy achieved with this approach was 74%.

Another relevant paper is the study conducted by Munguia Tapia et al. [4], which focuses on the real-time recognition of physical activities using wireless accelerometers and a heart rate monitor. The goal was to identify 30 different activities during gym sessions in a controlled environment. Involving 21 participants, the dataset covered 16.6 hours of data.

Using a different approach, the authors attempted to identify activities based on individual subjects, achieving an accuracy of 94.6%. However, when subject information was excluded, the accuracy dropped to 80.6%.

An important result was the increase in accuracy (1–2%) when heart rate information was included.

The study conducted by Tryon and Trejos [5] aimed to develop models that classify task weight using EEG–EMG fusion. To achieve this goal, 32 healthy subjects (mean age 24.9 ± 5.4 years) performed elbow flexion–extension motions under different conditions of speed and weight. The experimental setup included EEG sensors and encoders, and an accuracy of 83.01 ± 6.04% was reported in classifying three weight levels.

By contrast, the present study introduces an innovative use of multiple sensors in a real industrial environment, without control over external perturbations or noise typical of real-world manufacturing conditions. The objective is to classify different tasks and predict the operator's fatigue level.

III. ARCHITECTURE DESCRIPTION

The most important factor for effective data analysis is the dataset quality. The initial phase of the project is focused on constructing a well-structured dataset.

The system architecture includes wearable sensors, a server, and a Raspberry Pi. After initial lab testing, the components were installed at the partner company.

During assembly tasks, operators wear the following devices:

- Three quaternions:
  - One on the back to measure posture variations
  - One on the dominant hand to track movements
  - One on the non-dominant hand to track movements
- One Bitalino with four channels:
  - ECG to measure heart rate
  - EMG to monitor muscle activity in the arm
  - EMG to monitor muscle activity in the trapezius
- UWB bracelet to track the operator's position
- RFID glove on the dominant hand to detect touched objects

The quaternions and Bitalino send data to the Raspberry Pi; RFID gloves send data to a smartphone, and the UWB bracelet sends data to the POE. The POE is connected via Ethernet to the server, while Raspberry Pi and smartphones communicate through Wi-Fi.

The system includes a simple interface that allows the operator to start and stop data collection. The same interface also displays sensor status in real time. A more technical interface supports remote monitoring of sensor functionality, detecting disconnections or temporary malfunctions.

IV. MONITORED PROCESS DESCRIPTION

The project involves collaboration with a company based in the province of Trento (Italy) that produces items using a C-shaped production layout. Product weight ranges from 30 to 180 kg, and heights vary between 500 and 1300 mm. Assembly involves tools, manual handling, and movement. Each shift typically includes 6 to 12 items, with model variation affecting the process.

The company shared production cycles for all products, breaking them down into 20 specific assembly tasks. This segmentation helps identify tasks that contribute most to perceived fatigue.

Training sessions guided operators to wear sensors correctly and manage data acquisition autonomously. During the learning period (7–10 days), support was provided to resolve any operational challenges.

The participant group included operators of varying age and handedness. Despite defined production cycles, task execution varies due to operator habits and tool use, creating process variability. The use of precedence diagramming is under consideration to reduce variability and complexity.

A minimum of 100 assembled items has been set to ensure a consistent dataset for analysis. Human error and potential disconnections during data gathering are also considered.

The power of this innovation lies in the use of multiple sensors. The post-processing rules automatically filter out noise and irrelevant information, increasing the algorithm's accuracy by focusing on meaningful signals.

VI. ALGORITHM DEVELOPMENT

The data preprocessing phase focuses on methods for effectively cleaning and integrating the collected information, as sensors recorded continuous hours of activity without interruption, including short pauses or routine operator duties. The first step involves distinguishing valuable data from irrelevant segments. It is also necessary to isolate data sections affected by brief sensor disconnections, as these periods prevent complete tracking of the operator's actions.

During the assembly process, variations may occur due to human errors or defective components that require substitutions or corrective actions. Data related to these exceptional cases are excluded from the analysis to reduce variability and maintain consistency.

Several machine learning models have been developed, including deep learning architectures, recurrent neural networks (RNNs), and long short-term memory (LSTM) networks. In addition, intelligent algorithms have been implemented to identify and prioritize value-added data within the dataset.

The analysis of sensor signals aims to identify the tasks most strongly correlated with operator-reported fatigue. From a managerial perspective, this information is highly valuable, as it enables tracking the evolution of fatigue across tasks and entire shifts. This, in turn, allows for the adaptation of production plans to the physical capabilities of individual operators, enabling recovery periods after more demanding activities.

Such insights support the optimization of assembly processes and product design with the objective of minimizing operator strain.

The overarching goal of the project is to implement an algorithm capable of predicting, in real time, the perceived fatigue level of operators following the execution of specific tasks.

VIII. CITATIONS AND REFERENCES

References

[1] Eurostat, "Demography report 2023," European Commission, 2023.

[2] INAIL, "Reports of occupational injuries and diseases - INAIL data for January 2025," INAIL Press Release, March 2025.

[3] A. Bonakdar et al., "Fatigue assessment in multi-activity manual handling tasks through joint angle monitoring with wearable sensors," Univ. of Alberta, 2025.

[4] E. Munguia Tapia et al., "Real-Time Recognition of Physical Activities and Their Intensities Using Wireless Accelerometers and a Heart Rate Monitor," MIT, Stanford, Boston Med. Center.

[5] J. Tryon and A. L. Trejos, "Classification of Task Weight During Dynamic Motion Using EEG–EMG Fusion," IEEE Trans. on Neural Systems and Rehabilitation Engineering, vol. 31, pp. 1071–1080, 2023.

[6] M. Calzavara, M. Faccio, and I. Granata, "Multi-objective task allocation for collaborative robot systems with an Industry 5.0 human-centered perspective," The International Journal of Advanced Manufacturing Technology, 2023. [cite: 3461, 3462, 3463, 3090, 3091, 3092]

[7] Z. Wang and J. Yan, "Multi-sensor fusion based industrial action recognition method under the environment of intelligent manufacturing," Journal of Manufacturing Systems, vol. 74, pp. 575-586, 2024. [cite: 4267, 4268, 4269, 3809, 3810, 3811]

[8] J. Yang, T. Liu, Y. Liu, and P. Morgan, "Review of Human-Machine Interaction Towards Industry 5.0: Human-Centric Smart Manufacturing," in Proceedings of the ASME 2022 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, 2022. [cite: 5023, 4694]

[9] M.M. Falerni, V. Pomponi, H.R. Karimi, M. Lavit Nicora, L.A. Dao, M. Malosio, and L. Roveda, "A framework for human-robot collaboration enhanced by preference learning and ergonomics," Robotics and Computer-Integrated Manufacturing, vol. 89, p. 102781, 2024. [cite: 5697, 5327, 5568]

[10] C.B. Erdaş and S. Güney, "Human Activity Recognition by Using Different Deep Learning Approaches for Wearable Sensors," Neural Processing Letters, vol. 53, pp. 1795-1809, 2021. [cite: 7617, 7618, 7340, 7341]

[11] B. Gaffinet, J.A.H. Ali, Y. Naudet, and H. Panetto, "Human Digital Twins: A systematic literature review and concept disambiguation for industry 5.0," Computers in Industry, vol. 166, p. 104230, 2025. [cite: 8366, 8367, 8368, 8369, 8370, 8371, 8372, 8373, 7872, 7890, 7953, 7990, 7991, 8057, 8133, 8179, 8236, 8237, 8331]

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Calibration of storage tanks in the oil and gas sector based on the concept of digital metrology and big data analytics – Part I

Ankica Milinković[1], Sanja Tucikešić[2]

[1] *1 Laboratory Vekom Geo d.o.o, Todora Dukina 61, Belgrade, Serbia, ankica.milinkovic@vekom.com*
[2]*PhD Geodesy, assistant professor, Faculty of Architecture, Civil Engineering and Geodesy, University of Banja Luka, Bosnia and Herzegovina, sanja.tucikesic@aggf.unibl.org*

***Abstract* – In response to the digital transformation of metrology spearheaded by national and international quality infrastructure bodies, this paper lays the theoretical and methodological foundation for a new paradigm in the calibration of storage tanks within the oil and gas sector. Part I presents a structured framework that integrates the concepts of digital metrology, Industry 4.0, and distributed sensor systems into the domain of volumetric calibration. Emphasis is placed on the role of digital twins, metrological traceability in digital environments, and uncertainty quantification through software automation. The work systematically identifies the building blocks of this paradigm shift, ranging from intelligent infrastructure and real-time monitoring to semantic data integration and predictive analytics, establishing a basis for consistent, interoperable, and automated calibration practices. This foundational perspective serves as a precursor to the practical implementations and applied strategies elaborated in Part II.**

Keywords: digital metrology, 3d laser scanning, storage tank calibration, measurement uncertainty, digital twin, industry 4.0

I. INTRODUCTION

In technical and economically demanding sectors such as the oil and gas industry, accurate volume measurement of storage tanks plays a key role in efficient inventory management and in maintaining metrological integrity. The calibration of these tanks represents one of the critical processes within the quality assurance chain, as it directly affects the accuracy of commercial quantity calculations, as well as compliance with national and international standards.

Until recently, calibration procedures were based on standardized methods utilizing discrete distance and geometry measurements through electro-optical devices, which allowed only a limited representation of the tank's actual geometrical structure. Such traditional methods present several challenges: the duration of the process, limited data density, sensitivity to human factors, and limited capabilities for deformation detection and long-term monitoring of the tank's condition.

With the global trend of digitalization in metrology, led by major national metrology institutes and supported by initiatives such as EMPIR and EURAMET, there arises a need to transform the calibration process into an integral part of a broader framework of digital management of physical infrastructure. Here, the concept of digital metrology emerges, applying advanced sensors, numerical modeling, and automated data processing to improve the accuracy, repeatability, and transparency of measurement results.

Through the integration of virtual object modeling and the use of massively collected data (Big Data), a transition is enabled from static, isolated measurements to dynamic, digitally monitored, and adaptive systems. Big Data refers to the data or sets of records that are too large in volume to be operated using the existing database management tools and techniques [1]. The application of 3D laser scanning and the generation of high-density point clouds make it possible to create a detailed digital representation of the tank's geometry. These datasets are then analyzed using advanced software tools and machine learning algorithms to detect deformations, determine surface areas and volumes, and assess components of measurement uncertainty arising from environmental influences, object characteristics, sensors, and, most notably, the software components of the calibration system.

The motivation for this transformation stems from the need to ensure greater efficiency and scalability of the calibration process, higher reliability and traceability of results, the capability for continuous monitoring of tank condition during operation, and alignment with digital industry strategies such as Industry 4.0 and future Industry 5.0 models. Industry 4.0 addresses and solves some of the challenges facing the world today such as resource and energy efficiency, urban production and demographic

change [2].
The aim of this paper is to present the potential of modern technologies for enhancing the calibration process of storage tanks, with a particular focus on the application of laser scanning, measurement data analytics, and the quantitative management of measurement uncertainties resulting from the complex interaction between physical and digital factors. The proposed approach lays the groundwork for the development of intelligent, adaptive, and standards-compliant metrological systems within next-generation industrial environments.

II. STATE OF THE ART

In accordance with the International Vocabulary of Metrology (VIM), 3rd edition (JCGM 200:2012) [3], in a general sense, the calibration process is defined as an operation that, under specified conditions, in a first step, establishes a relation between the quantity values with measurement uncertainties provided by measurement standards and corresponding indications with associated measurement uncertainties, and, in a second step, uses this information to establish a relation for obtaining a measurement result from an indication. Specifically in the context of storage tanks, as defined by the International Recommendation OIML R 71 [4], calibration of a tank refers to a set of operations carried out to establish, under specified conditions, the relationship between the liquid level in the tank and the volume of that liquid. This specific calibration process ensures the traceable and accurate determination of liquid volume based on the measured liquid level, which is of particular importance in the oil and gas industry where tank volume measurements have a direct impact on commercial transactions, inventory control, and regulatory compliance.
The tank calibration shall be carried out by the applicable ISO Standards, or national standards as required, or if appropriate standard is not available, according to another relevant and competent method. Taking into account different needs and purposes of measuring tank information, there are different ways of measuring the liquid level and other properties of the liquid, depending on the type of tank, the type of liquid, and the way the tank is used [5].
The calibration of storage tanks in the oil and gas industry is currently governed by a series of well-established international and national standards, which primarily define geometric and volumetric methods based on discrete measurements or volumetric transfer. These standards provide robust methodologies for conventional calibration but were developed in an era prior to the widespread adoption of 3D scanning technologies, digital twins, and advanced data analytics.
ISO 7507-4 [6] specifies the methodology for the calibration of vertical cylindrical tanks with diameters greater than 5 meters, using internal measurements with an electro-optical distance ranging (EODR) device. The standard also defines procedures for the subsequent generation of tank volume tables. This method is applicable to tanks inclined up to 3% from the vertical, provided that the inclination correction is applied according to [7], ISO 7507-1:2003, Clause 11. Additionally, it is applicable to tanks with floating roofs or internal floating covers.
ISO 7507-5 [8] extends this approach to external measurements of uninsulated vertical cylindrical tanks with diameters greater than 5 meters, using EODR techniques. This standard also covers the generation of tank capacity tables and is applicable to tanks with conical bottoms (oriented upwards or downwards), as well as tanks with flat bottoms.
For horizontal cylindrical tanks, ISO 12917-2 [9] defines a calibration method based on internal EODR measurements, applicable to tanks with diameters greater than 2 meters. The procedure includes the preparation of tank capacity tables and is suitable for tanks inclined up to 10% from the horizontal, provided that appropriate inclination correction is applied.
In addition to geometric methods, ISO 4269 [10] defines a volumetric method for tank calibration based on the addition of batches of liquid. In this method, liquid serves as the volume transfer medium and is accurately measured using a certified flow meter. While effective for determining the total capacity of a tank, this method does not support the detection of local deformations or provide a detailed geometric model of the tank digital metrology and big data analytics.
While these standards form a solid basis for traditional calibration practices, they exhibit significant limitations in the context of modern digital metrology. This clearly highlights the need for either a revision of existing standards or the development of supplementary guidelines that can formally integrate modern digital technologies into tank calibration practices, thereby improving accuracy, efficiency, and traceability in line with the evolving requirements of the industry. The critical analysis of current standards clearly demonstrates that traditional calibration methods, while robust for their intended scope, are fundamentally limited in their ability to fully leverage modern digital technologies. The lack of integration with 3D laser scanning, automated data processing, digital twins, and real-time monitoring leaves a significant gap between conventional practices and the emerging capabilities of Industry 4.0/5.0 ecosystems.
In parallel, the broader field of metrology is undergoing a profound digital transformation, driven by initiatives such as EMPIR, EURAMET, and national programs led by key European National Metrology Institutes (NMIs). These efforts aim to develop digitally enabled, interoperable, and intelligent metrological processes that align with modern industrial needs. In this context, several key technological pillars are becoming increasingly relevant for tank calibration:

- the deployment of digital infrastructure to support traceable and transparent calibration workflows;
- the use of digital twins to enable continuous monitoring and predictive analysis of tank behavior;
- automated measurement systems to improve data quality and reduce human error;
- the adoption of software-driven solutions and Big Data analytics to enhance the modeling of tank geometry and the management of measurement uncertainties.

Furthermore, these developments are supported by strategic trends within European metrology, with ongoing projects actively exploring the harmonization of digital metrological frameworks across industries. The following sections provide an overview of these emerging approaches, highlighting their potential for transforming tank calibration and bridging the gap between conventional standards and next-generation digital methodologies.

A. *Application of Digital Infrastructure in Metrology*

According to various National Metrology Institutes (NMIs), the future of metrology can be shaped by focusing on data science, the development of new metrological services, and the emerging opportunities enabled by innovative technologies. We are amid a digital revolution that is challenging the metrology communities' working practices and paradigms for metrological traceability and reproducibility [11]. Leading institutes have redirected their research and development efforts toward the digitalization of measurement systems, data processing, software validation, and the implementation of Digital Calibration Certificates (DCCs), thereby opening new avenues for the calibration and control of complex structures such as storage tanks. It is also important to realize that DCC is not the only one pillar of digitalization in metrology so it must fit into the bigger concept of metrology digital for transformation. This means that the DCC concept must be designed to ensure interoperability, compatibility, modularity, and scalability while also establishing clear boundaries to ensure that metrology is always expressed clearly and comprehensibly. For this reason, several working groups focus on the development, implementation, and harmonization of DCC [12].

With the accelerated advancement of digital metrology, there is a growing need to redefine calibration methodologies in the direction of digitalization, automation, and integration with virtual models.

The following sections of this paper present the key elements expected to shape the future of digitally oriented calibration of storage tanks, in alignment with the strategic directions outlined by contemporary metrological developments across Europe.

B. *Digitally oriented calibration of storage tanks*

One of the fundamental principles of digital metrology is the establishment of a reliable and verifiable digital infrastructure. In the context of storage tank calibration, this implies that all stages, ranging from data acquisition (e.g., via 3D laser scanning), filtering, and point cloud analysis to volume computation, must be fully documented, validated, and digitally traceable. The implementation of Digital Calibration Certificates (DCCs), featuring digitally signed metadata on measurement conditions, instruments, and data processing algorithms, becomes a key prerequisite for maintaining confidence in digital measurement results. Digital twins enable the creation of virtual replicas of physical objects which are used to simulate structural behavior under various conditions [13]. The application of digital twins enables the modeling of tank behavior under real operational conditions, with continuous updates based on newly acquired scan data. These models integrate geometric features with parameters such as deformations, temperature gradients, and mechanical stresses. The development of a tank's digital twin supports dynamic monitoring of capacity changes, detection of localized structural anomalies, and the optimization of condition-based maintenance strategies. Digital metrology requires the development of novel methodologies for managing measurement uncertainty in environments where physical, software, and algorithmic variability coexist. In the context of digital tank calibration, it is particularly important to:

- consider environmental influences (e.g., temperature, hydrostatic effects),
- analyze uncertainty in point cloud processing,
- quantify software-induced error propagation.

It is essential to establish formal uncertainty models that encompass the entire digital calibration chain, from the measuring instruments to the final outputs.

Automated data processing, standardized protocols, and interoperability with laboratory and enterprise information systems represent foundational pillars of the digital calibration process. The calibration workflow must be designed to incorporate automated procedures for scanning, segmentation, volume computation, and seamless integration with platforms such as LIMS, ERP, and other monitoring systems. This approach ensures scalability, reduces human intervention, enhances operational efficiency, and minimizes the risk of human error.

For large-scale or geometrically irregular tanks, there is an increasing demand for modular and distributed measurement systems. A combination of multiple scan positions and sensor networks (e.g., for monitoring temperature, shell thickness, ultrasonic and visual inspection) enables a comprehensive understanding of the tank's structural condition.

Hybrid measurement architecture combining 3D geometric data, ultrasonic inspection results, and

environmental parameters, form the foundation for the next generation of digitally driven calibration systems. The future of storage tank calibration will be fundamentally shaped by the principles and technologies of digital metrology. The integration of digital twins, automated data processing algorithms, certified digital outputs, and system-level interoperability provides a clear pathway for transforming traditional calibration practices into a fully digital process.

To support this transition, it is essential to:

- develop harmonized standards and interoperable frameworks among metrology institutions,
- define digital protocols for verification, traceability, and documentation,
- extend calibration methodologies to encompass virtual and simulation-based environments.

Digital calibration of storage tanks thus becomes not merely a technical operation, but an essential component of intelligent digital infrastructures for measurement, decision-making, and operational control, fully aligned with the vision of Industry 4.0. In Chapter III, practical examples will be presented demonstrating the application of currently available Industry 4.0 concepts that are already in use and have the potential to significantly elevate calibration and inspection processes of storage tanks to the level of highly automated solutions, with considerable potential for further implementation and advancement.

III. CASE STUDIES

The calibration of storage tanks using 3D laser scanning technology produces high-density spatial data by capturing up to several million measurement points per second. The systematic exploitation of this "massive" dataset enables high-resolution reconstruction of internal surfaces, precise identification of localized deformations, and highly accurate volumetric computations. This study presents real-world applications in which terrestrial laser scanning has been successfully implemented in the metrological calibration of storage tanks, including both vertical and horizontal configurations. The quality of the results demonstrates that high-density spatial sampling not only enhances measurement accuracy but also facilitates complete digital traceability of the calibration process.

A total of three horizontal and three vertical tanks were included in the experimental campaign. The three vertical tanks are constructed from concentric belts of steel plates of varying thicknesses, a structural feature of critical relevance for future evaluations of deformation behavior under operational loads and thermal gradients. For all tanks included in the study, the elastic modulus corresponding to structural-grade steel was adopted as a reference parameter in subsequent numerical modeling procedures. The horizontal tanks are also made of steel, featuring flat circular end caps. In contrast to the above-ground design of the vertical tanks, the horizontal tanks are installed underground.

The objective of the calibration process is to support high-precision volumetric quantification for commercial transactions and regulatory compliance, including routine verification and correction volume calibration tables.

A. Instrumentation and Setup

All 3D scans were conducted using a Leica RTC360 terrestrial laser scanner (Figure 1). The acquisition resolution was configured to 12 × 12 mm at distances up to 10 meters, with a single-point spatial accuracy of ±1.9 mm in three-dimensional space. These parameters have a direct influence on the uncertainty budget, as they are explicitly considered in the estimation of registration uncertainty through the positioning accuracy of signal marker centers. Due to insufficient lighting conditions inside the underground tanks, the Visual Inertial System (VIS) functionality was not utilized. Instead, scan registration was performed using target-based referencing, with strategically placed black-and-white reflective markers inside the tanks.



Figure 1. Measuring systems, 3D laser scanner, Leica RTC360

Each horizontal tank was scanned from two interior scanner positions, strategically selected to ensure full surface coverage, including end caps and longitudinal shell curvature. In order to avoid shadowing effects caused by circumferential joints or weld seams, scanning was conducted from two scanner positions aligned along the vertical axis through the tank opening. The first scan was performed with the scanner positioned near the opening, while the second scan was conducted by vertically lowering the scanner by 1.5 meters relative to the initial position. The overlap between the two scans exceeded 90%, ensuring robust registration and full surface coverage. Despite the presence of an onboard Inertial Measurement Unit (IMU), environmental conditions were confirmed to be vibration-free during acquisition, thereby ensuring optimal data integrity. In the case of vertical tanks, the presence of internal elements such as heaters and measurement systems required a more careful selection of scanning stations, with the mandatory requirement that a minimum of three identical signal markers be captured from each scanner position. However, in all cases, the

number of measurement positions did not exceed four, considering the sufficient level of coverage of the tank bottom. The scans overlap was approximately 70%. This ensured the acquisition of a reliable measurement dataset suitable for automated processing of the densely captured point clouds.

The temperature monitoring of the tank shell was carried out using a spot-type thermal imaging sensor, enabling localized temperature readings across the wall surface. Measurements were performed in a structured grid pattern, where the density of measurement points was determined based on the sensor's resolution (e.g., 80 × 60 pixels for the FLIR TG165), the distance between the sensor and the tank surface (approximately 2 m for horizontal tanks and 10 m for vertical tanks), the field of view (FOV) of the thermal camera, and the desired spatial resolution, aligned with expected temperature gradients on the steel surface.

The measurement point density (points/m²) was calculated by projecting the sensor's pixel matrix onto the tank wall, with each pixel representing a single temperature data point. To ensure a representative thermal map of the tank shell, particularly in areas of critical relevance for volumetric analysis (e.g., near the tank bottom or in regions with pronounced temperature gradients), the distribution of measurement points was either:

- uniformly spaced, with a fixed interval between points (typically every 1 to 1.5 meters), or
- adaptively densified in zones where preliminary scans indicated high thermal variability.

As all tanks had original engineering drawings available containing information on shell thickness, these values were not measured in the field but were adopted directly from the design documentation. Their contribution to the measurement uncertainty was evaluated in accordance with ISO 7507-1, Annex D.

The inclination of the tanks was automatically incorporated into the geometric reconstruction process by utilizing scan-derived orientation data and fitted geometric surfaces.

IV. CONCLUSIONS

This paper establishes the theoretical, methodological, and conceptual foundation for a new paradigm of digital calibration of storage tanks in the oil and gas industry, through the integration of modern concepts of digital metrology. By applying 3D laser scanning, digital twins, and automated analysis of large datasets, the framework defines the pathways for high-resolution reconstruction of tank geometry, quantification of measurement uncertainty, and full digital traceability of the calibration process.

Through critical analysis of existing international standards, the study highlights their limitations in the context of modern digital technologies, emphasizing the need for their revision and/or extension. In this regard, the development of interoperable Digital Calibration Certificates (DCCs), standardized protocols, and formal uncertainty models in digital environments represents a key step towards the full digitalization of calibration processes, in line with the evolving requirements of next-generation industrial ecosystems.

Digital calibration of storage tanks is no longer viewed merely as a technical operation, but as an integral component of intelligent digital infrastructures for measurement, decision-making, and operational control, fully aligned with the vision of Industry 4.0 and the forthcoming Industry 5.0.

In order to open the discussion for experimental investigation, this study also describes the data acquisition concept and technical resources employed in the calibration of six tanks—specifically, three horizontal and three vertical storage tanks—as representative examples. The experimental data, technical reports, specific measured tank characteristics, and case study examples are presented in a separate publication that directly extends this study by demonstrating the practical application of the proposed methodological framework under real industrial conditions.

The continuation of this study is directly linked to the paper titled Calibration of Storage Tanks in the Oil and Gas Sector Based on the Concept of Digital Metrology and Big Data Analytics – Part II.

REFERENCES

[1] Walunj, S. K., Yadav, A. H., & Gupta, S. (2016). Big data: Characteristics, challenges and data mining. *International Journal of Computer Applications*, *(0975–8887)*, International Conference on Advances in Information Technology and Management (ICAIM – 2016).

[2] Aktan, A. E., Catbas, F. N., Grimmelsman, K. A., & Tsikos, C. J. (2000). Issues in infrastructure health monitoring for management. *Journal of Engineering Mechanics*, *126*(7), 711–724. https://doi.org/10.1061/(ASCE)0733-9399(2000)126:7(711)

[3] Joint Committee for Guides in Metrology (JCGM). (2012). *International vocabulary of metrology – Basic and general concepts and associated terms (VIM)* (3rd ed.). BIPM.

[4] International Organization of Legal Metrology (OIML). (2008). *OIML R 71: Fixed storage tanks – General requirements*.

[5] Milinković, A., García-Balboa, J. L., & Ruiz-Armenteros, A. M. (2020, October 20–22). An overview of methods applied to quality control of storage tanks volume. *17th IMEKO TC 10 and EUROLAB Virtual Conference: Global Trends in Testing, Diagnostics & Inspection for 2030*.

[6] International Organization for Standardization (ISO). (2010). *ISO 7507-4:2010 Petroleum and liquid petroleum products — Calibration of vertical cylindrical*

tanks — Part 4: Internal electro-optical distance-ranging method. ISO.

[7]International Organization for Standardization (ISO). (2003). *ISO 7507-1:2003 Petroleum and liquid petroleum products — Calibration of vertical cylindrical tanks — Part 1: Strapping method*. ISO.

[8]International Organization for Standardization (ISO). (2005). *ISO 7507-2:2005 Petroleum and liquid petroleum products — Calibration of vertical cylindrical tanks — Part 2: Optical-reference-line method*. ISO.

[9]International Organization for Standardization (ISO). (2002). *ISO 12917-2:2002 Petroleum and liquid petroleum products — Calibration of horizontal cylindrical tanks — Part 2: Internal electro-optical distance-ranging method*. ISO.

[10]International Organization for Standardization (ISO). (2001). *ISO 4269:2001 Petroleum and liquid petroleum products — Tank calibration by liquid measurement — Incremental method using volumetric meters*. ISO.

[11]BIPM. (2022, September). *Report on the actions taken by the CIPM towards a "CIPM Strategy 2030+"*. International Committee for Weights and Measures (CIPM).

[12]Koval, M., Tesař, J., Schönhals, S., Sharma, A., Balslev-Harder, D., Galvan-Hernandez, C., Baumann, H., & Mahovský, D. (n.d.). Digital Calibration Certificates (DCC): Benefits and challenges to overcome for implementation.

[13]Fuller, A., Fan, Z., Day, C., & Barlow, C. (2020). Digital twin: Enabling technologies, challenges and open research. *IEEE Access*, *8*, 108952–108971. https://doi.org/10.1109/ACCESS.2020.2998358

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Calibration of storage tanks in the oil and gas sector based on the concept of digital metrology and big data analytics – Part II

Ankica Milinković[1], Sanja Tucikešić[2]

[1] *1 Laboratory Vekom Geo d.o.o, Todora Dukina 61, Belgrade, Serbia,*
ankica.milinkovic@vekom.com
[2]*PhD Geodesy, assistant professor, Faculty of Architecture, Civil Engineering and Geodesy, University of Banja Luka, Bosnia and Herzegovina, sanja.tucikesic@aggf.unibl.org*

***Abstract* – Part II of this study presents the practical implementation and evaluation of the digital paradigm for storage tank calibration in the oil and gas sector, as conceptually and methodologically established in Part I. This section demonstrates the practical application of digital methods for calibration, including 3D laser scanning, the analysis and description of the final results obtained through the processing of collected measurements, the establishment of the metrological traceability chain, and a brief overview of key measurement uncertainty components. The findings indicate the potential for a shift from reactive to predictive calibration strategies, in alignment with Industry 4.0 principles and contemporary requirements for interoperability and digital metrology.**

Keywords: digital metrology, 3d laser scanning, tank calibration, digital twin, industry 4.0

I. INTRODUCTION

This paper represents a continuation of the previous study described in Calibration of storage tanks in the oil and gas sector based on the concept of digital metrology and big data analytics – Part I, in which the conceptual and methodological foundations of digital metrology applied to the calibration of industrial tanks were established. Within this second part, the focus is placed on the experimental application and technical evaluation of the digital calibration paradigm, with the aim of quantitatively verifying the accuracy, data quality, and traceability of measurements under real industrial conditions.

Through the scanning of three horizontal and three vertical steel tanks, factors influencing the quality of the collected data, geometric accuracy of the models, and reliability of the volumetric calibration tables were analyzed.

The scanning was conducted from multiple positions, with point cloud registration carried out in the Leica Cyclone REGISTER 360 software environment, and further processing in Cyclone 3DR, using Python scripts for volume correction, temperature influence simulation, and digital deformation modeling. The resulting models were used for geometric analysis in accordance with API 650/653 standards and for the calculation of measurement uncertainty in line with GUM and EA-4/02 guidelines.

The aim of this part of the study is to demonstrate the technical feasibility, benefits, and challenges of digitizing the calibration process for large industrial tanks, as well as to provide foundation for the future integration of interoperable digital solutions into accredited calibration protocols.

II DATA PROCESING AND GEOMETRIC RECONSTRUCTION

During the scanning of horizontal storage tanks, it was observed that the interior did not contain structural elements capable of generating additional signal noise (e.g., heaters, mechanical obstructions). This favorable condition enabled the acquisition of homogeneous and high-fidelity point clouds (Figure 1), without the need for additional post-processing aimed at filtering spurious reflections. In contrast, vertical tanks were equipped with internal components, such as heaters, instrumentation, and piping systems, that introduced localized artifacts into the point cloud by reflecting the laser signal. To mitigate their influence on volume estimation, an analytical correction procedure was applied, in which the volumes of these internal elements were identified, geometrically modeled, and subtracted from the total reconstructed tank volume. This procedure ensured a metrologically valid result. In this context, there is scope for integrating machine learning algorithms into the software environment, enabling the automated identification of critical internal structures based on their volumetric impact on total tank volume. These elements could then be retained or excluded from the final model according to predefined thresholds. Such internal components fall under the definition of “dead volume,” as further described in ISO 7507-1. Based on Figure 2, it can be concluded that the models for the horizontal tanks (R1, R2, R3) exhibit a high degree of

symmetry, with clearly defined end caps and no visible deformations along the cylindrical body. The absence of "cold" zones (indicated by blue and green tones) suggests high registration accuracy and good positional stability during point cloud processing.

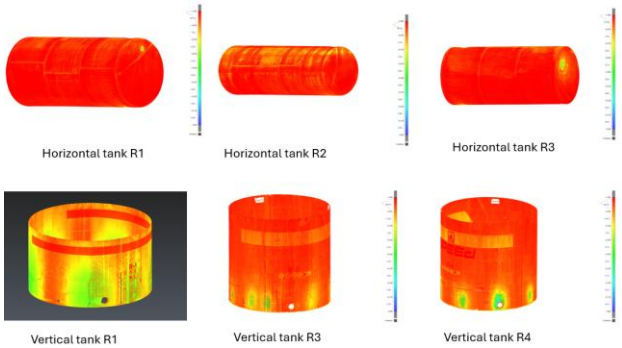


Figure 1. Registrated poit clouds of scanned tanks

In contrast, the models for the vertical tanks (R1, R3, R4) display greater color variation, particularly in the lower regions (blue/green areas), indicating the presence of localized deformations or internal structures such as heaters, sensors, and piping systems. Vertical tank R1 shows a distinctly multicolored distribution, suggesting more complex internal geometry and potential signal disturbances during scanning. Vertical tanks R3 and R4 exhibit pronounced horizontal bands, resulting from the joints between steel plates of varying thicknesses. The color-coded deviation patterns are more prominent (green and blue tones), indicating a broader range of deviations, potentially reaching up to ±5 mm in certain localized areas. Data acquisition was performed from multiple scanner positions to ensure full coverage of the internal surface.

Point cloud registration was carried out in the Leica Cyclone REGISTER 360 software environments using an automated workflow. Alignment of individual scan coordinate systems was performed via a seven-parameter Helmert transformation, ensuring accurate spatial integration of all points into a unified global reference frame. A prerequisite for this method was the availability of spatial coordinates of signal targets, distributed such that at least three common reference targets were visible within the overlap area of any two adjacent scans. Thanks to recent research on scan planning optimization (including the number of scan positions, distribution of reference targets, and required overlap), individual scan pairs exhibited a high degree of spatial overlap, facilitated by the geometrically constrained cylindrical shape of the tanks.

The number of reduced points, following the application of filtering and point cloud optimization, was approximately 440,000 per individual point cloud. This represents a significant reduction compared to the initial datasets, which contained several tens of millions of points. The reduction was carried out while preserving spatial resolution in critical areas, such as the transitions between the tank walls and bottom, as well as in regions where deformations were identified. This approach enabled more efficient numerical processing without compromising the geometric accuracy or the metrological relevance of the reconstruction. After the filtering stage, a three-dimensional surface mesh was generated (Figure 2). In the context of laser scanning, a mesh constitutes a continuous geometric approximation of discrete measurement points, whereby adjacent points are connected via triangular or quadrilateral facets to form a coherent digital surface. The resulting 3D model (Figure 3) served as the analytical foundation for deformation analysis, ovality detection, numerical volume computation, and visualization of deviations from ideal geometry.

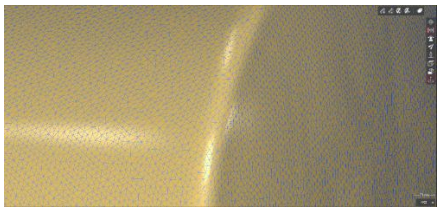


Figure 2. Meshed Surface of the Storage Tank

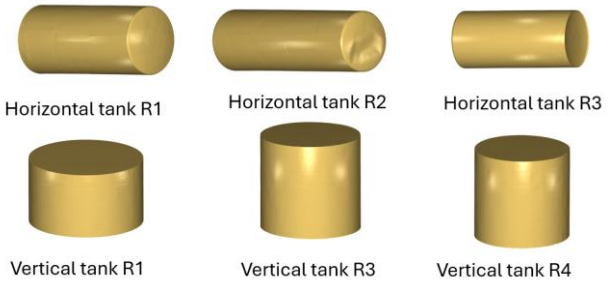


Figure 3. 3D models of tanks

Based on the 3D models of the scanned storage tanks (Figure 3), it can be concluded that all analyzed structures, three horizontal and three vertical tanks, exhibit topologically continuous and coherent surfaces, confirming successful point cloud registration and mesh generation. Horizontal tanks show a high degree of longitudinal symmetry and minimal artifacts, indicating effective scan planning with optimal overlap that mitigated shadowing and loss of data density. Vertical tanks R3 and R4 display mild geometric irregularities, such as local ovality and deviations from ideal cylindrical form, likely caused by internal components (e.g., heaters or piping), occluded scan areas, or meshing limitations. In horizontal tank R2, a slight deformation of the front-end cap and discrete horizontal surface variations suggest minor geometric imperfections potentially arising from operational stress, pressure variations, or fabrication

inaccuracies. Despite localized anomalies, all models maintain sufficient geometric fidelity to support advanced analyses, including ovality assessment, corrected volume computation, and digital twin development. For final verification, comparison with CAD reference models and extraction of sectional profiles is recommended, in line with the applicable standards for dimensional tolerances.

A. *Geometric Analysis and Deformation Assessment*

As part of a modern digital metrological approach, once a three-dimensional model of the tank is generated, a comprehensive geometric analysis is conducted to identify deviations and deformations relative to the design geometry. This analysis is performed in accordance with the requirements of API 650 [1] and API 653 [2].

One of the key steps is the analysis of shell ovality, which is carried out by evaluating horizontal cross-sections along the height of the tank. In each cross-section, deviations in radius from an ideal circular form are measured, enabling the identification of potential distortions resulting from non-uniform loading, thermal gradients, or mechanical damage to the structure. Ovality testing for vertical tanks was conducted using cross-sectional samples at 1-meter height intervals, while for horizontal tanks, the analysis was performed at 0.5-meter intervals along the horizontal cylindrical axis. Figure 4 presents representative examples of ovality profiles for Vertical Tank R4 at heights of 1 m, 5 m, and 10 m, and for Horizontal Tank R1 at longitudinal positions of 0.5 m, 3.5 m, and 5.5 m.

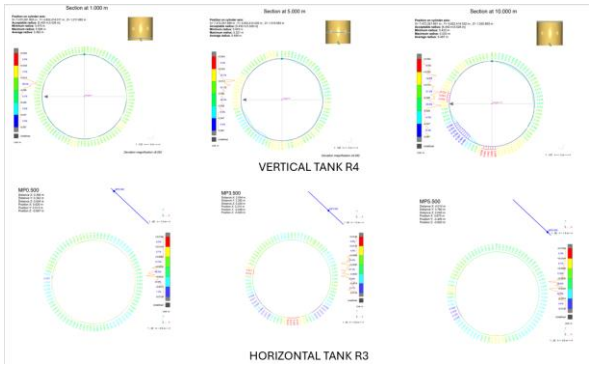


Figure 4. Representative Samples of Ovality Testing: Vertical Tank R4 and Horizontal Tank R1

In addition to ovality analysis, deviation maps (Figure 5) are generated, which provide a visual representation of local discrepancies between the reconstructed model and the reference geometric shape (in our case, the best cylinder shape). These maps allow for the efficient detection of localized anomalies, including dents, bulges, and other irregularities that may indicate structural weaknesses, material erosion, or damage sustained in prior operational cycles.

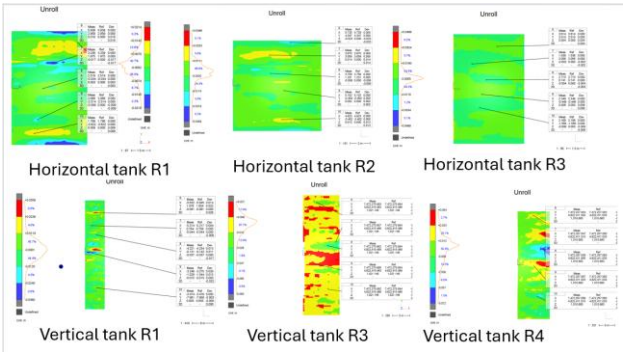


Figure 5. Deviation maps of the tank shell

From Figure 5, shows that for the horizontal tanks (R1, R2, R3), the colors are predominantly green, yellow, and partially blue and orange. Green means small deviations; the surface is very close to an ideal cylinder. Yellow/Red means bulges (positive deviation). Blue means indentations (negative deviation). The horizontal tanks are generally in good geometric condition. Maximum deviations are below ±5 mm. Vertical tanks (R1, R3, R4), the colors include a wider range of shades, including red and dark blue. Vertical tank R3 shows deviations (red and yellow zones) – deviations reach ±7 mm or more. Tank R1 displays a localized deviation zone (dark red), which may indicate a deformation or dent, but of a very localized and isolated nature. Tank R4 shows a localized bulge deviation (blue and dark blue zones).

The analysis of the spatial orientation of the vertical axis of the tank enables detection of tilt, curvature, or eccentricity in the structure, which is especially critical for the safe operation of tanks equipped with floating roofs or subject to dynamic loading conditions. This is achieved by determining the orientation of the tank's central axis based on multiple horizontal cross-sections taken along the tank's height. Aross 10 cross-sections, each spaced 18 degrees apart, the ovality assessment was performed circumferentially around the tank shell in a clockwise direction. An example of a single ovality evaluation is shown in Figure 6 corresponding to Vertical Tank R4, specifically for the sector between 0° and 180°.

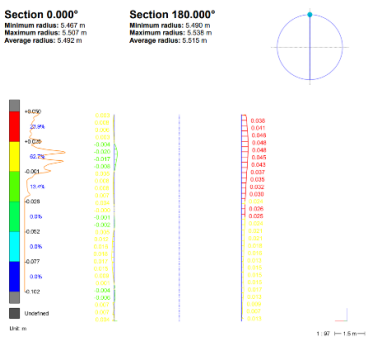


Figure 7. The spatial orientation of the vertical axis of the tank R4 Vertical, position 00°-180°

Furthermore, evaluation of the flatness and roundness of the tank bottom and roof structure could also be performed, particularly in designs featuring flat or conical foundations. Deviations in these zones can lead to uneven stress distribution, sediment accumulation, leakage, or structural vulnerabilities, and in accordance with API 653 guidelines, they require continuous monitoring and, if necessary, implementation of corrective measures. For the purposes of this study, this segment was not taken into consideration.

All the analytical procedures can be performed within the Leica Cyclone 3DR software environment, utilizing its built-in tools for sectional analysis, deviation map generation, axis transformation, and geometric evaluation of roundness and flatness. Alternatively, the same analyses can be automated and further adapted to specific industrial needs using the Python programming environment, supported by point cloud processing libraries (e.g. *Open3D*, *PCL*, *NumPy*, *Matplotlib*) and customized numerical algorithms. This enables a flexible and scalable application of the described procedures, both in commercial software platforms and in open-source systems that support full automation, integration with industrial databases, and real-time connectivity with monitoring systems.

In this phase of the inspection analysis of tank geometry, which precedes the computation of the geometric and subsequently corrected volume tables adjusted to reference conditions, a broadening of the analytical concept is proposed through the integration of additional sensor data. In this context, the use of advanced UAV systems designed for ultrasonic, thermal and visual inspection is considered highly relevant. By integrating these multi-source datasets with the point cloud obtained through laser scanning, it becomes possible to construct a unified informational model that supports a comprehensive analysis of both the geometric integrity and operational status of the tank. This approach not only extends the scope of metrology within inspection procedures but also facilitates intelligent decision-making in the context of preventive maintenance, as well as the development of digital twins based on multi-validated input data.

B. *Volume Calculation*

As part of an advanced digital metrology approach, following the generation of a high-density 3D point cloud model of the storage tank, a numerical computation of geometric volume as a function of filling height was conducted. This calculation is based on an accurate reconstruction of the tank's internal surface, with a volumetric table generated at a vertical resolution of 1 cm (layer-by-layer integration). The resulting table serves as the fundamental numerical structure for subsequent analyses and simulations within the digital twin environment.

To correct the nominal geometric volume and more accurately approximate the actual behavior of the tank under operational conditions, a custom Python script was developed to automatically apply multifactor corrections to each volumetric entry, in accordance with the specifications of ISO 7507-1:2003. The correction model includes the following physical influences:

- Hydrostatic pressure correction, implemented according to Annex A of the standard.
- Correction to the reference shell temperature, as defined in Section 16.1 of the standard,
- Deadwood volume correction, as described in Section 17.1 of the standard;
- Inclination correction, following the guidelines in Section 16.2 of the standard.

Additionally, particular emphasis was placed on modeling hydrostatic wall deformation, calculated as a function of the liquid column height, fluid density (ρ), Young's modulus (E), and local wall thickness (t). For vertical tanks, these parameters were defined across elevation segments, enabling a layer-wise structural compliance analysis (*segment-wise compliance model*). Resaults afre presented at Table 1 and Table 2.

The Python script ingests input data from a CSV table of geometric volumes per elevation layer and applies the defined correction expressions to each vertical unit. The corrected values are then aggregated into a final volume table, representing the numerical output of the digital twin, ready for integration into asset management systems, structural health monitoring platforms, and decision-making processes related to preventive or corrective interventions.

Beyond correction, the system also enables simulation of tank volume behavior under varying operational scenarios, including filling with different fluids possessing distinct physical properties (e.g., crude oil, water, chemicals), exposure to temporal and seasonal temperature variations, and estimation of volume differences before and after service cycles, to quantify deformation effects under real-world conditions.

This approach supports dynamic modeling and digital tracking of tank behavior throughout its entire life cycle. It is fully grounded in traceable measurement data and algorithmically consistent processing procedures, adheres to the methodological principles of contemporary digital metrology, and aligns with the operational philosophy of the digital twin concept in industrial applications.

At this stage of the process, a clear opportunity is identified for integration with permanently installed temperature sensors on the tank, which would enable real-time modeling of volumetric capacity as a function of ambient temperature. Furthermore, the development of a simulation model is proposed to analyze the behavior of the scanned tank geometry in relation to the spatial distribution of atmospheric temperature, derived from the generation of heat islands using satellite imagery of the

area where the tanks are located. This approach would allow for the identification and quantification of the impact of climate-induced temperature variations on storage capacity.
This analysis opens a promising field for further research, particularly in the direction of linking the digital twin of the tank with the spatial distribution of urban heat islands derived from satellite observations. Specifically, by integrating surface temperature data obtained from remote sensing platforms (e.g., Landsat, Sentinel-3, MODIS), it becomes possible to incorporate external climatic parameters into the simulation model of the digital twin. This integration would enable predictive modeling of temperature effects on tank volume under real operational and environmental conditions.
Such integration represents a significant step toward the spatiotemporal coupling of large-volume metrology with geo-environmental parameters and opens new avenues for optimizing storage systems in the context of climate change and energy analyses.

Table 1 Correction-adjusted volumetric tables for vertical storage tanks

Tank ID	Design Volume [m^3]	Volume at Max Height [m^3]	Max Filling Height [cm]	Reference Height [cm]
R1	1000	**1045.295**	735	777.5
R3	1000	**930.386**	975	1081.6
R4	1000	**929.193**	975	1081.7

From Table 1, following the applied volume corrections, it was observed that Tank R1 exceeds its nominal design capacity, while Tanks R3 and R4 exhibit slightly lower effective volumes compared to their respective design specifications. This discrepancy may be attributed, in the case of R1, to the removal of an internal membrane that is present in the original design drawings but absent in the as-built configuration. Conversely, the reduced effective volumes of R3 and R4 may result from the subsequent installation of internal components, such as heaters, which are physically present in the tanks but not reflected in the original design documentation. It is important to emphasize that the available design drawings appear to be outdated. Therefore, it is strongly recommended that the tank owner/operator undertake a formal review and validation of the technical documentation considering these structural modifications introduced during the operational lifespan of the tanks.

Table 2 Correction-adjusted volumetric tables for vertical storage tanks

Tank	Nominal Volume	Measured Volume	Max Filling Height	Deviation from the Design Volume [%]
R-1	40 m^3	40.105 m^3	275 cm	0.263
R-2	60 m^3	58.999 m^3	276 cm	-1.668
R-3	40 m^3	39.949 m^3	278 cm	-0.128

As shown in Table 2, the measured volumes are very close to the nominal design specifications, which indicates a high degree of manufacturing precision and confirms the reliability of the volumetric tables generated through the calibration process.

C. *Traceability of measurement*

As part of the calibration campaign, all storage tanks were measured using key metrological equipment a 3D laser scanner that had been previously calibrated in an accredited laboratory, in accordance with the requirements of ISO/IEC 17025. The accredited laboratory includes within its scope of accreditation a specific method for the calibration of 3D laser scanners, thereby formally confirming its competence to verify the spatial accuracy of optical measurement systems.
This calibration method is based on the use of highly precise and traceable reference instruments, including a laser interferometer for the verification of length measurements and a precision theodolite for the validation of angular measurements. Both instruments had previously been calibrated in a Swiss accredited laboratory, thereby ensuring the traceability of measurement results to the national standards for the SI units of length and angle, maintained by METAS. Length and angle represent fundamental physical quantities in 3D metrology and, for the purposes of this study, form the basis for the reconstruction of the three-dimensional position of points on the object surface, the generation of the geometric 3D model of the tank, and the volumetric computation through digital surface integration. Through this measurement chain, the entire sequence of results, from raw data to the final volume computation, remains fully metrologically traceable, providing a robust foundation for technical validity, regulatory compliance, and international acceptance of the calibration results.
A useful point to note is that ISO 17123-9 is also available, defining a method for field calibration of terrestrial laser scanners, thereby further enhancing measurement reliability and metrological traceability in accordance with internationally recognized procedures.

D. *Uncertainty Contributions*

In accordance with the principles of the Guide to the Expression of Uncertainty in Measurement (GUM) [3], [4], and relevant ISO standards (e.g., ISO 7507 -1 Annex D), the overall measurement uncertainty in the calibration of storage tanks, particularly when using 3D laser scanning technologies, arises from multiple sources. These sources can be grouped into distinct categories:
Instrumentation-Related Uncertainty (Type B): The intrinsic accuracy of the laser scanner, in terms of both angular and distance measurements;Traceability and calibration of the reference standards used in scanner calibration, such as laser interferometers and precision theodolites, which ensure alignment with SI units of length

and angle; Systematic errors in the scanner itself, including scale factor deviations, angular misalignments, and instrumental drift.

Environmental Influences (Type A & B): Temperature fluctuations during data acquisition affect both the thermal expansion of the tank walls and the physical behavior of the stored fluid; Vibrations, wind, and platform instability may compromise scanner stability;Atmospheric pressure and humidity impact the refractive index of air, potentially altering laser path lengths.

Object-Specific Uncertainty (Type B): Deviations from ideal geometry, such as ovality, out-of-roundness, and eccentricity;Non-uniform wall thickness, which affects the elastic deformation under hydrostatic loading; Unknown or unaccounted deadwood volume, including internal fixtures such as heaters or pipelines, particularly when these are absent from design drawings.

Data Processing and Software Uncertainty (Type B): Surface fitting, segmentation, and interpolation routines can introduce errors if the point cloud quality is compromised; Numerical integration techniques for volumetric calculation and temperature correction algorithms must be validated;Software limitations, including floating-point rounding and processing precision, may affect results.

Procedural and Human Factors (Type A & B): Scanner positioning and number of scan stations influence data completeness and registration accuracy; Manual interpretation of structural features and validation steps may introduce operator-dependent variability.

Modeling and Correction Uncertainty (Type B): Thermal expansion correction, based on assumed coefficients of expansion for the tank material and the fluid; Inclination correction, as per ISO 7507-1:2003, clause 16.2.; Hydrostatic deformation modeling, requiring accurate inputs for modulus of elasticity, local wall thickness, and fluid density.

A comprehensive uncertainty evaluation in tank calibration encompassed all identified sources, using both statistical analysis (Type A) and expert-based estimation (Type B). The combined standard uncertainty is aggregated through root-sum-square methods and expanded using a coverage factor (typically k=2 for a 95% confidence level). Taking into account the significance, distribution, and spatial propagation of the identified uncertainty sources, the combined measurement uncertainty was evaluated also in accordance with the methodology prescribed in the EA-4/02 guideline [3]. For the vertical storage tanks included in this study, the resulting expanded uncertainty was estimated at 0.2% of the calibrated volume, while for horizontal tanks, the corresponding value was 0.3% of the total volume. These values reflect the cumulative effect of instrumental, environmental, object-specific, and processing-related influences, and are considered compliant with internationally recognized practices for large-volume calibration.

IV CONCLUSIONS

This study demonstrates the application and further full potential of digital metrology for the calibration of both vertical and horizontal storage tanks using high-density terrestrial laser scanning. All analyzed tanks were accurately reconstructed as 3D digital models with topologically coherent meshes and geometric deviations well within the permissible limits defined by API 653 (≤0.105 m), confirming structural compliance and axial symmetry. Each tank was modeled from over 440,000 filtered spatial points, ensuring high geometric fidelity. The measurement process adhered to full metrological traceability, based on ISO/IEC 17025-accredited calibration procedures for laser scanners and reference instruments (interferometer and theodolite). Volume calibration tables were generated and corrected using custom-developed Python scripts, incorporating hydrostatic pressure, thermal expansion, inclination, and deadwood volume effects, in line with ISO 7507-1. The estimated expanded uncertainties were ≤0.2% for vertical and ≤0.3% for horizontal tanks, in accordance with OIML R 71. The study emphasizes the value of digital twins for real-time volume simulation and structural monitoring. It also recommends the integration of machine learning algorithms for automated detection of internal components, data fusion with visual and ultrasonic inspection systems, and coupling with real-time temperature monitoring platforms to model capacity variations due to environmental conditions. These findings underscore the transformative potential of digital calibration in improving traceability, operational efficiency, and predictive maintenance, contributing to the evolution of Industry 4.0-compliant metrological infrastructures.

REFERENCES

[1]American Petroleum Institute (API). (2020). *API Standard 650: Welded tanks for oil storage* (13th ed.).

[2]American Petroleum Institute (API). (2014). *API Standard 653: Tank inspection, repair, alteration, and reconstruction*.

[3]Joint Committee for Guides in Metrology (JCGM). (2023). *JCGM GUM-1:2023 – Guide to the expression of uncertainty in measurement — Part 1: Introduction*.

[4]Joint Committee for Guides in Metrology (JCGM). (2020). *JCGM GUM-6:2020 – Guide to the expression of uncertainty in measurement — Part 6: Developing and using measurement models*.

[5]European Cooperation for Accreditation (EA). (2022). *EA 4/02 M:2022 – Evaluation of the uncertainty of measurement in calibration*.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Structural Connection Parameter Measurement for Precision Motion Systems Using Accelerometers: A Preliminary Study

Li Li[1], Zheng Yang[1], Jia Ziqing[1]

[1] *Center of Ultra-precision Optoelectronic Instrument Engineering, Harbin Institute of Technology, Harbin, China, hitlili@hit.edu.cn, 24s001063@stu.hit.edu.cn, 120L011718@stu.hit.edu.cn*

***Abstract* – Accurate modelling is important for both the performance control and monitoring of high-precision motion systems in advanced manufacturing machines. It is not difficult to build dynamic models for motion systems based on the multi-rigid-body system assumption. However, it is usually not feasible to acquire the real connection parameters between rigid bodies like stiffness and damping coefficients especially after mechanical assembly which can be significantly different from their designed values. In this paper, a measurement scheme using multiple accelerometers is presented with discussion of the data processing methods. Simulation results on a three-rigid-body motion system and experimental results on a two-mass motion system preliminarily verify the scheme. Finally, existing problems are put forward for further study.**

I. INTRODUCTION

High precision motion systems are indispensable parts in advanced manufacturing machines, which play important roles in improving the machine accuracy and productivity [1-2]. Modeling is the primary step for the performance control and monitoring during the life-cycle of the motion system [3-4]. In practice, the high-precision motion systems can usually be described as multi-rigid-body systems [5]. A typical example is the lithography machine used for wafer scanning, in which the wafer stage can be simplified as an open-loop chain lumped mass–spring–damper model as shown Fig.1 [6]. Based on this assumption, it is not difficult to derive the mathematical dynamic model which can be used for performance evaluation [7]. However, despite the derived model structure can grasp the major features of the system dynamics, the accuracy of the model parameters will affect the authenticity of the evaluation results [8]. Since the mass values of mechanical parts can be known exactly before assembly, the structural connection parameters such as the stiffness and damping coefficients are the unknown model parameters which are required to be determined [9]. These parameters can only be determined after assembly, and their real values are usually significantly different from the design values.

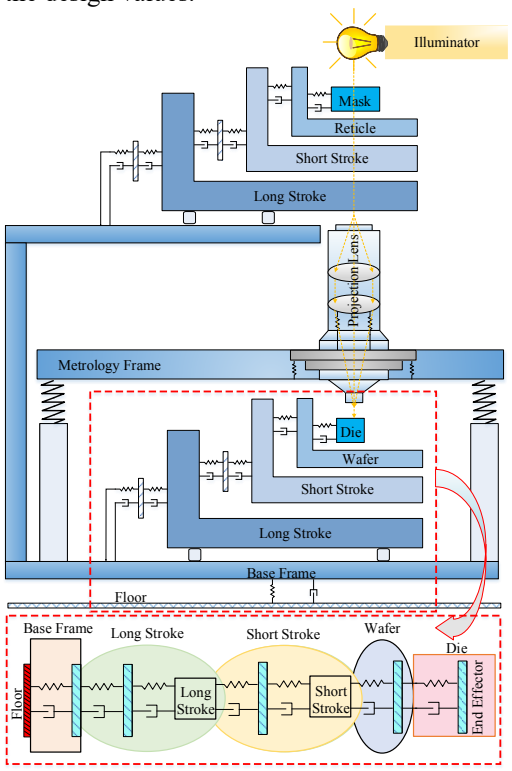


Fig.1 Schematic diagram of lithography machine and open-loop chain lumped mass–spring–damper model of wafer stage [6].

In this paper, the modeling of one translational DOF (degree of freedom) of the multi-rigid-body motion system is considered as an illustrated example. Based on the multi-rigid-body model structure and the already known mass values of the mechanical parts, multiple accelerometers are utilized to measure the structural connection parameters. The measurement scheme is presented and the data processing methods are discussed. Simulation results can verify the above theoretical results. In addition, the first version of the wireless measurement device is developed and preliminary experimental results are obtained. Finally, existing problems are discussed which are required to be solved in the future study.

II. MEASUREMENT SCHEME

Without loss of generality, the n-rigid-body system is considered here and can be shown in Fig.2. Denote x_i as the output (displacement [m]) of the ith rigid body. Denote m_i, k_i and b_i as the mass, stiffness and damping coefficients. Then it can be derived that

$$\sum_{j=i}^{n} m_j \ddot{x}_j = b_i(\dot{x}_{i-1} - \dot{x}_i) + k_i(x_{i-1} - x_i) \quad (1)$$

For the n-rigid-body system, n equations like Eq.(1) can be built, which are summarized as follows:

$$\begin{cases} m_n \ddot{x}_n = b_n(\dot{x}_{n-1} - \dot{x}_n) + k_n(x_{n-1} - x_n) \\ \vdots \\ \sum_{j=i}^{n} m_j \ddot{x}_j = b_i(\dot{x}_{i-1} - \dot{x}_i) + k_i(x_{i-1} - x_i) \\ \vdots \\ \sum_{j=2}^{n} m_j \ddot{x}_j = b_2(\dot{x}_1 - \dot{x}_2) + k_2(x_1 - x_2) \\ \sum_{j=1}^{n} m_j \ddot{x}_j = -b_1 \dot{x}_1 - k_1 x_1 \end{cases} \quad (2)$$

It should be noted that the known control input signals are neglected in above equations for the simplicity of the presentation, which will be considered and discussed in the future study.

Denote $\ddot{x}_i$ as a_i, $i = 1, \cdots, n$ which can be obtained by using accelerometers. The measurement value is denoted as $\tilde{a}_i$, which is corrupted with the measurement noises v_i. It means that $\tilde{a}_i = a_i + v_i$.

Then, the measurement scheme using wireless accelerometers to obtain the structural connection parameters for the multi-rigid-body motion system can be shown in Fig.2. It is clear that the data processing method will be the core of the scheme, the aim of which is to obtain the estimated values of b_i and k_i, denoting as $\hat{b}_i$ and $\hat{k}_i$, if $m_i, m_{i+1}, \cdots, m_n$ are supposed to be known values.

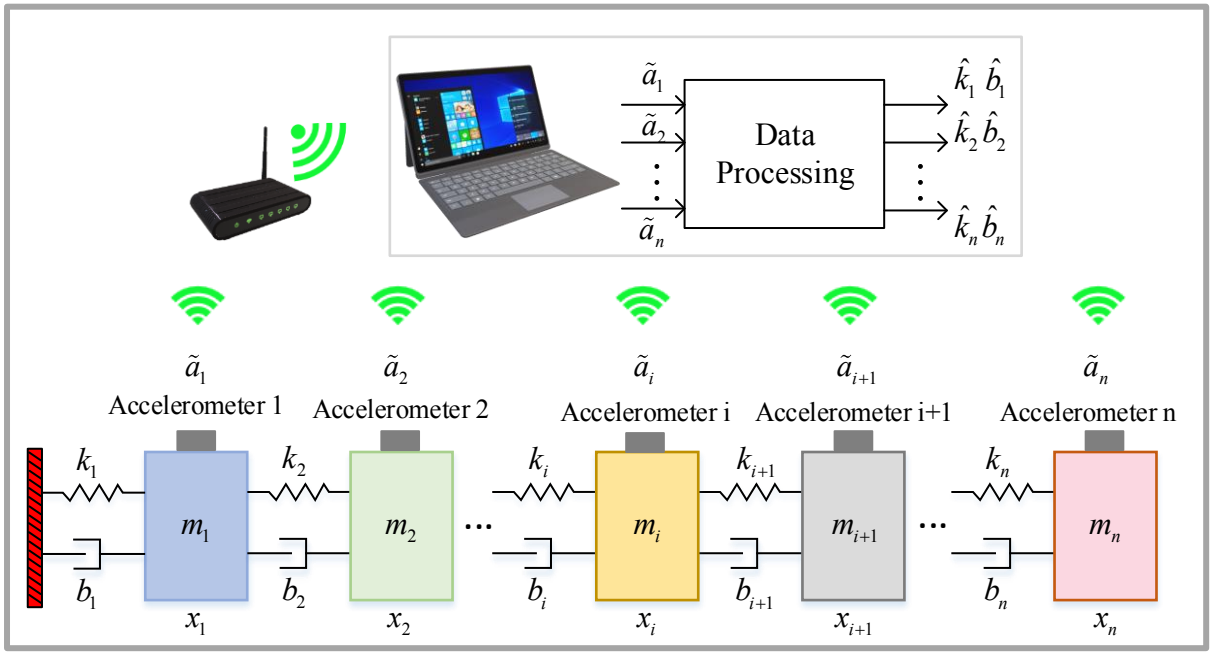


Fig.2 Schematic diagram of n-rigid-body system and its structural connection parameter measurement scheme.

III. DATA PROCESSING METHOD

In this section, the following two candidate methods that can be used for data processing are discussed:

A. Least-square method

Transform Eq.(1) into the following linear regression equation

$$y_i = \boldsymbol{\varphi}_i \boldsymbol{\theta}_i \quad (3)$$

where

$$y_i = \sum_{j=i}^{n} m_j a_j$$

$$\boldsymbol{\varphi}_i = [\int a_{i-1} - \int a_i \quad \int\int a_{i-1} - \int\int a_i]$$

$$\boldsymbol{\theta}_i = \begin{bmatrix} b_i \\ k_i \end{bmatrix}$$

Define the following vectors and matrices

$$\boldsymbol{Y}_i = [y_i(t_1) \quad \cdots \quad y_i(t_N)]^T$$

$$\boldsymbol{\Phi}_i = [\boldsymbol{\varphi}_i^T(t_1) \quad \cdots \quad \boldsymbol{\varphi}_i^T(t_N)]^T$$

where $y_i(t_1)$, $\cdots$, $y_i(t_N)$ represent the N sample values of the signal y_i at time instants t_1, $\cdots$, t_N, where $t_i = (i-1)T_s$, T_s is sampling period, $i = 1, \cdots, N$.

Then, Eq.(3) can be transformed into that

$$\boldsymbol{Y}_i = \boldsymbol{\Phi}_i \boldsymbol{\theta}_i \quad (4)$$

Using the least-square method, it can be obtained that

$$\boldsymbol{\theta}_i = (\boldsymbol{\Phi}_i^T \boldsymbol{\Phi}_i)^{-1} \boldsymbol{\Phi}_i^T \boldsymbol{Y}_i$$

Considering the noise effect, the following equation can provide the estimate of $\boldsymbol{\theta}_i$:

$$\hat{\boldsymbol{\theta}}_i = \left(\tilde{\boldsymbol{\Phi}}_i^T \tilde{\boldsymbol{\Phi}}_i\right)^{-1} \tilde{\boldsymbol{\Phi}}_i^T \tilde{\boldsymbol{Y}}_i \quad (5)$$

where $\tilde{\boldsymbol{\Phi}}_i$ and $\tilde{\boldsymbol{Y}}_i$ are constructed with the measurement values of the accelerometers.

B. Frequency-domain identification method

Transform Eq.(1) into the following equation:

$$m_i(\ddot{x}_i - \ddot{x}_{i-1}) + b_i(\dot{x}_i - \dot{x}_{i-1}) + k_i(x_i - x_{i-1}) = -\sum_{j=i+1}^{n} m_j \ddot{x}_j - m_i \ddot{x}_{i-1}$$

Define $z_i = \ddot{x}_i - \ddot{x}_{i-1}$, $r_i = \sum_{j=i+1}^{n} \frac{m_j}{m_i} \ddot{x}_j + \ddot{x}_{i-1}$, both of which are linear combinations of accelerations. The above equation can be rewritten as

$$m_i \ddot{z}_i + b_i \dot{z}_i + k_i z_i = -m_i \ddot{r}_i \quad (6)$$

Denote $Z_i(s)$, $R_i(s)$ as the Laplace transformations of z_i and r_i, respectively. Then, Eq.(6) further gives

$$\frac{Z_i(s)}{R_i(s)} = \frac{-m_i s^2}{m_i s^2 + b_i s + k_i}$$

Let $\tilde{z}$ and $\tilde{r}$ be the measured values of z_i and r_i that contain noises. Using the Fourier transformations of $\tilde{z}_i$ and $\tilde{r}_i$ which are denoted as $\tilde{Z}_i(j\omega)$ and $\tilde{R}_i(j\omega)$, the frequency response $\frac{\tilde{Z}_i(j\omega)}{\tilde{R}_i(j\omega)}$ can be calculated. By fitting the measured frequency response with model structure shown as the right side of Eq.(7), $\hat{k}_i$ and $\hat{b}_i$ can be obtained.

$$\frac{\tilde{Z}_i(j\omega)}{\tilde{R}_i(j\omega)} = \frac{m_i\omega^2}{\hat{k}_i - m_i\omega^2 + \hat{b}_i\omega j} \quad (7)$$

Remark: Both methods are presented with the example of how to obtain $\hat{k}_i$ and $\hat{b}_i$ considering Eq.(1). Actually, the methods can be applied to all equations of Eq.(2) to obtain all structural connection parameters. Considering the two methods, the drawback of the least-square method is easy to be discovered, that is $\tilde{\boldsymbol{\Phi}}_i$ is constructed by using the integration and double integration of the measured acceleration values which are corrupted with the noises. As N becomes a large value, the accumulation effect of the noise will become significant and degrade the accuracy of the estimation results.

IV. SIMULATION ANALYSIS

In this section, the two data processing methods will be discussed further via simulation analysis.

A. Simulation setup

In the simulation, the three-rigid-body motion system is considered as an example, that is $n = 3$, where the displacement of the second rigid body can be measured in real applications using the displacement sensor such as the grating ruler or the laser interferometer. By using the feedback control, the control force can be obtained and imposed on the first rigid body with the actuator such as the linear motor or the voice coil motor. For the tracking task, the fourth-order S-curve can be adopted as the reference trajectory. The simulation setup can be seen from Fig.3, where the three accelerometers are corrupted with white Gaussian noise with the variance of 0.01m/sec^2. The mass, stiffness and damping parameters used in the simulation are shown in Table 1. Since the parameters associated with the controller and the reference trajectory are not related with the main topic of this paper, they are omitted here. Finally, it should be noted that the sampling frequency of both the accelerometers and the controller is 5kHz.

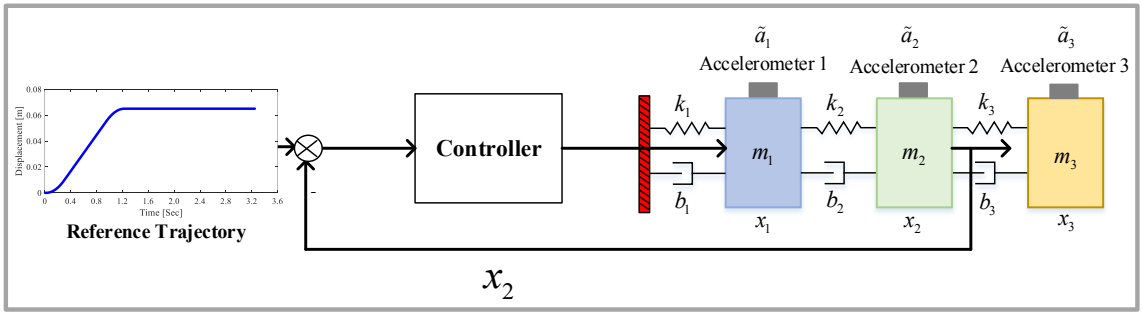


Fig.3 Schematic diagram of simulation setup.

B. Simulation results

First, the least-square method is tested, where the results can be shown in Fig.4. It can be seen that the estimated values have large uncertainties. This is primarily because that $\tilde{\boldsymbol{\Phi}}_i$ is constructed by using the integration and double integration of the measured acceleration values which are corrupted with the noises. As N becomes a large value, the accumulation effect of the noise will become significant and degrade the accuracy of the estimation results. This can be confirmed by comparing the results of $N = 500$ and $N = 100$. The estimated values for $N = 100$ has clearly lower uncertainty.

Second, the frequency-domain method is tested, where the results can be shown in Fig.5. Although the measured frequency responses $\frac{\tilde{Z}_i(j\omega)}{\tilde{R}_i(j\omega)}$ also present uncertainties, by manually fitting one frequency response with the model $\frac{m_i\omega^2}{\hat{k}_i - m_i\omega^2 + \hat{b}_i\omega j}$, it is feasible to obtain very accurate

estimated values. Of course, how to achieve high-accuracy estimated values from the measured frequency responses with the automatic software requires further investigation.

Table 1. Simulation Parameters.

	mass [kg]	stiffness [N· **m**]	damping [N· **sec/m**]
1st rigid body	$m_1 = 50$	$k_1 = 0$	$b_1 = 0$
2nd rigid body	$m_2 = 10$	$k_2 = 1.8950 \times 10^7$	$b_2 = 9.4748 \times 10^3$
3rd rigid body	$m_3 = 2$	$k_3 = 7.8957 \times 10^5$	$b_3 = 63.1655$

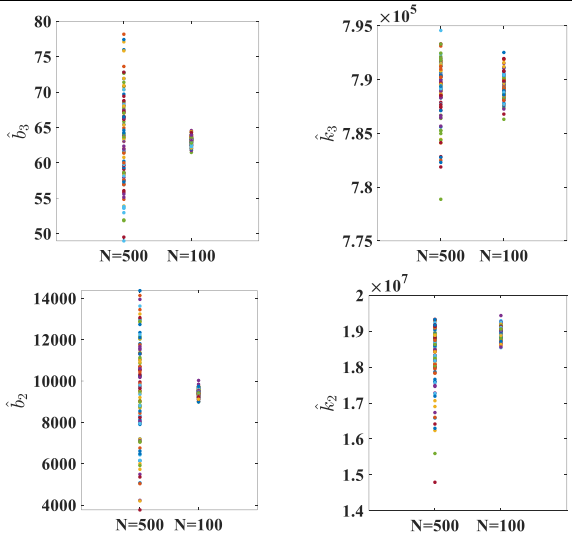


Fig.4 Parameter estimation results (100 times) with least-square method.

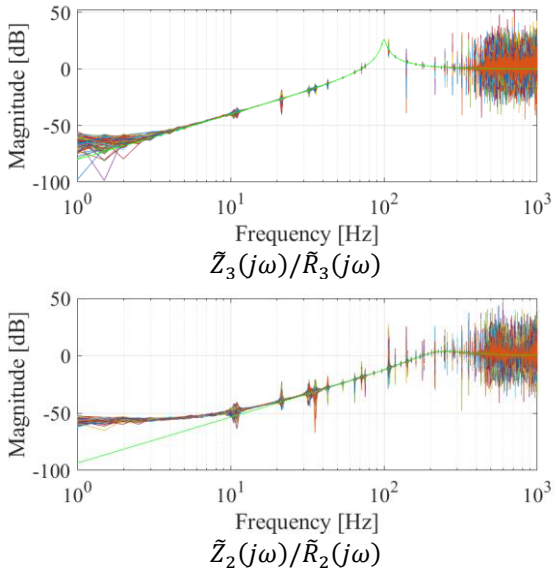


Fig.5 Measured frequency responses (100 times) compared to real frequency responses (green).

V. PRELIMINARY EXPERIMENTAL RESULTS

The 1st version of the wireless device has been developed, where the hardware components can be seen in Fig.6 and the system architecture is described as Fig.7.

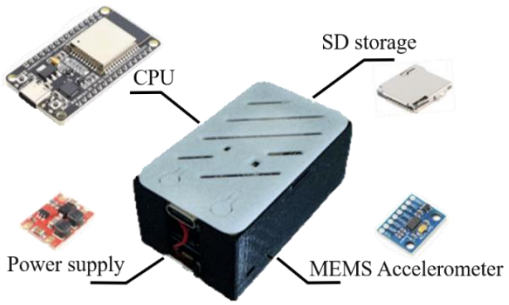


Fig.6 Hardware components of 1st version wireless device.

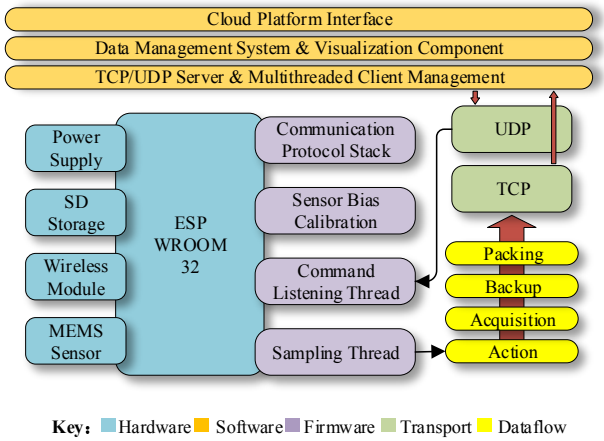


Fig.7 System architecture of 1st version wireless device.

Although there are still serious performance insufficiencies in the 1st version of the device, such as the low sampling frequency (100Hz) and the low resolution, the preliminary function testing has been conducted on a two-mass motion system shown in Fig.8. Let the 1st rigid body track a S-curve, the measurement values of the two accelerometers and the measured frequency response $\tilde{Z}_2(j\omega)/\tilde{R}_2(j\omega)$ are presented in Fig.9. The basic functional requirements have been satisfied, the development of the hardware in the following versions should consider the improvement of the sampling frequency (up to 5KHz) and the resolution.

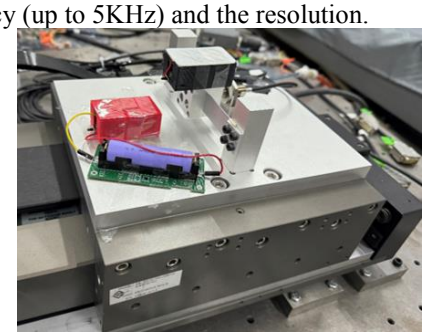


Fig.8 Preliminary function testing on a two-mass motion system.

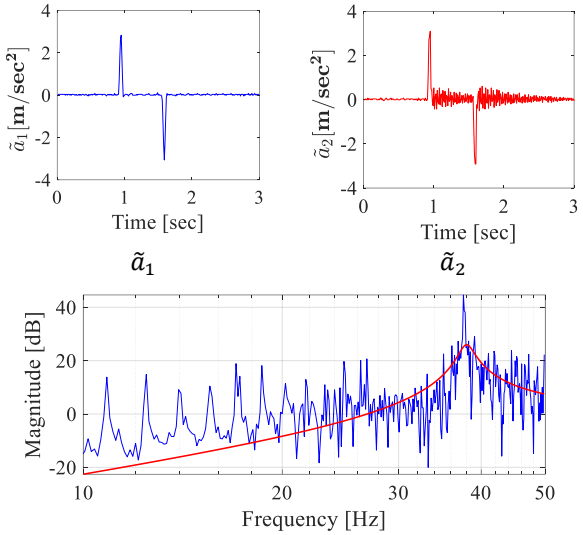


Fig.9 Preliminary experimental results

VI. CONSLUSIONS

In this paper, considering the measurement problem for the structural connection parameters of the multi-rigid-body motion system, this paper proposes a measurement scheme using multiple wireless accelerometers and discusses the candidate data processing methods. As a conclusion, the frequency-domain identification method is more preferable than the least-square method for real applications with sensor noises, since the integration of the measurement values of the accelerometers required in the least-square method causes the significant noise accumulation effect.

Through the verification on the first-version device developed recently, it has been identified that the sampling frequency and the resolution are two major factors for the single device. In addition, the lightweight design of the device and the calibration of the used accelerometers should further be considered [9]. For networked measurement for the structural connection parameters, the time synchronization of data is also crucial, which should be seriously considered in the future [10]. Finally, the uncertainty evaluation method of the frequency-domain identification method should also be studied by referencing the existing beneficial results [11-12].

REFERENCES

[1] T.Oomen, “Advanced motion control for next-generation precision mechatronics: Challenges for control, identification, and learning”, Proc. of 3rd IEEJ International Workshop on Sensing, Actuation, Motion Control, and Optimization, 2017, pp.1-12.

[2] M.F.Heertjes, H.Butler, N.J.Dirkx, S.H.van der Meulen, et al., “Control of wafer scanners: Methods and developments”, Proc. of 2020 American Control Conference, 2020, pp.3686-3703.

[3] K.Classens, W.M.Heemels, T.Oomen, “Digital twins in mechatronics: From model-based control to predictive maintenance”, Proc. of 2021 IEEE 1st International Conference on Digital Twins and Parallel Intelligence, 2021, pp.336-339

[4] K.Classens, M.Mostard, J.Van De Wijdeven, W.M.Heemels, et al., “Fault detection for precision mechatronics: Online estimation of mechanical resonances”, IFAC-PapersOnLine, vol.55, No.37, 2022, pp.746-751.

[5] N.Wengert, M.Nefzi, P.Eberhard, B.Geuppert, “Dynamics in lithographic projection objectives”, Multibody Syst. Dyn., vol.30, February 2013, pp. 233-245.

[6] Y.M.Al-Rawashdeh, M.Al Janaideh, M.Heertjes, “On characterization of a generic lithography machine in a multi-directional space”, Mech. Mach. Theory, vol.170, April 2022, pp.104638.

[7] Y. M. Al-Rawashdeh, M. Al Janaideh, M. F. Heertjes, “A suppress-excite approach for online trajectory generation of uncertain motion systems”, Mech. Syst. Signal Pr., vol. 186, March 2023, pp. 109769.

[8] G.Maculotti, M.Marschall, G.Kok, B.A.Chekh, et al., “A Shared Metrological Framework for Trustworthy Virtual Experiments and Digital Twins”, Metrology, vol.4, No.3, July 2024, pp.337-363.

[9] T.Bruns, H.Volker, S.Eichstädt, “Using opensource software tools for data analysis in high intensity shock calibration of accelerometers”, Proc. of IMEKO 23rd TC3, 13th TC5 and 4th TC22 International Conference, 2017.

[10] S.Eichstädt, A.P.Vedurmudi, M.Gruber, D.Hutzschenreuter, “Fundamental aspects in sensor network metrology”, Acta IMEKO, vol.12, No.1, March 2023, pp.1-6.

[11] S.Eichstädt, C.Elster, I.M.Smith, T.J.Esward, “Evaluation of dynamic measurement uncertainty–an open-source software package to bridge theory and practice”, J. Sens. Sens. Syst., vol.6, No.1, February 2017, pp.97-105.

[12] T.Esward, S.Eichstädt, I.Smith, T.Bruns, et al., “Estimating dynamic mechanical quantities and their associated uncertainties: application guidance”, Metrologia, vol.56, No.1, November 2018, pp.015002.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Interoperability of DCCs for weights and weighing instruments through standardization

Julian Haller[1], Gisa Foyer[2]

[1] *Sartorius Lab Instruments GmbH & Co. KG, Göttingen, Germany, julian.haller@sartorius.com*
[2] *Physikalisch-Technische Bundesanstalt, Braunschweig, Germany, gisa.foyer@ptb.de*

***Abstract* – This paper presents various efforts undertaken to standardize DCCs for weights and weighing instruments (balances) that aim to ensure interoperability between DCCs from different issuers. It is shown that, in addition to conventions regarding the structuring of information, "mini-taxonomies" and clear, unambiguous definitions for the identifiers of certain information are essential. Finally, it will be shown how "interoperability challenges" can be used to check the interoperability between DCCs from different issuers on the one hand and the unambiguity and clarity of the conventions on the other.**

I. INTRODUCTION

The digital transformation of calibration processes has necessitated the development of Digital Calibration Certificates (DCCs) to facilitate machine interpretability and interoperability of calibration data. The digitalization of calibration processes is not merely a technological advancement but a necessity driven by the need for efficiency, accuracy, and traceability. As calibration laboratories transition from traditional paper-based certificates to digital formats, the importance of standardization becomes evident. The Physikalisch-Technische Bundesanstalt (PTB) has spearheaded efforts to create a flexible DCC schema [1], which serves as a foundational framework for digital calibration documentation. The schema has been continuously developed for improvement and the current version 3.3.0 [2] is believed to be stable for a longer time except for minor improvements. However, the flexibility of this schema introduces potential problems concerning interoperability, as varied implementations may not seamlessly integrate across different laboratories.

The need for further harmonization with regard to the specific application of the schema for certain applications was therefore recognized and identified by the industry already at an early stage [3] - only through appropriate harmonization can it be ensured that no proprietary solutions are formed, i.e. that DCCs from different laboratories are interoperable in a way that importing software works regardless of who issued the DCC.

Fortunately, the “Deutscher Kalibrierdienst” (DKD, German calibration service) has taken on this task, so that various measures for the harmonization of DCCs were initiated in a committee consisting of a large number of accredited calibration laboratories from a wide range of branches under the leadership of PTB.

II. HARMONIZATION EFFORTS

To address the above-mentioned challenges, several harmonization approaches are currently undertaken in committees on different levels – an overview can be found in [4].

Within the DKD, several expert groups were formed – on the one hand, specific groups were formed by the several technical sections of the DKD to address specific harmonization topics for their specific community and additionally, a cross-sectional group was formed to address general harmonization topics.

The expert groups for “mass” and “weighing instruments” were two of the first respective groups formed and have undertaken several harmonization efforts that will be explained in detail in the next sections.

Most of these measures have been published in corresponding "DKD Expert Reports" and can be read in detail there [5-6].

A. Structural conventions

The first considerations concern structural conventions, i.e. how and for which results exactly the different elements available in the DCC schema should be used. This is essential for ensuring that DCCs are created in a consistent and standardized manner. The DKD expert reports provide detailed instructions on the use of the DCC schema.

For “mass”, the following general structural conventions have been formulated for weights and weight sets [5]:

- Each weight of a set is identified in a respective dcc:item element.
- Identification properties of the whole set can be given in the parent dcc:items element.

- For each weight and each measurement, a separate dcc:measurementResult is used to represent the measurement results.
- Therein, dedicated dcc:result elements can be used to represent the measurands like volume, density or mass and conventional mass.

For “weighing instruments”, similar structural conventions have been made as follows [6]:

- Each weighing instrument is identified in a respective dcc:item element of “first order” (see next point for explanation).
- Each range of a weighing instrument is identified in a respective dcc:item element of “second order”, i.e. as a child element of a dcc:subitems element being a child element to the “first order” dcc:item element.
- The same applies for modular parts (e.g. indicator, platform) of a weighing instrument – if they are to be identified dedicatedly, this should also happen in a respective dcc:item element of “second order”.
- For each range, a separate dcc:measurementResult is used to represent the measurement results.
- Therein, dedicated dcc:result elements should be used to represent different results of the individual measurements of repeatability, eccentricity and error of indication.

It should be noted that generally different ways of structuring the results would be possible (see [7] for descriptions of such different ways for weights), but these ways of structuring were chosen by the respective expert groups as the best mixture of high flexibility and low redundancy. By adhering to these structural conventions, calibration laboratories can ensure that their DCCs are compatible with other systems and platforms, facilitating seamless data exchange and integration.

B. Identifiers with clear definitions

Basically, specific information can be found via the known path to the corresponding element containing the information. However, at least if several elements of the same name exist at the same level and they then have the same path, it makes sense to explicitly label specific information. Examples are all those cases, where a “singular element” can be existing several times within a “plural element”, e.g. dcc:result in dcc:results or dcc:identification in dcc:identifications. Such explicit labelling is done via refType attributes that can be assigned to an element, e.g.: <dcc:result refType=”NAWI_repeatabilityMeasurement”>, which is clearly distinguishable from: <dcc:result refType=”NAWI_eccentricityMeasurement”>. As can be seen from the example, these refType attributes comprise specific prefixes denoting, whether they are considered generally applicable for all types of DCCs (via the prefix basic_) or only for specific communities (e.g. via the prefix mass_ for weights/weight sets or NAWI_ for non-automatic weighing instruments). Although usually the “suffix” (e.g. repeatabilityMeasurement) is given a descriptive name, it is nevertheless essential that for each harmonized refType, a clear definition (and ideally a bibliographic reference and rules/guidance on the usage) is given to ensure that every user can look up somewhere, what exactly can be expected in an element carrying a particular refType attribute.

In both DKD expert reports, these harmonized names are given along with clear definitions for a certain number of refType attributes (namely 55 for “mass” and 25 for “weighing instruments). They are also available in a TemaTres database [8] so that it is possible to address them in respective software.

The importance of clear definitions and standardized identifiers cannot be overstated. These elements are essential for ensuring that the data within DCCs is interpretable and usable across different laboratories. By providing clear definitions and structured identifiers, the DKD aims to improve data management and retrieval, ensuring that DCCs are compatible and interoperable. This approach helps to prevent misunderstandings and misinterpretations, ensuring that specific data is findable and interoperable.

One example of a refType and its definition is given below for the refType basic_measuredValue that is intended to distinguish (and make identifiable) a dcc:quantity element that represents a measured value from other types of quantities, like for a example a reference value or a nominal value:

Table 1 Example of a refType description (from [8])

Name	basic_measuredValue
Definition	quantity value representing a measurement result
Bibliographic source	JCGM 200:2012 (VIM) 2.10
Example	<dcc:result> <dcc:name> <dcc:content lang="en">Measuring result</dcc:content> </dcc:name> <dcc:data> <dcc:quantity refType="basic_measuredValue"> <dcc:name> <dcc:content lang="en">mass</dcc:content>

	</dcc:name> <si:real> <si:value>0.999997191</si:value> <si:unit>\kilogram</si:unit> <si:measurementUncertaintyUnivariate> <si:expandedMU> <si:valueExpandedMU>0.00000053</si:valueExpandedMU> <si:coverageFactor>2</si:coverageFactor> <si:coverageProbability>0.95</si:coverageProbability> </si:expandedMU> </si:measurementUncertaintyUnivariate> </si:real> </dcc:quantity> </dcc:data> </dcc:result>
Element	dcc:quantity

C. Descriptions for particular use cases

The standardization efforts include detailed descriptions of particular use cases, providing guidance on the implementation and use of the DCC schema. These use cases serve as examples of how the DCC schema can be adapted to various calibration scenarios, highlighting the flexibility and versatility of the schema. By providing clear guidelines and best practices, the DKD aims to ensure that DCCs are compatible and interoperable across different laboratories.

The use cases described in the two expert reports comprise (among others)

- Interventions like adjustment, repair or exchange
- Decision rules and conformity statements
- Marking of results outside a laboratory's accreditation scope

D. "Mini-taxonomies"

Some elements of the DCC schema allow referencing entries of taxonomies or databases, e.g. the element dcc:equipmentClass with child elements dcc:reference (to denote a reference database or taxonomy), dcc:classID (to denote the entry ID within the database or taxonomy) and dcc:link (to provide a direct link to the entry, if available). A taxonomy generally holds information for a categorization whereas databases supply lists of entries.

However, not for all desired ways of categorizing there might exist appropriate databases or taxonomies yet. For example, it is common to categorize weighing instruments into "single range", "multiple range" and "multi-interval" instruments, but there is no well-known reference for this categorization. A harmonized categorization however is necessary in order to avoid that entries from different laboratories meant to be the same are interpreted as different by machines due to different spellings (e.g. "single-range", "single range", "singlerange") or languages. Therefore, in the DKD report for weighing instruments [6], it was simply decided to directly provide a "mini-taxonomy" with 3 entries to be used in the dcc:classID element:

- NAWI-SR: single range instrument, having one scale interval d and one maximum capacity Max
- NAWI-MR: multiple range instrument with $i = 2 \ldots n$ ranges, each extending from 0 to Max_i with a respective scale interval d_i
- NAWI-MI: multi-interval instrument with $i = 2 \ldots n$ partial ranges, each one extending from Max_{i-1} (or 0 in case of $i = 2$) to Max_i with a respective scale interval d_i

With such, "mini-taxonomies" it is possible to harmonize entries in those elements that allow referencing databases or taxonomies and it as well helps to ensure that DCCs are interpretable and usable across different laboratories.

E. Example DCCs

Perhaps most helpful for beginners in the topic of DCCs are the sample DCCs that are included as attachments in both expert reports. In principle, it is possible to take these examples and adapt them to your own case.

The provision of example DCCs is an important aspect of the standardization efforts. They serve as valuable resources for calibration laboratories, providing practical guidance on the creation and use of DCCs. However, it should be mentioned that examples alone are typically ambiguous - i.e. there should always be an additional description (e.g. in the form of an expert report) so that an example is not misinterpreted in ambiguous cases or regarded as the only solution.

In the two expert reports, the following example DCCs are given as Annexes:

Mass:

- DCC for a single weight
- DCC for a set consisting of two weights

- DCC of a mass standard

Weighing instruments:

- DCC for a single range balance with two measurements (before/after adjustment)
- DCC for a multiple range balance
- DCC for a multi-interval balance consisting of two identifiable parts (indicator and platform)

III. TESTING INTEROPERABILITY

A. Background

It should be highlighted once again that interoperability must be a key focus concerning DCC: If users cannot rely on the fact that their import software, which works with DCCs from laboratory A, will also work with DCCs from laboratory B, any implementation of DCC usage would mean tying oneself to a specific laboratory - in the view of the authors, this would not only contradict the idea of the DCC, but would probably also prevent its widespread use.

The above-mentioned harmonization efforts and accordingly the respective expert reports are intended to ensure that such a scenario must not be feared by users. The intention clearly is that DCC users can reliably assume full interoperability of different laboratories as long as they all issue their DCCs in compliance with the expert reports.

In order to justify such an assumption, two questions have to be answered:

- Is the description in the expert reports complete, meaning that there are no cases where missing guidance or rules leads to different implementations concerning relevant information?
- Is the description in the expert reports sufficiently unambiguous so that different readers will not differently implement anything concerning relevant information?

B. Means for checking interoperability

In metrology the standard means of checking whether realizations fit together are comparison measurements. They basically ensure “interoperability” of measurement results. This is not transferable to the DCC, but the DCC is XML based. This enables some general means of validation:

Syntax of XML [9]: XML has syntax rules such as “All elements have to be opened and closed in the right order.” Which can be checked in most XML tools.

Example: <dcc:items> <dcc:item> </dcc:item> </dcc:items>.

Validation against schema file: The XML schema file in XSD format (here dcc.xsd [2]) can be used to check for general rules [10]. The applied XSD file is given in each XML file. The rules include the structure but also cardinality and data types.

Example: <xs:element name="item“ type= "dcc:itemType“ maxOccurs="unbounded"/> means that the DCC must have at least one “dcc:item” which must be of the “itemType”.

Validation against rules using Schematron: A Schematron file can be used to define rules as well as errors and warnings which are more flexible than the Schema file. This way, plausibility of inputs can be checked as well as additional structural rules that are not part of the XSD. Schematron is a standard XML means and has already been tested with the DCC [11].

Example: <sch:assert role="error" test="@id"> Id is missing </sch:assert> tests for an id e.g. at dcc:item which is expected if using the instructions from [5].

Using a combination of all these means ensures in general the highest level of interoperability (see Fig. 1).

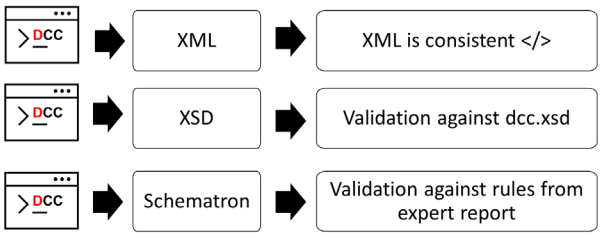


Fig. 1 Means for checking interoperability in XML in general and for the Interoperability Challenge in specific.

There is one other means which is less a formal checking tool but a means of showing information which can also be used:

Extraction and display of information using XSLT [12]: Using an XSLT file to ensure a harmonized human-readable

version can give a good idea about the overall interoperability for a human reader as missing information is easy to spot.

C. Interoperability Challenge

To test whether the expert report [4] actually leads to an interoperable DCC, a so-called interoperability challenge for mass DCCs was recently started, where a number of interested laboratories (comprising both NMIs and accredited calibration laboratories) formed three groups:

- Group 1 will provide 3 calibration scenarios, i.e. a description of calibration items and their calibration results, along with some additional information, e.g. on influence conditions
- Group 2 will create DCCs for these calibration scenarios and submit them
- Group 3 will analyze the results to determine if all implementations are the same or if not, if this is due to unclear/missing information in the expert report. Fig. 1 gives an overview of the planned interoperability checks. Furthermore, XSLT will be used to create a human readable result and further analyze the content of the created DCCs.

This interoperability challenge is currently in progress and results are expected in autumn 2025. It is believed that such interoperability challenges (also for other types of calibration objects) in the future might become common as they not only help to identify possible “authors’ white spots” concerning the harmonization conventions in expert reports, but a successful participation in such an interoperability challenge could also serve as some kind of “proof”. Just like the validity of a laboratory’s results can be controlled/proven via key comparisons, the interoperability of the DCCs issued by a laboratory could in the future be controlled/proven via such interoperability challenges.

IV. CONCLUSIONS

In conclusion, the digital transformation of calibration processes through the implementation of Digital Calibration Certificates (DCCs) marks a pivotal shift in the calibration industry, driven by the necessity for enhanced efficiency, accuracy, and traceability in data management. The transition from traditional paper-based certificates to digital formats is not merely a technological advancement but a fundamental requirement to meet the evolving demands of modern calibration laboratories. The efforts spearheaded by the Physikalisch-Technische Bundesanstalt (PTB) and the Deutscher Kalibrierdienst (DKD) have been instrumental in developing a robust and standardized framework that addresses the critical challenges of interoperability and harmonization across diverse laboratories.

The establishment of expert groups within the DKD has been a strategic move to tackle specific harmonization topics, ensuring that the DCC schema is applied consistently across various technical sections. The publication of DKD Expert Reports provides comprehensive guidelines and structural conventions, which are essential for creating DCCs in a standardized manner. These reports offer detailed instructions on the use of the DCC schema, ensuring that calibration laboratories can produce certificates that are compatible with other systems and platforms, thereby facilitating seamless data exchange and integration.

Moreover, the introduction of identifiers with clear definitions and structured naming conventions plays a crucial role in enhancing the interpretability and usability of DCCs. By providing explicit labels and harmonized refType attributes, the DKD aims to improve data management and retrieval, ensuring that DCCs are interoperable across different laboratories. This approach helps prevent misunderstandings and misinterpretations, making specific data findable and usable across various applications.

The standardization efforts also include detailed descriptions of particular use cases, offering practical guidance on the implementation and adaptation of the DCC schema to various calibration scenarios. These use cases highlight the flexibility and versatility of the schema, providing clear guidelines and best practices to ensure compatibility and interoperability across different laboratories. Additionally, the provision of example DCCs serves as valuable resources for calibration laboratories, offering practical insights into the creation and use of DCCs.

The concept of "mini-taxonomies" further contributes to the harmonization efforts by providing structured categorization for elements within the DCC schema. They help ensure that entries are interpretable and usable across different laboratories, avoiding discrepancies due to varied spellings or languages. By referencing these taxonomies, calibration laboratories can maintain consistency and reliability in their digital documentation. However, a more comprehensive taxonomy for measuring instruments and calibration terms is needed. One promising, yet not complete approach is the Measurement Information Infrastructure (MII) that has been set up and is still expanded by NCSLI [13].

Finally, the ongoing interoperability challenge underscores the importance of ensuring that DCCs from different laboratories can be reliably used across diverse applications. This challenge aims to identify potential gaps in the expert reports and ensure that the descriptions are complete and unambiguous, promoting widespread adoption and trust in digital calibration documentation. As industry continues to evolve, these harmonization efforts will play a crucial role in advancing the digitalization of calibration processes, paving the way for future innovations and improvements in the field. Through these collaborative efforts, the calibration industry is poised to embrace the full potential of digital

transformation, ensuring that DCCs become a cornerstone of modern calibration practices. Hopefully, it will also support the global effort in harmonizing DCCs.

V. OUTLOOK

It is important to point out that all of the harmonization efforts mentioned are voluntary – laboratories may or may not adhere to them. Nevertheless, it is of course hoped that as many laboratories as possible will adhere to one (or at least only a few) harmonization concepts in order to keep the variety of variants and thus the maintenance efforts small for both issuers and recipients of DCCs.

Perhaps in the future it will become an integral part of contract review for the customer and laboratory to agree on a certain harmonization before the calibration. Successful participation in an interoperability challenge could then serve as a proof of competence for a laboratory that it really does issue its DCCs according to this harmonization concept.

REFERENCES

[1] S. Hackel, S. Schönhals, L. Doering, T. Engel, R. Baumfalk, “The Digital Calibration Certificate (DCC) for an End-to-End Digital Quality Infrastructure for Industry 4.0”, Sci 2023, 5(1), 11; https://doi.org/10.3390/sci5010011

[2] https://www.ptb.de/dcc/dcc.xsd, accessed on 16.04.2025

[3] K. Banholzer, C. Müller-Schöll, J. Haller, „Need from industry on the DCC concerning harmonization and a regulating structure“, 2nd international DCC-Conference 01 - 03 March 2022 Proceedings, 56-62, 2022; http://dx.doi.org/10.7795/820.20220411

[4] S. Schönhals, M. Demir, L. Doering, B. Gloger, S. Hackel, W. Heeren, J. Jagieniak, M. Jordan, C. Keilholz, T. Krah, J. Loewe, K. Mienert, G. Öktem, „Harmonisation processes and practical implementation of machine-interpretable digital calibration certificates“, Measurement: Sensors 2025, 101470, https://doi.org/10.1016/j.measen.2024.101470

[5] G. Foyer, J. Haller, C. Müller-Schöll, S. Osang, A. Scheibner, „Expert report DKD-E 7-2 Instructions on how to use the DCC schema to create a digital calibration certificate for weights”, Edition 04/2022, Revision 0; http://dx.doi.org/10.7795/550.20220419B

[6] J. Haller, G. Foyer, K. Fritsch, S. Hackel, M. Häfner, D. Knopf, H. Lauer, C. Müller-Schöll, S. Osang, A. Scheibner, Z. Zelenka, „ Expert report DKD-E 7-3 Instructions on how to use the DCC schema to create a digital calibration certificate for non-automatic weighing instruments”, Physikalisch-Technische Bundesanstalt, Braunschweig and Berlin. Edition 03/2025, Revision 0: https://doi.org/10.7795/550.20250325

[7] G. Foyer, J. Haller, C. Müller-Schöll, „Data structures in calibrations of weights and mass standards”, Measurement: Sensors 2024, 101468, https://doi.org/10.1016/j.measen.2024.101468

[8] Metrology refType Database: https://digilab.ptb.de/dkd/refType/vocab/index.php, accessed on 14.07.2025

[9] W3 schools, „XML Syntax Rules”, https://www.w3schools.com/xml/xml_syntax.asp , accessed on 14.07.2025.

[10] W3C, “W3C XML Schema Definition Language (XSD) 1.1 Part 1: Structures”, https://www.w3.org/TR/2012/REC-xmlschema11-1-20120405/, accessed on 14.07.2025

[11] G. Söylev-Öktem, S. Hackel, J. Jagieniak, B. Gloger, L. Doering, “Using Schematron to verify DCCs”, 2nd International DCC Conference, https://oar.ptb.de/resources/show/10.7795/820.20220411.

[12] W3 schools, “XSLT Introduction”, https://www.w3schools.com/xml/xsl_intro.asp, accessed on 14.07.2025.

[13] NCSLI MII: https://github.com/NCSLI-MII, accessed on 14.07.2025

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Towards Seamless Digitalisation in TUBITAK UME Mass Laboratory

Emrecan Keskin[1], Beste Korutlu[2]

[1] *TUBITAK National Metrology Institute (UME), Kocaeli, Türkiye, email address: emrecan.keskn@gmail.com*
[2] *TUBITAK National Metrology Institute (UME), Kocaeli, Türkiye, email address: beste.korutlu@tubitak.gov.tr*

***Abstract* – We present automatic generation of Digital Calibration Certificate for a set of weights calibrated using a Robotic Mass Comparator operating in the mass range from 1 mg to 10 g. The tool features a Graphical User Interface that streamlines the processing of Digital Calibration Certificate from the automatically extracted and organized data in the calibration reports and the most recent calibration certificates of the reference standards and test weights adhering to relevant international recommendations, standards and guidelines. This user-friendly and highly automated software application represents a substantial advancement in the mass subfield of modern metrology in terms of handling large volumes of calibration data with minimal intervention.**

I. INTRODUCTION

Mass metrology plays an indispensable role for accuracy and fairness in trade, consumer protection, regulatory compliance, safety and quality control, innovation and technical advancement and sustainable practices. Therefore having a framework for Findable, Accessible, Interoperable, and Reusable (FAIR) data in compliance with international standards and regulations is foundational in modern mass metrology. In this context, Digital Calibration Certificate (DCC) provides a sophisticated solution for the global exchange of metrological information in Extensible Markup Language (XML) format [1 - 3]. The GEMIMEG tool is designed as a free-access software application with a user-friendly interface enabling users to create and standardize Digital Calibration Certificates (DCC) without the need for in-depth knowledge on XML [4]. The advantages of using XML are clearly stated in [5]. Accordingly, a fully automated pilot study of an International Virtual Mass Comparison is performed as proof-of-concept where the exchange format of artificial data mimicking a real world comparison scenario is derived from DCC schema [6]. The implementation of DCC in mass metrology requires subject-related specifications since it is developed as a template for all measurands. Instructions on how to effectively generate DCC for weights and weight sets are provided in DKD Expert Report [7]. In this paper, we describe the seamless path in generating DCC for a test weight set calibrated in TÜBİTAK UME Mass Laboratory aligned with the DKD Expert Report, the Physikalisch-Technische Bundesanstalt (PTB) DCC Schema in version 3.2.1 and with the recommendations of the International Organization of Legal Metrology (OIML) [8]. The tool incorporates a Graphical User Interface (GUI) that simplifies the creation of DCC by automatically extracting and sorting the data from the calibration reports and the most recent certificates of the reference standards and test weights. The study is conducted to lay the foundation for the seamless exchange of globally interoperable data in the era of modern metrology.

II. INFRASTRUCTURE

TUBITAK UME Mass Laboratory is equipped with six different automatic weighing systems to provide certification services for mass calibration in the range from 1 mg to 500 kg and with standard weights and weight sets of various OIML accuracy classes. The calibration reports by these automatic weighing systems are exported in different file formats such as Tanned Database (TDB), Portable Document Format (PDF), Microsoft Word Document (DOCX), Microsoft Excel Spreadsheet (XLS). The calibration range between 1 mg to 10 g is performed by a Robotic Mass Comparator (See Fig. 1). In this paper we demonstrate the seamless generation of DCC by calibration of a weight set in the range of 1 mg – 10 g.



Fig. 1. Robotic mass comparator (1 mg – 10 g).

The calibration procedure of a weight set is programmed by the operator via the control terminal of the comparator. Calibration results by this robotic comparator are reported in TDB format in a structured and organized way. The content of the calibration report is divided into three parts as "Administrative Data", "Measurement Results" and "Summary Results" (See Figs. 2 to 4). The "Administrative Data" and the "Summary Results" are used for the automatic generation of DCC. Measurement results contains raw data which is useful for checking outliers in the data set. A separate calibration report in TDB format is generated for each individual weight in the weight set. The database under the main menu stores the information on operators, reference and test weights, calibration procedures and customers. The operator is responsible for creating the database. The stored information in the database is used for regular customers getting periodic calibration services from TUBITAK UME. The database is updated for the new customers.

---------- Comparator ----------	
Operator	OZLEM PEHLIVAN YILDIRIM
Name and Surname	
Report number	C/13/02/24/08/43
Start date	2024.02.13 08:43:48
End date	2024.02.13 09:33:50
Test weight Set of Weights (10 mg - 2 g)	
Order number	2024.00168
Test weight number	TestCompany_159351_2_g_0_TestManufacturer
Test weight class	E1
Test weight position	B1
Reference weight	E0-02-2g
Real mass	0 g
Reference weight class	E0
Reference weight position	A23

Fig. 2. Administrative data in calibration report.

n	|A	|B	|B	|A	|D
1	|-0,0000012	|0,0000081	|0,0000077	|-0,0000028	|0.0000099
2	|-0,0000051	|0,0000062	|0,0000051	|-0,0000054	|0.0000109
3	|-0,0000063	|0,0000041	|0,0000034	|-0,0000058	|0.0000098
4	|-0,0000064	|0,0000034	|0,0000040	|-0,0000069	|0.00001035
5	|-0,0000063	|0,0000028	|0,0000034	|-0,0000077	|0.0000101
6	|-0,0000068	|0,0000044	|0,0000049	|-0,0000052	|0.00001065

Fig. 3. Measurement results in calibration report.

Mean difference	0,000010283333 g
Standard deviation	0,000000432049 g
Cycles quantity	6
Method	ABBA
Min temperature	19.993 °C
Max temperature	20.124 °C
Min humidity	41.26 %
Max humidity	41.38 %
Min pressure	986.066 hPa
Max pressure	986.603 hPa

Fig. 4. Summary results in calibration report.

III. CALIBRATION PROCEDURE

In mass metrology, test weight calibration is typically carried out by substitution weighing utilizing reference standards of higher accuracy class than the test weights. The mass value is assigned by the mean difference in weighing of test weight and the reference standard. The most commonly used weighing cycle is Reference-Test-Test-Reference (often referred as ABBA-cycle where "A" represents weighing of the reference standard and "B" represents weighing of the test weight). The ABBA-cycle is repeated at least by the minimum number specified in the relevant international recommendations, standards and guidelines for the particular accuracy class of the test weight [8]. The weighing difference Δm_{w_i} of ith cycle is determined by

$$\Delta m_{w_i} = \frac{B_{1_i} + B_{2_i}}{2} - \frac{A_{1_i} + A_{2_i}}{2}, \qquad (1)$$

where sub-index $\ll i \gg$ represents the number of cycle and sub-indices $\ll 1 \gg$ and $\ll 2 \gg$ indicate the number of times the reference and test weights are measured within that particular cycle. The average of weighing differences between the test weight and the reference standard for n-cycle is given by (See first line in Fig. 4)

$$\overline{\Delta m}_w = \frac{1}{n}\sum_{i=1}^{n} \Delta m_{w_i}. \qquad (2)$$

The weighing equation for the mass of the test weight m_t is written as

$$m_t = m_r + \rho_a(V_t - V_r) + \overline{\Delta m}_w, \qquad (3)$$

where m_r is the mass of the reference standard, ρ_a is the mean value of the air density, V_t and V_r are the volumes of the test and reference weights at 20 °C. The volume expansion and the center of gravity correction terms are not included in this study. The second term in Eq. 3 is the air buoyancy correction on the test mass. It appears because apart from the gravitational force, the buoyant force from the displacement of the air acts on the weights. CIPM-2007 formula is used for determining the density of moist air [9] from the mean values of air pressure, air temperature and relative humidity (See the rows 5- 10 in Fig. 4). The parameters m_r and V_r are extracted from the most recent certificate of the reference standard and V_t is taken from the latest certificate of the test weight.

For practical reasons, a conventional mass value is assigned to a test weight by [7]

$$m_{ct} = m_{cr} + (\rho_a - \rho_0)(V_t - V_r) + \overline{\Delta m}_w \qquad (4)$$

where the sub-index c represents the conventional mass, $\rho_0 = 1.2$ kg m^{-3} is the reference air density. m_{cr} is extracted from the most recent certificate of the reference standard as well.

In mass metrology, for each weight, the expanded uncertainty U of the conventional mass, shall be less than or equal to one-third of the the Maximum Permissible Error, δm of the corresponding test weight (See Table 1 in [8]). The calibration uncertainty is calculated according to the general rules and principles in "Evaluation of measurement data – Guide to the expression of uncertainty in measurement" document [10]. The general practice in reporting calibration uncertainty U is to quote $\delta m/3$ of the corresponding test weight (in case the calibration uncertainty is lower than the quoted value). The calibration results of the test weights are examined whether they comply with the following rules.

Rule 1. The annual drift of the mass and conventional mass values should be less than the expanded uncertainty.
Rule 2. For each weight the conventional mass should satisfy

$$m_0 - (\delta m - U) \leq m_{ct} \leq m_0 + (\delta m - U). \qquad (5)$$

where m_0 is the nominal mass of the test weight.

IV. DIGITAL CALIBRATION CERTIFICATE

The infrastructure in TUBITAK UME Mass Laboratory allows automatic generation of DCC for a weight set from the calibration reports of automatic weighing systems. The content of this paper is confined to the generation of the DCC by using the calibration reports from Robotic Mass Comparator working in the range of 1 mg – 10 g. The GUI of the DCC generator is given in Fig. 5. It is designed for users with no expertise in XML.

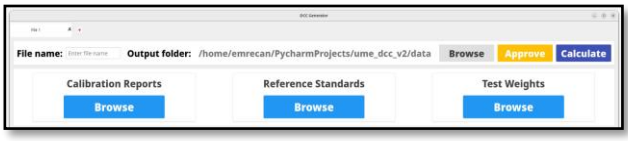


Fig. 5. GUI of automatic DCC generator.

Unified Modeling Language (UML) sequence diagram of the DCC generator is given in Fig. 6.

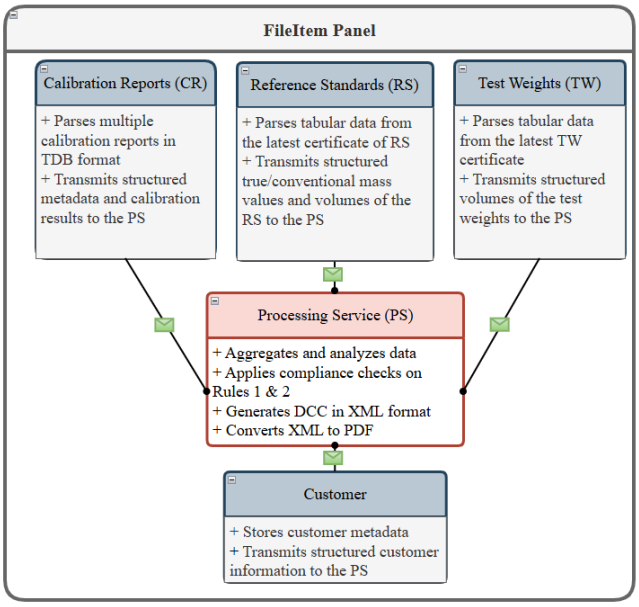


Fig. 6. UML sequence diagram.

FileItem Panel serves as the parent component for both uploads and services in the UML Sequence Diagram. The "Reference Standards" and "Test Weights" buttons are where the most recent calibration certificates of the reference standards and test weights are uploaded by the operator, typically in DOCX or XML formats. Currently, the calibration certificates issued by TUBITAK UME are in DOCX format and distributed to the customer as printed copies. Consequently, an archive of past certificates in DOCX format exists for our regular customers. If the previous certificate of the test weight set was not issued by TUBITAK UME Mass Laboratory, a PDF version of the certificate is provided by the customer; however, the format and the structure may vary significantly across different institutes. In such cases, the GEMIMEG tool is used for manually generating the DCC for the test weights as an intermediate solution [3]. As the DCC archive of certificates will be made available in the future, DCC generation tool currently allows both DOCX and XML formats as an interim solution during the transition phase. The "Calibration Reports" button is where the calibration reports of each individual test weight are uploaded by the operator. The sequence of file uploads is not critical. Notably, there is no communication between the top row buttons in the UML sequence diagram in Fig. 6. These buttons act as intermediaries between the operator and the "Processing Service". Once all the files are uploaded, the corresponding file paths are communicated to the "Processing Service". This service then automatically generates indexed dictionaries of the conventional mass, mass, volume/density data of the reference standards and test weights with their corresponding uncertainties, weighing difference, standard deviation and the environmental conditions. The "Calculate" button within the FileItem Panel triggers the routines in "Processing Service" that utilize these indexed dictionaries as input parameters for Eqs. 3 and 4 for each individual test weight. The calculated conventional mass value of each test weight is subsequently checked to ensure compliance with Rule 1 and Rule 2. Any test weights failing to meet these rules are displayed within the graphical user interface (GUI) as highlighted in red for the operator's attention (See Fig. 7).

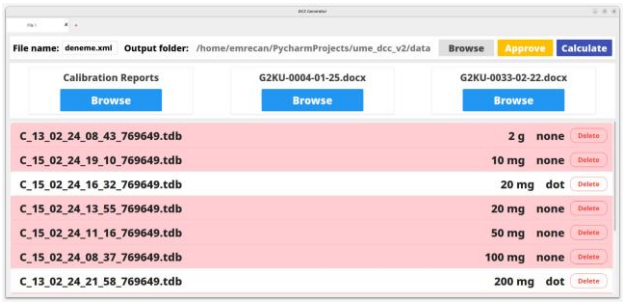


Fig. 7. Failing results summary in GUI.

The operator is then required to recalibrate the relevant test weights using an alternative reference standard and upload the updated files. The "Approve" button initiates the DCC generation routine for the test weight set. The generated DCC is in agreement with the version 3.2.1 of the DCC schema where five child elements dcc:administrativeData, dcc:measurementResults, dcc:comments, dcc:document and ds:Signature are under the main element dcc:digitalCalibrationCertificate [3], except dcc:document and ds:Signature which are presently under construction. The attributes such as *id*, *refType* and *refId* are attached to various elements as they are described in DKD Expert Report [7] for further interoperability of the data. Currently, the DCC is generated only in English. The dcc:measurementResults are partially displayed in Fig. 8, showing the mass value of 2 g with no marking. The number of digits in the quoted values are automatically

adjusted to match the resolution of the robotic mass comparator. The calibration uncertainty is checked to ensure that it is less than $\delta m/3$ of the corresponding test weight. If this condition is met, it is reported as $\delta m/3$ of the corresponding test weight. The conventional mass, volume and density tables are formed in a similar fashion in line with the PTB DCC Schema in version 3.2.1 [3].



```
<dcc:measurementResult refId="weight_159351_E1_2_g_none" refType="basic_isInCMC">
  <dcc:name>
  ...
  </dcc:name>
  <dcc:usedMethods>
  ...
  </dcc:usedMethods>
  <dcc:influenceConditions>
  ...
  </dcc:influenceConditions>
  <dcc:results>
    <dcc:result>
      <dcc:name>
        <dcc:content lang="en">Mass</dcc:content>
      </dcc:name>
      <dcc:data>
        <dcc:quantity refType="basic_nominalValue">
          <dcc:name>
            <dcc:content lang="en">Nominal value</dcc:content>
          </dcc:name>
          <si:real>
            <si:value>2</si:value>
            <si:unit>\gram</si:unit>
          </si:real>
        </dcc:quantity>
        <dcc:quantity refType="basic_measuredValue">
          <si:real>
            <si:value>2.0000091</si:value>
            <si:unit>\gram</si:unit>
            <si:expandedUnc>
              <si:uncertainty>0.0000040</si:uncertainty>
              <si:coverageFactor>2</si:coverageFactor>
              <si:coverageProbability>0.95</si:coverageProbability>
            </si:expandedUnc>
          </si:real>
        </dcc:quantity>
        <dcc:quantity refType="basic_measurementError">
          <si:real>
            <si:value>0.0000091</si:value>
            <si:unit>\gram</si:unit>
            <si:expandedUnc>
              <si:uncertainty>0.0000040</si:uncertainty>
              <si:coverageFactor>2</si:coverageFactor>
              <si:coverageProbability>0.95</si:coverageProbability>
            </si:expandedUnc>
          </si:real>
        </dcc:quantity>
```


Fig.8. Partial measurement results in DCC.

The administrativeData is automatically generated from the information in the calibration reports (See Fig. 9). The Unique Identifier is generated initially as UME G2KU-0000 Month-Year with the date derived from the day the DCC is generated.



```
<dcc:administrativeData>
  <dcc:dccSoftware>
    <dcc:software>
      <dcc:name>
        <dcc:content>UME-DCC</dcc:content>
      </dcc:name>
      <dcc:release>v0.0.1</dcc:release>
    </dcc:software>
  </dcc:dccSoftware>
  <dcc:refTypeDefinitions>
  ...
  </dcc:refTypeDefinitions>
  <dcc:coreData>
    <dcc:countryCodeISO3166_1>TR</dcc:countryCodeISO3166_1>
    <dcc:usedLangCodeISO639_1>en</dcc:usedLangCodeISO639_1>
    <dcc:mandatoryLangCodeISO639_1>en</dcc:mandatoryLangCodeISO639_1>
    <dcc:uniqueIdentifier>UME G2KU-0000 02-24</dcc:uniqueIdentifier>
    <dcc:identifications>
      <dcc:identification refType="basic_orderNo">
        <dcc:issuer>calibrationLaboratory</dcc:issuer>
        <dcc:value>2024.00168</dcc:value>
        <dcc:name>
          <dcc:content lang="en">Order number</dcc:content>
        </dcc:name>
      </dcc:identification>
    </dcc:identifications>
    <dcc:beginPerformanceDate>2024-02-13</dcc:beginPerformanceDate>
    <dcc:endPerformanceDate>2024-02-15</dcc:endPerformanceDate>
    <dcc:performanceLocation>laboratory</dcc:performanceLocation>
    <dcc:issueDate>2024-02-22</dcc:issueDate>
  </dcc:coreData>
```


Fig. 9. The coreData in administrativeData in DCC.

After the approval of the DCC by the operator via the "Approve" button in the FileItem Panel, the four digit 0000 is replaced by the number of the issued certificate for that specific year. The count function for the number of issued certificate is linked with the approval of the operator. The beginPerformanceDate and endPerformanceDate are extracted from the calibration reports of the weight set, representing the earliest and latest calibration dates, respectively. The issueDate is set as the approval date of the DCC by the operator.

The items in the DCC are imported from the information supplied while programming the calibration procedure (See Fig. 10). The input data by the operator on "Test Weight" and "Test Weight Number" are provided systematically so that the information in DCC on devices under test is filled routinely.



```
<dcc:items>
  <dcc:name>
    <dcc:content lang="en">1 OIML set of weights from 10 mg to 2 g</dcc:content>
  </dcc:name>
  <dcc:equipmentClass>
    <dcc:reference>OIML R111-1:2004</dcc:reference>
    <dcc:classID>E1</dcc:classID>
  </dcc:equipmentClass>
  <dcc:identifications>
    <dcc:identification refType="basic_serialNo">
      <dcc:issuer>manufacturer</dcc:issuer>
      <dcc:value>159351</dcc:value>
      <dcc:name>
        <dcc:content lang="en">Serial No.</dcc:content>
      </dcc:name>
    </dcc:identification>
  </dcc:identifications>
  <dcc:item id="weight_159351_E1_10_mg_none">
    <dcc:name>
      <dcc:content lang="en">10 mg</dcc:content>
    </dcc:name>
    <dcc:identifications>
      <dcc:identification refType="mass_shape">
        <dcc:issuer>manufacturer</dcc:issuer>
        <dcc:value>Wire</dcc:value>
        <dcc:name>
          <dcc:content lang="en">Form</dcc:content>
        </dcc:name>
      </dcc:identification>
      <dcc:identification refType="basic_marking">
        <dcc:issuer>manufacturer</dcc:issuer>
        <dcc:value>-</dcc:value>
        <dcc:name>
          <dcc:content lang="en">Marking on weight</dcc:content>
        </dcc:name>
      </dcc:identification>
    </dcc:identifications>
    <dcc:itemQuantities>
      <dcc:itemQuantity refType="basic_nominalValue">
        <dcc:name>
          <dcc:content lang="en">Nominal mass</dcc:content>
        </dcc:name>
        <si:real>
          <si:value>10</si:value>
          <si:unit>\milli\gram</si:unit>
        </si:real>
      </dcc:itemQuantity>
    </dcc:itemQuantities>
  </dcc:item>
```


Fig. 10. The items in administrativeData in DCC.

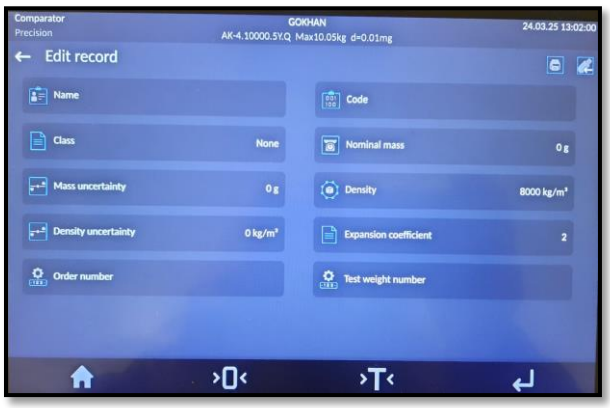


Fig. 11. Control terminal of robotic mass comparator.

The "test weight" is written as "1 OIML Set of Weights (from minimumNominalValue unit to maximumNominalValue unit)" and the test weight number as "Customer_SerialNumber_NominalValue_Unit_Mark_Manufacturer". The order number and class of the test weight, already specified in the calibration report, are entered during the calibration procedure programming. The customer's address is retrieved from a database within the FileItem Panel, matching the "Customer" listed in the "Test Weight Number" (See Fig. 11). The generated DCC is uploaded to GEMIMEG tool and it worked without errors in generating machine readable format in XML and human readable review in PDF [11]. The DCC generator developed by TUBITAK UME produces both XML and PDF files as machine readable and human readable formats, as well. The mass table generated in PDF is shown in Fig. 12.

Results: Mass

Nominal Mass	Marking	Measured Value	Unit	Uncertainty	Unit	Coverage Factor
10 mg	none	10.0042	mg	0.0010	mg	2.0
20 mg	dot	20.0077	mg	0.0010	mg	2.0
20 mg	none	20.0078	mg	0.0010	mg	2.0
50 mg	none	50.0052	mg	0.0013	mg	2.0
100 mg	none	100.0093	mg	0.0017	mg	2.0
200 mg	dot	200.0014	mg	0.0020	mg	2.0
200 mg	none	200.0099	mg	0.0020	mg	2.0
500 mg	none	500.0021	mg	0.0027	mg	2.0
1 g	none	0.9999936	g	0.0000033	g	2.0
2 g	dot	2.0000035	g	0.0000040	g	2.0
2 g	none	2.0000091	g	0.0000040	g	2.0

Fig. 12. The mass table generated in PDF.

V. CONCLUSION

The automatic generation DCC from the calibration reports of Robotic Mass Comparator working in the range of 1 mg – 10 g is demonstrated. The automatic generation of DCCs streamlines the calibration workflow by significantly reducing manual effort and the potential for human error in reporting the results. In order to adhere to the FAIR principles, version 3.2.1 of the PTB DCC schema is followed in generating DCC and the attributes are attached to various elements as they are described in DKD Expert Report. The use of standardized formats and automated tools ensures that each calibration certificate is consistent, accurate, and generated in real-time, without delays. This enhances the overall efficiency of the calibration workflow, allowing laboratories to handle large volumes of calibration data for various weight classes with minimal intervention.

REFERENCES

[1] "DCC Wiki documentation", https://dccwiki.ptb.de/en/home, (accessed March 2025).

[2] S.Hackel, S.Schönhals, L.Doering, T.Engel, R.Baumfalk, "The Digital Calibration Certificate (DCC) for an End-to-End Digital Quality Infrastructure for Industry 4.0", Sci., vol.5, No.11, March 2023, pp.1-13. DOI: 10.3390/sci5010011.

[3] S.Schönhals, L.Doering, B.Gloger, S.Hackel, F.Härtig, D.Hutzschenreuter, J. Jagieniak, T.Krah, J.Loewe, T.Schrader, G.S.Öktem, "Recent advances of the long-term available dcc schema version 3", IMEKO TC6 International Conference on Metrology and Digital Transformation September 19 – September 21, 2022, Berlin, Germany.

[4] "Browser-based version of the GEMIMEG tool", https://gemimeg-tool.ptb.de/v1/, (accessed March 2025).

[5] S.Hackel, S.Schönhals, B.Gloger, L.Doering, J.Jagieniak, "Digitization versus Digitalization & Digital Transformation in the field of calibrations and their subsequent use", Physikalisch-Technische Bundesanstalt (PTB), DOI: 10.7795/120.20240126.

[6] B.Rodiek, G.Á.Clara, C. Brown, S. Davidson, M-A.Demir, S.Eichstädt, W.E.Jaoua, Y.H.Fung, J-L.Hippolyte, T.Hoffmann, D.Hutzschenreuter, M.Jordan, B.Korutlu, N.Kuramoto, S.Cho, Z. Li, A.Malengo, N.Murnaghan, L.M.P.Pérez, S.Schönhals, I.Smith, "Project on a fully automated evaluation of a virtual comparison of mass using the Digital Calibration Certificate (DCC) schema", Measurement: Sensors, December 2024, DOI:10.1016/j.measen.2024.101361.

[7] G.Foyer, M.Häfner, J.Haller, C.Müller-Schöll, S.Osang, A.Scheibner, "Instructions on how to use the DCC schema to create a digital calibration certificate for weights and weight sets", Expert Report DKD-E 7-2: Edition 03/2025, pp.1-60. DOI:10.7795/550.20250325.

[8] International Organization of Legal Metrology OIML R 111-1 "Weights of classes E1, E2, F1, F2, M1, M1–2, M2, M2–3 and M3 Part 1: Metrological and technical requirements", Edition 2004 (E).

[9] A.Picard, R.S.Davis, M.Glaser, K. Fujii, "Revised formula for the density of moist air (CIPM-2007)", Metrologia, vol. 45, 2008, pp.149–155 DOI:10.1088/0026-1394/45/2/004.

[10] JCGM 100:2008 GUM 1995 with minor corrections "Evaluation of measurement data — Guide to the expression of uncertainty in measurement Évaluation des données de mesure — Guide pour l'expression de l'incertitude de mesure".

[11] M-A.Demir, M.Jordan, T.Krah, S.Schönhals, S.Hackel, F.Härtig, T.Schrader, J.Loewe, L.Doering, B.Gloger, J.Jagieniak, G.S.Öktem, "A Human readable form of DCC", IMEKO TC6 International Conference on Metrology and Digital Transformation Sep. 19 – Sep. 21, 2022, Berlin, Germany.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Automated Calibration System with Digital Calibration Certificates Support

Serhii Kursin[1], Oleh Velychko[2], Ihor Pototskyi[3], Andrii Bachynskyi[4]

[1] Institute of Electromagnetic Measurements
State Enterprise "Ukrmetrteststandard", Metrologichna str. 4, 03143 Kyiv, Ukraine
skursin@ukrcsm.kiev.ua
[2] Institute of Electromagnetic Measurements
State Enterprise "Ukrmetrteststandard", Metrologichna str. 4, 03143 Kyiv, Ukraine
velychko@ukrcsm.kiev.ua
[3] Institute of National Metrology Service
State Enterprise "Ukrmetrteststandard", Metrologichna str. 4, 03143 Kyiv, Ukraine
pototskiy@ukrcsm.kiev.ua
[4] Institute of Electromagnetic Measurements
State Enterprise "Ukrmetrteststandard", Metrologichna str. 4, 03143 Kyiv, Ukraine
bav-ivt-pbf23@lll.kpi.ua

***Abstract* – This paper presents the development of an automated system for the calibration of radio-frequency (RF) power meters, addressing the need for accurate, traceable, and efficient calibration in the RF domain. The system is based on a Rohde & Schwarz signal generator and reference power sensors, operating over a frequency range up to 43.5 GHz. A Python-based software solution enables full automation of the calibration process, including signal level auto-balancing and uncertainty estimation. Calibration results – such as calibration coefficients, linearity, and Voltage Standing Wave Ratio (VSWR) – are formatted into Digital Calibration Certificates (DCCs) using XML. DCCs offer a reliable, machine-readable solution for integration into Industry 4.0 systems, enhancing traceability and reducing human error. The proposed approach improves calibration efficiency and supports the global shift toward digital metrology. A user interface and example XML structure are also presented.**

I. INTRODUCTION

Power measurement is fundamental to metrological support in the radio frequency (RF) range. Measuring parameters of the electromagnetic field, signal transmission path attenuation, and antenna characteristics is not possible without RF power measurement. These measurements are essential in areas such as radio communication, mobile communication, control signal transmission, video signal transmission, and GPS navigation. Therefore, the calibration of power meters in the radio frequency range must be performed over a wide frequency and dynamic range.

The results of such wattmeter calibration – namely, the calibration coefficient or the deviation of the measured value and the voltage standing wave ratio (VSWR) – are directly used in corresponding calculations of other measurement parameters in the RF range. These power meters can be used in automated systems, continuous monitoring systems and more. Thus, there is a growing need to create DCCs for RF power meters, which is in line with the current trend towards digitalization of metrology and provides an opportunity to automate the calculation of measurement results [1].

Automation of such measurements significantly reduces the time required for calibration, improves accuracy through statistical processing of a larger number of results, and minimizes potential human errors.

Therefore, the development of an automated system for the calibration of radio-frequency power meters is a relevant and timely task. Such a system must ensure the traceability of the power unit across the entire frequency range and dynamic range of the power meters. Calibration results – specifically, the calibration coefficient (or deviation of the measured value), linearity across the dynamic range, and the voltage standing wave ratio (VSWR, or reflection coefficient) – must be represented in a DCC.

II. BACKGROUND FOR SYSTEM

At the National Metrology Institute of the State Enterprise "Ukrmetrteststandard", there is an extensive range of radio-frequency power meters that are used to ensure traceability of the power unit during the calibration of instruments such as spectrum analyzers, measuring receivers, field strength meters, and attenuators [2]. All

power meters are manufactured by the company Rohde & Schwarz. For the calibration of these power meters, dedicated calibrators were developed based on NRP-Z51 and NRP-Z55 power sensors, as well as 2-way resistive power splitters. This simplified procedure, unlike the classical example of calibrating a high-frequency wattmeter [3, 4], makes it possible do not take into account the reflection coefficients of the wattmeter – device under test (DUT). The errors cused by the mismatch are included in the total error [5].Such calibrators ensure the traceability of the power unit across a frequency range up to 44 GHz. The measurement signal is generated using a Rohde & Schwarz SMF100A signal generator operating in the range of 1 GHz to 43.5 GHz. For this calibration setup, the components of the uncertainty budget have been evaluated [6].

III. AUTOMATED SYSTEM

The automation of the proposed calibration system can be implemented using a Personal Computer. The block diagram of the system is shown in Figure 1.

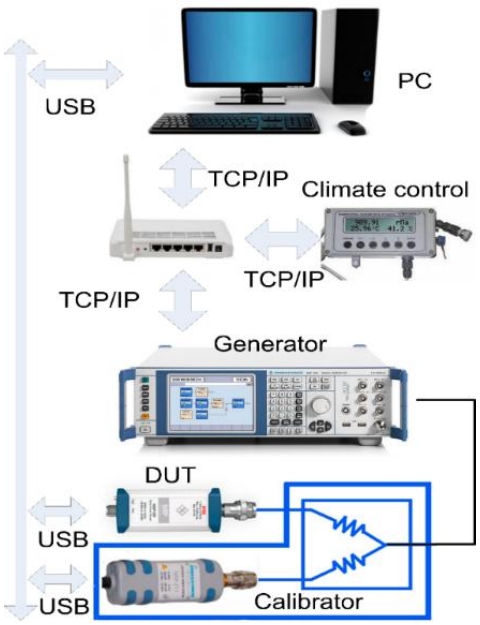


Fig. 1. Block diagram of the automated system for wattmeter calibration.

The software on Personal Computer controls the signal generator via an Ethernet interface and collects measurement data from both the Calibrator and DUT via USB connection. The system also allows for environmental condition monitoring, either directly from a sensor or via a network server over Ethernet.

Rohde & Schwarz provides instrument drivers for their devices. Python was chosen as the programming language for software development, as it offers robust libraries for data processing and visualization [7].

It is also possible to automate the measurement of the reflection coefficient of the DUT input. To do this, it is necessary to add a directional splitter and an additional reference wattmeter. Or, such measurements can be performed separately, using a Vector Network Analyzer.

IV. SIGNAL LEVEL AUTO-BALANCING PROCEDURE

The output signal level of the generator depends on the signal frequency, the characteristics of the input connectors of the connected power meters, and the frequency-dependent transmission coefficient of the power splitter. Therefore, to ensure proper calibration conditions – specifically, a predefined signal level across the entire frequency range – it is necessary to implement an signal level auto-balancing procedure [3]. The essence of the signal level auto-balancing procedure lies in setting the required signal level on the generator, measuring the signal at the output of the power splitter using a reference wattmeter, calculating the correction factor, and adjusting the generator's output accordingly. The block diagram of this algorithm is shown in Figure 2.

The decibel-milliwatt (dBm) is a unit of power level expressed using a logarithmic decibel (dB) scale respective to one milliwatt. The developed software provides the ability to choose between dBm and watts, depending on the units in which the signal level values for the calibration points are given. The input data of the procedure are the required signal level and the calibrated deviation for the Calibrator (reference wattmeter). The values of the number of measurements and the maximum deviation from the obtained average value are set, such a check is necessary to remove gross errors. The required signal level is set on the generator. 5 power measurements in watts are performed. It is checked whether the obtained value is greater than 0. Find the average value and discard values that differ from the average by more than the maximum permissible deviation (0.001). Find the average value of the filtered results.

Calculate the adjusted signal level value using formula

$$L_{corr} = L_{watt} - (Av_{watt} - L_{Cal} - L_{watt}), \qquad (1)$$

where, L_{corr} is the corrected signal level value, Av_{watt} is the average of the filtered values, L_{cal} is the correction value from the calibration certificate of the standard, L_{watt} is necessary level.

For correct conversion of all values into dBm units, the conversion is carried out according to formula

$$P_{dBm} = 10 \cdot log_{10} \cdot \left(\frac{P_w}{0.001}\right), \qquad (2)$$

where, P_{dBm} is the power value in dBm, P_w is the power value in watts.

The Python code implementation for this procedure with signal level in watts is presented in Figure 3.

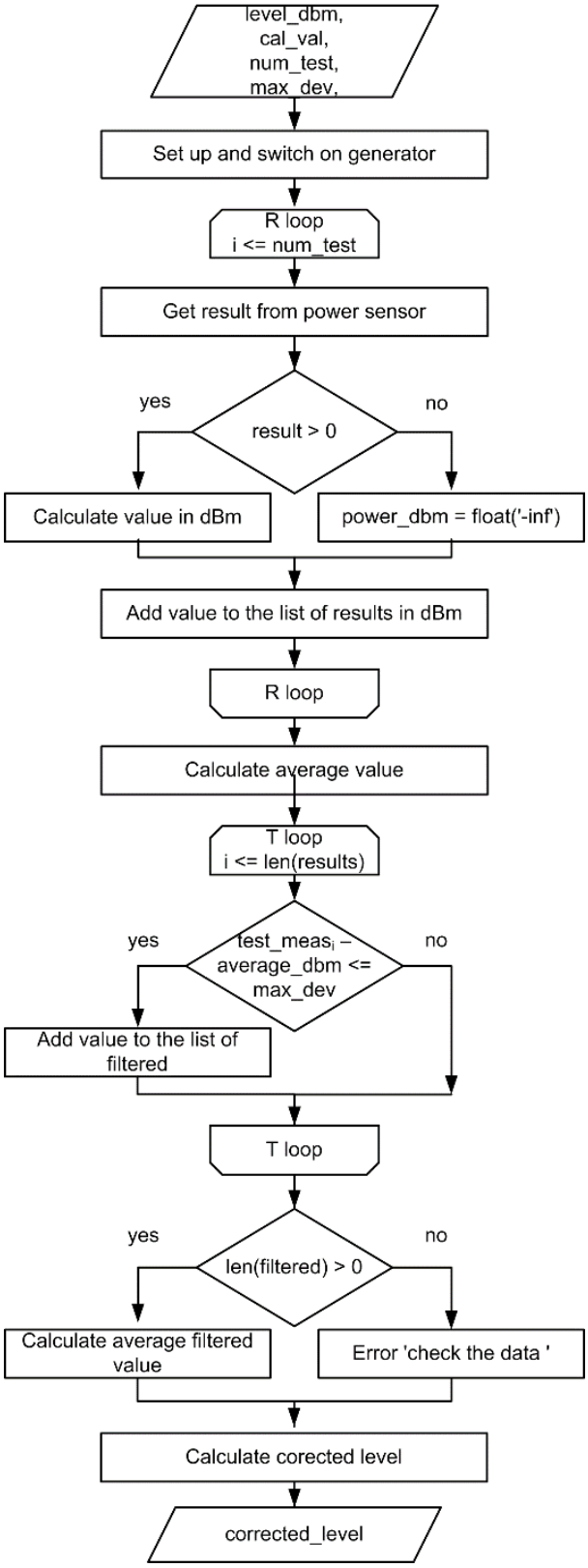


Fig. 2. Auto-balancing procedure algorithm.

```
def collect_test_measurements(self, sensor, level_watt,
                              cal_val, num_test, max_dev):
    test_measurements =
    set_output_level(self.smf100a, level_watt)
    rf_on(self.smf100a)
    for _ in range(num_test):
        sensor.write_str('INIT:IMM')
        result = sensor.query_str('FETCH?').split(',')
        power_watt = float(result[0])
        if power_watt > 0:
        test_meas.append(power_watt)
        else:
            power_watt = float('-inf')
    average_watt = sum(test_meas) / len(test_meas)
    filtered_measurements = [x for x in test_meas
                    if abs(x - average_watt) <= max_dev]
    if filtered_measurements:
        average_filtr_watt = sum(filtered_measurements)
                             / len(filtered_measurements)
    else:
        raise ValueError('check the data')
    corrected_level = level_watt - (average_filtr_watt
                                - cal_val - level_watt)
    rf_off(self.smf100a)
    return corrected_level
```

Fig. 3. Code fragment of the generator signal level auto-balancing procedure.

Using the developed software, it is possible to calculate the type A uncertainty for the measurements made at each calibration point using formula

$$u_A = \sqrt{\frac{1}{n\cdot(n-1)}\sum_{i=1}^{n}(x_i - \bar{x})^2}, \qquad (3)$$

where, n – number of measurements, x_i – i-th measurement result, $\bar{x}$ – average value n measurements.

The user interface of the developed software is shown in Figure 4.

The developed software provides automatic detection of wattmeters connected to the system and allows the operator to define calibration points, specifically the frequency and signal level. It also includes functionality for measuring the reflection coefficient and calculating the refkection coeficient.

V. DIGITAL CALIBRATION CERTIFICATES

DCCs are a key component of the digital transformation of metrology in the era of Industry 4.0. They provide a machine-readable format with a high level of reliability and ease of integration into automated quality management systems. Unlike traditional paper-based or PDF certificates, DCCs enable automated data processing, reducing the risk of human error and accelerating production workflows. They support digital signatures, ensuring the integrity and authenticity of the information [8, 9].

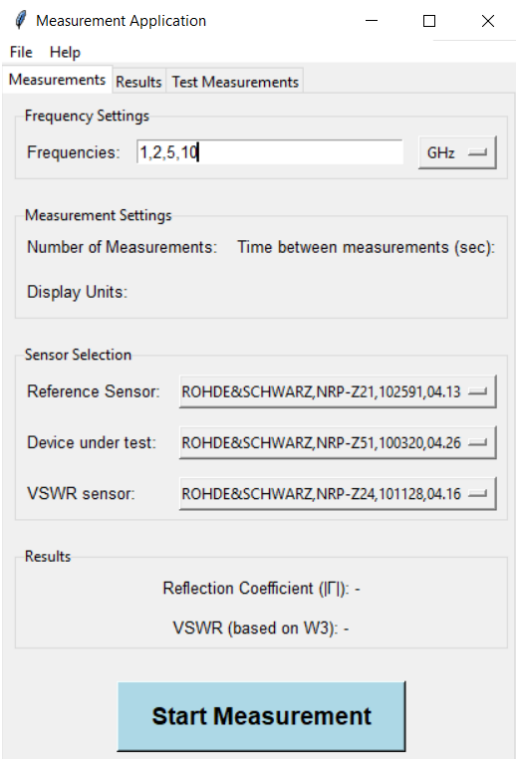


Fig. 4. User Interface.

The main formats of DCCs are based on XML data structures. The use of DCCs enhances transparency, traceability, and efficiency in industrial processes, enabling reliable exchange of calibration data at both national and international levels [10].

The GEMIMEG project website [11], dedicated to DCCs, provides examples of calibration results for temperature, and humidity measuring instruments.

At this stage, we are using the DCC v3.2.1 scheme version, as it is supported by the working version of GEMIMEG-Tool v1. Using the GEMIMEG-Tool provides the ability to translate DCC created during the development phase in XML format into human readable format. As a general example, the PTB Temperature example (simplified) Good Practice DCC with D-SI version 2.2.1 was used.

It was also necessary to resolve the issue of the units of measurement used. The dBm is not included in the International System of Units (SI) and is therefore not recommended for use in documents or systems that follow SI standards. The corresponding SI unit is the watt. However, the decibel (dB), without the 'm' suffix, is allowed for expressing relative quantities, though it is not approved for direct use alongside SI units[12, 13]. At the same time, the use of such units is an established practice and is convenient for calibrating RF power meters [14]. An example of calibration results from the certificate issued by the accredited in DKD Rohde & Schwarz Laboratory is shown in Figure 5.

Freq. /GHz	Power level Dev. /dB	Unc. /dB	Output Reflection of Po Real /lin.	Imag /lin.	Unc. /lin.
0.001	-0.001	0.0078	0.000	0.000	0.005
0.01	0.000	0.017	-0.001	0.000	0.005
0.03	0.001	0.017	-0.001	0.001	0.005
0.05	0.000	0.017	-0.002	0.001	0.005
0.1	0.000	0.017	-0.001	0.001	0.005

Fig. 5. Example of calibration results.

Therefore, we propose the following XML-based structure for representing the calibration results of power meters within a DCC, using <*si:hybrid*> data elements for watts and dBm. Also, since it is necessary to present the calibration results for a frequency range, the <*si:realListXMLList*> data structure is used. The fragment of the DCC structure without reflection results is shown in Figure 6.

```
<dcc:quantity refType="basic_measuredValue">
<dcc:name>
<dcc:content lang="ua">Середнє виміряне
значення
</dcc:content>
<dcc:content lang="en">Measured value
</dcc:content>
</dcc:name>
<si:hybrid>
<si:realListXMLList>
<si:valueXMLList>0.9982 0.9977 0.9972</
si:valueXMLList>
<si:unitXMLList>\milli\watt</si:unitXMLList>
</si:realListXMLList>
<si:realListXMLList>
<si:valueXMLList>-0.008 -0.010 -0.012</
si:valueXMLList>
<si:unitXMLList>dBm</si:unitXMLList>
</si:realListXMLList>
</si:hybrid>
</dcc:quantity>
<dcc:quantity refType="basic_measurementError">
<dcc:name>
<dcc:content lang="ua">Відхилення</dcc:content>
<dcc:content lang="en">Measurement error
</dcc:content>
</dcc:name>
<si:realListXMLList>
<si:valueXMLList>0.005 0.002 -0.006
</si:valueXMLList>
<si:unitXMLList>\decibel</si:unitXMLList>
<si:expandedUncXMLList>
<si:uncertaintyXMLList>0.020 0.036 0.036</
si:uncertaintyXMLList>
<si:coverageFactorXMLList>2 2 2</
si:coverageFactorXMLList>
<si:coverageProbabilityXMLList>0.95 0.95 0.95
</si:coverageProbabilityXMLList>
<si:distributionXMLList>normal normal normal
</si:distributionXMLList>
</si:expandedUncXMLList>
</si:realListXMLList>
</dcc:quantity>
```

Fig. 6. Example of wattmeter calibration result representation in XML format.

Presenting calibration results in units such as dB makes it possible to clearly evaluate the DUT measurement results, since to take into account the obtained corrections [15], it is enough to perform simple mathematical operations of addition or subtraction according to the expression:

$$P_{correct}\,(\mathrm{dBm}) = P_{meas}(\mathrm{dBm}) + P_{dev}(\mathrm{dB}) \quad (4)$$

However, calibration of reference wattmeters involves taking into account path mismatch, reflection coefficients, and dynamic range nonlinearity. Therefore, it is necessary to further develop a DCC structure for reference wattmeters with calibration coefficients or efficiency coefficients.

VI. CONCLUSION

An automated calibration system for radio-frequency power meters has been successfully developed and implemented. The system enables precise and efficient calibration across a wide frequency and dynamic range using reference sensors and a programmable signal generator. The software, developed in Python, supports signal level auto-balancing procedure, and user-defined calibration points. One of the key innovations is the integration of DCCs, structured in XML format, which aligns with current trends in digital metrology and supports automated data exchange. This approach enhances traceability, reduces calibration time, minimizes human errors, and ensures compliance with Industry 4.0 requirements. The system is suitable for use in national metrology institutes and industrial laboratories requiring reliable RF power measurements. Tasks for further research are identified.

REFERENCES

[1] Hackel, S.; Schönhals, S.; Doering, L.; Engel, T.; Baumfalk, R. The Digital Calibration Certificate (DCC) for an End-to-End Digital Quality Infrastructure for Industry 4.0. Sci 2023, 5, 11. https://doi.org/10.3390/sci5010011

[2] O. Velychko, V. Gaman and S. Kursin, "Metrological Traceability of Power Measurements at Microwave Frequencies," 2022 XXXII International Scientific Symposium Metrology and Metrology Assurance (MMA), Sozopol, Bulgaria, 2022, pp. 1-4, doi: 10.1109/MMA55579.2022.9993365

[3] Danaci, E. Analyzing the Automatic Power Level Control Effect of a Signal Generator in RF Power Sensor Calibration by a Direct Comparison Transfer Method and a Millimeter Wave Application. Sensors 2024, 24, 609. https://doi.org/10.3390/s24020609

[4] EA-04/02M. Evaluation of the Uncertainty of Measurement in Calibration. EA, 2021, 78 p.

[5] Oleh Velychko, Serhii Kursin, Valentyn Gaman. Calibration of High and Ultra-high Frequency Power Meters in the Range of up to 30 mHz. Measuring equipment and metrology. Vol. 83, No. 4, 2022. https://doi.org/10.23939/istcmtm2022.04.036

[6] Velychko O., Gaman V., Kursin S., "Calibration features for power meters of high and microwave frequencies". Ukrainian Metrological Journal, 2022, No. 2, pp. 9–14. DOI: 10.24027/2306-7039.1.2022.263724.

[7] RsInstrument 1.100.0 Online [Accessed 12 April 2025] - https://pypi.org/project/RsInstrument

[8] Mustapää, T., Nummiluikki, J., & Viitala, R. (2022). Digitalization of Calibration Data Management in Pharmaceutical Industry Using a Multitenant Platform. Applied Sciences (Switzerland), 12(15), Article 7531. https://doi.org/10.3390/app12157531

[9] Martin Koval, Jiří Tesař, Shanna Schönhals, Anjali Sharma, David Balslev-Harder, Carlos Galvan-Hernandez, Henri Baumann, David Mahovský. Digital Calibration Certificates (DCC): Benefits and Challenges to overcome for implementation. EPJ Web Conf. 323 01001 (2025) DOI: 10.1051/epjconf/202532301001

[10] Vashti Galpin, Ian Smith, Jean-Laurent Hippolyte, Tracking and viewing modifications in digital calibration certificates, Acta IMEKO, vol. 12, no. 1, article 11, March 2023, identifier: IMEKO-ACTA-12 (2023)-01-11

[11] GEMIMEG project. Online [Accessed 12 April 2025] https://www.gemimeg.ptb.de/en/gemimeg-home.

[12] Thompson and Taylor 2008, Guide for the Use of the International System of Units (SI), NIST Special Publication SP811

[13] Hutzschenreuter, D., Härtig, F., Heeren, W., Wiedenhöfer, T., Forbes, A., Brown, C., Smith, I., Rhodes, S., Linkeová, I., Sýkora, J., Zelený, V., Ačko, B., Klobučar, R., Nikander, P., Elo, T., Mustapää, T., Kuosmanen, P., Maennel, O., Hovhannisyan, K., … Paciello, V. (2019). SmartCom Digital System of Units (D-SI) Guide for the use of the metadata-format used in metrology for the easy-to-use, safe, harmonised and unambiguous digital transfer of metrological data (D-SI 1.3). Zenodo. https://doi.org/10.5281/zenodo.3522631

[14] NRP-Z51 THERMAL POWER SENSOR Calibration Certificate, Document No. 0001A300705550, Rohde & Schwarz Calibration Laboratory, 87700 Memmingen, Rohde-und-Schwarz-Str. 1. 2023-08-22.

[15] International Electrotechnical Commission, IEC 60027-3:2002, Letter symbols to be used in electrical technology – Part 3: Logarithmic and related quantities, and their units, Geneva, Switzerland: IEC, Jul. 2002.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Transition to DCC-based calibration management: findings from proof-of-concept investigations

Tuukka Mustapää[1]

[1] *Beamex Oy Ab, Jakobstad, Finland, tuukka.mustapaa@beamex.com*

***Abstract* – The introduction of standardized data formats for measurement and calibration data such as the digital calibration certificate (DCC) will enable further automation of calibration-related processes. This will open new possibilities for optimization of calibration processes in industrial applications. This paper describes some of the challenges and possibilities related to the transition to DCC-based calibration processes and management in the process industry. The findings presented are based on case examples and proof-of-concept projects from the process industry, in which automation of a calibration process is pursued using a combination of systems from multiple providers. The investigations and findings also suggest that parts of the DCC data structure could potentially be beneficial also for other applications, in which exchanging of calibration data is needed, but system integration is otherwise challenging due to lack of a suitable harmonized data structure.**

I. INTRODUCTION

Calibrations are an essential part of quality management of any industrial process that is dependent on measurements. Traditionally the results of a calibration have been documented in a calibration certificate that also contains detailed information on how the calibration has been performed. Due to their significance for quality management of any process that is dependent on measurements and thus the accuracy of the measurements, the contents of a calibration certificate have been standardized. As a part of the digitalization efforts around metrology and quality infrastructures, definition of a standardized format for a digital calibration certificate (DCC) and development of associated processes have been one of the key areas of development within the metrology community in recent years [1,2].

Digitalization of calibration related processes have many benefits but achieving them can require comprehensive changes in current processes and systems. Due to this expected that the transition will need to be done incrementally, e.g. as presented in [3]. A major benefit is improvement in cost-efficiency, as there are various direct and indirect costs associated with calibration, and reductions in the lead time of performing a single calibration quickly add up when the number of instruments requiring calibration increases. Availability of the calibration data in an easily usable format allows its use for analytics-based decision making and optimization of the processes, e.g., optimization of calibration intervals and tolerance analysis [4,5].

Digitalization also plays a big role in improving information security and data integrity, as manual work steps have a chance for human error, e.g., in data entries, and are thus considered a compromise for data integrity, which is a highly regulated topic, e.g., in the pharmaceutical industry [3]. A key factor in efficiency and security improvements is the possibility for better system integration as a standardized format for calibration data minimizes the need for bespoke system to system integration, which can become complex to maintain long term as the systems involved are being updated [6].

Ideally, the availability of unambiguous calibration data also allows more extensive network benefits within metrology ecosystems of customers, service providers, instruments manufacturers and other organizations [7]. This is one of the reasons for investigations on the establishment of metrology data spaces [8].

Achieving all the benefits of digitalization is still depending on more extensive harmonization and establishing more specifically defined ways of using the DCC in different applications [9]. This has led to efforts in defining good practices and guidelines for correct use of the DCC for different fields of metrology, e.g., by technical committees of the German Calibration Service (DKD) [10,11]. An important part of these efforts is the definition of XML attributes, i.e., refTypes, which are needed for ensuring correct interpretation of data. Such guidelines are already available, e.g., for weighing instruments [12,13]. The need for more specific semantics of the calibration data has also generated interest in the use of ontologies in more detailed descriptions of different data that are included in a DCC [14].

This paper presents an overview of technical investigations and testing based on use cases from the pharmaceutical industry. The presented part of the work focuses mostly on the end-user's side of the calibration process and calibration management. The work on the

proof-of-concept implementations has not been finalized by the time of writing this paper, so the paper focuses mostly on presenting the investigated use cases, challenges identified so far and possible solutions to these challenges. Good practices and comprehensive lists of refTypes for the quantities and instrument types were not yet available at the time the work started. To overcome this some additional refTypes are used for testing purposes, and in this regard the solutions used proof-of-concept implementations will likely have some limitations for more generalized use in other cases.

II. CALIBRATION PROCESSES

Calibrations of high-precision measurement instruments are traditionally performed by calibration laboratories, to which the customers send their instruments to be calibrated. However, in industries such as process industry, the number of instruments at a single facility can be very high meaning that it is more efficient to calibrate instruments on-site if the tolerances for the measurements allow it. It is also common practice that larger companies with more annual calibrations outsource on-site calibration services to external calibration service providers. When calibrations are due, a technician from the service provider comes to perform the calibration on-site at the customer's facility, either using equipment from the service provider or customer. In general, the calibration processes used in the industry tend to follow mostly the same steps regardless of whether the calibration is performed on-site or at a laboratory.

The process can be considered to begin when the customer, i.e., the instrument's owner, requests a calibration to be done by the service provider. Once a calibration has been requested, the service provider needs to confirm that the requested calibration can be performed, and the customer and service provider agree on the calibration procedure and other practical details.

When everything has been agreed on the actual calibration is then performed. If the calibration is to be performed at a laboratory, the instrument to be calibrated is sent to the service provider. After that the actual calibration is performed and a calibration certificate is issued and sent to the customer. If the calibration was performed at a laboratory, the instrument is also sent back to the customer.

Upon receiving the calibration certificate, the customer will need check that calibration procedure etc. match the agreed procedure. Typically, the calibration is then approved on the calibration management system and the results are imported into the system and the original certificate is archived.

III. OPERATIONAL AND INFORMATION TECHNOLOGY SYSTEMS

In the industry systems used at factories, power plants or other kind of facilities are categorized into operational technology (OT) systems and information technology systems (IT). The main difference between of the two is the connectivity. Typically, OT systems such as supervisory control and data acquisition (SCADA) are used in an isolated local environment. Many typical OT systems have been designed with the assumption that systems are isolated, and thus their security may heavily depend on that. In addition, OT environments are typically not as well suited for regular updating of systems. This can lead to the systems having severe vulnerabilities if they were exposed to a cyberattack. Respectively, IT systems are intended to be interconnected so the approaches used for ensuring security are very different.

Introduction of industry 4.0 and internet of things (IoT) has meant that nowadays there are more interest towards converging IT and OT systems partially to enable more flexible use of data within organizations. However, there are several challenges related to the converging of IT and OT systems, most of which are related to ensuring the security of the OT environment.

IV. PROOF-OF-CONCEPT OVERVIEW

The proof-of-concept projects used as the basis for these investigations are focusing on two different use cases. The first use case is related to conventional calibration processes involving a customer and a service provider or instrument manufacturer. The second case is related to a more experimental use case where the use of the DCC's data structure is used to exchange calibration results between an automated calibration system and calibration management system developed by different companies. In both cases there are two essential functionalities for the calibration management system: importing calibration data from a DCC and generating and sending a calibration request. The main difference of cases is that in the first case the DCC is received from outside the OT environment where the calibration management system is running, whereas in the second case the systems involved are OT systems.

A. Importing calibration data

The first step of the implementation has been developing a simple solution for importing data from a DCC into a calibration management system to enable further testing. As a first iteration, the importing functionality was based on a file manager view where the user can browse for the DCC file that is to be imported. While this kind of approach is simple to implement, it is by no means ideal or particularly practical, as the process still involves several manual steps. This kind of solution used alone can still save some time in the process, if the data would otherwise have to be entered manually into the calibration management system.

Ideally, the importing process is done without needing to specifically browse and select a file to be imported. This requires that the results to be imported are transitorily

saved in a specific location and are clearly identifiable so that a result can be automatically assigned to the correct instrument in the calibration management system's database. However, when the DCCs are to be received from a third-party IT system, e.g., via an Application Programming Interface (API) or a cloud service, into a calibration management system operating in an OT environment, additional security measures are needed to secure the OT environment.

B. Calibration request

Capability to create and send a digital calibration request (DCR) that is similarly structured file as a DCC will benefit different use cases. For industry, DCR is considered as a necessary part of the calibration process as it provides additional assurance in the calibration's conformity to the end-user's procedures. The most common use case for the DCR is requesting calibration from a service provider. In these cases, the DCR will be handled in a similar way whether the service provider is going to perform the calibration on-site or at an external laboratory.

One new opportunity that DCR can also bring is the ability to have the newly ordered instrument calibrated according to the customer's own procedure by the manufacturer. In industrial application instruments are often used in very specific operating ranges and have their calibration procedures defined accordingly. For these reasons, the calibration certificates provided by the manufacturers may not be that useful for the customers if the used calibration procedure uses different calibration points. For these reasons the customers may end up recalibrating brand-new instruments according to their own procedures. Thus, the DCR can in the best case help get rid of these kinds of seemingly unnecessary calibrations. When the calibration management system is considered, it is a relatively infrequent use case when compared to re-calibrations of existing instruments. In this case some of the data that is typically considered necessary for requesting calibration is also not available yet, so the requesting process must be possible also without it.

In the second use case presented, the requirements for calibration request are somewhat different, as the DCR is mainly needed to provide the instructions for the automated calibration system for performing the calibration, meaning that the process itself is much more straightforward. In this case the benefit from the use DCR is mainly making the operation of the calibration system less dependent on manual operation, which still is an important part of improving the efficiency of the calibration process as a whole.

C. Overcoming challenges in the IT-OT convergence

In the transition to a DCC-based calibration process, the IT-OT convergence quickly becomes a challenge when the exchange of the DCCs and DCRs is done using IT systems, as calibration management systems are typically OT systems. As such, multiple layers of security precautions are necessary. One of these approaches is implementing separate system between the interface and the calibration management system as a buffer between the IT and OT environments. This allows the validation of the received DCC before it is imported into the calibration management system, ensuring, e.g., that only DCCs corresponding to the original request are imported. Another important part of the security solution is access management for the interfaces, so that only the specified users or organizations can access the interface. Similarly, also electronic signing or sealing of the transferred files and verification of the signatures and seals is essential to ensure the integrity and authenticity of the files.

V. FINDINGS FROM THE INVESTIGATIONS AND PROOF-OF-CONCEPTS

Although the implementation is still at an early stage, there have already been some findings and challenges, for further investigations would be beneficial.

A. Adding instruments to calibration management system based on a DCC

DCCs contain some of the information that are typically also needed in calibration management systems. This means that there is a possibility to make adding new instruments easier if the instrument is delivered with a DCC from the manufacturer. Examples of such data include information for identifying the instruments such as the manufacturer name, model designation and serial number, or information on the technical specification or performance of the instrument.

B. Using partial DCCs for transferring calibration data

Since the DCC has been defined to be compliant with various standards and regulations regarding accredited calibrations, meaning that a lot of administrative data is considered as mandatory in the DCC schema. In industrial applications, in which a calibration would be performed on-site, some of the mandatory data in the DCC might not be relevant, but a standardized data format would still be needed to handle the communication between instruments or systems from different manufacturers or providers.

There are a few different approaches how the DCC, and possibly also the DCR, could be used to overcome the lack of a more suitable data format. The first way would be to include any data that the automated calibration system is not capable of generating itself into a partially filled DCC file so that the calibration system would be capable of filling the desired parts. Alternatively, the calibration result could be provided by the calibration system as a partially incomplete DCC, consisting of the minimum information required for successfully importing the data into the calibration management system. If a complete DCC would be required, any missing data could be

included by the calibration management system so that the result could be exported later as an official document.

C. Non-SI units and conversions

From the beginning, the definition of the DCC has been strongly tied to the parallel development of the D-SI. The main purpose of the D-SI is to cover the units included in the SI system as defined by the BIPM. Thus, the DCC is also primarily supporting the use of SI units. Following this requirement and other best practices can be very impractical if the existing instruments or systems used in the industry do not natively support SI units. Extending the scope of the D-SI to cover all other units as well is not really a realistic option either. Instead, some alternative approach would be needed as the DCC also needs to cover the use of non-SI units. Currently, it is recommended to be done by including a definition for the used non-SI units in the DCC, but so far there is not an established database containing for the definitions, or any other kind of a harmonized naming practice for the non-SI units for use with the DCC. A more harmonized way of doing it would be needed to make DCC really support all applications of metrology. A similar approach as with the definitions of refTypes could be used for the most common units for different quantities.

Another challenge with the current versions of the DCC and D-SI is that when a measurement result is presented using two different units, there is not a clear distinction of the which value is the original value, and which value is a conversion. This issue arises especially in cases where the conversion would require additional significant digits or a higher resolution than the originally measured or indicated value to maintain a comparable level of accuracy. For any further use of the data, the original measurement value and the converted value should be clearly distinguishable.

As the DCC is intended to be a globally applicable standard, it should cover all different fields of metrology and the associated industrial applications. There are plenty of use cases where the use of a non-SI unit has become a common practice and forcing a change to that is not going to be feasible with the introduction of the DCC alone.

VI. CONCLUSION

The development of the DCC has reached a stage where its use in industrial applications can be considered realistic as good practices are agreed on and being published, although there still some needs for refinement. The transition to a fully digitalized DCC-based calibration processes will require significant revising of the current calibration processes. Having a strict separation of OT and IT systems may require even organizational changes before the DCC can be taken into use in an ideal way, as the DCC-based processes can be considered to require some degree of convergence of IT and OT systems.

The DCC and D-SI have been defined for specific purposes, but at least from the industry's perspective there are also other potential uses for them in applications where calibration data is being transferred, and a corresponding standardized data format does not exist. However, using DCC data structure in a way that would be deviating from the general guidelines for the use of the DCC and D-SI. This suggests that there would be use for a more generic data structure for calibration data that would use the same structure as the DCC for the measurement results, but with less restrictions.

There are many industrial applications where non-SI units have become the de-facto standard measurement unit, and to make the DCC truly universally applicable standard for calibration data, some additional harmonization for non-SI units is also needed. This could be a guideline for labelling conventions, or more ideally a quantity specific database for commonly used non-SI units similarly as there already are for refTypes.

REFERENCES

[1] S.Hackel, F.Härtig, T.Schrader, A.Scheibner, J.Loewe, L.Doering, B.Gloger, J.Jagieniak, D.Hutzschenreuter, G.Söylev-Öktem, "The fundamental architecture of the DCC", Measurement: Sensors, Volume 18, 2021, 100354, ISSN 2665-9174, DOI: 10.1016/j.measen.2021.100354.

[2] S.Schönhals, M.-A.Demir, L.Doering, B.Gloger, S.Hackel, W.Heeren, J.Jagieniak, M.Jordan, C.Keilholz, T.Krah, J.Loewe, K.Mienert, G.Söylev-Öktem, "Harmonisation processes and practical implementation of machine-interpretable digital calibration certificates", Measurement: Sensors, 2024, 101470, ISSN 2665-9174, DOI: 10.1016/j.measen.2024.101470.

[3] H.Foldal, "Digitalization in the metrology quality infrastructure – perspectives from Novo Nordisk", 22nd International Metrology Congress (CIM2025), EPJ Web Conf. 323, 01002 (2025), DOI: 10.1051/epjconf/202532301002.

[4] J.Schroll-Fleischer, F. Paustian, "Calibration data mining – Perspectives from Novo Nordisk", Measurement: Sensors, 2025, 101737, ISSN 2665-9174, DOI: 10.1016/j.measen.2024.101737.

[5] T.Engel, "From a digital calibration certificate to a digital quality infrastructure", 22nd International Metrology Congress (CIM2025), EPJ Web Conf. 323, 01003 (2025), DOI: 10.1051/epjconf/202532301003.

[6] T.Mustapää, S.Koskinen, M.Sundfors, J.Jonsson, K.Riska, L.Löytynoja, J.-A.Broo, "Enabling the use of digital calibration certificates in industrial calibration management systems", J. Sens. Sens. Syst., 2024,13, 71–79, DOI: 10.5194/jsss-13-71-2024.

[7] J.Nummiluikki, T.Mustapää, K.Hietala and R.Viitala, "Benefits of network effects and interoperability for the digital calibration certificate management," 2021 IEEE International Workshop on Metrology for

Industry 4.0 & IoT (MetroInd4.0&IoT), Rome, Italy, 2021, pp. 352-357, DOI: 10.1109/MetroInd4.0IoT51437.2021.9488562.

[8] T.Sołtysiński, J.Niederhausen, S.Eichstädt, “Digital Calibration Certificate in a trusted quality infrastructure federated data space: A proof of concept”, Measurement: Sensors, 2024, 101484, ISSN 2665-9174, DOI: 10.1016/j.measen.2024.101484.

[9] M.Koval, J.Tesař, S.Schönhals, A.Sharma, D.Balslev-Harder, C.Galvan-Hernandez, H.Baumann, D.Mahovský, “Digital Calibration Certificates (DCC): Benefits and Challenges to overcome for implementation”, 22nd International Metrology Congress (CIM2025), EPJ Web Conf. 323, 01001 (2025), DOI: 10.1051/epjconf/202532301001.

[10] S.Schönhals, M.Jordan, M.Melzer, D.Röske, “Ideas for the transition from paper-based force and torque calibration certificates to machine-readable XML data”, Measurement: Sensors, 2024, 101334, ISSN 2665-9174, DOI: 10.1016/j.measen.2024.101334.

[11] M.Melzer, R.Deschermeier, L.Bünger, M.-A.Demir, J.Schüür, “Good practice definitions for the DCC on temperature and humidity quantities”, Measurement: Sensors, 2024, 101477, ISSN 2665-9174, DOI: 10.1016/j.measen.2024.101477.

[12] Expert report DKD-E 7-2, “Instructions on how to use the DCC schema to create a digital calibration certificate for weights and weight sets”, Edition 01/2024, Revision 1, Physikalisch-Technische Bundesanstalt, Braunschweig and Berlin. DOI: 10.7795/20240119B.

[13] Expert report DKD-E 7-3, “Instructions on how to use the DCC schema to create a digital calibration certificate for non-automatic weighing instruments”, Edition 03/2025, Physikalisch-Technische Bundesanstalt, Braunschweig and Berlin. DOI: 10.7795/550.20250325.

[14] M.Jordan, S.Schönhals, S.Auer, “Enhancing interoperability in digital calibration data exchange: A case for ontology development”, Measurement: Sensors, 2024, 101459, ISSN 2665-9174, DOI: 10.1016/j.measen.2024.101459.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

dcclib – Consolidation of Digital Calibration Certificate (DCC) tools into a unified Python library

Jan Loewe[1*], Justin Jagieniak[1], Shanna Schönhals[1]

[1]*Physikalisch-Technische Bundesanstalt (PTB), Bundesallee 100, 38116 Braunschweig, Germany*
[*]*Corresponding author: jan.loewe@ptb.de*

***Abstract* – The increasing adoption of Digital Calibration Certificates (DCCs) has led to the development of different tools and libraries, each addressing different aspects and needs of DCC creation and processing workflows. These tools and libraries differ in their functionality, interfaces and programming languages used, resulting in a fragmented landscape with an inconsistent user and developer experience. This work presents *dcclib*, a library that provides a standardized interface for basic DCC functionality by consolidating the functionalities from existing tools and libraries into a single, easy-to-use, and modular Python library. *dcclib* is designed to be extensible and can be used as the basis for more advanced DCC tools and libraries. Its functionality is accessible through an interface that can be used in Python scripts, a command line interface (CLI) and a REST API.**

I. INTRODUCTION

The Digital Calibration Certificate (DCC) is an XML-based document and data exchange format for calibration certificates. It has been developed since 2017 by an international consortium from the international metrology community coordinated by Physikalisch-Technische Bundesanstalt (PTB) in close collaboration with industry stakeholders. It aims to provide calibration results in a structured, machine-readable, and machine-actionable format, facilitating automated processing and increased interoperability across systems [1]. A DCC consists of four main areas: a regulated administrative data section, partially regulated measurement results, unregulated comments, and additionally a human-readable document [2]. Due to its XML-based structure, and because it is intended to eliminate manual human intervention of handling calibration data, the DCC workflow requires dedicated tools and libraries to create, validate, and process DCCs. Such middleware is essential for making full use of the benefits of the DCC and enabling integration into automated workflows.

A. *Current situation*

A range of tools has been developed to support the creation and processing of DCCs. These tools address specific use cases and have emerged from projects such as GEMIMEG [3] and GEMIMEG-II [4]. However, the landscape is characterized by fragmentation, with tools implemented in various programming languages, including Python, JavaScript, Java, and Rust, and exhibiting significant variations in software quality. In addition, most tools are limited in scope and functionality, meaning that different tasks often require different tools.

One example is the library *pyDCC*, which provides basic functionality for extracting administrative data and measurement results from DCCs [5]. This library was developed by partners of the GEMIMEG-II consortium and is available on GitHub under the MIT License [6]. By definition, it is a read-only library and it is out of its scope to provide other functionalities. For instance, it lacks support for other crucial aspects such as the generation of human-readable outputs or evaluating formulae. Graphical tools like the *GEMIMEG-Tool* offer broader features such as generating and modifying DCC, schema validation and XSLT-based rendering [7] but do not cover all aspects of the DCC lifecycle either. The result is an inconsistent and often inefficient tooling environment.

B. *Motivation*

The fragmented state of existing tooling has led to a clear need for consolidation. Multiple stakeholders, including national metrology institutes and commercial calibration laboratories, have expressed the desire for more integrated and robust tooling solutions that can be integrated in production environment workflows.
This work addresses this demand by designing and implementing a unified software library that consolidates selected core functionalities currently spread across various existing tools. In addition to functional consolidation, ensuring a consistently high software quality is a key motivation for this work.

II. DCCLIB

A. Overview

The library *dcclib* is designed to standardize the available functionalities from existing tools and libraries into a single, easy-to-use Python library. As shown in Figure 1, the functionality of *dcclib* is accessible through a Python interface, a command line interface (CLI) and a REST API. The CLI and REST API are built on top of the libraries public interface, allowing users to access its modules in other programming languages or environments. Scripts can use the CLI to automate tasks such as validating examples in a CI/CD pipeline, while the REST API allows for remote access and integration into other systems.

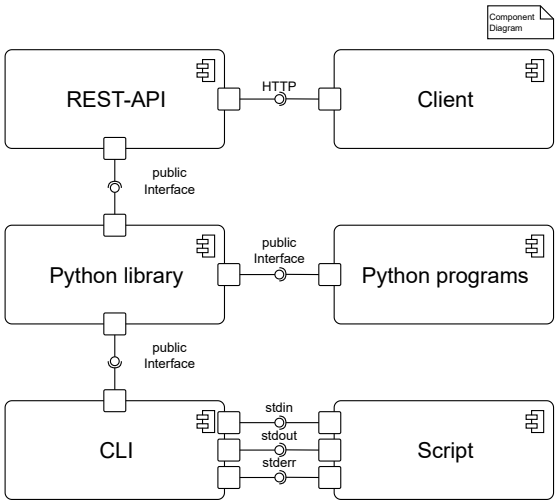


Fig. 1. Component diagram showing the library and its interfaces.

B. Modules

All functionality of the library is structured into modules, each of which contains classes for specific tasks. The five modules with a description are shown in Table 1. This design leaves room for extension and allows for the addition of new classes, for example to support new formats in the *conversion* module or to extract additional data from DCCs in the *extraction* module.

The *lxml* library is used by all modules to provide the `ElementTree` API for XML processing. *xmlschema* is used to convert XML to JSON. For XSLT transformations *saxonche* is used. Signatures are applied and checked using *signxml*.

In some cases the modules use each other to provide a more complex functionality. For instance, the *validation* module utilizes the XSLT class from the *transformation* module to execute Schematron validation using *SchXslt*.

Table 1. Module overview

Name	Description
conversion	Convert DCCs into different formats (e.g., JSON)
extraction	Extract key data from DCCs (e.g., formulae, attachments)
signature	Verify and apply digital signatures
validation	Validate structure and content of DCCs (e.g., against the XSD or Schematron)
transformation	Apply transformations (e.g., XSLT) to create human-readable outputs

C. Implementation

Python was chosen as the implementation language, due to its overall popularity and widespread use in data science [8]. Most of the existing tools and libraries that are considered for the consolidation are written in Python, making it easier to port existing code to the new library.

For the implementation of the CLI the library *click* was used, because it provides annotations to define the command line interface in a declarative way and is POSIX-compliant [9]. Each module is accessible via dedicated subcommands, for example validating a DCC against the XSD or Schematron with commands as shown in Figure 2. These commands take arguments like the file to transform (`<dcc_file>`) and flags such as the Schematron file to use (`--schematron <schematron_file>`). Each flag has a short and long version, for example `-o` and `--output` for the output file.

```
dcclib transform xslt [-o,
--output <output_file>] <xslt_file>
<dcc_file>
```

```
dcclib validate schematron [-s,
--schematron <schematron_file>]
<dcc_file>
```

Fig. 2. Example of the CLI usage.

The REST API is built using the library *APIFlask*, which extends *Flask* to provide a simple way to create RESTful APIs [10]. It automatically generates an OpenAPI specification, which documents the API and can be used to generate client and server libraries in different programming languages [11]. Similar to the CLI, the REST API provides endpoints for each module. The REST design rules by Masse (2011) [12] are followed to achieve a consistent and intuitive API design. Because of this and the stateless nature of REST, the DCCs need to be temporarily stored on the server to process them. Figure 3 exemplarily shows three endpoints together with their HTTP methods. The first endpoint is used to upload and temporarily store

a DCC on the server. With the second endpoint, DCCs can be validated against the XSD. Files that are attached to a DCC can be extracted using the third endpoint.

```
POST /dccs
POST /dccs/<id>/validate/xsd
GET  /dccs/<id>/extract/files
```

Fig. 3. Example of the REST API usage.

D. Formula Module

The formula module is the most sophisticated component of the library. Internally, it is divided into two parts, each handling a specific task: one part is responsible for extracting formulae and variables from the DCC, while the other handles the conversion of Content *MathML* formulae into *SymPy* expressions.

Formula extraction

In the first step, the formula is extracted from the XML document. A machine-interpretable formula is typically provided in Content *MathML* [13], which represents mathematical expressions in Polish notation. This format is algorithmically efficient and easily parsed by machines.

In a DCC, formulas may also appear in Presentation *MathML* or LaTeX. However, since these formats are primarily intended for visual rendering rather than computational processing, they are disregarded during extraction.

Variable extraction via IDs

Content *MathML* references variables using an `xref` attribute, which links to unique identifiers (IDs). These IDs correspond to quantities defined elsewhere in the DCC. Due to the DCC's structured quantity hierarchy, extracting the actual numerical values is straightforward.

If a variable is part of a list structure within the DCC's quantity definition, it is preserved in the formula object. This ensures that the program handles the data correctly. A validation check ensures that either only a single value is provided per variable, in which case it is applied to every element of other variable lists, and all lists have the same length (as required in a valid DCC).

Conversion from Content MathML to SymPy

The conversion process transforms Content *MathML* into a *SymPy*-compatible expression, primarily by parsing the Polish notation into a string that *SymPy* [14] can interpret. When mathematical errors, invalid formulae and other problems are encountered, exceptions are thrown. These can be handled with try-catch-clauses.
The current implementation does not support calculations involving units and uncertainties, as the *SymPy* library provides limited or no support for these features. However, due to the modular design of *dcclib*, it is feasible to use a different backend for mathematical computations in the future. One option is the *dccQuantities* library, which supports arithmetic operations with both units and uncertainties [15].

Recursive parsing approach

Due to the nested nature of Polish notation, the conversion employs recursion, starting with the deepest (innermost) elements:

- **Function mapping:** A dictionary of function pointers (e.g., plus, sin, log) maps *MathML* operators to their corresponding *SymPy* functions.
- **Variable handling:** Variables are passed as parameters to these functions.
- **List support:** If a variable is a list, the parser iterates over each element, recursively processing them individually. The results are aggregated into a list for output.

The recursion unwinds outward, combining intermediate results until the full *SymPy* expression is constructed. This method ensures accurate representation of complex mathematical operations while maintaining computational efficiency. Multi-dimensional arrays or matrices are not supported in the current implementation, but support for them in the future is feasible.

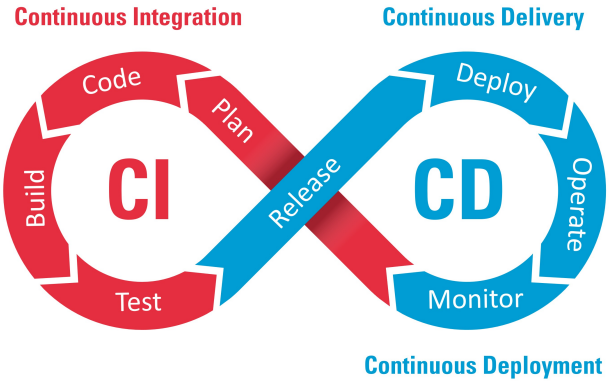


Fig. 4. Continuous integration and continuous delivery workflow.

E. Quality Assurance

To ensure the quality of the code, unit-, integration- and end-to-end tests are implemented using *pytest*. These tests cover all modules and functionalities of the library and are automatically run on every change using GitLab CI/CD, in the phase Test in Figure 4. For example, in the formula extraction and conversion process, the tests verify that the

extracted formulae are correctly converted into *SymPy* expressions and that the calculated values match the expected results.

F. Distribution

The library and its interfaces are distributed through several channels to provide broad accessibility to its functionalities. The library itself and the CLI are available on the GitLab package registry and can be installed via package managers like *pip* or *uv*.
Docker images are provided for the REST API and the CLI on the GitLab container registry. This allows users to run the REST API and CLI in isolated environments without the need for a local installation. Another use case is using the Docker image for the CLI in CI/CD pipelines for example, to validate examples against the XSD or Schematron.
For use cases where no Python installation is available, the CLI is also distributed as a standalone executable which is built using the library *PyInstaller*.
All of these distribution methods are performed after testing and building the library and its interfaces in the Deploy phase of the CI/CD pipeline, as shown in Figure 4.

G. Limitations

Although *dcclib* provides a consolidated approach to DCC processing, some limitations must be acknowledged.

Conversion limitations

Currently the conversion module only supports conversion to JSON. This can be extended in the future to support other formats, such as CSV or YAML, but is not yet implemented. More importantly, the XML to JSON conversion process has inherent limitations, particularly the loss of element ordering information. This only happens when differently named nodes are on the same level in the XML hierarchy, which are then grouped together in arrays in the JSON output.

REST API Security

Currently, the REST API does not implement any security measures such as authentication or rate limiting. This limitation is known and can be addressed by deploying the REST API behind a reverse proxy that handles these aspects. The REST API also requires temporary storage of the DCCs to process it. This is achieved by storing the DCCs in a temporary directory on the server, from which they are deleted after a configurable time period (five minutes by default). At present, there are no measures in place to stop denial-of-service attacks caused by the uploading of large files or a high volume of files in a short period of time. In its current state, the REST API is intended for non-public use cases, such as integration into internal workflows or systems.

III. CONCLUSION

A. Summary

This work presented the approach to a consolidated toolset that provides core functionalities to work with DCCs. After analysing the current state of available tooling options that revealed a fragmented landscape, a modular approach has been developed with the focus on functional consolidation as well as consistently high software quality. The result of the work is the library *dcclib* that is structured into functional modules that address at the moment five core functionalities required for processing DCCs. The work includes implementation aspects such as the choice of Python as programming language, testing as an important part of quality assurance as well as the interfaces that are built-in components of the libraries to enable integration in existing workflows and environments. The aspect of distribution is covered as well addressing different distribution pathways including GitLab, Docker images and executables.

B. Outlook

As the development and introduction of the DCC progresses, additional functionalities will come into focus to facilitate the work with these documents. Thus, future work will focus on the extension of *dcclib* to support additional DCC functionalities, such as enhanced formula evaluation and more extensive format conversion options. Also, functionalities and libraries to facilitate handling of tables in the DCC or working with quantities could be incorporated and provided as part of the consolidated library. Further developments could also include the integration of a graphical user interface (GUI) to make the library even more accessible to non-technical users which is important for better understanding and easier uptake of the DCC.
The library is planned to be released as open source software, enabling the community to contribute to its development and improvement. Their contributions are highly appreciated and will be facilitated through a public Git repository. The packages will be distributed via Python Package Index (PyPI) and the Docker images via *Docker Hub*, making the work available to a wider community through standard channels.

IV. CITATIONS AND REFERENCES

[1] Siegfried Hackel et al. “The fundamental architecture of the DCC”. In: *Measurement: Sensors* 18 (2021), p. 100354. ISSN: 2665-9174. DOI: 10.1016/j.measen.2021.100354.

[2] Siegfried Hackel et al. “The Digital Calibration Certificate”. In: *Metrologie für die Digitalisierung von Wirtschaft und Gesellschaft* (2018). DOI: 10.7795/310.20170403.

[3] *GEMIMEG-I*. URL: https://www.gemimeg.ptb.de/en/gemimeg-i/ (visited on 07/15/2025).

[4] *GEMIMEG-II*. URL: https://www.gemimeg.ptb.de/en/gemimeg-home/ (visited on 07/15/2025).

[5] Siegfried Hackel et al. "The Digital Calibration Certificate (DCC) for an End-to-End Digital Quality Infrastructure for Industry 4.0". In: *Sci* 5.1 (2023). ISSN: 2413-4155. DOI: 10.3390/sci5010011.

[6] Tim Ruhland Andreas Tobola and Benedikt Seeger. *Python library supporting automated processing of digital calibration certificates (DCC)*. URL: https://github.com/siemens/pydcc (visited on 07/15/2025).

[7] Héctor Laiz et al. "2nd international DCC-Conference 01-03 March 2022 Proceedings". In: *PTB-OAR*. Physikalisch-Technische Bundesanstalt (PTB). 2022. DOI: 10.7795/820.20220411.

[8] JetBrains. *The State of Developer Ecosystem*. Tech. rep. JetBrains s.r.o., 2023. URL: https://www.jetbrains.com/lp/devecosystem-2023/ (visited on 03/24/2025).

[9] *Why click? — Click Documentation*. URL: https://click.palletsprojects.com/en/stable/why/ (visited on 03/24/2025).

[10] *APIFlask — APIFlask Documentation*. URL: https://apiflask.com/ (visited on 03/24/2025).

[11] Darrel Miller et al. *OpenAPI Specification v3.1.0*. Feb. 2021. URL: https://spec.openapis.org/oas/v3.1.0.html.

[12] M. Masse. *REST API Design Rulebook: Designing Consistent RESTful Web Service Interfaces*. O'Reilly Media, 2011. ISBN: 978-1-4493-1990-8.

[13] Robert R Miner, Patrick D F Ion, and David Carlisle. *Mathematical Markup Language (MathML) Version 3.0 2nd Edition*. W3C Recommendation. https://www.w3.org/TR/2014/REC-MathML3-20140410/. W3C, Apr. 2014.

[14] Aaron Meurer et al. "SymPy: symbolic computing in Python". In: *PeerJ Computer Science* 3 (Jan. 2017), e103. ISSN: 2376-5992. DOI: 10.7717/peerj-cs.103. URL: https://doi.org/10.7717/peerj-cs.103.

[15] *dccQuantities*. URL: https://gitlab1.ptb.de/digitaldynamicmeasurement/dccQuantities (visited on 07/15/2025).

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Development and Application Study of a Calibration Certificate Anomaly Detection System

Dong-Hun Ryu[1], Chae-Wook Lim[1]

[1] *Korea Testing Laboratory, 723 Haean-ro Sangnok-gu Ansan-Si, Gyeonggi-do, 15588, South Korea, dhryu@ktl.re.kr, hicor1@ktl.re.kr*

This study focuses on the development of a system to effectively extract and utilize data from approximately 3,000 types of unstructured calibration certificates (e.g., paper, PDFs) widely used in the calibration industry. To process document-based data that are difficult for machines to recognize, the Korea Testing Laboratory (KTL) has developed a calibration certificate anomaly detection system called *K-Argos*, which applies object detection AI technology based on Convolutional Neural Networks (CNNs).
The core function of the system is to extract data from unstructured documents, convert them into structured data, and standardize and store the results.
Using the converted structured data, *K-Argos* provides statistical data management features such as performance analysis of calibrated equipment and anomaly detection in calibration data. Additionally, the system can be used to correct inaccurate identification information (e.g., manufacturer, model, serial number) and to analyze errors within the calibration data.
In the future, the *K-Argos* system will be expanded to include an online technical supervisor service that automatically determines anomalies in data during the certificate approval stage and provides relevant information to the approver (technical supervisor). This is expected to significantly improve the efficiency and reliability of calibration tasks and contribute to the digital transformation of the calibration industry through the implementation of Digital Calibration Certificates (DCCs)

I. INTRODUCTION

Korea Testing Laboratory (KTL) is a public institute under the ministry of trade, industry and energy of South Korea and in charge of test and certification for industries. Among the nine divisions, the industrial standards division is performing calibration of measuring instruments, legal metrology and reference materials. The division issues approximately 120k calibration certificates annually and recently developed a tool to digitalize and utilize the calibration data to enhance the quality of calibration service.

The main works can be summarized as follows.

1. DB accumulation using OCR
2. Grouping similar results with identifier (manufacturer, model, S/N) correction
3. Grouping similar results with nominal value estimation
4. Analysis using the accumulated results

Two main technologies are OCR and grouping.

Currently, it is planned to develop the following services further to improve the internal efficiency and customer experience.

1. Implementation of the developed anomaly detection system in the certificate approval process
2. The client service to digitalize paper certificate
3. Related terms search feature

II. DB ACCUMULATION USING OCR

In case of KTL, the first page of calibration certificates contains the information required by ISO 17025 [1] such as title, the name and address of the client, the location of the lab, etc. It is relatively easy to make a structured database because the format is similar in all fields. But it is difficult to accumulate the data and manage it as database from the second page because each tester uses his/her own format and software (word, excel, etc.) to express the results and test method.

To develop a service with calibration results, it is preferred to have enough data. But there was a lack of human resources to make a database of legacy certificates already issued, manually. As an alternative, we developed an OCR tool for data extraction from pdf certificates of the past, which utilizes the object detection AI based on convolutional neural network. 120k certificates can be categorized to 3000 formats and we trained an AI model for most common 100 formats and made a database of 100k+ certificates. The figure 1

shows that various formats are used to show the results. While the first page always uses similar format, the second page contains different number of data in different tabular forms. This causes difficulties in digitalizing the certificate using OCR, because the number and location of target data changes. Coding each case is also very inefficient because too many cases exist. The figure 2 shows the developed OCR tool.

The procedure of OCR tool can be divided into two steps. The first step is a table detection, and we used the Cascade TabNet model.[2] From the 100 formats, 1000 cases were generated by image distortion and used for learning. The table detection was very successful, and the image was converted to low resolution, black and white image to lower the cost of second step using LLM. The second step is the table structure recognition, and we used GEMMA 3[3] and EXAONE[4]. GEMMA 3 is open-source AI model developed by GOOGLE and EXAONE is open-source AI model developed by LG AI Research. EXAONE has competency in understanding Korean language and GEMMA 3 and EXAONE are used via LangChain.[5] The data extracted by GEMMA 3 are sent to EXAONE and recognized as nominal value, reference, indication, error and uncertainty.

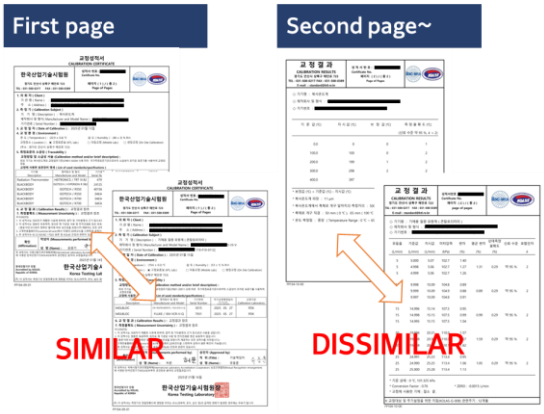


Fig. 1. Example of calibration certificates that shows the format difference of results page

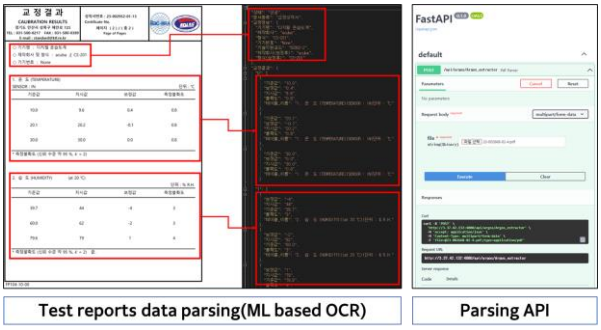


Fig. 2. Developed Machine Learning based OCR and API

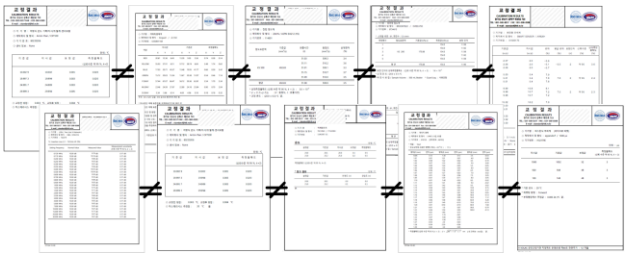


Fig. 3. Example of calibration certificates that shows the various formats depending on measurement field

Figure 3 shows that a very wide range of formats are used depending on the measurement field, and different labels are used quite often. For example, "the conventional mass" is the label used for the "reference value" in mass weight certificate and sometimes table has different structure and merged cells. We focused on realizing database establishment of past certificates using OCR and testing possible service with this DB. The success rate is difficult to know because we could not test all 3000 formats. As a first step of applying OCR and DB establishment, we focused on the most used formats, and it was successful for 100 formats, and we could make a database of 100k+ certificates. This required a lot of prompt engineering to make it understand that "conventional mass" can also mean "reference value" sometimes. We are planning to expand the applicable formats in future, and some formats are very difficult to recognize and discussion with field experts for standardization is needed.

III. IDENTIFIER CORRECTION

To analyze the results, it is essential to identify and group the results. The manufacturer, model and serial number are very good identifiers of measurement instruments, but we encountered some problems with grouping based on the forementioned identifier. There were many typos or different ways of expressing it. For example, a viscometer manufacture "BROOKFIELD" was sometimes expressed as "Brookfield", "brook-field" and model "QL200-MK" was "QL200 MK" or "QL200MK" This small difference of capital letter, dash and space made a trouble in identifying instruments and caused errors and inefficiencies.

So, we developed similarity search system to make a correction of identifier to make it possible to find the name even when no exact matches are found. This also could be applied when the number of strings is different, or there are typos in two words. The following table shows the correction examples.

Table 1. Correction example

As is	Corrected
BROOKF**EI**LD	BROOKFIELD
TRANSFO**MM**ER	TRANSFOMER
PO**V**VER D**II**AGNOSTIX	POWER DIAGNOSTIX

In developing this similarity search system, we used our database which has 70000 manufacturers and model names based on past certificates. After correcting the blank and capital letters, the data is categorized into the names and then the most used name is identified as the right manufacturer name. For example, if the most common name is “BROOKFIELD”, then “Brook-field”, “BROOKFIELD©”, “Brookfield™” will be recognized as the most common name and will not be recommended either.

The reason we mention “recommendation” is that we used this similarity search not only for grouping the calibration results data but also for recommending in self-query for customers. When customers request a query, this can recognize the misinput of manufacturer or model name, and shows the right results.

To accelerate the search, the algorithm works in sequential steps. Exact match is performed first which is the fastest, and then partial match (%like%) search comes next. It calculates cosine similarity, and we used the threshold value of 0.8. Average search time is 30 ms.

IV. NOMINAL VALUE ESTIMATION

After the grouping of the calibration certificates of the same manufacturer and model and serial number, it is required to group the values to compare the results. Reference value varies slightly, and it is sometimes difficult to make a group based on the value. Nominal value is more distinguishable and helpful when making a group of similar range results.

Additionally, it was found that each field uses different words in the certificate such as “reference value”, “conventional mass”, etc. for same meaning. For practical reasons, we decided to minimize the database category and standardized DB names. In case of 100 formats, we identified only “setting value”, “reference value”, “indication value”, “corrector or error” and “uncertainty”

V. ANOMALY DETECTION SYSTEM

We developed a calibration support tool with this data. The figure 4 is a screenshot of a system being developed by the end of 2025. From the database of calibration certificates of 5 years, annual reports for the same device or same model are sorted and analyzed statistically to detect anomalies. It shows which calibration certificate is suspected to have a typo in manufacturer, model or serial number. The right rectangle is for the case when measurement error seems to be too large when compared to the same model results. The technical manager or calibration staff can click a certificate number and check the details.

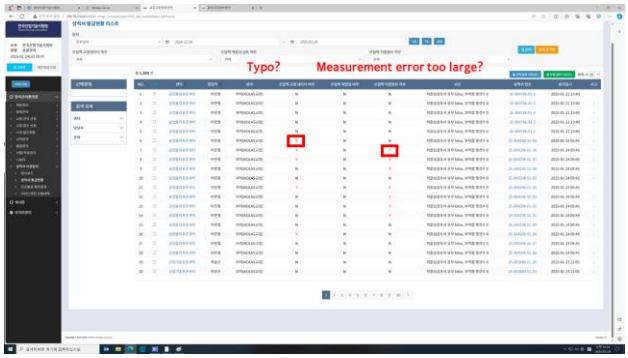


Fig. 4. Screenshot of anomaly detection system

When someone clicks a certificate number, the screen changes to the details screen, and shows the results as shown in the figure 5. This screen is an example of a torque wrench, and there were 125 results of same model in the database as shown in the figure 6. The bottom left graph shows the previous results of same serial number, which means, the same instrument. The bottom right graph shows the guideline range of the same model, and the dots in the middle show that the results are within the range.

This system is being developed for internal use and in the beginning stage. The guideline range was setup as 75 % and 25 % of previous same model results arbitrarily for now, and we plan to get feedback from the staff to find better ways in guideline set-up next year.

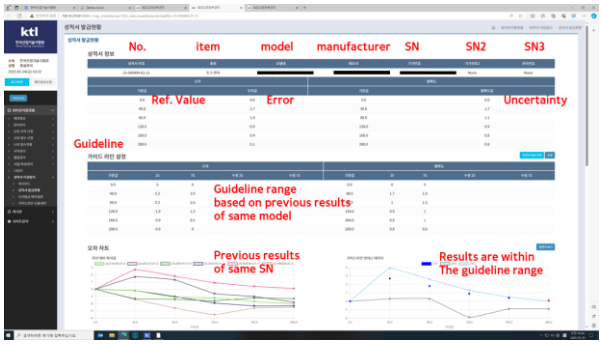


Fig. 5. Details screenshot of anomaly detection system

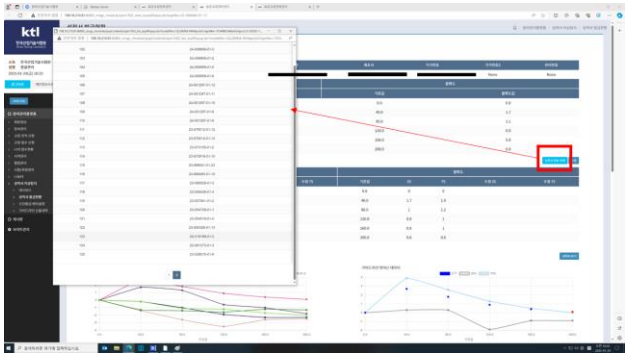


Fig. 6. The list of certificates used in the guideline setup

One more example is a micro pipette in figure 7. There is only one line in the left bottom, and it means there is no record of same pipette for the last 5 years. The bottom right graph is guideline based on the results of same model and the number of certificates used was 484. The red dot is outside the range and possibly it could be typo or test error or instrument failure. The technical manager can check it, and we think it would be a good start to give people a clue to detect anomaly. As mentioned before, the guideline range needs to be elaborated and currently manual input by each field expert is activated with default value of 25 % and 75 % value of previous results

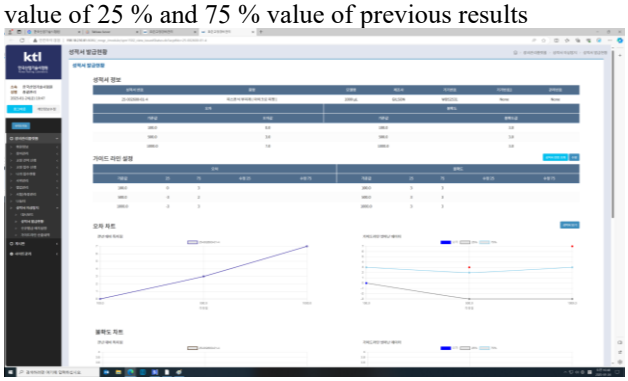


Fig. 7. Example for results outside of guideline

VI. FUTURE PLAN

A. Pre-check service

Usually, the calibration follows the (1 ~ 5) steps of the following figure. It is sometimes difficult to check all calibration certificates and every result and model and manufacturer in it because of workload. Our plan is the implementation of the anomaly detection system between step 3 and step 4 so that the calibration staff or the technical manager can easily detect the calibration certificate that seems suspicious and focus on it to check. The system is named K-argos because it is used to monitor errors in calibration certificates. Argos is a giant with 100 eyes in myth and often used for the name of monitoring system, and K is from KTL.

1) A calibration staff performs calibration
⇩
2) Prepares a calibration certificate.
⇩
3) Submit a calibration certificate for approval.
⇩
+ Certificate Pre-Check Service :
detects errors in calibration data (incorrect results, equipment information, certificate forms, etc.) and notifies key review points
⇩
4) The technical manager reviews the certificate and makes the final approval.
⇩
5) The approved certificate is provided to the customer.

Fig. 8. Pre-check service scheme

B. Related terms search feature

Though some different expressions were found in the certificates (QL200-MK, QL200 MK, etc.), much wider range of expressions are used in the field. This causes difficulty when a client wants to find the right category for his/her instrument to register or request a quote. We mentioned that the similarity search was useful not only in grouping results but also in the customer query process in section III. We are going to expand the function to the related terms search feature and the idea is that we are going to make an expanded database using LLM. We let KeyLLM analyze the database and extracts related keyword pairs. Then KeyLLM expands the database including related keywords, and this is used in the search process when exact or similar match is not found. For example, “nogisu” is often used in the construction field of Korea, and the related terms search feature can recognize conceptual similarity and recommend “vernier calipers”.

C. Certificate stream service

We are reviewing the plan to offer the clint the OCR technique used in the digitalization of historical calibration certificate. With the digitalized calibration certificates inside KTL, it will be possible to provide a certificate stream service to the customer. The following figure shows the schematic diagram of the service. Currently, the KTL network is isolated for security reasons, and we made “k-tools” website for customer service. The plan is to establish APIs, and a digital calibration certificate will be provided if authorized customers log in to the k-tools and request it. The customer can digitalize his/her own calibration certificates in pdf form and store them in k-tools for

convenience. However, it is difficult to guarantee the success of xml conversion because user created OCR pdf certificates will have very wide variety. One possible plan is to provide excel conversion of pdf for convenient data handling with limited liability.

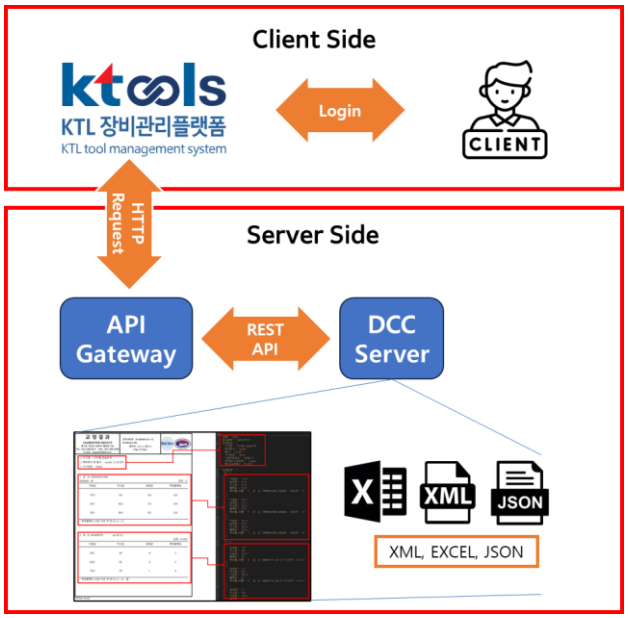


Fig. 9. Certificate stream service scheme

VII. CONCLUSION

KTL issues 120k+ calibration certificates every year and plan to digitalize certificates and make use of results. We trained AI to extract data from past pdf certificates and accumulated DB. Then, we made a correction module to identify manufacturer & model & SN. The data were sorted based on the nominal value and calibration support system was developed.

We plan to implement the calibration support system in the internal system for staff, and provide the certificate stream service and the related terms search service for external customers.

REFERENCES

[1] ISO/IEC 17025:2017 General requirements for the competence of testing and calibration laboratories

[2] https://github.com/DevashishPrasad/CascadeTabNet

[3] https://deepmind.google/models/gemma/gemma-3/

[4] https://www.lgresearch.ai/exaone

[5] https://www.langchain.com/

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Checking Traceability by Analyzing DCC Blockchain Transactions

Cristian Zet[1], Gabriel Constantin Dumitriu[2], Foșalău Cristian[1]

[1,3] *Gheorghe Asachi Technical University of Iasi, D Mangeron, Blvd. 23, Iasi, Romania, czet@tuiasi.ro*
[2] *Individual Enterprise GCD, Hlincea str. 27, Iasi, Romania, dumitriu.gabi@gmail.com*

***Abstract* – Traceability is essential in metrology, linking measurement results to primary standards through documented calibration chains. While blockchain ensures secure, immutable storage of Digital Calibration Certificates (DCCs), retrieving traceability paths remains complex. This paper proposes a method to reconstruct calibration traceability by analyzing blockchain transactions recorded in a private, permissioned network. We describe an architecture combining PHP-based asset management, LabVIEW DCC generation, and recursive graph analysis to extract calibration dependencies. Realistic scenarios illustrate the approach, demonstrating improved auditability, reduced reliance on physical certificates, and secure, decentralized traceability management that aligns with modern digital metrology requirements.**

I. INTRODUCTION

Metrological traceability, defined as the unbroken and documented chain of calibrations that links a measurement result to a reference standard, is a cornerstone of confidence in measurement systems [1]. Ensuring traceability is fundamental, not only for scientific reproducibility but also to industrial quality control, trade, legal metrology, and healthcare. Traditionally, this traceability chain has been maintained using physical or semi-digital calibration certificates stored in local metrology labs, or more often, in isolated, non-interoperable databases.

The recent digital transformation of metrology is oriented toward Digital Calibration Certificates (DCCs) and machine-interpretable representations of measurement metadata. Initiatives such as SmartCom [2], EMPIR 17IND02 (SmartCom project) [3], and the Digital SI framework proposed by the BIPM [4] emphasize the importance of using structured data standards (e.g., XML, JSON, RDF) to increase the automation, portability, and auditability of calibration information.

Despite these advances, the verifiability and long-term integrity of DCCs remain a problem, especially when data must be exchanged across organizational or national boundaries. In this context, blockchain technology, with its properties of distributed consensus, cryptographic immutability, and decentralized trust, has gained significant attention as a mean to prove calibration data in a much safer and transparent way [5]. Prior work, including our own system for blockchain-enabled calibration data management [6], has shown that calibration certificates can be securely registered on a proprietary blockchain network by referencing their content via cryptographic hashes and metadata, providing immunity to data manipulation and timestamping.

However, immutability alone does not equate to traceability. While blockchain ensures that data cannot be altered later on, it does not provide mechanisms to semantically connect and reconstruct the traceability chain from an instrument under test to its ultimate reference standard. This interpretive layer is essential in metrological contexts, where traceability must be demonstrated, verified, and validated, for example, during audits or regulatory inspections.

Several researches have explored the use of blockchain networks to record calibration activities in a tamper-resistant manner, thereby reinforcing the traceability of measurement results. Takatsuji et al. [7] introduced a blockchain-based system for visualizing the entire calibration chain, facilitating end-to-end traceability from measurement instruments to primary standards. Building on this concept, Takegawa et al. [8] proposed a traceability management system leveraging Ethereum smart contracts to allow users instant access to calibration data, thus improving transparency and reliability. Miličević et al. [9] further developed the idea of a digitally transformed traceability pyramid, underlining the role of blockchain in maintaining trust in hierarchical calibration structures. In a previous paper [6] we presented a system that automates the calibration workflow and records the DCCs in a private blockchain, claiming the possibility for ensuring the traceability. These advancements demonstrate the possibility of integrating blockchain technology into metrological infrastructures. The present paper addresses the problem of tracing and validating calibration dependencies across our private blockchain network, thereby enabling interpretable and verifiable metrological traceability.

II. BLOCKCHAIN NETWORK ARCHITECTURE

In a previous paper we implemented a dedicated blockchain network which integrates Digital Calibration Certificates (DCCs) into transactions.

The architecture employed is a private, permissioned blockchain network composed of several nodes, an application programming interface (API) layer and a web-based smart asset management platform. Each node functions as a decentralized server that stores a synchronized copy of the blockchain ledger, validates new transactions, and participates in the consensus protocol. Once a transaction is validated, it is recorded into a data block, which contains a hash of the previous block, a timestamp, and the transaction data. These blocks are chained together sequentially, forming a verifiable and unchangeable timeline of calibration events. The consensus mechanism is based on the X15 algorithm, which combines features from both Proof-of-Work (PoW) and Proof-of-Stake (PoS). This dual mechanism provides a balance between energy efficiency and security and the use of 15 chained hashing algorithms in the X15 model increases resistance to cryptographic attacks while allowing deployment even on less powerful platforms as Raspberry Pi 2.

The security and the traceability are further enhanced by the use of native blockchain wallets. These wallets provide full access to transaction history, monitoring, and asset management. The network operates over designated ports—peer-to-peer communication on port 10218 and remote procedure calls via port 20208—adding an additional layer of operational control.

For the integration of instrument calibration, a web-based asset management tool was developed and a LabVIEW-based calibration application. Through a dedicated function, a unique blockchain address is generated for each instrument using the "*getnewaddress*" command. This address, along with DCC's metadata is used to create a digital identity that is securely identifes the device. During calibration, the LabVIEW application formats the DCC data into an XML structure, which is then embedded into a Blockchain transaction. The transaction is initiated using the "*sendtoaddress*" command, while the metadata is stored within the optional comment field.

Each calibration event results in a transaction that is permanently recorded on-chain. The traceability chain is possible to be retrieved by querying the blockchain through wallet functions, which enables users to search for previous calibration records of the same instrument. This mechanism allows the system to retrieve the complete history of verifications and measurements associated with a specific device.

III. METHODOLOGY AND SCENARIOS

This present work introduces a methodology designed to reconstruct and verify the metrological traceability chain of a measuring instrument by analyzing calibration events that have been digitally recorded and stored in a private blockchain network. Each Digital Calibration Certificate (DCC) corresponds to a calibration event, which is stored as blockchain transaction and cryptographically signed by an accredited calibration laboratory. The central hypothesis of this approach is that the relationships between measuring instruments and their calibration references, as recorded through these transactions. Each DCC serves as a directed connection between two instruments: the instrument under test (IUT) and the reference standard used during calibration. Each calibration transaction is thus formalized as a record containing the identifiers of both the IUT and the reference instrument, the timestamp of the calibration event, and associated metadata defining a directed dependence in the traceability graph, pointing from the calibrated instrument toward the reference standard, indicating a dependency of metrological quality.

The traceability application is built in PHP as a modular architecture consisting of five functional components that work together to provide the complete traceability verification. For testing it several scenarios were possible, as described in the following.

A. Simple linear scenario

The simplest scenario is a simple calibration chain from the working instrument (VM1) to the national standard (REF0), via a secondary standard (REF1). For this case, 2 transactions are recorded in the network (Table 1).

Table 1. Simple linear scenario

DCC ID	Instrument ID	Standard ID	Date
T 001	REF1	REF0	12.12.2024
T 002	VM1	REF1	15.02.2025

The software should trace from VM-A to REF-1 to REF-0, building a traceability graph with depth 2.

B. Multiple calibration scenario with diverging standards

This scenario is presented in Table 2.

Table 2. Diverging standard scenario

DCC ID	Instrument ID	Standard ID	Date
T 001	REF1	REF0	12.12.2023
T 002	VM1	REF1	15.02.2024
T 003	REF2	REF0	15.09.2024
T 004	VM1	REF2	1.02.2025

In this case, the instrument has been calibrated with 2 different standards for last 2 times. First time VM2 was

calibrated with secondary standard REF1 and second time with REF2. Both standard instruments were calibrated with national standard REF0. Both chains trace back to the same national standard REF0. The last calibration must be filtered by the date.

C. Branching Tree with Intermediate Levels

In this scenario an intermediate standard is used for multiple instruments calibration. Several instruments VM1, VM 2 and VM 3) were calibrated with diferent intermediate standards (REF2 and REF3). These were calibrated with an secondary standard REF1, which was at its turn calibrated with the national standard REF0. Table 3 presents this scenario.

Table 3. Branching tree scenario

DCC ID	Instrument ID	Standard ID	Date
T 001	REF1	REF0	12.12.2024
T 002	REF2	REF1	15.12.2024
T 003	REF3	REF1	1.01.2025
T 004	VM1	REF2	1.02.2025
T 005	VM2	REF3	5.04.2025
T 006	VM3	REF2	7.06.2025

All 3 instruments trace back to REF0. The standard REF1 acts like a shared note. If we want to check its backward traceability we will find 2 branches, but in real applications could be many branches.

D. Full scale metrological chain

The working instrument (VM1) is calibrated with an working standard (REF3). This was calibrated several times with secondary standards (REF2 and REF1). These are, at their turn, calibrated with the national standard REF0. The national stadard was calibrated based on an international travelling standard ITS. At its turn, this was part of an interlaboratory comparison campaign ILCC. This is presented in Table 4.

Table 4. Full scale metrological scenario

DCC ID	Instrument ID	Standard ID	Date
T 001	ITS	ILCC	12.12.2024
T 002	REF0	ITS	15.12.2024
T 003	REF1	REF0	1.01.2025
T 004	REF2	REF1	1.02.2025
T 005	REF3	REF1	5.03.2025
T 006	VM1	REF2	7.06.2025
T007	VM1	REF3	10.12.2025

IV. IMPLEMENTATION OF PHP BLOCKCHAIN TRACEABILITY VERIFICATION SYSTEM

This chapter presents the implementation and architectural details of the PHP-based blockchain traceability verification system designed to validate and track Digital Calibration Certificates (DCCs) stored on distributed ledger infrastructure. The system serves as a critical component in the proposed blockchain-enabled metrology framework, providing comprehensive verification capabilities for calibration hierarchies and certificate authenticity.

The PHP Traceability Checker implements a modular architecture consisting of five core functional components that work synergistically to provide complete traceability verification. The system employs a recursive traversal algorithm to construct calibration hierarchies while maintaining security through comprehensive input validation and error handling mechanisms.

Each module serves specific roles in the traceability validation process. The primary processing engine coordinates between transaction retrieval mechanisms, chain construction algorithms and device certification validation protocols. The application flow diagram is presented in Figure 1.

The main processing function (*processTraceabilityCheck()*) serves as the system's orchestration layer, accepting blockchain addresses, maximum traversal depth parameters, and network configuration objects. This component implements comprehensive address validation protocols before initiating transaction retrieval and chain construction processes. The function returns structured data containing complete traceability information and certified device listings.

The transaction retrieval component (*getAddressTransactions()*) implements an optimized data access layer capable of processing up to 2000 transactions per blockchain address. The module incorporates intelligent caching mechanisms to minimize redundant network requests while maintaining data freshness. Error handling protocols ensure system resilience when encountering network failures or malformed transaction data.

The chain construction module (*buildTraceabilityChain()*) implements a sophisticated recursive algorithm designed to traverse calibration hierarchies from individual devices to primary measurement standards. The algorithm employs depth-limited traversal with cycle detection mechanisms to prevent infinite recursion while ensuring complete hierarchy mapping. The function utilizes advanced pattern matching to identify calibration relationships through transaction comment analysis, specifically targeting "DCC:calibrated by" patterns that indicate hierarchical relationships.

The recursive implementation maintains a visited

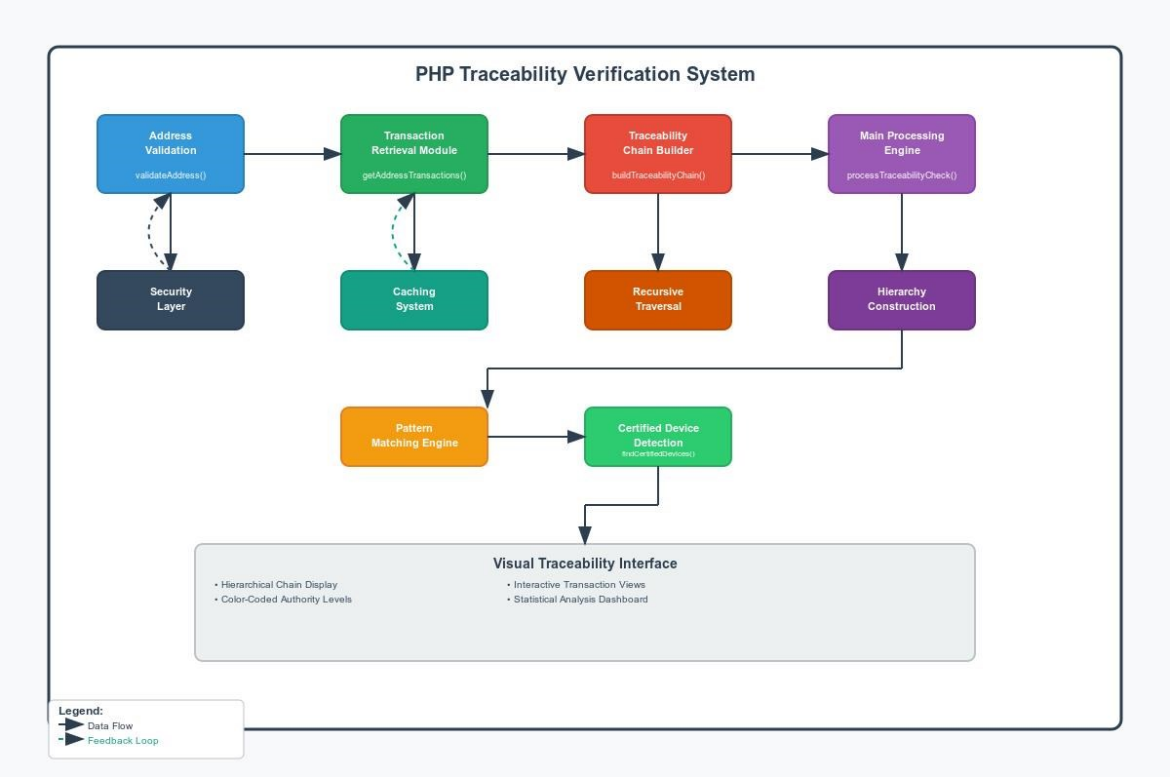


Fig. 1. Flow diagram.

address registry to prevent circular references while constructing a comprehensive authority hierarchy. Each iteration processes transaction data to identify upstream calibration authorities, creating a complete genealogy of measurement traceability that extends to primary national standards.

The device certification detection module (*findCertifiedDevices()*) implements enhanced logic for identifying measurement instruments that have been certified by specific calibration authorities. The system applies multi-criteria validation protocols that evaluate transaction amounts, comment pattern matching, and confirmation status to determine certification validity.

Certification criteria include transaction amount thresholds exceeding unity values, presence of "DCC:SerialNo" patterns in transaction metadata, and satisfaction of minimum blockchain confirmation requirements. The module processes both outbound and inbound transaction types to provide comprehensive coverage of certification activities.

The address validation component (*validateAddress()*) implements comprehensive security protocols through regex-based format validation and input sanitization mechanisms. This module serves as the primary defense against injection attacks and malformed input data, ensuring system integrity throughout the verification process.

The system incorporates sophisticated regular expression engines for transaction analysis, implementing specialized patterns for DCC serial number identification and calibration relationship detection. These patterns enable precise extraction of metrology-specific information from blockchain transaction metadata.

The verification system provides detailed transaction analysis capabilities including statistical summaries, confirmation status monitoring, temporal analysis, and quantitative calculations. This comprehensive approach ensures complete visibility into calibration activities and their associated blockchain records.

The system implements a visualization framework featuring hierarchical chain displays with color-coded authority levels distinguishing between primary standards and intermediate calibration devices. Interactive elements provide detailed transaction examination capabilities while statistical dashboards offer quantitative analysis of calibration activities (Figure 2).

V. CONCLUSIONS

The PHP verification system serves as a critical component in the proposed blockchain-enabled metrology framework, providing essential capabilities for DCC authenticity verification, calibration authority tracing, certified device monitoring, and comprehensive audit trail generation. This integration ensures regulatory compliance and quality assurance in metrology applications.

The implementation incorporates comprehensive error

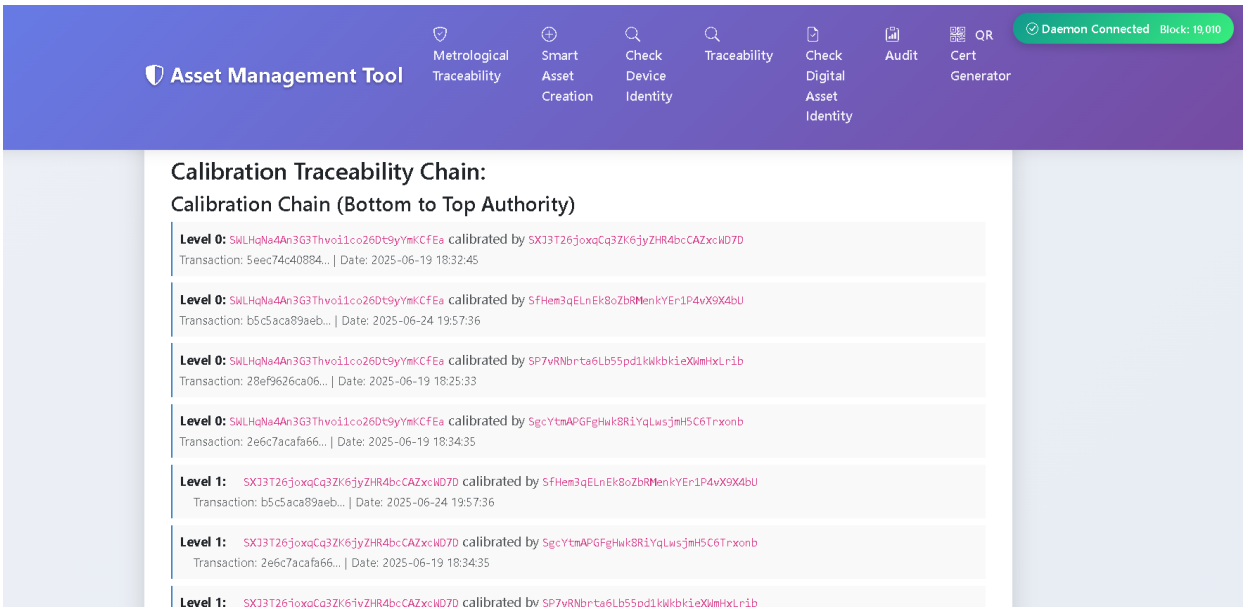


Fig. 2. The web application – traceability function

handling, intelligent caching mechanisms, confirmation-based validation and detailed logging systems to ensure robust operation in production environments. These features collectively provide the reliability and performance characteristics required for critical metrology applications.

The PHP Blockchain Traceability Verification System represents a comprehensive solution for validating Digital Calibration Certificates within distributed ledger environments. Through its modular architecture, advanced algorithms, and robust security framework, the system provides the verification infrastructure necessary to support blockchain-enabled metrology applications while maintaining the highest standards of reliability and regulatory compliance.

REFERENCES

[1] BIPM, International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM), 3rd Edition, 2008.

[2] Rauscher, E., & Hoffmann, J. (2020). "The SmartCom Project – Towards Digital Calibration Certificates." IMEKO TC6 Proceedings.

[3] EMPIR 17IND02 SmartCom, "SmartCom Final Report," EURAMET, 2021.

[4] BIPM, "Digital SI Framework," Digital Transformation Working Group (DTWG), 2022

[5] Nakamura, T., et al. "Blockchain for digital calibration certificates." IEEE Sensors Journal, 20(5), 2020

[6] C. Zet, G. C. Dumitriu, C. Fosalau, G.C. Sarbu (2022). "A software tool to support digital calibration certificates based on blockchain." *ACTA IMEKO 12(1):1-7, 2023*

[7] Toshiyuki Takatsuji, Hiroshi Watanabe, Yuichiro Yamashita, Blockchain technology to visualize the metrological traceability, Precision Engineering, Volume 58, July 2019, Pages 1-6

[8] Takegawa, T., & Furuichi, Y. (2023). Traceability Management System Using Blockchain Technology. Sensors, 23(3), 1673

[9] Miličević, K.; Tolić, I.; Vinko, D.; Horvat, G. Blockchain-Based Concept for Digital Transformation of Traceability Pyramid for Electrical Energy Measurement. Sensors 2022, 22, 9292.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Automatic Method of Assessment and Grading of Students in a Measurement Laboratory

Anamaria Hariton[1], Cristian Zet[1], Cristian Foşalău[1]

[1] *"Gheorghe Asachi" Technical University of Iasi*
Faculty of Electrical Engineering
Iasi, Romania
anamaria.hariton@tuiasi.ro, czet@tuiasi.ro, cfosalau@tuiasi.ro

***Abstract* – High activity levels in measurement laboratories can often exceed the time allocated for individual laboratory sessions. These laboratories involve a wide range of measurement tasks—from simple to complex—while students operate under strict time constraints. Consequently, the manual evaluation of student performance by instructors becomes labor-intensive and time-consuming. The digitization of the assessment process offers an effective solution. This paper presents a method that integrates LabVIEW with an SQL database to automatically record, verify, and grade student performance in laboratory sessions. The proposed method includes automated data saving for each student, validation of measurement accuracy, and objective grading mechanisms. Various tools are employed for real-time data acquisition, verification, and automatic evaluation of student activities.**

I. INTRODUCTION

Technical universities distinguish themselves through specialized laboratory sessions for experimentally proving the theoretical knowledge. The diversity and complexity of laboratory exercises are crucial in academic programs, but they also increase the difficulty of the assessment process for instructors. As technology evolves, so does the complexity of laboratory tasks—especially in measurement laboratories—requiring educators to:

a) verify the measurement data collected by students;

b) check the correctness of parameters calculated by students;

c) assess the conclusions formulated at the end of each measurement task.

Digital tools for learning assessment support the structured documentation of student performance. This organization enhances the instructor's ability to analyze stored data for planning future teaching activities and improving student outcomes. As outlined by Pellegrino & Quellmalz (2010) [1], the key contributions of digital technologies to learning assessment include: immediate feedback, enhancement of student-centered learning, support for collaborative learning, broader measurement capabilities, improved response flexibility, increased efficiency, reduced instructor workload, and better student performance. Moreover, they facilitate the integration of formative and summative assessment strategies [2].

Several platforms have already been developed to assist educators with grading. For instance, the "BetterExaminations" platform optimizes traditional exam workflows with diverse question types and advanced management features, enabling students to take exams remotely [3]. Crowdmark provides a grading toolkit that allows educators to deliver rich feedback and has demonstrated up to a 75% productivity gain in independent studies [4]. Additionally, automated tools for both formative and summative assessment in introductory programming courses are detailed in [5], while other approaches offer automation without compromising evaluation quality [6].

II. DESCRIPTION OF THE METHOD

In measurement laboratories that employ both interfaced and non-interfaced instruments, verifying the correctness of measurement processes is inherently challenging. This paper proposes a solution involving the acquisition, verification, and grading of measurement data using digital tools.

After completing the laboratory worksheet with the results obtained from measurements, the student inputs the data into a virtual instrument developed in LabVIEW. This tool allows for immediate verification of the measured values by comparing them to expected ranges. Regardless of correctness, all data are saved in a .txt file, enabling subsequent instructor review. To demonstrate the method, a virtual instrument was created for a specific laboratory task titled "Study and Verification of a Data Acquisition Board." In this experiment, students use a National Instruments USB-6008 data acquisition (DAQ) board, which integrates the ADS7870 system. The block diagram of the system is shown in Figure 1.

This system includes voltage reference, input multiplexer, precision programmable amplifier, 12-bit

analog-to-digital converter, internal oscillator with programmable frequency divider, configuration registers for multiplexer, amplifier, and ADC and an SPI interface.

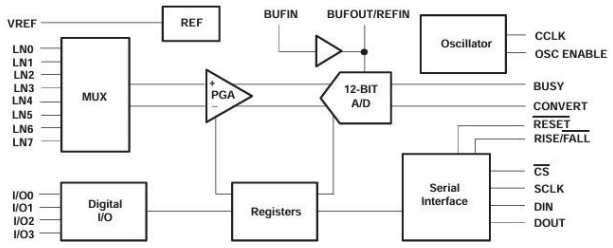


Figure 1. ADS7870 data acquisition system

To evaluate the accuracy of this board, students perform two setups: one for DC voltage measurement and another for AC voltage RMS value measurement.

Figure 2 shows the first setup.

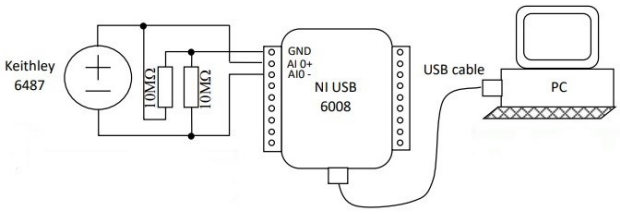


Figure 2. Measurement circuit for USB6008 accuracy checking in DC

A Keithley 6487 voltage source, configured as a standard voltage generator, is connected to the differential analog input (AI0+, AI0−) of the USB-6008 board. The board is interfaced with a PC running the LabVIEW virtual instrument. The student applies voltages from −10 V to +10 V with 2 V increments and records the corresponding readings from the USB-6008. These values are entered into the virtual instrument for validation. The student also calculates the following parameters: the absolute error, the relative error and the allowable error of the USB-6008 and the allowable error of the Keithley 6487. All these values are entered into the virtual instrument tables for verification.

Figure 3 shows the second setup.

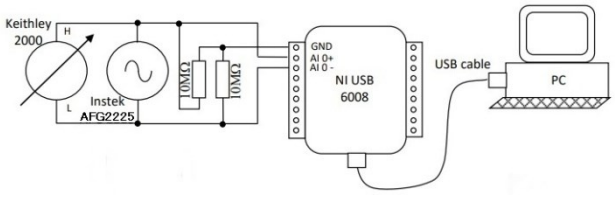


Figure 3. Measurement circuit for USB6008 accuracy checking in AC

An Instek AFG2225 signal generator is connected to the USB-6008's differential input (AI0+, AI0−), providing a 50 Hz sinusoidal signal with variable amplitude. A Keithley 2000 digital multimeter is also connected to measure the RMS voltage. The signal amplitude varies from 1 V to 7 V (1 V steps), which corresponds to RMS values up to 7.07 V, within the board's ±10 V input range.

The student must apply voltages from 1 V to 7 V RMS, in 1 V steps and read these fixed values from the USB6008 card. He will add the values indicated by the Keithley 2000 digital multimeter in the virtual instrument that checks their correctness. After completing the measurement process, the student must calculate the other parameters such as: the absolute error and the relative error of USB6008 and the allowable error of the reference instrument. These parameters will be input in the tables (Figure 4) available in the virtual instrument for verification.

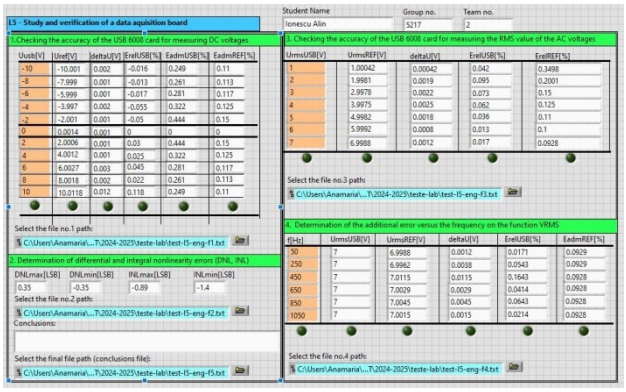
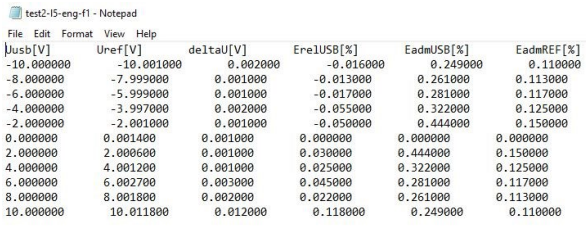


Figure 4. LabVIEW instrument front panel

The instrument includes four main sections corresponding to the four subsections of the student's lab worksheet: two for DC analysis, two for AC analysis. A dedicated section for student conclusions is also included. The fixed reference values are displayed in orange, while students enter their measured/calculated data in white fields. Blue fields indicate file paths for saving the .txt outputs. Figures 5 and 6 show examples of the .txt generated files.



test2-I5-eng-f1 - Notepad

File Edit Format View Help

Uusb[V]	Uref[V]	deltaU[V]	ErelUSB[%]	EadmUSB[%]	EadmREF[%]
-10.000000	-10.001000	0.002000	-0.016000	0.249000	0.110000
-8.000000	-7.999000	0.001000	-0.013000	0.261000	0.113000
-6.000000	-5.999000	0.001000	-0.017000	0.281000	0.117000
-4.000000	-3.997000	0.002000	-0.055000	0.322000	0.125000
-2.000000	-2.001000	0.001000	-0.050000	0.444000	0.150000
0.000000	0.001400	0.001000	0.000000	0.000000	0.000000
2.000000	2.000600	0.001000	0.030000	0.444000	0.150000
4.000000	4.001200	0.001000	0.025000	0.322000	0.125000
6.000000	6.002700	0.003000	0.045000	0.281000	0.117000
8.000000	8.001800	0.002000	0.022000	0.261000	0.113000
10.000000	10.011800	0.012000	0.118000	0.249000	0.110000

Figure 5. Sample Output File: DC Measurement Section

To interface LabVIEW with a SQL database, the following steps were required: SQL installation, ODBC server setup, and installation of the LabVIEW SQL Toolkit.

The most complex task was the ODBC configuration and the database structure, shown in Figure 7.

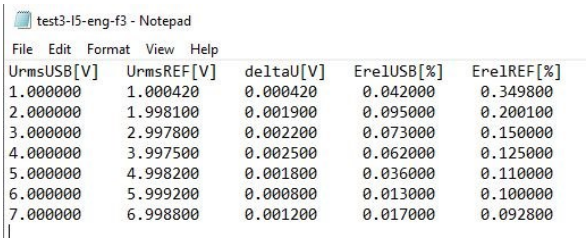


Figure 6. Sample Output File: AC Measurement Section

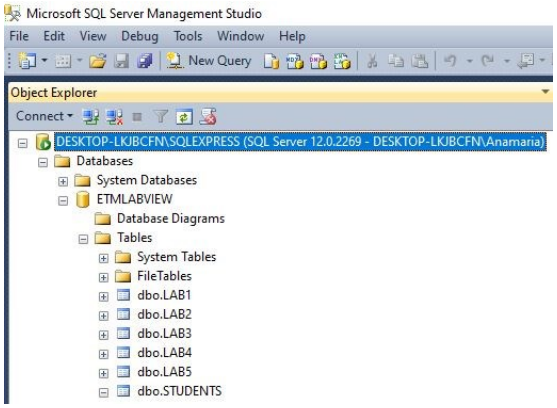


Figure 7. Database structure

LabVIEW transfers the acquired data into an SQL database named ETMLABVIEW, which includes a table with STUDENTS and five additional tables corresponding to different lab assignments. This paper focuses on Lab Assignment 5, for which the data flow was described above.

After the SQL Toolkit is installed, the blocks for database read/write operations become available in LabVIEW. The LabView application is created using a sequence type structure that includes 7 frames. The operating diagram from Figure 8 presents the 7 frames.

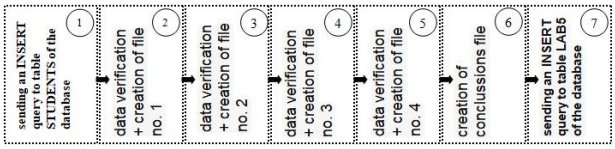


Figure 8. Virtual instrument frames in LabVIEW

The 7 frames perform the following functions:

1. The automatic input of the student's identification data into the database.
2. The verification of the measurement data and caomputed parameters and the creation of the first txt file.
3. The verification of the measurement data and calculated parameters and creation of the second txt file.
4. The verification of the measurement data and calculated parameters and the creation of the third txt file.
5. The verification of the measurement data and calculated parameters and creation of the fourth txt file.
6. The creation of the file that contains the conclusions of the 4 measurement processes.
7. The automatic input of all five files into the database.

The block diagram, shown in Figure 9, handles the insertion of measurement data and calculations from Lab Assignment 5, by importing the .txt files generated by each student. This action is carried out in the last frame of the virtual instrument.

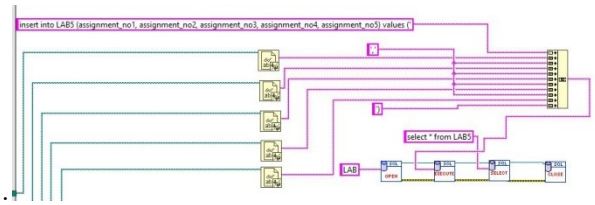


Figure 9. LabVIEW Block diagram for inserting text files into the database

These files populate the LAB5 table, making the data accessible for grading and analysis.

III. RESULTS AND DISCUSSIONS

To verify the correct integration between the .txt files generated in LabVIEW and the SQL database, data from three students, each belonging to a different lab team, were introduced into the virtual instrument. The first stage of verification consists of analyzing the recorded measurement data. If discrepancies are detected between the results of different students in the same team, these inconsistencies are noted in the database, as illustrated in Figure 10.

As seen in Figure 10, for the first laboratory assignment, three different files were uploaded. The first six columns of each entry list the names of the measured and calculated parameters. The subsequent six columns contain the actual numerical values. Notable differences are observed among the three students in parameters such as the voltage measured by the reference source, and in the computed absolute and relative errors.

After data validation, the calculated parameters are compared against the acceptable ranges predefined by the instructor. Based on this comparison, a final grade is determined and saved in the database for each laboratory sheet submitted by students.

student_id	assignment_no1											
1	Uusb[V]	Uref[V]	deltaU[V]	ErelUSB[%]	EadmUSB[%]	EadmREF[%]	-10.000000	-10.001600	0.002000	-0.016000	0.249000	0.110000	-8.000000
2	Uusb[V]	Uref[V]	deltaU[V]	ErelUSB[%]	EadmUSB[%]	EadmREF[%]	-10.000000	-10.001800	0.002000	-0.018000	0.249000	0.110000	-8.000000
3	Uusb[V]	Uref[V]	deltaU[V]	ErelUSB[%]	EadmUSB[%]	EadmREF[%]	-10.000000	-9.992900	0.007000	-0.071000	0.249000	0.110000	-8.000000

Figure 10. Example Result of a Database Selection Query

For example, 3 students were entered the data as they appear in the database, in figure 10, by running the virtual tool. We analyzed the first file sent by the students, which was loaded in the column assignment_no.1. It is observed that the information of interest is separated by spaces and thus the data analysis will be performed following those spaces.

The actual data analysis begins after the sixth row of the column, as the initial entries contain either the names of the measured or calculated parameters or predefined values. The first verification step involves the reference voltage corresponding to the initial reading from the data acquisition card (−10.000 V), which must fall within the acceptable range of [−10.009 V to −9.900 V]. Subsequently, the absolute error is evaluated and must lie within [−0.09 V to 0.09 V], while the relative error must fall within [−0.09% to 0.09%].

The next phase consists of validating the permissible error limits for the two devices involved in the measurement process—namely, the data acquisition card and the direct current (DC) source. For the acquisition card, the accepted error margin is between 0.2% and 0.3%, whereas for the reference DC source, the error must be within the interval of 0.09% to 0.19%.

If all recorded values in the database conform to the specified stability intervals, the student is awarded the maximum score for that section of the laboratory sheet. Finally, the mean value of the scores obtained across the five evaluation sections on the lab sheet is computed, which constitutes the final grade for the laboratory assignment.

IV. CONCLUSIONS

Assessing students' understanding of laboratory material is a crucial component of engineering education. Traditionally, this evaluation is performed through a final test administered at the end of the laboratory cycle. However, continuous laboratory assessment, where students receive regular evaluations throughout the term, is increasingly being adopted by university instructors. This method is especially beneficial in online or distance learning environments. Continuous assessment aligns well with the objectives of measurement education, as it mirrors real-world engineering contexts in which professionals must routinely meet short deadlines. It encourages students to consistently engage with course material and provides ongoing feedback on their understanding. Furthermore, this approach contributes to the development of students' metacognitive skills, which are essential for lifelong learning and adaptability in modern engineering practice [7].

In the present work, the proposed assessment method was tested on a single laboratory exercise. However, the framework can be extended to include all laboratory assignments in a course dedicated to electrical measurements. The grading process focused primarily on the measurement and calculation sections, but the laboratory activity also includes a conclusions section, which allows students to reflect on their experience and outcomes.

In future developments, we aim to apply artificial intelligence techniques to automatically analyze and evaluate student conclusions. This would facilitate an objective assessment of the qualitative aspects of lab work, further supporting fair and comprehensive grading.

REFERENCES

[1] James W. Pellegrino, "Perspectives on the Integration of Technology and Assessment", JRTE Vol. 43, No. 2, 2010, pp. 119–134.

[2] Mohd Isha Awang, "The Digitalization of Learning Assessment", The Proceeding Book of the 4th International Conference on Multidisciplinary Research 2021, Volume 04, No.1, 2021, ISSN: 2808-6229

[3] https://www.betterexaminations.com/solutions/higher-education/

[4][4] https://crowdmark.com/k12/

[5] Felipe Restrepo-Calle, J.J. Ramirez-Echeverry, Fabio A. González, "Using an interactive software tool for the formative and summative evaluation in a computer programming course: an experience report", Global Journal of Engineering Education, Volume 22, Number 3, 2020.

[6] Matija Novak, Dragutin Kermek, "Assessment Automation of Complex Student Programming Assignments", Education Sciences 2024.

[7] Yeray Rodriguez Rincon, Ana Munarriz, Alberto Magrenan Ruiz, "A new approach to continuous assessment: Moving from a stressful sum of grades to meaningful learning through self-reflection", The International Journal of Management Education, Volume 22, Issue 3, November 2024.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Quantum Communications for Distributed Measurement Systems: Current Situation and Research Trends

Arman Neyestani, Ioan Tudosa, Luca De Vito, Fatemeh Khalesi†, Sergio Rapuano*

Department of Engineering, University of Sannio, Benevento, Italy
{neyestani, ioan.tudosa, devito, rapuano}@unisannio.it
† f.khalesi@studenti.unisannio.it

***Abstract* – This survey analyses how five key mechanisms—Quantum Key Distribution (QKD), Quantum Secure Direct Communication (QSDC), entanglement-enhanced sensing, Quantum Clock Synchronization (QCS), and quantum teleportation/entanglement swapping—map onto the core metrological attributes of accuracy, stability, sensitivity, synchronization, security, and traceability. The paper identifies the principal research gaps— in particular, scalable entanglement distribution, hybrid classical–quantum integration, and SI-traceable calibration of quantum devices—and the technical advances required to translate laboratory prototypes (e.g. QKD links, QSDC networks, entangled-clock arrays, and QCS demonstrators) into field deployable quantum-enabled metrology.**

Keywords: Quantum Metrology, QKD, Quantum Clock Synchronization, Distributed Sensing, Traceability

I. INTRODUCTION

Quantum communication technologies are emerging as transformative tools in the field of measurement and metrology. Modern Distributed Measurement Systems (DMSs) demand extreme accuracy, stability, and reliability, while also needing robust security and synchronization in distributed environments. Classical DMSs are based on fundamental limitations: noise limits the precision of sensor data and cyber threats endanger data integrity [1]. By incorporating quantum communication resources, these limitations can be overcome – quantum techniques enhance the achievable measurement accuracy and provide unconditional security guaranteed by physical laws [1]. Indeed, quantum entanglement and related phenomena enable measurement precisions beyond classical limits and fundamentally secure data transmission. Quantum communication, alongside quantum sensing and computing, is thus poised to elevate measurement systems to unprecedented performance levels. Recent advances illustrate the impact: quantum networks have been demonstrated to enable applications beyond the reach of classical networks, including provably secure communications, high-precision clock synchronization, and distributed quantum sensing [2]. This survey examines how various quantum communication mechanisms (e.g. quantum key distribution, entanglement-based networks, quantum clock synchronization, etc.) influence core metrological attributes of measurement systems. This survey reviews the current state-of-the-art and key case studies, analyzes comparative benefits and remaining challenges, and emerging research directions and standardization efforts. The overarching research question is how quantum communications can improve measurement systems' attributes (accuracy, stability, sensitivity, synchronization, security, and traceability) and what new capabilities or considerations they bring. By synthesizing recent findings, the aim is to provide researchers with a fresh perspective on integrating quantum communications into next-generation DMSs. This paper begins by recalling the well-known core metrological attributes of measurement systems, then examines in detail the impact of each quantum communication mechanism—supported by illustrative case studies—and concludes with a comparative analysis of remaining research gaps and a conclusion on future developments.

II. QUANTUM COMMUNICATION MECHANISMS AND THEIR IMPACT ON MEASUREMENT ATTRIBUTES

Measurement systems are judged by six interdependent attributes—accuracy, stability, sensitivity, synchronization, security, and traceability—that together define metrological quality and suitability [3, 4]. Quantum communication furnishes a complementary toolkit—QKD, QSDC, entanglement-based sensor networks, QCS, and quantum teleportation—that can elevate these attributes beyond classical limits: QKD and QSDC deliver information-theoretic security and maintain traceability, entanglement-enhanced sensing improves sensitivity and accuracy, QCS achieves picosecond-level timing alignment, and teleportation enables long-range transfer of quantum standards. The remainder of this paper reviews each mechanism, surveys representative demonstrations, and assesses the outstanding technical limitations to practical deployment.

*This work was supported by PNRR – Mission 4, Component 2, Investment 1.5 – ECS_00000024 – Rome Technopole – Call CODBAN_000378 – Spoke 3 – Project N. UR301-2023-000090 – "MES&QT: Distributed Measurement Systems based on Quantum Technologies for Secure Communications".

A. QKD and Secure Keying

QKD is a quantum communication protocol that enables two parties to generate a shared secret key with security based on the principles of quantum mechanics. By transmitting quantum states (such as polarized photons), QKD ensures that any eavesdropping attempt will be detectable due to unavoidable quantum disturbances. This guarantees information-theoretic security, making QKD particularly valuable for critical measurement infrastructures—such as power grids and aerospace systems—where secure data transmission is essential [5].

The primary impact of QKD on DMS attributes is a significant enhancement of the security of measurement data: QKD delivers provably secure encryption keys, protecting data confidentiality and integrity even against adversaries equipped with quantum computers. This, in turn, reinforces traceability and trust in remote measurements, as data authenticity can be verified throughout transmission. However, QKD does not improve attributes like accuracy or sensitivity, and its deployment introduces requirements for specialized hardware and integration with classical networks. Despite these challenges, QKD represents a potential solution for securing measurement data, already demonstrated in real-world fiber and satellite networks [6, 7].

B. Quantum Secure Direct Communication

Quantum Secure Direct Communication (QSDC) transmits confidential data via entangled quantum states—bypassing separate key exchange—and guarantees that eavesdropping yields only random noise. and inter-city links have reached 100 km, with free-space and satellite trials underway. QSDC significantly enhances the *security* attribute by delivering information-theoretic confidentiality and integrity and bolsters *traceability* by making tampering immediately detectable [8]. Its practical uptake is limited by low data rates, the complexity of reliable state discrimination, and the need for specialized hardware (entangled photon sources, quantum memories), but ongoing advances—and prospective satellite deployments—point to scalable, quantum-secure DMSs.

C. Entanglement-Based Distributed Sensing and Enhanced Sensitivity

Entanglement-based distributed sensing leverages quantum-correlated probe states to surpass classical precision limits—scaling as $1/N$, where N is the number of entangled probes (Heisenberg limit) rather than $1/\sqrt{N}$ Standard Quantum Limit (SQL) [1]—as exemplified by a four-node entangled atomic clock network achieving a 4.5 dB improvement in timing precision—equating to a $\tilde{2}.8\times$ reduction in phase variance—over independent clocks, and an 11.6 dB decrease in projection noise (nearly a fourfold reduction) while preserving high entanglement fidelity [9]. By sharing entangled probes across sensors, quantum correlations suppress noise to boost sensitivity and accuracy, align phases for sub-SQL synchronization, anchor remote devices to common SI standards, and secure data by limiting access to parties performing the required joint quantum measurement.Scaling beyond a few nodes is difficult because entanglement decays with channel loss and environmental noise, and high-order multipartite states are generated only probabilistically; chaining more resilient Bell-pair link (which is usually a two-node entanglement channel that, owing to its facile generation and swapping, underpins scalable teleportation, QKD, and distributed sensing at a minor sensitivity cost) can alleviate these issues, though at the expense of ultimate sensitivity [10, 11]—but ongoing advances in entangled photon sources, quantum memories, and standardized protocols suggest a forthcoming redefinition of metrological capabilities beyond classical bounds.

D. Quantum Clock Synchronization and Timing Links

Quantum clock synchronization (QCS) harnesses entangled photons to synchronize remote clocks with femtosecond to picosecond precision, immune to classical delay asymmetries and capable of detecting timing spoofing through broken quantum correlations [12]. Satellite-based QCS, demonstrated by the Micius mission, targets sub-100 fs alignment across global distances. These advances dramatically boost the synchronization attribute—enabling high-fringe-contrast interferometry, coherent multi-node sensing, and sub-picosecond event timestamping—and reinforce security, since any intercept or delay breaks entanglement and is immediately evident. Key challenges include photon loss and decoherence over extended links and scaling to multiple users; potential solutions involve wavelength-division multiplexing, quantum repeaters, and error-corrected quantum memories to extend range and network capacity.

E. Quantum Teleportation and Remote State Transfer

Quantum teleportation transfers an unknown quantum state to a distant node using pre-shared entanglement and two bits of classical information, with no residual copy at the source and immunity to eavesdropping on the teleported state. In metrology, this facilitates remote calibration by conveying quantum standards without physical transport, thereby enhancing traceability and *accuracy*. State fidelity quantifies the statistical overlap between the teleported and target quantum states, interpreted as the probability that the recovered qubit is correct [13]. A mean value of 0.80 over 1,200 km surpasses the 0.67 classical ceiling and approaches about the 0.91 achieved on a 64 km fibre link, demonstrating strong long-range performance. Heralding rate is the number of successful, signalled teleportation events per second; 32 events is unusually high for satellite experiments, which typically operate at hertz scale [14]. Teleporting a phase-encoded calibration tone over 64 km of fiber produced a phase deviation of only 0.3° ($\approx 5 \times 10^{-3}$ rad), corresponding to a $15\ \mathrm{mHz}$ uncertainty on a $10\ \mathrm{MHz}$ carrier. Chained entanglement swapping on Micius extended entanglement to 1,000 km at ap-

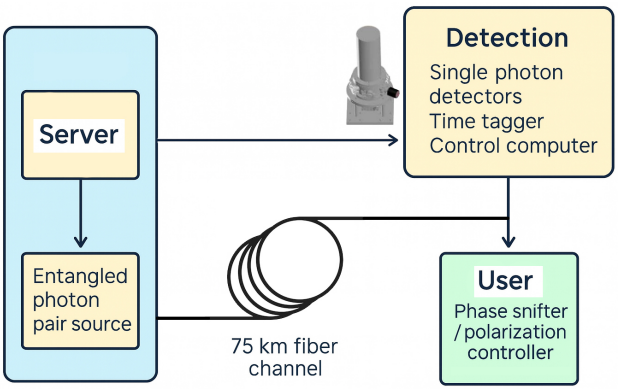


Fig. 1. Experimental setup flowchart of a single-user QCS protocol over a 75 km optical fiber link [2]

.

proximately $1\ \mathrm{s}^{-1}$ pair rates, boosting synchronization and sensitivity in networked sensors. Security is built in the classical feedforward bits, which reveal no state information without the entangled resource. in other hand, Bell-state measurement projects a two-qubit system onto the entangled basis [15]:

$$|\Phi^{\pm}\rangle = \frac{1}{\sqrt{2}}\left(|00\rangle \pm |11\rangle\right), |\Psi^{\pm}\rangle = \frac{1}{\sqrt{2}}\left(|01\rangle \pm |10\rangle\right).$$

where Φ labels the even-parity pair, Ψ labels the odd-parity pair, and is implemented by interfering two photons and detecting coincidences to identify the Bell state—an essential step for quantum teleportation. However, key challenges remain low success probabilities ($< 10^{-4}$ at continental scales), the need for quantum memories with millisecond coherence, and Bell-state measurement efficiencies below 80%. Progress in high-brightness sources, multiplexed memories, and loss-tolerant detection will be pivotal for teleportation-enabled metrology networks.

III. CASE STUDIES

To ground the discussion in practical terms, a selection of case studies is reviewed where quantum communication technologies have been implemented in measurement systems or testbeds. These examples highlight the affected performance attributes and report key metrics, demonstrating the current situation in this research area.

A. *Case Study 1: Quantum-Enhanced Clock Synchronization over 75 km Fiber*

In 2023, researchers demonstrated a multi-user quantum clock synchronization network linking a server and remote client over a 75 km optical fiber using entangled photons. In this experiment [2] as shown in Fig. 1, a polarization-entangled photon source at the server generated pairs of photons that were distributed to the remote user. By measuring arrival times of entangled photons (one traveling to the user and back, one measured locally), the clock offset between the user and server was determined with extremely high precision. The results were impressive: they achieved a measured clock time difference uncertainty of 4.45 ps, and a timing stability characterized by a TDEV of 426 fs after 4000s of averaging. For comparison, classical network time synchronization (e.g. Network Time Protocol (NTP) or Precision Time Protocol (PTP) over that distance) would be in the nanoseconds to microseconds range at best, and GPS typically offers nanoseconds precision under ideal conditions. This quantum method inherently avoided errors from asymmetric channel delays, since only one-way travel of quantum signals was used (with entangled timing correlations). Synchronization was directly improved to ps-level, vastly better than classical tech. This also improves accuracy for any measurements dependent on the clocks (e.g. time-of-flight measurements or comparing remote frequency standards). The scheme was secure against certain attacks (like delay manipulation), as noted by the authors, because any attempt to tamper would be evident in the correlation data [2]. The case study illustrates quantum communication enabling a new regime of synchronization useful for e.g. distributed radio telescopes, financial transaction timestamping, and fundamental constant measurements (where remote clocks need to be compared). It also showed scalability: the use of WDM (wavelength multiplexing) means multiple users could be synchronized off one entangled source, each getting their own wavelength channel.

B. *Case Study 2: Entangled Sensor Network for Distributed Quantum Sensing*

A notable experiment in 2020 (K. Malia et al.) [9] involved an entangled network of atomic clocks or sensors demonstrating improved measurement precision across nodes. In this setup, up to four atomic ensembles (acting as atomic clocks or interferometric sensors) were entangled via a shared quantum non-demolition measurement. The network was used to measure a common phase (or frequency) with enhanced precision. The reported gain was significant: the entangled network achieved a 4.5 dB better precision than a network with only local (separate) entanglement at each node, and about 11.6 dB better precision than a comparable classical network limited by independent quantum projection noise because phase noise maps directly onto fractional-frequency error, the result corresponds to tightening clock comparisons from the typical 10^{-12} level to the mid-10^{-13} range after a single second of averaging, with continued τ^{-1} Heisenberg scaling thereafter.. In other words, the entanglement between nodes gave a multi-fold reduction in measurement uncertainty beyond the usual $1/\sqrt{N}$ scaling, where N is the number of Nodes. The experiment demonstrated both atomic clock comparison and interferometric phase sensing, showing the generality of the approach. This directly relates to sensitivity/accuracy – it shows that distributing entanglement can improve the signal-to-noise ra-

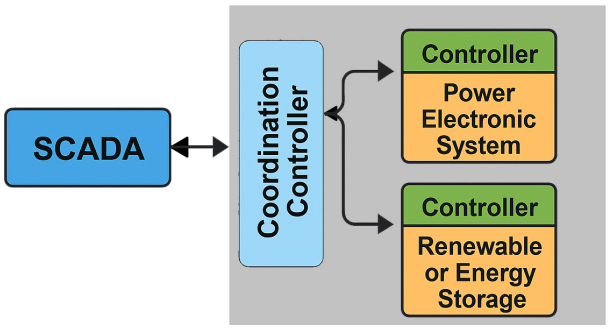


Fig. 2. A supervisory SCADA layer exchanges data with a central coordination controller, which issues commands to two local controllers—one governing the power-electronic interface and the other the renewable/energy-storage unit [17].

tio of measurements across a network. Each node individually has certain sensitivity, but together (when entangled), the network functions like a single larger sensor with lower noise. This has implications for, say, coordinating arrays of quantum sensors to detect tiny gravitational perturbations or comparing distant optical clock frequencies for geodesy (measuring gravitational redshift differences). By improving precision, it also strengthens traceability between standards: e.g., measuring frequency ratios to 18-digit accuracy as cited by the authors (beyond what was previously possible). The experiment did not focus on security, but if one can distribute entanglement for precision, one might also integrate QKD in parallel for secure classical channels among the sensor nodes (the Phys. Rev. A 2025 paper by Moore et al. indeed considered secure quantum-enhanced sensing protocols [16]). The challenge highlighted in this case was technical: entangling multiple macroscopic atomic systems and maintaining that entanglement was non-trivial and required careful collective measurements.

C. *Case Study 3: QKD-Secured Smart Grid Sensor Communications*

In 2022, a team from Oak Ridge National Lab and partners demonstrated the integration of Quantum Key Distribution into a smart grid communication network for securing measurement data [17]. They deployed a QKD system (entanglement-based QKD using a commercial Qubitekk unit) on a realistic utility fiber network between an electrical substation and a control center as shown in Fig. 2. The purpose was to authenticate and encrypt SCADA communications (telemetry from sensors like voltage/current meters and control signals to equipment) using the quantum-generated keys. The case study showed that QKD could operate alongside the utility's existing data traffic and continuously provide fresh secret keys for cryptographic use. They specifically implemented the QKD keys in the Message Queuing Telemetry Transport (MQTT) messaging protocol for machine-to-machine SCADA data, achieving quantum-enhanced authentication of commands and measurements. The feasibility was confirmed in a real-world environment (Chattanooga Electric Power Board (EPB) utility), not just a lab. The primary attribute in this case is security – the measurement system (power grid sensors and actuators) was secured against eavesdropping and tampering by virtue of QKD's unbreakable encryption keys. This improves the trustworthiness of grid measurements (voltage readings, frequency measurements, etc.), which is crucial for grid stability and safety. A power system operator can be confident that data has not been spoofed or altered in transit if it's protected by QKD-based authentication. Indirectly, this contributes to the stability and reliability of the measurement system (not the sensor reading itself, but the overall infrastructure), because a cyber attack attempting to falsify measurements can be thwarted. The experiment also underscores the importance of standards and interoperability – they had to use specific networking setups to integrate QKD, and they highlight challenges like key management, latency, and scalability for more nodes. Nevertheless, this case proved that quantum communication (QKD) can be added to critical measurement networks today to elevate security without harming the real-time operation (the key rate was sufficient for frequent re-keying of messages, and the classical control traffic coexisted with quantum signals on the same fiber via wavelength multiplexing). Even a modest QKD key rate of $\sim$1 kbps is ample: phasor measurement units (PMUs) transmit about 4 kbps of sensor data, so 1 kbps of fresh key material allows subsecond rotation of the AES key, well above the security requirements of North American Electric Reliability Corporation (NERC) and Critical Infrastructure Protection (CIP). Time alignment itself still relies on existing GPS or emerging quantum-synchronization links; QKD secures the data path but does not affect the timing budget. Thus, throughput is sufficient for remote smart-grid measurements, and synchronization performance remains unchanged while confidentiality and integrity are raised to information-theoretic levels.

D. *Case Study 4: Multi-User Quantum Secure Direct Communication Network*

Another illustrative example, though more futuristic, is the 15-user QSDC network demonstrated in China in 2021 [18]. While this was primarily a communication network experiment, one can envision its application in a measurement context for highly secure collaborative DMSs. In the demonstration, 15 users formed a fully connected network where any pair could directly exchange secret messages through entangled photon pairs and sum-frequency generation (a technique to perform Bell state measurements). They achieved >95% entanglement fidelity between any two users over 40 km fiber links and an information transmission rate above 1 kbit/second. From a metrology perspective, imagine these 15 nodes are distributed sensing

Mechanism	Key Attribute(s)	Notable Achievements	Limitations
QKD	Security, traceability	Utility fiber QKD secures SCADA; proven unconditional key security [17]	∼100 km range; low key rate; needs single-photon hardware; standardization needed
QSDC	Security, authenticity	15-user, 40 km, >95% fidelity; direct quantum messaging [18]	Very low data rates; experimental; range limited; complex state management
Entanglement-Enhanced Sensing	Sensitivity, accuracy, (sync, security)	>11 dB precision gain in 4-node clock network [9]; quantum-secure metrology proposed [1, 16]	Entanglement fragile; small demos; quantum memory and specialized algorithms needed
QCS	Synchronization, timing accuracy, (security)	4.5 ps sync over 75 km fiber [2]; secure against delay attacks	Range/robustness limits; ultrafast detectors; not widely deployed; classical comm. still needed
Teleportation or Swapping	Enabler for sensitivity, sync, traceability	>1000 km entanglement via satellite; quantum internet backbone [13]	High loss; low success probability; quantum memory needed; complex coordination

Table 1. Compact summary of quantum communication mechanisms in measurement systems.

stations (for environmental monitoring, for example) that need to share data or aggregate results securely. QSDC would allow them to send not just keys but the measurement data itself in quantum form, with eavesdroppers gaining no information. Security is again the focus – data confidentiality and integrity across the network were guaranteed by physics. There is also a resilience benefit because QSDC doesn't rely on stored keys, the system is less vulnerable to key compromise or man-in-the-middle attacks (any attempt would break the entanglement and be noticed). The trade-off is the modest data rate. This network doesn't yet improve the accuracy or sensitivity of the measurements being shared, but it ensures the secure collaboration of multiple measuring units. This is relevant for distributed measurement projects (like an array of air quality sensors sending readings to each other and a central hub without fear of tampering, or distributed timekeeping where clocks exchange correction data securely). The case study also provided insight into scalability: the network used innovative multiplexing to connect many users without a trusted node. This hints at how future quantum networks might simultaneously cater to multiple functions – e.g., a single quantum network could handle QKD, QCS, and QSDC between many parties, serving as a multipurpose metrology backbone.

IV. DISCUSSION AND RESEARCH GAPS

Having examined various quantum communication mechanisms and their demonstrated impacts, a comparison of their relative advantages and a discussion of remaining research gaps is presented. Table 1 summarizes the core quantum communication mechanisms covered, the measurement attributes they principally improve, examples of achievements, and associated challenges or limitations.

From the table 1, no single quantum communication technology addresses all performance attributes, but each has a niche: For security and data integrity, QKD (and to a more experimental degree, QSDC) are the go-to solutions. They dramatically enhance the security attribute with proven theoretical security, something classical methods cannot match. However, they do not inherently improve measurement precision or accuracy – they ensure the trustworthiness of the measurement process rather than the quality of the raw measurement itself. Their deployment, therefore, is often a question of infrastructure and engineering (deploying fibers or satellites, handling keys) and aligning with industry standards, rather than altering the sensor physics.

For sensitivity, accuracy, and precision, entanglement-based approaches (quantum metrology, sensor networks) are unparalleled. They tackle the core physics of measurement by reducing noise floors and enabling correlated measurements. The potential here is transformative for high-end metrology (like clock networks, gravitational wave detection, etc.). The drawback is complexity – these improvements often require pristine laboratory conditions, careful state preparation, and low-noise environments. Real-world sensors (in industrial settings, for example) might find it hard to utilize these exotic states without more robust technology. There is also a diminishing returns vs complexity tradeoff: achieving a small additional precision gain might require exponentially more complex entangled states or error correction.

For synchronization, QCS is a clear quantum advantage solution. It directly outperforms classical synchronization in principle. If one's application needs timing beyond what GPS or fiber two-way can do, quantum is the way forward. But if classical synchronization is satisfies the design specifications, then QCS will be a tough sell. Right now, only very demanding applications (e.g. experimental tests of fundamental physics, or possibly future telecom networks at single-picosecond bit alignment) truly need what QCS offers. That being said, as other parts of technology advance (like analog-to-digital converter speeds, high-frequency trading, etc.), picosecond synchronization might become more relevant [19].

Interestingly, some attributes are interrelated, and quantum technologies can target them together. For instance, entanglement helps both sensitivity and synchronization (an entangled clock network yields better precision in time offset measurement, effectively synchronizing more accurately). Security can be layered on top of entanglement-

based sensing, meaning one can design a system that is simultaneously more precise and more secure than its classical counterpart. These multi-faceted quantum systems are very appealing but also highlight research gaps: integrating different quantum protocols together is complex. The Moore [16] work exemplifies this by mixing entangled and separable states to keep security in a large sensor network – essentially a hybrid approach to maintain both precision and security at scale. This is a new direction, blending QKD concepts with quantum sensing.

Despite rapid progress, key research gaps limit the broad impact of quantum communication on metrological attributes such as security, sensitivity, synchronization, and traceability [20]. Scaling quantum networks beyond current point-to-point or small multi-node demonstrations (e.g., QKD or QCS over tens of kilometers, entanglement-enhanced sensing in 2- 4 node labs) is hindered by the fragility of entanglement, resource demands, and a lack of operational quantum repeaters. Without advances in error-corrected quantum memory and efficient multi-node protocols, network-wide secure synchronization or entanglement-based precision remains unattainable. Integration with classical infrastructure also presents challenges: QKD devices must support existing IT standards and failover schemes, while QCS outputs require seamless interfacing with time distribution systems. Moreover, metrological calibration and standardization of quantum systems are underdeveloped—for instance, establishing SI-traceable standards for single-photon detectors or certifying quantum time transfer is ongoing, with efforts like ISO/IEC 23837-1 (2023) [21] and NIST-led detector standards [22]. Practicality remains an issue as many protocols need high photon rates or cryogenic components, and translating laboratory results (e.g., picosecond synchronization over 75 km fiber, $>$11 dB sensitivity gains in entangled clocks) into real-world, resource-efficient, telecom-compatible devices is not yet realized. Ultimately, widespread adoption will require research into application-specific benefits, hybrid quantum-classical solutions, and robust metrics to quantify and certify the measurement improvements enabled by quantum communication.

V. CONCLUSION

Quantum communication already can deliver concrete benefits: QKD and QSDC provide provably secure data channels that reinforce traceability; entanglement-based sensing demonstrably surpasses classical signal-to-noise ratios by double-digit decibels; and QCS realises picosecond-accurate timing links that eclipse GPS-class solutions. Yet these gains remain confined to point-to-point or few-node testbeds. The path to widespread adoption hinges on three advances: (i) scalable, resource-efficient entanglement distribution via quantum repeaters and multiplexed network protocols; (ii) seamless hybridisation with classical digital network and timing infrastructure, backed by emerging standards such as ISO/IEC 23837-1; and (iii) robust, SI-traceable calibration methods for single-photon sources, detectors, and quantum timing links. Addressing these challenges will enable an "internet of quantum sensors" in which security, synchronization, and sensitivity are elevated simultaneously, catalysing next-generation measurement capabilities across power grids, telecommunications, fundamental physics, and beyond.

REFERENCES

[1] M. T. Rahim, A. Khan, U. Khalid, J. u. Rehman, H. Jung, and H. Shin, "Quantum secure metrology for network sensing-based applications," *Scientific Reports*, vol. 13, no. 1, p. 11630, 2023.

[2] B.-Y. Tang, M. Tian, H. Chen, H. Han, H. Zhou, S.-C. Li, B. Xu, R.-F. Dong, B. Liu, and W.-R. Yu, "Demonstration of 75 km-fiber quantum clock synchronization in quantum entanglement distribution network," *EPJ Quantum Technology*, vol. 10, no. 1, pp. 1–10, 2023.

[3] A. Menditto, M. Patriarca, and B. Magnusson, "Understanding the meaning of accuracy, trueness and precision," *Accreditation and quality assurance*, vol. 12, pp. 45–47, 2007.

[4] D. Matsakis, J. Levine, and M. Lombardi, "Metrological and legal traceability of time signals," in *"Proceedings of the 49th annual precise time and time interval systems and applications meeting"*, pp. 59–71, 2018.

[5] W. Yan, X. Zheng, W. Wen, L. Lu, Y. Du, Y.-Q. Lu, S. Zhu, and X.-S. Ma, "A measurement-device-independent quantum key distribution network using optical frequency comb," *npj Quantum Information*, vol. 11, no. 1, p. 97, 2025.

[6] F. Xu, X. Ma, Q. Zhang, H.-K. Lo, and J.-W. Pan, "Secure quantum key distribution with realistic devices," *Reviews of modern physics*, vol. 92, no. 2, p. 025002, 2020.

[7] A. Meda, A. Mura, S. Virzì, A. Avella, F. Levi, I. P. Degiovanni, A. Geraldi, M. Valeri, S. Di Bartolo, T. Catuogno, *et al.*, "Qkd protected fiber-based infrastructure for time dissemination," *Scientific Reports*, vol. 15, no. 1, p. 13419, 2025.

[8] H. Zhang, Z. Sun, R. Qi, L. Yin, G.-L. Long, and J. Lu, "Realization of quantum secure direct communication over 100 km fiber with time-bin and phase quantum states," *Light: Science & Applications*, vol. 11, no. 1, p. 83, 2022.

[9] B. K. Malia, Y. Wu, J. Martínez-Rincón, and M. A. Kasevich, "Distributed quantum sensing with mode-entangled spin-squeezed atomic states," *Nature*, vol. 612, no. 7941, pp. 661–665, 2022.

[10] S.-R. Zhao, Y.-Z. Zhang, W.-Z. Liu, J.-Y. Guan, W. Zhang, C.-L. Li, B. Bai, M.-H. Li, Y. Liu, L. You, *et al.*, "Field demonstration of distributed quantum sensing without post-selection," *Physical Review X*, vol. 11, no. 3, p. 031009, 2021.

[11] Z. Zhang and Q. Zhuang, "Distributed quantum sensing," *Quantum Science and Technology*, vol. 6, no. 4, p. 043001, 2021.

[12] J. Shi and S. Shen, "A clock synchronization method based on quantum entanglement," *Scientific Reports*, vol. 12, no. 1, p. 10185, 2022.

[13] J.-G. Ren, P. Xu, H.-L. Yong, L. Zhang, S.-K. Liao, J. Yin, W.-Y. Liu, W.-Q. Cai, M. Yang, L. Li, *et al.*, "Ground-to-satellite quantum teleportation," *Nature*, vol. 549, no. 7670, pp. 70–73, 2017.

[14] S. Shen, C. Yuan, Z. Zhang, H. Yu, R. Zhang, C. Yang, H. Li, Z. Wang, Y. Wang, G. Deng, *et al.*, "Hertz-rate metropolitan quantum teleportation," *Light: Science & Applications*, vol. 12, no. 1, p. 115, 2023.

[15] S. Welte, P. Thomas, L. Hartung, S. Daiss, S. Langenfeld, O. Morin, G. Rempe, and E. Distante, "A nondestructive bell-state measurement on two distant atomic qubits," *Nature Photonics*, vol. 15, no. 7, pp. 504–509, 2021.

[16] S. W. Moore and J. A. Dunningham, "Secure quantum-enhanced measurements on a network of sensors," *Physical Review A*, vol. 111, no. 1, p. 012616, 2025.

[17] M. Alshowkan, P. G. Evans, M. Starke, D. Earl, and N. A. Peters, "Authentication of smart grid communications using quantum key distribution," *Scientific Reports*, vol. 12, no. 1, p. 12731, 2022.

[18] Z. Qi, Y. Li, Y. Huang, J. Feng, Y. Zheng, and X. Chen, "A 15-user quantum secure direct communication network," *Light: Science & Applications*, vol. 10, no. 1, p. 183, 2021.

[19] C. G. Song and Q. Y. Cai, "Excessive precision compromises accuracy even with unlimited resources due to the trade-off in quantum metrology," *npj Quantum Inf*, vol. 11, p. 115, July 2025.

[20] K. Azuma, S. E. Economou, D. Elkouss, and Others, "Quantum repeaters: From quantum networks to the quantum internet," *Reviews of Modern Physics*, vol. 95, p. 045006, Dec 2023.

[21] "Information security — security requirements, test and evaluation methods for quantum key distribution — part 1: Requirements," International Standard ISO/IEC 23837-1:2023(E), International Organization for Standardization and International Electrotechnical Commission, Aug 2023.

[22] J. C. Bienfang, T. Gerrits, P. S. Kuo, A. Migdall, S. Polyakov, and O. Slattery, "Single-photon sources and detectors dictionary," NIST Interagency/Internal Report NIST IR 8486, National Institute of Standards and Technology, Gaithersburg, MD, Sep 2023.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Quantum Channel Characterization in QKD: A Metrological Perspective

Fatemeh Khalesi[1], Ioan Tudosa[1], Francesco Picariello[2], Arman Neyestani[1], Sergio Rapuano[1*]

[1]*Department of Engineering, University of Sannio, Benevento, Italy*
[2]*Department of Engineering and Science, Mercatorum University, Rome, Italy*
f.khalesi@studenti.unisannio.it, {ioan.tudosa, neyestani, rapuano}@unisannio.it, francesco.picariello@unimercatorum.it

***Abstract* – Quantum Key Distribution (QKD) offers secure communication by leveraging core principles of quantum mechanics. Characterization of QKD system is fundamental for comparing the experimental results provided by researchers. Unfortunately, there is no standardized or unified approach; therefore, different researchers present results that provide heterogeneous information, making comparisons difficult. A systematic analysis of the literature is necessary to identify common approaches from a metrological perspective. This paper provides an overview of measurements used to characterize quantum channels, with a focus on parameters such as attenuation, polarization effects, and timing stability. To support this analysis, real-world case studies are examined.**

Keywords: Quantum Channel, QKD, Metrological Characterization.

I. INTRODUCTION

Quantum Key Distribution (QKD) is a cryptographic technique that leverages quantum mechanics to securely generate and share encryption keys between two parties. Unlike classical cryptography, which relies on computational complexity, QKD offers *information-theoretic security* by encoding information in quantum states—typically modifying properties of single photons transmitted over optical links [1]. The emergence of quantum computing poses a significant threat to classical cryptographic systems, as quantum algorithms like Shor's and Grover's can efficiently break commonly used classical encryption schemes [2]. In contrast, QKD remains secure even in the presence of quantum adversaries [1]. Security is guaranteed by quantum principles: the *Heisenberg indeterminacy principle* prevents the precise simultaneous measurement of conjugate variables, while the *no-cloning theorem* prohibits copying unknown quantum states [2, 3].

Practical QKD systems are intrinsically affected by environmental noise, signal attenuation, and decoherence within quantum channels [3]. While most QKD implementations emphasize system design and functional performance, the thorough characterization of quantum channels—particularly incorporating principles such as traceability, repeatability, and reproducibility—is often overlooked [4]. This study investigates how quantum channel parameters influence key metrological aspects of QKD systems. It provides a review of current characterization approaches and analyzes real-world case studies to identify benefits, challenges, and gaps in practice. The focus is on how improved quantum channel characterization can enhance core measurement attributes such as accuracy, repeatability, reproducibility, and traceability, with attention to emerging needs for standardization.

To support this objective, Section II introduces a typical QKD system. Section III presents the metrological characterization of quantum channels, while Section IV provides supporting real-world case studies. Section V offers a comparative analysis and highlights research gaps and future directions. Finally, Section VI concludes the study.

II. QKD SYSTEM OVERVIEW

The process of QKD between Alice (Sender) and Bob (Receiver) is illustrated in Fig. 1, in which a secure key is shared using a quantum channel to transmit qubits—or optionally, entangled qubits—and a separate classical channel to perform basis reconciliation. A qubit is the basic unit of quantum computing, like a classical bit but able to exist in a superposition—being both 0 and 1 at once. Qubits can also be entangled, meaning the state of one instantly influences another, even at a distance. These properties allow detection of eavesdropping due to the delicate nature of quantum states [5].

To begin, Alice encodes single photons or photon pairs in a chosen quantum basis and sends them through the fiber to Bob. Optical loss during transmission in the fiber critically impacts QKD performance and is measured using standard tools such as an Optical Power Meter (OPM), which assesses input/output power to calculate loss during setup, and an Optical Time-Domain Reflectometer (OTDR), which locates loss points along the fiber but is used offline due to its high-intensity pulses [6, 7].

At Bob's side, sensitive Single-Photon Detectors (SPDs) register the incoming photons. Devices such as Superconducting Nanowire SPDs (SNSPDs) offer ultra-low jitter, high efficiency, and low dark counts, ideal for long-

*This work was supported by PNRR – Mission 4, Component 2, Investment 1.5 – ECS_00000024 – Rome Technopole – Call CODBAN_000378 – Spoke 3 – Project N. UR301-2023-000090 – "MES&QT: Distributed Measurement Systems based on Quantum Technologies for Secure Communications".

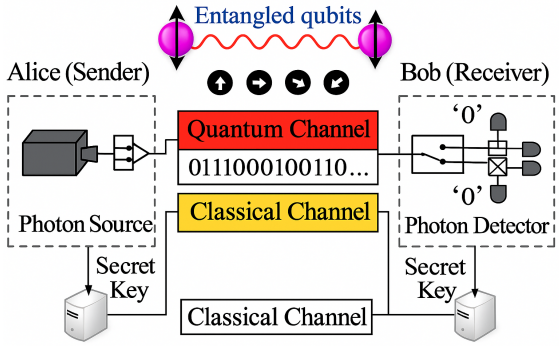


Fig. 1. A Schematic of QKD, qubits are transmitted over a quantum channel, while bits are exchanged via a classical channel for basis reconciliation and key verification [5].

distance QKD [8, 9], while Single-Photon Counting Modules (SPCMs) provide compact, integrated detection with built-in electronics [10]. Other similar detectors are also used depending on system requirements.

Timing synchronization is critical to ensure correct qubit detection. This is achieved using such tools as: (I) Time-to-Digital Converters (TDCs) and Time Interval Analyzers (TIAs) at Bob's side for high-precision timestamps [11, 12]; (II) GPS-Disciplined Oscillators (GPSDOs) and atomic clocks at both ends for stable long-term timing [13]; and (III) Phase-Locked Loops (PLLs) and frequency synthesizers embedded in hardware to generate and stabilize clock signals [9].

Finally, Alice and Bob perform basis reconciliation by comparing which bases they used— typically rectilinear ($|0\rangle, |1\rangle$) or diagonal ($|+\rangle, |-\rangle$). They keep only the bits where their bases match [5]. Then, in key verification, a subset of bits is compared to detect eavesdropping. If the quantum bit error rate (QBER) is below a set threshold, the rest of the bits are used to form a secure key [5].

III. QKD CHANNEL CHARACTERIZATION

Accurate quantum channel characterization is essential for optimizing QKD system performance and ensuring its security. This section reviews key channel parameters, the corresponding measurement methods, and the physical components involved in a quantum communication link. Although various methods are available, measurement uncertainty is rarely quantified or reported, which limits the reliability and comparability of results. The selected parameters are influenced by physical effects—such as absorption, thermal noise, mechanical stress, and signal jitter—that degrade signal integrity and contribute significantly to the overall uncertainty in QKD measurements.

A.1 Channel Attenuation (Optical Loss)

Photon loss in quantum communication channels primarily results from absorption, scattering, and imperfect coupling in optical fibers or free-space links [14, 15]. Attenuation refers to the reduction of signal strength during transmission and is typically expressed in *decibels per kilometer (dB/km)* and can be measured by devices like OPM or OTDR. It is defined as the ratio of output power (P_{out}) after transmission through a link of length L to the input power (P_{in}) [16]:

$$\text{Attenuation } \alpha = \frac{-10}{L} \log \left(\frac{P_{\text{out}}}{P_{\text{in}}} \right) \tag{1}$$

Attenuation reduces the number of detected photons, weakening the signal relative to background noise, and thereby increasing the QBER. The QBER is dimensionless and typically expressed as a percentage (%) and is calculated as [17]:

$$\text{QBER} = \frac{R_{\text{error}}}{R_{\text{error}} + R_{\text{sift}}} \tag{17}$$

where R_{error} is the rate of erroneous bits, and R_{sift} is the rate of sifted bits correctly received when basis choices match [17, 18].

The attenuation also directly affects the Secure Key Rate (SKR), R, which quantifies the number of secret key bits generated per second [19].

$$R = \nu \cdot s \cdot p \cdot q \cdot (d \cdot (1 - H(\text{QBER})) - f \cdot H(\text{QBER})) \tag{2}$$

where: ν is the *pulse repetition rate* (Hz), indicating how many light pulses (e.g., from a laser) are sent per second, set by the hardware, s is the *sifting coefficient*, the fraction of raw key bits retained after basis reconciliation (typically ≈ 0.5 for BB84, one of the possible QKD protocols), p is the *parameter estimation efficiency*, representing the portion of sifted bits used for key generation after estimating the QBER, q is the *transfer/measurement factor*, a dimensionless parameter reflecting losses in the quantum channel and detector inefficiency, d is the *decoy-state coefficient*, estimating the fraction of secure single-photon events via intensity-varying pulse analysis, f is the *error correction inefficiency*, quantifying the overhead beyond the Shannon limit during classical error correction, $H(\text{QBER})$ is the *binary Shannon entropy*, defined as $H(x) = -x \log_2 x - (1-x) \log_2 (1-x)$, measuring the information leakage due to errors in the key. QBER and SKR can be measured by devices like SPDs and SNSPDs.

In the characterization of QKD systems, measurement uncertainty is significantly influenced by channel attenuation. As attenuation increases, the photon detection rate decreases, which amplifies statistical fluctuations and raises the uncertainty in the QBER. These uncertainties have a direct effect on the SKR, since the SKR is highly sensitive to both detection efficiency and error rates.

A2. Dark Count Rate / Background Noise: A dark count is a false detection event in the absence of a photon, caused by thermal noise, electronic fluctuations, defects, or ambient light. The Dark Count Rate (DCR), measured in counts per second (cps), is evaluated by blocking the

input and counting false detections over a known time interval [20].

$$N_{0.5p.e}(T) \approx AT^{\frac{3}{2}} \exp\left(-\frac{E_g}{2kT}\right) \tag{3}$$

The detector measures the DCR N, defined at a 0.5 photoelectron threshold within the pulse window, and is temperature-dependent. The DCR can be modeled as a function of temperature T, where A is a constant related to charge trap formation during avalanche events, E_g is the semiconductor bandgap energy, and k is Boltzmann's constant [20].

To model the relationship between QBER and DCR [18]:

$$\mathrm{QBER} = \frac{R_{\mathrm{fACF}}}{R_{\mathrm{shift}} + R_{\mathrm{error}}} + \frac{R_{\mathrm{dc}}}{R_{\mathrm{shift}} + R_{\mathrm{error}}} \tag{4}$$

where R_{fACF}: error rate from frequency/polarization misalignment; R_{dc}: error rate due to dark counts ; R_{error}: total error detection rate ($= R_{\mathrm{fACF}} + R_{\mathrm{dc}}$). Lowering the DCR is key to keeping QBER low, especially over long distances or high loss. One of the devices that directly measures DCR and helps evaluate both QBER and SKR is the SPCM. Regarding measurement uncertainty, [21] reported a DCR uncertainty of 1% under stable conditions, combining Type A (0.1%) from Poisson statistics and Type B (0.9%) from temperature ($\pm 0.25\,^{\circ}\mathrm{C}$) and threshold variations.

A3. Polarization Drift: The gradual, often unpredictable change in an optical signal's polarization due to temperature, stress, or fiber bending, causing misalignment between sender and receiver, increases QBER, and reduces SKR. The polarization misalignment angle θ is estimated using the following formula [22]:

$$\theta \approx \arcsin\left(\sqrt{\frac{N_{\text{erroneous}}}{N_{\text{max}}}}\right) \tag{5}$$

where $N_{\text{erroneous}}$ is the number of erroneous single-photon detections ($N_{\text{erroneous}} = R_{\text{error}} \cdot \Delta t$ in the measurement interval Δt), N_{max} is measured during calibration under ideal alignment and represents the maximum expected single-photon detections for a given intensity. It serves as a reference for estimating polarization misalignment.

In QKD systems, polarization drift is addressed by two separate devices with distinct functions: the *polarimeter* for measuring and the *Polarization Controller (PC)* for correction. The polarimeter measures the state of polarization (SOP) of incoming photons and detects deviations. Its output is used to estimate $N_{\text{erroneous}}$. Then, the PC can actively adjusts the SOP.

Due to polarization drift, the measurement uncertainty observed in the paper [23] is approximately 2–3% in terms of QBER. This is considered a Type A uncertainty, as it is derived from statistical analysis of repeated experimental measurements over thousands of compensation cycles.

A4. Timing and Frequency Synchronization: Accurate synchronization is crucial in QKD to minimize QBER [24]. *Timing synchronization* ensures that Alice and Bob agree on when photons arrive. Optical clock pulses are sent at intervals ΔT_s, which are divided into N_w equal time windows of length τ_w [25]:

$$\Delta T_s = N_w \cdot \tau_w \tag{6}$$

This allows Bob to correctly identify photon arrival windows, minimizing errors (QBER) by aligning detection times with transmission. *Frequency synchronization* is achieved by locking to the pulse repetition rate $f = 1/\Delta T_s$. Precision instruments like TDCs and GPSDOs are used to ensure accurate synchronization. Time measurement uncertainty mainly comes from detector jitter and synchronization, typically ranging between 10 ps and 100 ps, as stated in [26]. Another study [27] reports a timing resolution of 12.6 ps (Type B) using a Field-Programmable Gate Array (FPGA) device, and a stability uncertainty of 8.4% (Type A), based on the standard deviation of the interference amplitude over time [27].

A.5 Environmental Effects: Changes in temperature, vibrations, and humidity increase uncertainty in QKD systems. These factors can disrupt alignment, detector efficiency, and timing, which leads to higher QBER and lower SKR [28]. To assess environmental impacts on quantum communication systems, a test chamber and various instruments can be used. (I) *Environmental Test Chambers* are used for offline lab testing to simulate controlled temperature and humidity variations, helping evaluate system stability under extreme conditions [29]. (II) *Accelerometers and Vibration Sensors* are deployed in both lab and field settings for real-time monitoring of mechanical disturbances that may affect optical alignment or system stability [30]. (III) *Fiber Bragg Grating Sensors* offer distributed, real-time measurements of strain and temperature along optical fibers, making them particularly suitable for field diagnostics [31]. (IV) *Infrared Thermometers and Thermocouples* provide continuous temperature readings of critical components, supporting both lab characterization and in-field operation [32].

IV. CASE STUDIES

This section summarizes four practical implementations of QKD. While much of the existing research focuses on designing and testing QKD systems, many papers lack key details needed to evaluate how reliable or accurate the results are.

A. Case Study 1: A Room-Temperature QKD System

The case study [23] presents a metrological evaluation of a room-temperature molecular single-photon QKD system in a free-space laboratory. Key quantum channel parameters were measured, with uncertainties interpreted based on experimental context. Channel loss was emulated

using up to 27 dB variable attenuators and quantified by measuring detector efficiency, reported as $30\% \pm 2\%$. This value was obtained through calibration with power meter measurements and spectral averaging, corresponding to a Type B uncertainty. The QBER was computed from photon detection statistics using single-photon avalanche detectors (SPADs), resulting in a value of $3.4\% \pm 0.2\%$ with Type A uncertainty based on statistical variation. The SKR was derived from the measured QBER and photon count rates, yielding up to 0.5 ± 0.05 Mbits/s in a back-to-back setup. This includes both Type A and Type B uncertainties. The DCR was extracted by fitting the QBER-loss relationship and was found to be in the range of 0.4 to 4×10^{-6} cps (Type A). Polarization drift was inferred from the detection error probability, which was $3.9\% \pm 0.5\%$ and reduced to 2.0% when a spectral bandpass filter was used. These effects are attributed to limitations in polarization control across a broadband emission spectrum, reflecting Type B uncertainty. The setup showed good repeatability under laboratory conditions and traceability through calibration-based measurements. Reproducibility was limited due to manual alignment and lack of environmental control. Despite these limitations, the study achieved a significant milestone by demonstrating a compact, room-temperature QKD system with performance comparable to cryogenic implementations [23].

B. Case Study 2: QKD Deployment Over 78 km Fiber

The paper [8] presents a QKD system over a 78 km fiber link between Braunschweig and Hannover, Germany. Entangled photon pairs were generated using a 532 nm laser via spontaneous parametric down-conversion in temperature-controlled ppKTP crystals, achieving entanglement fidelities of up to 95%. The signal photon is detected locally at Alice using a Single-Photon Avalanche Diode (SPAD)—a semiconductor detector sensitive to individual photons in the visible to near-infrared range. The idler photon travels through the fiber to Bob, where it is detected using an SNSPD that is ideal for telecom wavelengths [8]. The QKD system was characterized in both laboratory and field settings. Channel attenuation was measured using OTDR, revealing losses up to 29 dB over 78 km. QBER and SKR were derived from photon arrival times recorded by SPADs (Alice), SNSPD (Bob), and time taggers. SKR reached 1200 bps at 52 km and 30 bps at 102 km. Background photon counts of up to 5000 cps were observed due to crosstalk, mitigated by a 0.1 nm fibered bandpass filter. Polarization drift was compensated internally. From a metrological perspective, uncertainties arose from Raman noise (which is caused by scattered light from classical signals in optical fibers). The results were repeatable under laboratory conditions and reproducible when equivalent losses were introduced using long fiber spools or calibrated optical attenuators. While the spools simulated real-world transmission effects, the attenuators allowed for controlled testing without introducing additional fiber noise. Repeatability is supported by 24-hour performance averages and consistent trends in SKR and QBER. However, reproducibility remains only partially addressed, as no independent or cross-laboratory validation was conducted. Traceability is lacking-there is no reference to calibrated instruments, measurement standards, or SI-traceable procedures. Similarly, the study does not report any measurement uncertainty. Overall, the study's emphasis is on practical system performance rather than metrological analysis.

C. Case Study 3: QKD-TFD Integration on the NQL

The paper [9] presents a dual-purpose optical fiber infrastructure for both time and frequency dissemination (TFD) and QKD over an 78 km segment of the Niedersachsen Quantum Link (NQL), using separate fibers for each function. Link attenuation was $\sim$ 25 dB, characterized with an industrial OTDR and cross-validated using OPMs in both field and laboratory environments. Polarization-entangled photon pairs were generated via spontaneous parametric down-conversion in ppKTP crystals, pumped by a 532 nm laser. Detection was performed using SPADs at Alice and SNSPDs at Bob. The system achieved a QBER of 3% and a SKR of up to 30 bits/s over 102 km. Background photon rates due to Raman scattering were measured with SNSPDs, revealing levels around 10^{10} cps in the 10 nm QKD passband. Polarization control is achieved without direct SOP measurement by exploiting the use of a Faraday mirror, which passively compensates for polarization fluctuations in the fibre link. Synchronization was achieved using the *ELectronic STABilization (ELSTAB)* system, a commercial unit (OSTT-4L by PIKTime Systems [33]), which provides a time uncertainty of approximately 3 ps (1 PPS) over a 10 s averaging period and a frequency uncertainty below 3×10^{-13} at 1 s for a 10 MHz signal. Delay fluctuations were actively compensated. System performance was validated using a K+K FXE frequency counter (FXE-19 by K+K Messtechnik [34]), with a phase uncertainty of 12.2 ps and a frequency uncertainty of 1.2×10^{-11} at 1 s. The system achieved fractional frequency instability as low as 1×10^{-21} at 10^5 s averaging, it means the system's frequency changes almost not at all. Environmental influences were modeled rather than directly sensed, using Python-based simulations and veri-

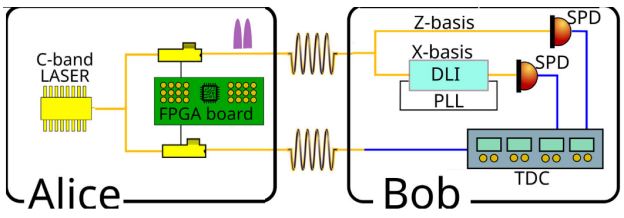


Fig. 2. A schematic of the QKD implementation, including the transmitter-receiver setup and synchronization [11].

Case Study	Key Attributes	Achievements	Limitations
Case 1 [23] Room-Temp QKD	QBER $3.4\% \pm 0.2\%$ (Type A); Channel loss $30\% \pm 2\%$ (Type B); DCR $\sim 10^{-6}$ cps (Type A)	SKR 0.5 ± 0.05 Mbits/s (Type A, B); Room-temp source; calibrated optical loss meaurements; defined uncertainties; traceable measurements	Lab-only; manual alignment limits reproducibility
Case 2 [8] 78 km Fiber QKD	26-29 dB loss (OTDR); SPAD + SNSPD detection; QBER 11%; DCR 5000 cps	Field deployment; SKR up to 1200 bps; 24-hr stability	No uncertainty or calibration traceability
Case 3 [9] QKD-TFD on NQL	$\sim$ 25 dB loss; QBER 3%; ELSTAB sync $\sim$ 3 ps uncertainty	Integrated QKD and TFD; SKR up to 30 bps; frequency instability $\sim 10^{-21}$	No uncertainty budget or traceability; polarization control indirect; environmental effects modeled
Case 4 [11] Inter-EU QKD Network	Link loss 14–25 dB; SPDs with 20% efficiency, DCR $\sim 2.5kcps$; QBER 0.82-7%; 1 ps timing resolution	SKR up to 3130 bps; stable 7-hr operation; some reproducibility across three field links	No uncertainty or polarization analysis; limited calibration details

Table 1. Concise metrological comparison of the case studies of QKD implementations.

fied via filtered SNSPD measurements [9]. This paper investigates critical elements of quantum channel characterization. Although the consistency between laboratory and field measurements demonstrates repeatability, the lack of traceability, standardized calibration procedures, and comprehensive uncertainty budgets constrains the overall metrological reliability of the results.

D. Case Study 4: The Inter-European QKD Network

The paper [11] describes a real-world deployment of a QKD network linking Italy, Slovenia, and Croatia via fiber optics, as shown in Fig. 2. Pulses were generated using a C-band laser. The system employed SPDs from ID Quantique, MPD, and Aurea, with a detection efficiency of 20% and a count rate in the range of 2500 cps. However, the detectors experienced dark counts of approximately 2.5 kcps. Timestamping was performed using TDCs from Qutools (QuTAG) and Swabian Instruments (Ultra), offering picosecond-level precision with 1 ps resolution. Synchronization was implemented using an intensity modulator controlled by an FPGA, in combination with TDCs, achieving timing accuracy better than one trillionth of a second. A Delay-Line Interferometer (DLI) was used to compare two time-separated light pulses and measure their phase difference. Combined with a PLL, it enabled phase adjustment through a feedback loop. Channel attenuation was estimated based on power loss across the fiber links, with reported values of 14 dB for the Trieste–Postojna and Ljubljana–Postojna links, and 25 dB for the Trieste–Rijeka link. SKRs reached up to 3130 bps, and QBERs ranged from 0.82–2.9% in the Z-basis and up to 7% in the X-basis. While dark counts were monitored, no calibration method was described, and polarization drift was not addressed. Environmental effects were mitigated using adaptive temporal filters (60–200 ps), which improved signal quality despite the absence of calibrated sensors From a metrological perspective, the study demonstrated good repeatability—evident in stable 7-hour operation—and some level of reproducibility across three field links with different configurations.

V. DISCUSSION AND RESEARCH GAPS

A comparative evaluation of the selected QKD implementations (Case Studies 1–4) highlights notable differences in how metrological principles are integrated into both experimental and field deployments. As summarized in Table 1, only Case Study 1 [23] provides a comprehensive metrological treatment including traceability, quantified uncertainties, and reproducibility under laboratory conditions. Case Study 1 demonstrates the most important metrological approach, incorporating both Type A and Type B uncertainty evaluations for critical parameters such as optical loss (reported as $30\% \pm 2\%$), QBER ($3.4\% \pm 0.2\%$), and SKR (0.5 ± 0.05 Mbps). The use of calibrated power meters and explicit uncertainty reporting establishes a traceable measurement foundation. However, its applicability is limited by the laboratory setting and manual alignment that restrict reproducibility. In contrast, Case Studies 2 [8], 3 [9], and 4 [11] emphasize practical system deployment, synchronization advances, and environmental integration but largely omit formal uncertainty quantification and traceability. For example, Case Study 2 reports fiber attenuation via OTDR (up to 29 dB) and SKR values (up to 1200 bps) without uncertainty budgets or calibration traceability. Similarly, Case Study 3 achieves sub-picosecond synchronization and frequency stability down to 10^{-21} but does not provide an uncertainty budget for QKD parameters nor direct SOP measurement. Case Study 4 demonstrates multi-country QKD operation with SKRs up to 3130 bps and timing resolution at the picosecond scale, yet lacks calibration details and uncertainty evaluation, especially regarding polarization effects.

From a metrological perspective, the following research gaps are evident: (I) *Traceability:* Most field-deployed systems lack references to SI-traceable standards or national metrology institutes, limiting confidence in measurement comparability. (II) *Uncertainty Quantification:* Formal error analyses and uncertainty budgets are generally absent, undermining the reliability of reported QKD performance metrics. (III) *Reproducibility:* While some temporal repeatability is documented (e.g., 24-hour stability in Case Study 2, 7-hour operation in Case Study 4), cross-system or cross-site reproducibility through independent validations or round-robin tests remains unaddressed. (IV) *Environmental Effects:* Temperature, vibration, and other environmental disturbances are acknowledged but not systematically modeled or incorporated into uncer-

tainty estimates, particularly in field conditions. (V) *Standardized Synchronization:* Timing and frequency synchronization methods are often undocumented without benchmarking against established metrological references, despite their critical role in minimizing QBER.

Bridging the gap between quantum communication engineering and metrology is essential to support reliable deployment and certification of QKD systems. Future research should prioritize integrating SI-traceable instruments, adopting uncertainty evaluation protocols consistent with the *Guide to the Expression of Uncertainty in Measurement* (GUM) [35], and developing standardized test procedures to ensure interoperability and comparability across QKD platforms.

VI. CONCLUSION

This paper examined QKD from a metrological perspective, highlighting the importance of traceability, repeatability, and uncertainty analysis in the characterization of key quantum channel parameters, including channel loss, DCR, QBER, SKR, and synchronization. To support this analysis, real-world case studies are examined. The detailed case studies showed that laboratory experiments tend to address metrological considerations more than field implementations. Notably, significant shortcomings were identified in the standardization of measurement practices, particularly in field settings where environmental effects introduce substantial measurement uncertainty and reduce comparability across systems.

REFERENCES

[1] Y.Cao, Y.Zhao, Q.Wang, J.Zhang, S.X.Ng, and L.Hanzo. The evolution of quantum key distribution networks: On the road to the qinternet. *IEEE Communications Surveys & Tutorials*, 24(2):839–894, 2022.

[2] P.Sharma, A.Agrawal, V.Bhatia, S.Prakash, and A.K.Mishra. Quantum key distribution secured optical networks: A survey. *IEEE Open Journal of the Communications Society*, 2:2049–2083, 2021.

[3] R.Steijl. Quantum information science-recent advances and computational science applications.

[4] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. International Vocabulary of Metrology — Basic and general concepts and associated terms (VIM). Joint Committee for Guides in Metrology, JCGM 200:2012. (3rd edition).

[5] B.Zhao, X.Zha, Z.Chen, R.Shi, D.Wang, T.Peng, and L.Yan. Performance analysis of quantum key distribution technology for power business. *Applied Sciences*, 10(8):2906, 2020.

[6] VIAVI Solutions Inc. What is an optical power meter?, n.d. Accessed: 2025-08-05.

[7] SISCO. What is an optical time domain reflectometer and how does it work?, n.d. Accessed: 2025-08-05.

[8] A.Hreibi, A.K.Kniggendorf, A.Kuhl, J.Kronjäger, H.Schnatz, and S.Kück. Entanglement-based intercity quantum key distribution: metrology and implementation. *Measurement: Sensors*, page 101777, 2025.

[9] A.K.Kniggendorf, A.Hreibi, J.Kronjäger, and S.Kück. Time and frequency dissemination and quantum key distribution on the niedersachsen quantum link–a practical approach. *Measurement: Sensors*, page 101776, 2025.

[10] G.Q.Ramzi and A.I.Khaleel. Performance evaluation of underwater bb84 protocol using various single photon counting modules. *Journal of Optics*, pages 1–9, 2025.

[11] D.Ribezzo, M.Zahidy, I.Vagniluca, N.Biagi, S.Francesconi, T.Occhipinti, L.K.Oxenløwe, M.Lončarić, I.Cvitić, M.Stipčević, et al. Deploying an inter-european quantum network. *Advanced Quantum Technologies*, 6(2):2200061, 2023.

[12] R.Nock, N.Dahnoun, and J.Rarity. Low cost timing interval analyzers for quantum key distribution. In *2011 IEEE international instrumentation and measurement technology conference*, pages 1–5. IEEE, 2011.

[13] L.Bacsardi, A.Mihaly, and K.Olah. From free-space quantum communication systems to satellite-based networks. In *2025 International Conference on Quantum Communications, Networking, and Computing (QCNC)*, pages 100–104. IEEE, 2025.

[14] D.Bolotov, M.Kuzmin, A.Nasaraia, N.Pchelkina, and S.Kuznetsov. A method for estimating losses in a quantum channel for implementing quantum key distribution technology for atmospheric laser communication terminals. In *2022 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF)*, pages 1–5. IEEE, 2022.

[15] G.Keiser. *Fiber Optic Communication Networks*, pages 507–575. Springer Singapore, Singapore, 2021.

[16] Losses in optical fiber. `https://www.scribd.com/document/485321971/Losses-in-optical-fiber`. Accessed: 2025-07-21.

[17] P.Kumar and A.Prabhakar. Bit error rates in a frequency coded quantum key distribution system. *Optics Communications*, 282(18):3827–3833, 2009.

[18] N.J.Muga, M.F.S.Ferreira, and A.N.Pinto. Qber estimation in qkd systems with polarization encoding. *Journal of Lightwave Technology*, 29(3):355–361, 2011.

[19] G.Bebrov. On the (relation between) efficiency and secret key rate of qkd. *Scientific Reports*, 14(1):3638, 2024.

[20] M.Rahmanpour, M.Khaje, A.Erfanian, M.Fahimifar, and A.Afifi. Afterpulse and dark count simulator for single photon avalanche detector. 2023.

[21] M.Li, N.Rodphai, C.Liu, Z.Wang, Z.Peng, J.Wang, N.Anfimov, D.Korablev, T.Lachenmaier, A.G.Olshevskiy, et al. Dark count rate stability of juno 20-inch pmts in mass testing. *arXiv preprint arXiv:2506.15164*, 2025.

[22] O.Bedroya, C.Li, W.Wang, J.Hu, H.Lo, and L.Qian. Resource-efficient real-time polarization compensation for mdi-qkd with rejected data. *Quantum*, 8:1435, 2024.

[23] V.Mayboroda, N.Rudavin, P.Kupriyanov, O.Fat'yanov, and R.Shakhovoy. Analytical solution to the problem of polarization drift compensation in an all-fiber qkd system. *Optics Express*, 32(26):45421–45435, 2024.

[24] R.D.Cochran and D.J.Gauthier. Qubit-based clock synchronization for qkd systems using a bayesian approach. *Entropy*, 23(8):988, 2021.

[25] A.Pljonkin, K.Rumyantsev, and P.Kumar Singh. Synchronization in quantum key distribution systems. *Cryptography*, 1(3):18, 2017.

[26] A.Czerwinski and S.Haddadi. Optimizing qkd efficiency by addressing chromatic dispersion and time measurement uncertainty. *Physics Letters A*, 526:129972, 2024.

[27] N.Chandra and P.K.Krishnamurthy. Fpga based synchronization of frequency domain interferometer for qkd. *IEEE Transactions on Quantum Engineering*, 2024.

[28] G.Granados, W.Velasquez, R.Cajo, and M.Antonieta-Alvarez. Quantum key distribution in multiple fiber networks and its application in urban communications: A comprehensive review. *IEEE Access*, 2025.

[29] M.A.Butt, G.S.Voronkov, E.P.Grakhova, R.V.Kutluyarov, N.L.Kazanskiy, and S.N.Khonina. Environmental monitoring: A comprehensive review on optical waveguide and fiber-based sensors. *Biosensors*, 12(11):1038, 2022.

[30] Y.Lei, R.Li, L.Zhang, P.Hu, and Q.Huang. Optical accelerometers for detecting low-frequency micro-vibrations. *Applied Sciences*, 12(8):3994, 2022.

[31] A.Theodosiou. Recent advances in fiber bragg grating sensing. *Sensors*, 24(2):532, 2024.

[32] C.Xu, X.Hao, P.Pei, T.Wei, and S.Feng. Research on time constant test of thermocouples based on qnn-pid controller. *Sensors*, 25(12):3819, 2025.

[33] Ostt-4 fiber-optic. `https://piktime.com/offer/ostt-4/`, 2025. Accessed: 2025-08-01.

[34] K+k messtechnik gmbh. `https://kplusk-messtechnik.de/`, 2025. Accessed: 2025-08-01.

[35] JCGM. Guide to the expression of uncertainty in measurement, 1995. JCGM 100:2008 (corrected version of first edition 1995).

Machine Learning-based Classification and Identification ofArabic Dialects through Acoustic Feature of Guttural Sounds.

Dr Hamed Al-Tairi
School of Information Technology
Whitecliffe College
New Zealand
20232006@mywhitecliffe.com

Dr. Seyed Ebrahim Hosseini
School of Information Technology
Whitecliffe College
New Zealand
seyedh@whitecliffe.ac.nz

Dr Shahbaz Pervez Chattha
School of Information Technology
Whitecliffe College
New Zealand
shahbazp@whitecliffe.ac.nz

***Abstract*— This study addresses the challenges of accurately identifying Arabic dialects due to their phonetic complexity and diversity. By developing a robust machine learning model for accurate Arabic dialect identification, the research focuses on guttural sounds and analyzes their acoustic features within Yemeni, Palestinian, Saudi, and Egyptian Arabic dialects. Utilizing a dataset of guttural sounds, we applied various machine learning models, including CatBoost, LightGBM, and Random Forest, to classify these dialects. Among the models tested, CatBoost demonstrated strong performance and generalization capabilities, effectively distinguishing between the dialects. This study contributes to the improvement of speech recognition technologies and offers valuable insights for linguistic research and communication technology.**

Keywords— ***Arabic dialect identification; Machine learning; Guttural sounds; Acoustic features; Speech recognition***

I. INTRODUCTION

Speech recognition in natural language processing (NLP) has revolutionized human-computer interaction, offering practical applications such as voice-activated assistants like Siri and Alexa, communication aids for the hearing impaired, and enhanced smart home device interactions. De-spite these significant advancements, accurately identifying and processing Arabic dialects remains a formidable challenge due to their phonetic complexity and diversity [1]. This paper aims to address these challenges by developing a robust machine learning model for accurate Arabic dialect identification. Arabic is a macrolanguage with many dialects that can differ significantly in phonetics, vocabulary, and grammar [2]. This diversity poses a unique challenge for speech recognition systems, which need to accurately identify and process these variations to function effectively. Accurate dialect identification is essential for improving the performance of voice-activated assistants, aiding forensic investigations, enhancing socio-linguistic research, optimizing media content localization, and supporting language learning applications.

There have been several recent studies, including those by [3-6], in the domain of Arabic dialect identification that have leveraged machine learning models to tackle the challenge of distinguishing between Arabic dialects. The studies primarily used audio recordings of speech, with specific features extracted such as Mel-Frequency Cepstral Coefficients (MFCC), Temporal Frequency Cepstral Coefficients (TFCC), segmental, and prosodic features. Researchers such as [3-6], have demonstrated the application of various machine learning techniques, from Hidden Markov Models (HMM) and Deep Neural Networks (DNN) to Support Vector Machines (SVM) and hybrid models. [3] utilized a combination of HMM and DNN, achieving an accuracy of 79% to 86% in accent recognition, although they faced challenges with recording quality variability and distinguishing closely related accents. [4] employed SVM to analyze segmental and prosodic features, effectively capturing dialectal variations without detailing specific limitations. [2] used features such as MFCC and TFCC with various classifiers, finding K-Nearest Neighbors (KNN) using TFCC features to be the most accurate at 76%. [6] explored a hybrid approach combining KNN and Decision Trees (DT), improving speech recognition accuracy by over 10%, though the study used a relatively small dataset.

Other notable studies include [7], who introduced the ADI17 dataset for fine-grained Arabic dialect identification, highlighting its value despite dialect diversity limitations. [8] combined phonetic and lexical features with a multi-class SVM, achieving 59.2% accuracy in dialect identification but faced challenges with code-switching and recording setup variability. [9] proposed a fusion approach with Convolutional Neural Networks (CNN), i-vectors, and transformer models, achieving 82.44% classification accuracy, though data augmentation was not employed. [10] utilized a combination of Gaussian Naïve Bayes, SVM, Recurrent Neural Networks (RNN), and CNN-RNN classifiers, with the CNN-RNN classifier achieving the highest precision, recall, and F1- score. [11] focused on Arabic dialect identification using CNN, achieving 83% accuracy, but noted limitations due to the small dataset size. [12] explored intonation patterns with hybrid LSTM networks, achieving state-of-the-art results but faced potential issues with generalizing to other datasets. [13] applied motif discovery in time series analysis, achieving 62.75% accuracy, while [14] introduced a transformer-based system, achieving 86.29% accuracy on the ADI17 dataset.

Ashifur and his team researcher [15] emphasized the critical need for extensive, high-quality, and publicly available Arabic speech datasets, identifying nine available corpora and discussing their applications in various challenging and realistic environments. They highlighted the importance of incorporating diverse features such as variations in input devices, dialects, age groups, environments, noise levels, and speech disabilities to create more robust Arabic ASR systems. Despite these advancements, existing research primarily focuses on broad dialectal differences without delving into finer phonetic nuances, particularly at the level of phonemes or sounds. This gap in the literature highlights the need for more

detailed exploration of specific phonetic features that can serve as critical markers for dialect identification. By concentrating on guttural sounds and analyzing their acoustic features, namely the first three formants of the vowels that are adjacent to the guttural sounds within Yemeni, Palestinian, Saudi, and Egyptian Arabic dialects, we can investigate how these acoustic features contribute to the classification and identification of Arabic dialects.

Guttural sounds in Arabic, including emphatic sounds such as /tQ/ and /sQ/ (also known as Tafkheem in Arabic), uvulars such as /X/ and /K/, and pharyngeals such as /è/ and /Q/, as well as laryngeals like /P/, are central to its phonetic system. These sounds exhibit significant variance in articulatory and acoustic properties across different dialects. Based on various investigations into Arabic gutturals in different dialects, previous studies have debated whether these sounds should be classified as one group or multiple groups due to their varied articulatory strategies. For example, researchers [16, 17] consider emphatics, pharyngeals, and uvulars as one group produced with the retraction of the tongue root, while [18, 19] argue that only pharyngeals involve tongue root retraction.

These divergent views suggest that guttural sounds could provide distinctive acoustic features that are crucial for dialect identification. This study focuses on the acoustic features of guttural sounds within Yemeni, Palestinian, Saudi, and Egyptian Arabic dialects. Our primary objective is to develop a machine learning model that classifies these dialects based on guttural sound patterns, representing a novel approach in the literature. By analyzing the first three formants (F1, F2, F3) of these sounds, we aim to improve the accuracy of speech recognition technologies. This research contributes to information technology with potential applications in communication technology, linguistic research, and beyond.

Following [20], we propose a comprehensive framework designed to tackle the complexities of Arabic dialect identification. Our approach includes the extraction of phonetic features and the application of various machine learning models such as CatBoost, LightGBM, and Random Forest. Each model is rigorously evaluated using cross-validation techniques to ensure generalizability and reliability. Feature importance analysis and hyperparameter tuning are conducted to optimize model performance. This paper is organized as follows: Section 1 provides a brief background of the study with the aim. Section 2 details the methodology, including data collection, preprocessing, and model training. Section 3 presents the results and Section 4 discusses the results and their implications.

II. Materials and Methods

A. Dataset Description

The data utilized in this study were derived from a previous PhD research by [21], which featured eight adult male participants from four different dialects. The study included two Egyptians (E1 from Upper Egypt and E2 from Cairo), one Palestinian from Jerusalem, two Saudis (S1 from Jeddah and S2 from Hail), and three Yemenis (Y1 from Dhamar, Y2, and Y3 from Hadhramout). At the time of the study, all participants were students at the University of British Columbia in Vancouver, Canada, aged between 24 and 35 years. Their time residing in Vancouver was less than five years, and none reported any hearing or speech impairments.

The dataset used in this research, derived from [21], is particularly relevant because it includes a diverse range of guttural sounds from different dialects, providing a rich source of phonetic variation that is critical for our analysis. These sounds are central to the phonetic systems of the dialects under study, making the dataset ideal for examining the unique acoustic features that distinguish these dialects. The study encompassed twenty-one test conditions, as detailed in Tables 1 and 2, with a focus on the non-round vowels /a:/, /i:/, and /u:/ across various conditions. Monosyllabic words featuring these long vowels were selected for their pronounced and extended formant trajectories relative to their shorter counterparts. The selected words, introduced to participants, contained either a plain coronal, emphatic (coronal), uvular, or pharyngeal at the onset. Subjects were instructed to articulate each target word in their native Arabic dialect within the carrier phrase [ga:lli. . . .] “he said to me. . . ”.

A total of 1,691 tokens were collected, with 640 tokens each for /a:/ and /i:/ across eight conditions, and 400 tokens for /u:/ across five conditions. Data collection involved simultaneous acoustic and articulatory (ultrasound) recordings at UBC using a Sennheiser MKH 416 P48 microphone and a desktop computer setup. Each participant was shown test words on a MacBook laptop through PowerPoint slides, with each word repeated ten times across two sessions.

Table 1: Test Conditions

Environment	CV:C
Vowels	(/a:/, /i:/) after (/t^{Ω}/, /t/, /s^{Ω}/, /s/, /X/, /K/, /è//, /Q/)
	(/u:/) after (/t^{Q}/, /t/, /X/, /è//, /Q/)
Total	21 test conditions

Table 2: Word List

Plain	Emphatics	Uvulars	Pharyngeals	Velar
ta:b ‘repented’	tQa:b ‘recovered’	Xa:b ‘failed’	èa:b ‘like’	ka:s ‘glass’
sa:b ‘left’	sQa:b ‘targeted’	Ka:b ‘was absent’	Qa:b ‘didn’t like’	ki:s ‘bag’
si:b ‘leave’	sQi:b ‘touch’	Xi:b pseudoword	èi:d ‘deviate’	ku:b ‘cup’
ti:b pseudo word	tQi:b ‘perfume’	Ki:b ‘be absent’	Qi:d ‘holiday’	

III. RESULTS

The training results reveal that the CatBoost model achieved the highest validation accuracy of 79.32%, followed closely by LightGBM at 75.64% and XGBoost at 74.22%. The Random Forest and Decision Tree models also performed well on the training data, with accuracies of 78.19% and 71.67%, respectively. Support Vector Machine (SVM) and Logistic Regression lagged behind, with accuracies of 56.66% and 37.68%. In terms of precision, recall, and F1-scores, CatBoost consistently outperformed other models across most classes, indicating its robustness. For instance, CatBoost achieved precision scores of 0.77, 0.68, 0.84, and 0.81 across different classes, which were higher than those of other models. Similarly, its recall and F1-scores were consistently strong, demonstrating its ability to balance precision and recall effectively.

On the test data, there was a noticeable drop in the performance of all models, indicating potential overfitting. CatBoost still led the pack with a validation accuracy of 66.45%, showing better generalization compared to other models. XGBoost and LightGBM followed with accuracies of 63.16% and 61.84%, respectively. The Random Forest and Decision Tree models also experienced performance drops, with test accuracies of 65.79% and 61.18%. The precision, recall, and F1-scores for CatBoost, while lower than on the training data, remained relatively high, reflecting its strong performance. For instance, CatBoost achieved a recall of 0.72, 0.33, 0.68, and 0.72, and F1- scores of 0.69, 0.39, 0.64, and 0.74.

This consistency indicates that CatBoost managed to mitigate overfitting better than other models, maintaining balanced precision, recall, and F1-scores across different class In summary, CatBoost emerged as the best-performing model, achieving the highest vali- dation accuracy on both training (79.32%) and test (66.45%) datasets. It consistently maintained high precision, recall, and F1-scores across different classes, demonstrating strong generalization capabilities. XGBoost and LightGBM also performed well but were slightly less effective in generalization compared to CatBoost. The significant drop in performance for models like Random Forest, Decision Tree, SVM, and Logistic Regression highlights the challenges of overfitting and the importance of robust model selection and tuning.es.

Table 3: Training Results for Various Models

Model	Val Acc	Prec (avg)	Recall (avg)	F1-Score (avg)
Random Forest	0.7819	0.77	0.75	0.76
Logistic Regression	0.3768	0.39	0.39	0.37
Support Vector Machine	0.5666	0.55	0.56	0.55
Decision Tree	0.7167	0.70	0.69	0.69
XGBoost	0.7422	0.73	0.71	0.72
LightGBM	0.7564	0.74	0.74	0.74
CatBoost	0.7932	0.78	0.77	0.77

Table 4: Test Results for Various Models

Model	Val Acc	Prec (avg)	Recall (avg)	F1-Score (avg)
Random Forest	0.6579	0.63	0.57	0.58
Logistic Regression	0.3684	0.35	0.33	0.33
Support Vector Machine	0.4408	0.46	0.49	0.44
Decision Tree	0.6118	0.56	0.56	0.56
XGBoost	0.6316	0.61	0.58	0.59
LightGBM	0.6184	0.60	0.59	0.59
CatBoost	0.6645	0.62	0.61	0.61

A. Hyperparameter Tuning for Models

The tables provided summarize the performance of CatBoost and Random Forest models on both training and test datasets. For the training data, the CatBoost model was optimized with parameters including a learning rate of 0.1, l2 leaf regularization of 3, 500 iterations, and a depth of 8, achieving a best score of 0.7844 and a validation accuracy of 0.8017. The precision, recall, and F1-score per class were [0.79, 0.67, 0.85, 0.82], [0.84, 0.69, 0.77, 0.84], and [0.82, 0.68, 0.81, 0.83], respectively. The Random Forest model, with parameters such as 100 estimators, a minimum samples split of 5, a minimum samples leaf of 1, 'auto' max features, a max depth of 20, and bootstrap set to True, achieved a best score of 0.7674 and a validation accuracy of 0.7819. Its precision, recall, and F1-score per class were [0.79, 0.67, 0.80, 0.80], [0.77, 0.64, 0.77, 0.85], and [0.78, 0.66 0.79, 0.82], respectively.

For the test data, the CatBoost model retained its optimized parameters and achieved a best score of 0.6817 and a validation accuracy of 0.6579. The precision, recall, and F1-score per class were [0.66, 0.47, 0.57, 0.77], [0.72, 0.39, 0.62, 0.72], and [0.69, 0.42, 0.59, 0.74], respectively. Nm Similarly, the Random Forest model, with its best parameters, achieved a best score of 0.6789 and a validation accuracy of 0.6579. Its precision, recall, and F1-score per class were [0.66, 0.57, 0.62, 0.68], [0.72, 0.22, 0.53, 0.82], and [0.69, 0.32, 0.57, 0.74], respectively. An additional test was performed on the CatBoost model with selected features, resulting in a validation accuracy of 0.6436. The precision, recall, and F1-score per class were [0.64, 0.47, 0.54, 0.79], [0.64, 0.47, 0.59, 0.75], and [0.64, 0.47, 0.57, 0.77], respectively.

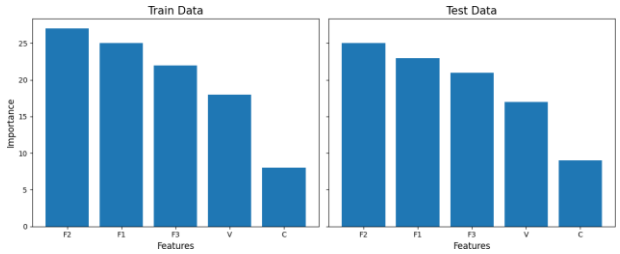


Figure 4: Feature Importance of CatBoost Model

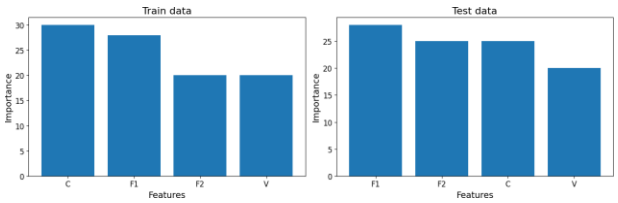


Figure 5: Selected Features in CatBoost Model

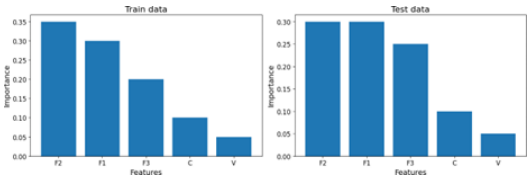


Figure 6: Feature Importance in Random Forest Model

Table 5: Summary of Model Parameters and Score

Data Type	Model	Best Parameters	Best Score	Validation Accuracy
Train Data	CatBoost	{learning_rate: 0.1, l2_leaf_reg: 3, iterations: 500, depth: 8}	0.7844	0.8017
Train Data	Random Forest	{n_estimators: 100, min_samples_split: 5, min_samples_leaf: 1, max_features: auto, max_depth: 20, bootstrap: True}	0.7674	0.7819
Test Data	CatBoost	{learning_rate: 0.1, l2_leaf_reg: 3, iterations: 500, depth: 8}	0.6817	0.6579
Test Data	Random Forest	{n_estimators: 100, min_samples_split: 2, min_samples_leaf: 1, max_features: auto, max_depth: None, bootstrap: True}	0.6789	0.6579

Table 6: Summary of Validation and Test Metrics

Data Type	Model	Precision	Recall	F1-score	Accuracy
2*Train Data	CatBoost	0.79, 0.67, 0.85, 0.82	0.84, 0.69, 0.77, 0.84	0.82, 0.68, 0.81, 0.83	0.7532
	Random Forest	0.79, 0.67, 0.80, 0.80	0.77, 0.64, 0.77, 0.85	0.78, 0.66, 0.79, 0.82	-
2*Test Data	CatBoost	0.66, 0.47, 0.57, 0.77	0.72, 0.39, 0.62, 0.72	0.69, 0.42, 0.59, 0.74	0.6436
	Random Forest	0.66, 0.57, 0.62, 0.68	0.72, 0.22, 0.53, 0.82	0.69, 0.32, 0.57, 0.74	-

IV. Discussion

The results of this study highlight the efficacy of various machine learning models in classifying Arabic dialects based on guttural sound features, with a particular focus on CatBoost, LightGBM, and Random Forest classifiers. The CatBoost model demonstrated superior performance, achieving the highest validation accuracy of 79.32% on the training data and 66.45% on the test data. This indicates CatBoost's robustness in generalizing well to unseen data, making it a suitable choice for complex tasks such as Arabic dialect identification where subtle phonetic variations are crucial.

The success of the CatBoost model can be attributed to its efficient handling of categorical features and its ability to manage overfitting through hyperparameter tuning and cross-validation. The model's precision, recall, and F1-scores were consistently higher than those of other models across different classes, emphasizing its balanced performance. The feature importance analysis revealed that formant frequencies (F1, F2, F3) were the most significant features for dialect classification, corroborating the linguistic hypothesis that guttural sounds play a pivotal role in distinguishing Arabic dialects.

The Random Forest model also performed well, achieving a validation accuracy of 78.19% on the training data and 65.79% on the test data. However, its performance was slightly lower than that of CatBoost, indicating that while Random Forest is effective, it may not capture the intricate patterns in the data as efficiently as CatBoost.Despite these promising results, the study has several limitations. The dataset used was relatively small and homogeneous, with only eight male participants from specific regions. This limited sample size might not fully represent the extensive variability within Arabic dialects, potentially affecting the generalizability of the findings. Additionally, the focus on male speakers could introduce gender bias, limiting the applicability of the models to female speakers.

The implications of these results are significant for both technological and linguistic fields. For speech recognition technology, the findings suggest that incorporating machine

learning models like CatBoost, which effectively handle phonetic features, can substantially improve the accuracy of dialect identification. This can lead to better performance in voice-activated assistants, enhancing user experience and accessibility. In forensic investigations and socio-linguistic research, accurate dialect identification can provide more precise insights into speaker origins and regional language variations. Furthermore, in media content localization and language learning applications, these advancements can facilitate more accurate and culturally relevant translations and educational materials.

Future research should address these limitations by incorporating a larger and more diverse participant pool, including female speakers and individuals from a wider range of Arabic-speaking regions. Furthermore, exploring more advanced deep learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), might enhance the ability to capturecomplex acoustic features and improve classification accuracy. In conclusion, this study presents a novel approach to Arabic dialect identification using ma- chine learning models focused on guttural sound features. The findings underscore the importance of feature selection and model tuning in achieving high classification accuracy. While CatBoost emerged as the best-performing model, future work should aim to enhance the dataset's diversity and explore more sophisticated modelling techniques to further improve performance and generalizability. The implications of these findings extend to improving speech recognition technologies, forensic linguistics, socio-linguistic research, media localization, and language learning tools.

V. Conclusion

This study has demonstrated the potential of machine learning models, particularly CatBoost, for the classification of Arabic dialects based on the acoustic features of guttural sounds. By focusing on formant frequencies of emphatics, uvulars, and pharyngeals, we have highlighted the significance of these phonetic features in distinguishing between Yemeni, Palestinian, Saudi, and Egyptian Arabic dialects. The CatBoost model achieved the highest accuracy, showcasing its robustness and ability to generalize well to unseen data. This underscores the model's capability to handle the complexity and variability inherent in Arabic dialects, providing a reliable tool for enhancing speech recognition technologies.

VI. Limitations and Future Work

One of the dataset's primary limitations is its relatively small sample size, which might not fully capture the extensive variability within each represented dialect. Additionally, the participant pool is exclusively male, potentially limiting the generalizability of findings across genders. Another challenge arises from the inherent complexity of dialectal Arabic phonetics, where subtle acoustic differences may be difficult to capture and classify accurately.

To mitigate these limitations, future studies are suggested to:

- Include a larger, more diverse participant pool, incorporating female speakers and a wider range of dialects to enhance the dataset's representativeness.
- Employ sophisticated acoustic analysis techniques and machine learning models capable of capturing subtle phonetic variations. This includes the potential use of deep learning approaches that can learn complex patterns from limited data.

References

[1] H. A. Alsayadi, A. A. Abdelhamid, I. Hegazy, and Z. T. Fayed, "Arabic speech recognition using end‐to‐end deep learning," IET Signal Processing, vol. 15, no. 8, pp. 521-534, 2021.

[2] S. Jamal et al., "In the identification of arabic dialects: A loss function ensemble learning based-approach," in International Conference on Model and Data Engineering, 2022: Springer, pp. 89-101.

[3] E. Alsharhan and A. Ramsay, "Robust automatic accent identification based on the acoustic evidence," International Journal of Speech Technology, vol. 26, no. 3, pp. 665-680, 2023.

[4] G. Brown and S. Hellmuth, "Computational modelling of segmental and prosodic levels of analysis for capturing variation across Arabic dialects," Speech Communication, vol. 141, pp. 80-92, 2022.

[5] K. M. Nahar, O. M. Al-Hazaimeh, A. Abu-Ein, and M. A. Al-Betar, "Arabic dialect identification using different machine learning methods," 2022.

[6] B. Mouaz, C. Walid, B.-H. Abderrahim, and E. Abdelmajid, "A new framework based on KNN and DT for speech identification through emphatic letters in Moroccan dialect," Indonesian Journal of Electrical Engineering and Computer Science, vol. 21, no. 3, pp. 1417-1423, 2021.

[7] S. Shon, A. Ali, Y. Samih, H. Mubarak, and J. Glass, "ADI17: A fine-grained Arabic dialect identification dataset," in ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020: IEEE, pp. 8244-8248.

[8] A. R. Ali, "Multi-dialect Arabic speech recognition," in 2020 International Joint Conference on Neural Networks (IJCNN), 2020: IEEE, pp. 1-7.

[9] M. A. Humayun, H. Yassin, and P. E. Abas, "Dialect classification using acoustic and linguistic features in Arabic speech," IAES International Journal of Artificial Intelligence, vol. 12, no. 2, p. 739, 2023.

[10] M. Abdelazim, W. Hussein, and N. Badr, "Automatic Dialect identification of Spoken Arabic Speech using Deep Neural Networks," International Journal of Intelligent Computing and Information Sciences, vol. 22, no. 4, pp. 25-34, 2022.

[11] M. Alrehaili, T. Alasmari, and A. Aoalshutayri, "Arabic speech dialect classification using deep learning," in 2023 1st International Conference on Advanced Innovations in Smart Cities (ICAISC), 2023: IEEE, pp. 1-5.

[12] A. A. Alvarez and E. S. A. Issa, "Learning intonation pattern embeddings for arabic dialect identification," arXiv preprint arXiv:2008.00667, 2020.

[13] M. Moftah, M. Fakhre, and S. El-Ramly, "Spoken Arabic Dialect Identification Using Motif Discovery," The Egyptian Journal of Language Engineering, vol. 5, no. 1, pp. 25-36, 2018.

[14] W. Lin, M. Madhavi, R. K. Das, and H. Li, "Transformer-based Arabic dialect identification," in 2020 International Conference on Asian Language Processing (IALP), 2020: IEEE, pp. 192-196.

[15] A. Rahman, M. M. Kabir, M. F. Mridha, M. Alatiyyah, H. F. Alhasson, and S. S. Alharbi, "Arabic Speech Recognition: Advancement and Challenges," IEEE Access, 2024.

[16] S. Davis, "Emphasis spread in Arabic and grounded phonology," Linguistic inquiry, pp. 465-498, 1995.

[17] S. Rose, "Variable laryngeals and vowel lowering," Phonology, vol. 13, no. 1, pp. 73-117, 1996.

[18] M. S. Bin-Muqbil, Phonetic and phonological aspects of Arabic emphatics and gutturals. The University of Wisconsin-Madison, 2006.

[19] S. A. M. Shar, "Arabic emphatics and gutturals: A phonetic and phonological study," 2012.

[20] A. Géron, Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. " O'Reilly Media, Inc.", 2022.

[21] H. Al-Tairi, "Tongue Retraction in Arabic: An Ultrasound and Acoustic Study," ResearchSpace@ Auckland, 2018.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Measuring Privacy: Critical Reflections and Directions for a Metrology-Based Approach

Eulalia Balestrieri[1], Ilaria Amelia Caggiano[2], Francesco Picariello[3], Ioan Tudosa[1]

[1] *University of Sannio, Benevento, Italy, {balestrieri, ioan.tudosa}@unisannio.it*
[2] *University Suor Orsola Benincasa, Naples, Italy, ilaria.caggiano@unisob.na.it*
[3] *Universitas Mercatorum, Rome, Italy, francesco.picariello@unimercatorum.it*

***Abstract* – Privacy measurement in digital systems lacks a standardised metrological framework to ensure reliable and comparable assessments. A metrological approach ensures that privacy measurements are reliable, reproducible, and comparable over time and across different contexts. In this work most common privacy metrics, including k-anonymity, ℓ-diversity, t-closeness, differential privacy, and mutual information, are critically evaluated, identifying their strengths and limitations from a metrological perspective. Initial directions and open challenges toward a metrology-based approach to privacy measurement are outlined, too.**

I. INTRODUCTION

The increasing reliance on personal data in modern digital systems has made privacy protection a central concern across multiple domains. Although regulatory frameworks have established strong principles for data protection, the quantitative assessment of privacy remains an open challenge. Numerous privacy metrics have been proposed, aimed at attempting to quantify the level of protection afforded to individuals against data disclosure, but they often lack the foundational properties expected of reliable measurement tools. The term "metric" is often used loosely, without a consistent reference to measurement theory.

Metrology provides a well-established framework for evaluating the quality of metrics. Key concepts to do that are included in the International Vocabulary of Metrology (VIM) [1] and the Guide to the Expression of Uncertainty in Measurement (GUM) [2]. Although developed for physical measurements, they are increasingly being adapted to digital and informational domains. Efforts by institutions such as PTB (Physikalisch-Technische Bundesanstalt) through its “Metrology for Digital Transformation” initiatives (e.g., D-SI metadata model and European Metrology Cloud projects) [3], and NIST (National Institute of Standards and Technology), through its extension of legal metrology principles such as the “chain of trust” to digital measuring systems, are beginning to explore the intersection of metrology and digital trust [4]. Privacy is governed by legal frameworks like the European General Data Protection Regulation (GDPR) [5], which emphasise principles such as data minimisation, purpose limitation, and accountability. However, these regulations do not prescribe specific quantitative metrics, further highlighting the need for rigorous and standardised measurement approaches. This paper investigates the applicability of metrological principles to the most commonly used privacy metrics, aiming to identify gaps, promote rigour, and suggest directions for future standardisation. The paper is structured as follows. First, a brief overview of existing privacy regulations and standards is provided, followed by a description of the most commonly used privacy metrics. Next, the relationship between these privacy metrics and the metrological approach is examined, with particular attention to the challenges and requirements for building a metrology-based framework to measure privacy. Finally, conclusions are drawn and directions for future work are outlined.

II. EXISTING PRIVACY REGULATIONS AND STANDARDISATION

Modern data protection regulations emphasise accountability, that is, the obligation of data controllers and processors to demonstrate compliance with legal and ethical obligations. This entails not only the implementation of appropriate technical and organisational measures but also the ability to prove their effectiveness. Common themes concern emphasis on proactive measures and continuous assessment, encouragement (or requirement) of risk-based evaluations, focus on documentation, and demonstrable safeguards.

However, these requirements often lack concrete guidance on how to measure or verify the effectiveness of privacy measures.

Regulatory frameworks and standardisation efforts worldwide have established the legal and procedural basis for privacy protection, but often without clearly defined, quantitative metrics. This lack of metrological rigour hinders the objective assessment and comparability of privacy-preserving techniques.

The European Union’s GDPR (Regulation (EU) 2016/679) [5] is among the most influential data protection

regulations. While it emphasises principles such as data minimisation, purpose limitation, and accountability, it does not prescribe specific measurement methods. However, Recital 78 encourages the use of privacy-enhancing technologies and data protection by design, indirectly requiring accessible levels of privacy.

ISO/IEC 29100 [6] provides a high-level privacy framework but lacks operational metrics. ISO/IEC 27701 [7] extends ISO 27001 with controls for Privacy Information Management Systems (PIMS), introducing risk-based approaches. Nevertheless, privacy risk is typically evaluated qualitatively or semi-quantitatively, based on likelihood and impact, without standardised units of measure or uncertainty.

The NIST Privacy Framework [8] adopts a structured, flexible, and voluntary model to help organisations identify and manage privacy risks. Although it includes control categories like “Minimize Data Collection and Retention” or “Manage Data Processing,” it does not define universal or standardised privacy metrics. However, related NIST documents (e.g., NISTIR 8062 [9]) discuss the concept of privacy engineering and encourage the development of “measurable privacy goals”, leaving implementation to specific domains.

The International Telecommunication Union has published several technical standards (e.g., ITU-T X.1149 [10]) on privacy risk management and X.1205 [11] on cybersecurity, which include references to impact assessments and, in some cases, probabilities of leakage. However, they fall short of providing reproducible and standardised measurement protocols.

Overall, while these frameworks recognise the importance of assessing privacy risk, few provide formalised, metrology-compliant metrics, nor do they establish traceable, uncertainty-aware measurement practices.

III. EXISTING PRIVACY METRICS

Over the past two decades, a variety of privacy metrics have been developed across academic and applied settings. There have been so many different and complex metrics, making the choice among them very challenging, that rather than choosing, it is often preferred to propose and create new metrics. Consequently, the comparison among different privacy systems, in terms of the assured degree of privacy to the users, as well as the level of protection assured by privacy-enhancing technologies, becomes impossible to carry out.

There is not the same approach in defining the privacy metric and its requirements and functioning. Not all proposed metrics have a mathematical base [12]. Multiple areas are of concern regarding assured privacy levels, as communication systems, databases, location services, smart metering, social networks, etc. [12]. Therefore, many privacy metrics are context-dependent and often lack comparability and interpretability [12].

The most common and used metrics for privacy are k-anonymity, ℓ-diversity, t-closeness, Differential Privacy (DP) and entropy-based metrics.

k-anonymity ensures that each individual in the dataset is indistinguishable from at least k-1 other individuals with respect to Quasi-Identifying attributes (QI) (e.g., age, gender). The dataset is transformed (generalisation, suppression) so that each combination of QI appears at least k times. It does not protect against attacks on homogeneous sensitive attributes within a group [12],[13],[14],[15].

ℓ-diversity extends k-anonymity by introducing the concept of diversity in the sensitive attributes within each anonymous group. Each anonymous group must have at least ℓ different sensitive attribute values (e.g., different disease categories). It can be bypassed if the sensitive values are semantically similar (e.g., "severe" vs. "very severe"). Although it provides a level of heterogeneity, the definition of diversity can be subjective [12],[13],[15].

t-Closeness requires each anonymous group to have a distribution of sensitive attributes that is "close" (within a threshold t) to the distribution of the entire dataset. It uses statistical distance metrics (e.g., Earth Mover's Distance) to quantify proximity. It requires more complex calculation, dependent on the choice of threshold t [12],[15].

DP defines a formal limit on the amount of information that can be inferred from an output, regardless of prior knowledge. It adds calibrated noise (e.g., Laplacian noise) to statistical query responses to ensure that the presence or absence of an individual has a limited impact [12],[15]. The scale of the noise added to outputs is controlled by the parameter ε; a smaller ε requires injecting more noise, making it harder to infer whether any individual’s data was included. A larger ε allows less noise, improving utility but reducing privacy protection.

Entropy-based metrics quantify indeterminacy about private information (e.g., Shannon entropy, conditional entropy); higher entropy generally implies higher privacy. They are sensitive to the granularity of data representation and may overestimate privacy if the attacker's knowledge is not uniform. Mutual Information (MI) is a special case of entropy-based metrics that measures information leakage. It quantifies the reduction in indeterminacy (entropy) of one variable due to knowledge of another. Lower MI indicates better privacy. It requires accurate estimation of joint probability distributions, which is challenging for large or continuous datasets, is sensitive to assumptions about data distribution and computationally intensive for high-dimensional data [12],[15]. It is worth noting that k-anonymity, ℓ-diversity, t-closeness and DP are in fact privacy models or mechanisms, originally conceived as techniques to protect data. MI, instead, is an evaluation metrics that quantify information leakage.

IV. METROLOGY AND PRIVACY METRICS

Measuring privacy is complex because privacy is a

multidimensional concept and often context-dependent. A metrological approach for privacy metrics ensures that privacy measurements are reliable, reproducible, and comparable over time and across different contexts. When metrics are based on metrological principles, other people or systems can replicate the measurements and obtain similar results, which is essential for verifying and validating privacy levels. Standardising metrics allows comparing the results of different privacy solutions or technologies, making it easier to choose the best strategy or technology. Accurate and standardised measurements help companies, regulators, and developers make informed decisions and correctly assess privacy risks and protections. Moreover, a metrological method enables monitoring how metrics change over time, useful for evaluating the effectiveness of implemented measures and for continuous improvement.

A metrological compliant privacy metric is required to have a well-defined measurand (the quantity being measured) and a meaningful estimate of uncertainty.

A valid measurement also requires repeatability (the ability to obtain consistent results under the same conditions) and reproducibility (the ability to replicate results across different settings and analysts). In the following subsection, the measurands of the most common privacy metrics and their repeatability and reproducibility are analysed from a metrological point of view.

A. Privacy Metrics measurands

Table 1 provides concise definitions of these measurands, summarising the quantities that each metric is intended to measure according to current practice in the literature. k-anonymity is easy for people to interpret, while DP and MI require a technical background. DP and MI are theoretically precise, while k-anonymity, ℓ-diversity, and t-closeness rely on contextual definitions of QI and sensitive attributes, which can vary.

For k-anonymity and ℓ-diversity, the measurand is not a continuous quantity but a cardinality, and the scale is discrete.

For t-closeness, the measurand is a distance between distributions and depends on the chosen distance function; the unit is conventional (dimensionless real value).

For DP, ε is unitless (log of the odds ratio). It is a dimensionless quantity on a ratio scale.

For MI, the amount of shared information is measured in bits.

The interpretation of metric values differs across methods. For k-anonymity and ℓ-diversity, higher values correspond to greater privacy. Conversely, t-closeness achieves stronger privacy with lower values.

In practice, the choice of parameter values depends on the dataset and the desired privacy level.

The k-anonymity threshold is most commonly set at 3 or 5 [13]. Values above 5 are uncommon, and values exceeding 15 are rare [14]. However, in other cases, it can be found that considering values of 10 or lower are insufficient to ensure adequate data protection [15]. It is worth noting that the selection of k must carefully balance privacy protection with data utility, adapting to the specific context and objectives of the study.

Similar to k-anonymity, for ℓ-diversity, there is no ideal value for ℓ, although it is typically less than or equal to k and greater than 1 [15],[16].

Values of t are typically between 0 and 1 [17],[18]. Lower values of t (for example, 0.1 or 0.05) mean that the distributions within each group closely match the overall distribution, resulting in stronger privacy [17],[18],[19]. Higher values of t (for example, 0.2 or 0.3) permit greater variation, which can improve data utility but reduce privacy [17],[18]. The suitable value of t depends on the specific dataset, the sensitivity of the attribute, and the desired privacy level. It is typically determined through experimentation, aiming to balance data privacy with usefulness [17],[18].

Concerning DP, smaller values of ε offer greater privacy protection but may reduce data accuracy. Larger values of ε may improve accuracy but reduce privacy protection. The choice of ε remains a debated issue in the community. While theoretical works generally recommend small values ($\varepsilon < 1$) to guarantee strong privacy, in practice, larger values are often used. Values in the range 0–5 are considered conservative and provide robust protection in most contexts [20]. However, higher values ($5 < \varepsilon < 20$) may still be acceptable in some practical settings [20].

For MI, small values (>0, but low) express a small dependency, so a minimal risk of privacy leak. Large values instead express strong correlation and so a greater risk of inferring sensitive data.

Table 1. Metrics measurands

Metric	Measurand
k-anonymity	The minimum equivalence class size (number of records with identical QI after generalisation/suppression).
ℓ-diversity	The minimum number of distinct sensitive attribute values in any equivalence class.
t-closeness	The maximum statistical distance between the sensitive attribute distribution in any equivalence class and the global distribution.
Differential Privacy	The maximum change in the output distribution of a randomised algorithm when any one individual's data is changed, expressed by the parameter ε.
Mutual Information	The amount of information that observable attributes reveal about sensitive data.

Most privacy metrics are defined in terms of parameters chosen a priori, k in k-anonymity, ℓ in ℓ-diversity, t in t-

closeness, and ε in DP. These parameters are fixed by design and do not directly reflect the actual, realised protection of user data in a given dataset. For DP, ε represents a theoretical bound on leakage, not an observable quantity. Similarly, k, ℓ, and t are design thresholds rather than measurements of effective risk.

MI quantifies information leakage but can vary depending on dataset realisation and noise mechanisms, making it partially suitable for statistical treatment.

Existing privacy metrics are largely based on design parameters that do not directly quantify actual information leakage. They are often context-dependent, rely on abstract or discrete scales, and can have non-intuitive interpretations. By defining an observable measurand, privacy can be assessed through a direct, consistent, and quantitative indicator, enabling rigorous evaluation, reproducibility, and meaningful comparison across datasets and methods.

B. Repeatability and Reproducibility of Privacy Metrics

Existing privacy metrics were not originally conceived with requirements concerning repeatability and reproducibility in mind, which raises significant challenges for their adoption as standardised measures.

k-anonymity, ℓ-diversity, and t-closeness are formally well-defined metrics whose reported values are fully repeatable when applied to the same dataset. Being deterministic metrics, they are reproducible for definition [18],[21],[22],[23]. This doesn't mean that if the same k, ℓ, or t is achieved, there are no variations in the resulting dataset in the case of applying randomisation or heuristic tie-breaking, common in greedy or optimisation-based anonymisation methods [18].

Beyond algorithmic factors, the practical utility of these metrics as indicators of privacy is limited by external conditions. The actual risk of re-identification depends heavily on the adversary's background knowledge, access to external datasets, and data characteristics such as skewed distributions of sensitive attributes or attribute correlations [18],[21],[22],[23]. Empirical measures of re-identification risk can vary significantly across datasets, implementations, or even repeated applications, limiting the metrics' reliability as indicators for real-world privacy protection [18],[21],[22],[23].

DP introduces a formal guarantee based on the privacy parameter ε, ensuring that the influence of any single individual on the output is strictly bounded. However, different mechanisms calibrated to the same ε can produce substantially different empirical outcomes in practice, due to variations in function sensitivity, noise distribution, or implementation details. While ε defines a theoretical upper bound on privacy loss, it does not directly quantify the actual risk or information leakage in a specific dataset or execution. Moreover, the intrinsic randomness of DP mechanisms means that repeated executions on the same dataset may yield different outputs, affecting the operational reproducibility of reported results. Consequently, transparent and standardised evaluation protocols are essential to assess the effective privacy and ensure comparability and reproducibility of empirical measurements across studies [24].

Empirical estimation of Mutual Information (MI) can exhibit high variability due to factors such as sample size, underlying data distribution, and the choice of estimator. Methods like k-nearest neighbours (kNN) or neural-based approaches such as MINE may yield different values even on the same dataset, depending on parameter settings [25],[26]. Unlike the theoretical MI, which is well-defined mathematically, empirical estimates carry intrinsic uncertainty that must be quantified and reported. Standardised protocols for MI estimation, including consistent data preprocessing, estimator selection, and uncertainty evaluation, are therefore essential to ensure reproducible and comparable measurements across studies and datasets.

Overall, these limitations highlight that, despite their solid theoretical foundations, existing privacy metrics exhibit substantial variability in empirical estimation and limited metrological reproducibility, and therefore cannot be considered fully consistent or universally comparable measures of real-world privacy risk.

V. TOWARD A METROLOGY-BASED FRAMEWORK

Privacy should be defined as a physical informational quantity, that is, something that can be observed, simulated, or calculated in a replicable way. Traditional metrics such as k-anonymity, ℓ-diversity, t-closeness, DP, or MI are relative definitions: they describe properties of data with respect to a dataset or algorithm, but they do not rely on a universally accepted physical or informational unit. Each of these metrics measures a different aspect of privacy: k-anonymity expresses indistinguishability between records; ℓ-diversity ensures protection against inference of sensitive attributes; t-closeness preserves the statistical distribution of attributes; DP captures the probability of inference for a single record; MI quantifies the amount of information revealed. However, there is no way to combine them into a single, consistent number.

Different metrics capture different dimensions of privacy, but without a common unit or scale, it is not possible to rigorously compare one solution with another. If the value of a metric is just a parameter set upstream (for instance, k, ℓ, t, or ε), it makes no sense to speak about uncertainty regarding it. Uncertainty only makes sense with respect to quantities that are actually measured or observable, that is, with respect to the effective privacy that a system guarantees. Rather than referring to fixed parameters, it would be more useful and effective to measure observed privacy or actual risk. All systems should therefore be evaluated with the same observable quantity and the same protocol. Comparisons between systems must be based on

observable measurement ± uncertainty, and this requires a metric that is an observable quantity on real data, reflecting the actual risk or the actual amount of information revealed.

The measurement must represent how effectively a system protects user data, regardless of the nominal parameter chosen. In order to be scientifically rigorous, such a metric must be: observable on real data (measurable); comparable across different systems; and quantifiable with uncertainty. Building this metric by combining the existing ones can have some advantages: it leverages results that are already consolidated and recognised by the community, it is easier to be accepted by standardisation bodies such as ISO or NIST, and it covers various attack scenarios. On the other hand, each metric has different semantics, making them difficult to compare. Adding uncertainty becomes complicated, and there is the risk of creating a "patchwork index" without true metrological consistency.

By inventing a new metric from scratch, defined metrologically, it could be designed from the outset as an observable measurand with associated uncertainty. This would make it easier to build a common scale and would allow direct comparisons between different systems and technologies. However, such a new metric would require significant effort in theoretical modelling and experimental validation, and it could be perceived as too new, struggling to become standardised, unless the community is convinced to converge toward it.

A combination of approaches may be appropriate to integrate different measurement perspectives into a composite indicator; multiple risk dimensions can be joined into a single observable measure. A standard measure of effective privacy could be derived from a weighted or aggregated combination of existing metrics, after transforming them into observable and normalised quantities. Such a composite measure would provide several advantages: different types of attacks or inferences are covered; the risk of overestimating protection based on a single metric is reduced; heterogeneous systems can be compared with a single index.

Nevertheless, several common metrological challenges remain. First, the definition of the measurand: current metrics do not directly measure the observable phenomenon of interest, such as the probability of re-identification or the extent of information loss. It is therefore necessary to clearly define what is being observed and in what units it is measured. Second, the problem of common units of measurement: each metric currently uses different scales, such as number of records, probabilities, distances, or bits of information. To compare different systems, a normalised scale is required, for example, an index ranging from 0 to 1. Third, the problem of associated uncertainty: traditional privacy metrics do not provide a way to estimate uncertainty. A metrological approach instead requires the identification of uncertainty sources and repeatable procedures to quantify the reliability of the measurement. Fourth, the issue of traceability: standardised test datasets and attack models must be created to achieve traceability. Finally, universal applicability: each metric has been designed for a specific context (tabular data, query releases, information-theoretic channels), while a privacy metric must be adaptable to different scenarios while still remaining interpretable.

Traditional privacy metrics, therefore, cannot be directly used as measurands in a metrological sense. By defining observable quantities that reflect real privacy risk and combining them into a standardised metric, we can achieve a measurement-based, uncertainty-aware framework for evaluating privacy systems. This approach would bridge the gap between theoretical guarantees and practical, quantitative assessment. Attempting to stitch together metrics born in different eras and contexts, such as k-anonymity for relational databases, DP for streaming or query systems, and MI for communication channels, risks creating an inelegant compromise that is difficult to standardise. Instead, a composite metric of effective privacy, grounded in observable risk and accompanied by an uncertainty estimate, could provide a quantitative, comparable, and reproducible measure of privacy protection. Such an approach could facilitate the integration of privacy metrics into existing regulatory frameworks and standards, enabling the benchmarking of privacy-preserving systems according to a measurable and standardised criterion, evidence-based compliance assessment for data protection regulations such as GDPR or ISO/IEC standards, and consistent evaluation across heterogeneous systems and datasets, improving transparency and accountability. By bridging the gap between theoretical guarantees and empirically measured privacy, the combined metric could serve as a normative tool, guiding both technical design and regulatory verification of privacy-preserving technologies. Initial steps for an interdisciplinary effort to bring metrological thinking into privacy evaluation may include defining theoretical measurands aligned with privacy goals, creating protocols for uncertainty evaluation in privacy computations, and aligning privacy measurement methods with international metrology standards such as the GUM and the VIM.

VI. CONCLUSIONS AND FUTURE WORK

Quantifying privacy remains a complex challenge that requires a solid measurement foundation. This paper outlines the essential requirements for metrologically compliant privacy metrics and critically examines the limitations of current approaches. Future work will focus on empirical studies and case analyses, paving the way for proposals aimed at standardisation and integration with existing measurement science frameworks. In particular, observable quantities for each base metric, k-anonymity, ℓ-diversity, t-closeness, DP, and MI, will be investigated

to identify which observables most accurately capture effective privacy under realistic threat models. Various strategies for combining these observables into a single index will be explored, including weighted schemes, conservative minima, and multi-criteria optimisation. Subsequently, procedures to estimate the uncertainty of the composite metric will be developed, considering how dataset characteristics, mechanism randomness, and attacker models influence uncertainty estimates. By applying a metrology-based framework, privacy can be transformed into a comparable and reproducible quantity, enabling both practitioners and regulators to quantify, compare, and enforce privacy in a consistent and scientifically grounded manner.

REFERENCES

[1] JCGM 200:2012, *International Vocabulary of Metrology*, 2012. doi:10.59161/JCGM200-2012

[2] JCGM 100:2008(E), *Evaluation of Measurement Data – Guide to the Expression of Uncertainty*, 2008. doi:10.59161/JCGM100-2008E

[3] F.Thiel et al., "A Digital Quality Infrastructure for Europe: The European Metrology Cloud," *PTB-Mitteilungen*, vol.127, no.4, 2017, doi:10.7795/310.20170404

[4] NIST, "Legal Metrology Meets the Digital Age," 2025. https://www.nist.gov/news-events/news/2025/07/legal-metrology-meets-digital-age

[5] European Union, *Regulation (EU) 2016/679 – GDPR*, 2016.

[6] ISO/IEC 29100:2011, *Information Technology – Security Techniques – Privacy Framework*, ISO, 2011.

[7] ISO/IEC 27701:2019, *Extension to ISO/IEC 27001 and ISO/IEC 27002 for Privacy Information Management*, ISO, 2019.

[8] NIST, "NIST Privacy Framework: A Tool for Improving Privacy through Enterprise Risk Management", 2020.

[9] NIST, "An Introduction to Privacy Engineering and Risk Management in Federal Systems," *NISTIR 8062*, 2017.

[10] ITU-T, "Risk Management and Risk Assessment Techniques in Information Security Management," *ITU-T X.1149*, 2011.

[11] ITU-T, "Overview of Cybersecurity," *ITU-T X.1205*, 2008.

[12] I.Wagner, D.Eckhoff, "Technical Privacy Metrics: A Systematic Survey," *ACM Comput. Surv.*, vol.51, no.3, 2018, doi:10.1145/3168389

[13] T.Benschop, M.Welch, "Statistical Disclosure Control for Microdata: A Practice Guide", 2025. https://sdcpractice.readthedocs.io/en/latest/

[14] K.El Emam, F.K.Dankar, "Protecting Privacy Using k-Anonymity," *J. Am. Med. Inform. Assoc.*, vol.15, no.5, 2008, doi:10.1197/jamia.M2716

[15] R. Aufschläger et al., "Anonymization Procedures for Tabular Data," *Information*, vol.14, no.9, 2023, doi:10.3390/info14090487

[16] Utrecht University, *Data Privacy Handbook*, May 2025. https://utrechtuniversity.github.io/dataprivacyhandbook/disclaimer.html

[17] D.Roy, S.K.Jena, "Determining t in t-Closeness Using Multiple Sensitive Attributes," *Int. J. Comput. Appl.*, vol.70, 2013, doi:10.5120/12179-8291

[18] N.Li et al., "t-Closeness: Privacy Beyond k-Anonymity and l-Diversity," *ICDE*, 2007, pp.106–115, doi:10.1109/ICDE.2007.367856

[19] H.Liang, H.Yuan, "On the Complexity of t-Closeness Anonymization and Related Problems," *Data Structures and Algorithms*, 2013. https://arxiv.org/abs/1301.17511

[20] J.Near, D.Darais, "Differential Privacy: Future Work & Open Challenges," *NIST Cybersecurity Insights Blog*, 2022. https://www.nist.gov/blogs/cybersecurity-insights/differential-privacy-future-work-open-challenges

[21] J.Domingo-Ferrer, V.Torra, "A Critique of k-Anonymity," *ARES*, 2008, pp.990–993, doi:10.1109/ARES.2008.97

[22] K.El Emam et al., "Re-identification Attacks on Health Data," *PLoS ONE*, vol.6, no.12, 2011, doi:10.1371/journal.pone.0028071

[23] S. Li et al., "Reidentification Risk in Panel Data," *Inf. Syst. Res.*, vol.34, no.3, 2022, doi:10.1287/isre.2022.1169

[24] NIST, *Special Publication 800-226: Guidelines for Evaluating Differential Privacy Guarantees*, 2025. doi: 10.6028/NIST.SP.800-226

[25] M. I. Belghazi et al., "Mutual Information Neural Estimation," *ICML*, 2018, arXiv:1801.04062

[26] P. Zhao, L. Lai, "Analysis of KNN Information Estimators," *IEEE Trans. Inf. Theory*, vol.66, no.6, 2020, doi:10.1109/TIT.2019.2945041

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Digital Metrological Expert – design of a software for automated key comparison data analysis in a digital world

Daniel Hutzschenreuter[1], Clifford Brown[1], Wafa El Jaoua[1] Moritz Gafert[1], David Urban[2]

[1] *Physikalisch-Technische Bundesanstalt, Bundesallee 100, 38116 Braunschweig, Germany,*
daniel.hutzschenreuter@ptb.de
[2] *Physikalisch-Technische Bundesanstalt, Abbestraße 2-12, 10587 Berlin, Germany*

***Abstract* – Key and supplementary comparisons are a core component of the Mutual Recognition Arrangement. Digitalisation of data processing steps has the potential to reduce workloads of metrologists and to improve consistency of outcomes. The Digital Metrological Expert is an open-source software tool for automating the comparison data analysis and report creation. It is designed for use by metrology experts and for an integration into end-to-end digital workflows for machines. Automation is supported by applying FAIR principles for machine-actionable data and APIs that are based on the SI. The tool can also utilise digital standards from the quality infrastructure such as Digital Calibration Certificates and Smart Standards to fulfil its work.**

I. INTRODUCTION

Key and supplementary comparisons are a core component of the Mutual Recognition Arrangement (MRA) and almost 2000 have been carried out since 1999 [1]. All National Metrology Institutes (NMIs) participate in such comparisons and many have acted as comparison pilots. However, calculating a key comparison reference value (KCRV) and the degrees of equivalence (DoE) for participants can be difficult. While tools and guidelines exist for some straightforward comparisons, many real comparisons have complexities making them not easily applicable. Further challenges arise when data are inconsistent and decisions must be made about how to handle these inconsistencies. This represents a barrier to small and emerging National NMIs and Designated Institutes (DIs) in particular, who are aiming to pilot comparisons, but struggle with the mathematical practicalities of the data analysis.

Providing software for an automated key comparison (KC) analysis will lead to better and efficient data processing, enabling more NMIs and DIs to pilot comparisons and it will provide benefits in a wide range of scientific disciplines. Such improvements will impact the metrology community in several ways. First, faster comparison analysis will accelerate the delivery of this core component of the MRA, allowing NMIs and DIs to obtain approval for Calibration Measurement Capabilities (CMCs), and hence the ability to offer recognised measurement services, more readily. Second, improved understanding of comparison models will enable working groups of technical committees responsible for comparisons to make decisions more promptly about handling statistically inconsistent results.

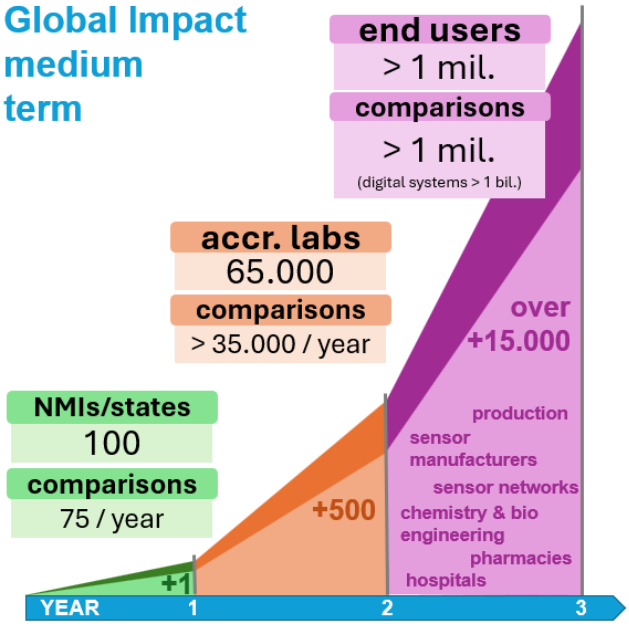


Fig. 1. Estimation of the medium-term impact of comparison data analysis software.

Improving the quality and availability of validated evaluation methods (digital standards) for comparison data analysis on the level of key- and supplementary comparisons also has an exponential impact on secondary (accredited) laboratories and in applications of end-users. Figure 1 outlines this impact. The comparison software will transfer the expertise of 100 metrology organisations to around 65.000 accredited laboratories and over one million end-user organisations operating in important

areas of the quality infrastructure in industry, health, and other fields. It will increase the capabilities of these stakeholders to conduct consistent comparison analysis resulting in reliable outcomes beyond their existing capabilities. Increasing the number of international comparisons by a single number a year will also lead to thousands of subsequent comparisons addressing multiple related and aligned measurands.

Moreover, the application of comparison methods is becoming increasingly relevant for wider digitalisation. Comparison of sensor values is used to establish traceability of sensor quality and in-situ calibration in large homogeneous networks. Establish reference data by an application of KCRV calculations is also a research topic in the field of testing Artificial-Intelligence (AI) methods.

The development of the Digital Metrological Expert (DME) is an internationally driven project which started in 2023 aiming to establish a reliable open-access software tool for comparison data analysis and lifting the value of digitalisation. A proof-of-principle use-case was implemented for a virtual comparison of mass measurements values by Rodiek et al. [2] and a real-world comparison of radiation temperature data from the CCT-K11 EURAMET loop was established by Müller et al in 2024 [3]. Both applications provided valuable insights into common structures and requirements for comparison data analysis tools. It led to the design of an initial version of a universal software architecture of the DME which is able to accommodate multiple kinds of comparisons comprising different metrological areas and measurands.

This paper will show details of the architecture design. Section II gives an overview on the operational principles of the DME in the context of tools and services that are developed by organisations belonging to the quality infrastructure. The generic framework and software-architecture for automated key-comparison evaluation is outlined in Section III. Conceptual models for representing data within the process of comparison evaluation are presented in Section IV. Our work will be concluded with a discussion of the current state of implementation in Section V and an outline of future work in Section VI.

II. TESTBED FOR INTEROPERABLE QUALITY INFRASTRUCTURE

Brown et al. [4] define the DME as an autonomous operating software with the ability of performing data and result processing as requested for standard work usually done by human metrology experts. The design and implementation of the DME are derived from the following core considerations for efficient digitalisation:

§1 DMEs are provided for standard work such as evaluation of intercomparison, calculation of model based measurands and quantities, and the verification of metrological data. They take over a larger amount of tedious human work.

§2 DMEs exchange machine-actionable information with tools and services from organisations belonging to the quality infrastructure through SI-based data [5] and services that follow the FAIR (findable, accessible, interoperable, reusable) data principles [6].

§3 DMEs can assess data and, depending on the intended application, can propose suitable ways of processing. This includes verification and transformation of data, an application of filters, outlier detection, propagation of measurement uncertainty and more.

§4 DMEs provide their results by machine-actionable reports that disclose metrological traceability of all outputs to the input data. Resolvable, unique, and persistent identifiers (PIDs) are interlinking all data and metadata for this purpose.

§5 DMEs that are operated by authoritative organisations ensuring the quality of services and maintenance are themselves digital standards in the quality infrastructure.

Metrology, accreditation, standardisation, and other players in the quality infrastructure have started to establish digital formats for their standards and products. The DME will utilise these developments to reach higher levels of automation and to establish software that is future-proof in the sense of end-to-end digital workflow. In this respect, the DME is one of the first applications in metrology interoperating machine-actionable data, services and tools from multiple providers. Thus, our software is not only a tool for comparison data analysis but also a testbed for the wider digital transformation of the quality infrastructure.

The set-up of the DME in a digital workflow is shown in Figure 2. Reports containing the measurement results of participants are provided through an online data storage accessible for machines. The DME will read the participant data and provide a machine-readable report of analysis results through Application Programming Interfaces (APIs). Digital standards (data, services, and tools) for quality infrastructure are used to define data exchange protocols for the APIs and to facilitate data analysis. More background information is disclosed in subsections II.A to II.D.

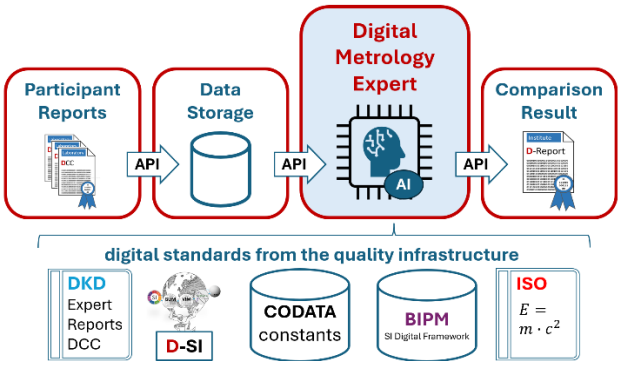


Fig. 2. Basic set-up of the DME in a digital workflow.

A. SI Digital Framework and D-SI

The SI Digital Framework (SIDF) denotes a collection of digital references and services for the definition of the International System of Units (SI) which is provided by the International Bureau of Weights and Measures (BIPM) [7, 8]. The DME will use the SIDF as the foundational layer to establish interoperation of units of measurement and related core knowledge from the SI across all data coming from the quality infrastructure. An effective application of the SIDF in software typically requires the use of data exchange and representation formats that are compatible with the SI and its underlying concepts for the representation of quantity values comprising knowledge on the kind of quantity, a number, and the unit of measurement. In the case of the DME, the D-SI metadata model for core-metrological data was adopted [9] to implement all relevant quantity values that are related to comparison data. It is closely interlinked with the SIDF and practical applications such as calibration. It adheres to FAIR data principles, considers semantic interoperability needs, also allowing the integration of non-SI units, and provides common models for measurement uncertainty data based on the Guide to the Expression of Uncertainty in Measurement [10]. Furthermore, a web service is provided with the D-SI that can be used within software to directly interoperate units with the BIPM services [11].

B. Smart Standards

Digitalisation of normative documents provides Smart Standards. These are machine-actionable representations of knowledge and requirements for products, services, and procedures, that have been agreed-upon by experts. The DME project assumes interoperation with the following five pillars of knowledge that are delivered by Smart Standards: (i) machine-actionable numbers which are interoperable through the SI, (ii) tables which define common ways to represent structured data, (iii) formulas defining machine-actionable mathematical models and equations, (iv) textual information that can be understood by humans and machines, and (v) knowledge in the form of images such as plots of outcomes of comparisons.

The International Organisation of Standardization (ISO) and the International Electronical Commission (IEC) are establishing fundamental concepts and formats for Smart Standards [12]. The DME considers the automated application of guidelines for comparison analysis and methods from the GUM which will be fostered through the availability of Smart Standards that are harmonised with the joint work of ISO and IEC.

C. DCC and DKD Expert reports

Digital Calibration Certificates (DCCs) serve as digital formats for the exchange of machine-readable measurement data and results of calibrations. They typically implement a broad range of measurements of services that are subject to quality control through interlaboratory comparisons. Against this background, the DME is using DCCs to represent all data reported by participants of comparisons and the report containing the results of the data analysis. In particular, the internationally developed and widely used XML implementation of the DCC is used [13]. This format includes the D-SI and follows FAIR principles. The XML DCC is supported by a broad user community worldwide which is advancing the harmonisation of community standards for interpretable and interoperable measurement data. It is also establishing useful tools for the application of DCCs such as the GEMIMEG tool for creating and reading of XML for non-IT experts [14].

Good practice guidelines for using the XML DCC are issued in the form of pre-normative Expert Reports by the German Calibration Service (DKD). DKD represents over 400 calibration laboratories who agree upon these Expert Reports in twelve metrological, area-specific working groups and one cross-domain working group. These Expert reports, e.g., guidance for DCCs for weights [15] and gauge blocks [16], are the basis for the design of domain-specific measurement data in the DME. In the future, measurement devices (respectively software) used in laboratories will be able to directly produce such DCCs and therefore increase the capabilities of digital workflows for comparison. Furthermore, the DKD is also establishing a database of various terminology that is used in calibration [17]. The terms are called refTypes and are organised within a FAIR and machine-actionable thesaurus [18]. It has the capability to introduce semantic references across terminology from different providers in the quality infrastructure which will facilitate interoperability beyond the scope of the SIDF.

D. CODATA

CODATA provides and regularly updates internationally agreed values of fundamental physical constants. Such constants are part of equations for computationally derived quantity values at comparison. Besides constants from CODATA, various community agreed reference values and similar pre-set data can be required to perform comparison data analysis.

III. DME SOFTWARE ARCHITECTURE

Computer Science recommends the design of a software architecture (SA) when implementing complex software systems [19]. SAs are high-level definitions of the components of a software disclosing the responsibilities of each of the components, the interaction between components and interactions with external resources. They are also independent from the programming language.

A proper SA is indispensable for universal comparison software such as the DME due to the comprehensiveness of applications in general and the technical demands arising from the digital quality infrastructure. The SA increases the efficiency of feature implementation, usage,

and maintenance of the DME tool as a whole.

Figure 3 shows the major components of the DMEs SA. It assumes that the software is operated (deployed) as a web-service. This solution offers a high level of flexibility as the operation of the software is scalable from running it on office computers, to company servers, or even in cloud-based environments. The SA structures the software into various layers which are separated from each other by clearly defined interfaces.

API-layer: This layer defines the API of the DME. Humans and machines run the tool exclusively through the APIs. The interfaces are based on the RESTful HTTPs requests [20]. Documentation and examples are provided through an (FAIR) open API documentation based on Swagger UI [21].

Service-layer: All data analysis is performed within the software modules of the Service-layer. A main service receives all requests coming from the API-layer and evaluates which domain specific services need to be operated. Domain specific services are implementations of KC data analysis for specific metrological areas and quantities, such as the before mentioned virtual mass comparison and comparison of radiation temperature measurements. Data processing of domains is implemented based on a library of modules for the calculation of KCRVs, DoEs, data and consistency checks, filters, rounding and other relevant methods. These modules have common interfaces which simplify the creation of new modules and automation of the selection of data analysis methods.

Model-layer: A comprehensive amount of data and metadata is acquired and created by the software and stored within a data model for comparison. Section IV outlines this model. The modules of the Service-layer can read from and write to the model.

In addition, the SA defines a **Reader Service** which provides a common interface for implementing software that can read data such as DCCs containing inputs for the comparison data analysis. The Reader Services will write data to the data model. In the future, different readers can be implemented to access data from different versions of DCCs and other machine-readable data exchange formats for comparisons. Finally, the **Writer Service** is an interface to extract data from the comparison data model into a machine-readable output report format. The first Writer Service implementation of the DME creates DCC-based reports. It is also considered to allow different outputs, e.g., in formats that can directly be integrated into BIPM's Key Comparison Data Base [22].

Non-IT experts can operate the DME through a simple **web user interface (UI)** [23]. The UI allows to fill in information on participants and contributions to the comparison into a simple form. The form converts the inputs into the operations that run the API. Outputs from the API, such as the DCC report, are provided back to the users as file downloads. Any web-browser and a connection to the internet is sufficient to run the UI. Management platforms for comparisons such as the EURAMET Comparison Toolbox [24] could also add in future user-interfaces to operate the DME.

The DME has been implemented in the JAVA programming language using the Spring Boot framework for the web-services backend [25]. Especially, the separation into the API-, Service-, and Model-layer is a common design pattern for web-services recommended within the Spring Boot framework [26]. The DME front-end is implemented using Angular [27]. Links to the source code are provided at the UI webpage [23].

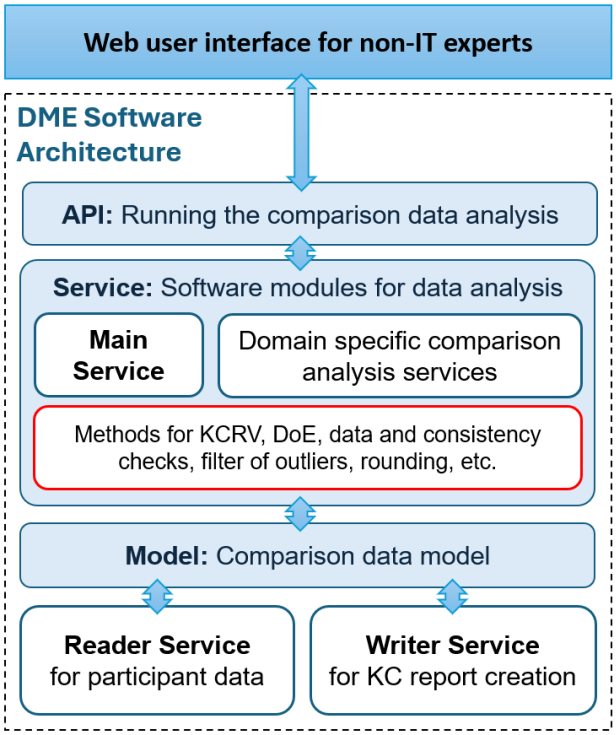


Fig. 3. Outline of the major components of the universal DME software architecture for comparison data analysis web-service.

IV. COMPARISON DATA MODEL

Comparisons typically comprise significant amounts of information. This information is, for example, protocols, management data, schedules, participant information, contributions, stability studies of transfer standards, actual measurement data and metadata reported by participants, data analysis methods and results, plots of data, descriptions and more. Developing data models for all the above-mentioned properties is certainly a longer-term effort. Hence, the DME project prioritises the development of early-stage data models for the core functionality of computing KCRVs and DoEs which is shown in Figure 4. This model provides data on the domain specific service, contributions, and the entities under comparison.

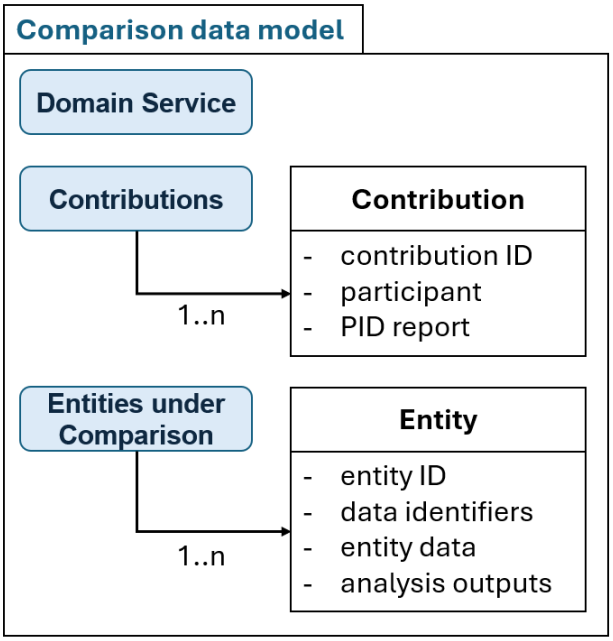


Fig. 4. Outline of the data model within the DME.

The **Domain Service** contains information on the application, such as a virtual comparison of mass or a radiation temperature comparison, which is used to process the data.

Contributions are a list of inputs subject to the comparison. Each contribution has a unique contribution identifier (ID), a participant laboratory providing the contribution and a PID pointing to a machine-readable DCC which contains the reported measurement data. Depending on the structure of the comparison, e.g., round robin, petal structure or bilateral, participants can hand in one or more contributions which are going into the data processing.

Entities under comparison elements contain the actual measurement data (entity data) and analysis outputs. Each comparison can have multiple entities under comparison depending on the used transfer standards and measurement plans. In the case of a comparison of a set of weights for example, each of the weights would provide one entity. In another case such as the comparison of temperature measurements, each nominal set point of measurement, would provide one entity. All entities have a unique identifier.

In practical comparisons, one or more measurands are reported for each entity. These are typically measurements of properties of the transfer standards and measurements of influence conditions such as environmental conditions within laboratories. The DME provides various data identifiers for each measurement. These identifiers comprise unique IDs of the measurements, unique IDs of transfer standards, specific properties of measurands (DKD refTypes), and other metadata, that could be used to clearly describe the measurements and to allow an automated access to this data in DCCs. The entity data element lists the measurement values and associated uncertainties for all measurements and all contributions. Entity data are created by reading data from the contributing DCCs or can be calculated as part of the data analysis process.

Finally, the entities under comparison provide a list with analysis outputs. These are the calculated KCRVs and DoEs, but also equivalence values between contributions and the KCRV, bilateral equivalence values and plots of data. Having a list of outputs allows to represent various data analysis steps. Types of outputs are for example results of hypothesis testing, such as test of the homogeneity of input data, interim KCRV and DoE values before removing outliers, or simply results of different KCRV-calculation-approaches to compare their performance.

V. DISCUSSION

A. Capabilities of generic structure

The DME implementation established a robust level of automation by the initial generic structures of the software and data model. Especially the strict linear structuring of comparisons by entities and contributions is capable to automate the data flow from input DCCs into core data analysis procedures and to create common output reports.

A challenge is the provision of identifiers that are suitable to properly identify and read the right data out of inputs and to also properly identify results in output reports. Our concept for identification will probably require some semantic improvements in the future to allow the handling of a wider number of different measurands. The use of Large Language Models (LLM) is also considered as one of the future tools to overcome shortcuts of domain specific input data. A LLM could be trained to use the semantics of identifiers of measurands to perform “metrology informed” queries on arbitrary DCCs as inputs.

Moreover, a generic concept for data pre-processing is not yet implemented. Future efforts will for example add consistency checks of data and units of measurement reported by participants as well as data normalisation before calculations. Methods for statistical testing of data properties such as testing the homogeneity of data and testing for normal distribution of data are also required. These tests are a prerequisite for a selection of suitable KCRV- and DoE-methods yielding better consistency of results.

Finally, additional data needs to be included such as administrative data on the supervising committees for comparisons, pilots and participants. In an early development phase, this data would be provided in form of human-readable names and addresses. In following development phases, additional metadata suitable for machines could be added such as unique PIDs for each

entity that are registers by authoritative organisations. BIPM for example is using Research Organization Registry (ROR) identifiers to digitally link NMIs and DIs within machine-readable database entries for CMC within the SI Digital Framework [7]. As today in the analogue world, pilots who are responsible for the comparison data analysis are also ensuring the integrity of the data related to all participating organizations and entities in the process.

B. Capabilities of API

The current API is very simple in respect to the data that is exchanges. The inputs are the list of contributions and the selection of the domain specific KC service. Its output is the comparison report. This is sufficient to operate various automated comparisons as long as they are based on fixed domain specific service implementations. A future advancement of the API considers transmitting additional information from the protocol of the comparison. This protocol comprises for example links to stability tests of transfer standards and instructions on using specific methods for comparison data analysis. The later, would also enable a more flexible use of the software by non-IT experts. Pilots will be able to simply select methods by pre-defined tags and the DME will automatically use the software modules implementing these methods.

A future API may need nothing more but a machine-readable comparison protocol to operate the automated comparison. Developing such protocols will have significant impact not only on progressing the automation of the data analysis, but also on the capabilities of participants. It could help to automate parts of the processes at the site of laboratories. Clear semantic definitions of measurands and underlining terms are provided. Thus, it supports laboratories to better understand required measurements, which will lead to more consistent data.

VI. CONCLUSION AND OUTLOOK

The adoption of digital tools and services supporting key- and supplementary comparisons is promising to save significant amounts of time for data analysis and reporting. It also allows to share expertise on data evaluation with a wider number of users and helps to improve the quality and consistency of outcomes of comparisons.

The DME project has established the basis for a generic software tool to handle comparisons from different metrological areas. The tool is implemented following a modern software architecture design and it is also fostering the integration of digital products and services from organisations of the wider quality infrastructure such as the SI Digital Framework, DCCs, and Smart Standards. The software and its source code are also available open source to support the metrological community in digitalisation.

Our project will continue to evaluate the practical needs of pilots and participants of key- and supplementary comparisons to improve the useability of the tool. The first two applications will be completed in summer 2025. In addition, research and development will be continued regarding machine-readable protocols for comparisons and integrating additional data analysis methods.

VII. ACKNOWLEDGEMENS

We like to thank all contributors from the use-cases for their valuable inputs, discussions, and thoughts on the DME. In Particular, we like to thank Shanna Schönhals, Beatrice Rodiek, Stuard Davidson, Ingmar Müller, Claudia Swart, David Newton, and Frank Härtig for their dedicated support. The work is funded by Physikalisch-Technische Bundesanstalt.

REFERENCES

[1] CIPM, ‘Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes, Technical Supplement revised in October 2003’. BIPM, Oct. 14, 1999. Accessed: Apr. 13, 2025. Sdf

[2] B. Rodiek et al., ‘Project on a fully automated evaluation of a virtual comparison of mass using the Digital Calibration Certificate (DCC) schema’, Measurement: Sensors, p. 101361, 2024, doi: 10.1016/j.measen.2024.101361.

[3] I. Müller et al., ‘CCT K11 blackbody temperature from 34.5 °C to 41.5 °C – Reporting and evaluation in the EURAMET loop using digital calibration certificates’, Measurement: Sensors, p. 101626, 2025, doi: 10.1016/j.measen.2024.101626.

[4] C. Brown, W. El Jaoua, T. Hoffmann, D. Hutzschenreuter, B. Rodiek, and S. Schönhals, ‘A digital expert for a digital key comparison report processor’, Measurement: Sensors, p. 101463, 2024, doi: 10.1016/j.measen.2024.101463.

[5] BIPM, ‘The International System of Units (SI), 9th edition’. 2019. [Online]. Available: www.bipm.org /documents/20126/41483022/SI-Brochure-9-EN.pdf

[6] M. D. Wilkinson et al., ‘The FAIR Guiding Principles for scientific data management and stewardship’, Sci Data, vol. 3, p. 160018, 2016, doi: 10.1038/sdata.2016.18.

[7] BIPM, SI Digital Framework, Digital references for FAIR measurement data. (2024). [Online]. Available: https://si-digital-framework.org/

[8] CIPM, ‘CIPM Vision - Transforming the International System of Units for a Digital World’. BIPM, 2023. [Online]. Available: https://tinyurl.com /cipmVision

[9] D. Hutzschenreuter et al., ‘SmartCom Digital System of Units (D-SI) Guide for the use of the metadata-format used in metrology for the easy-to-use, safe, harmonised and unambiguous digital transfer of metrological data - Second Edition’. Zenodo, Juli 01,

2020. doi: 10.5281/zenodo.3816686.
[10] JCGM, ‘JCGM 100:2008(E), Evaluation of measurement data - Guide to the expression of uncertainty in measurement (2008)’, 2008. [Online]. doi: 10.59161/JCGM100-2008E.
[11] D. N. Coppa, M. Gruber, W. El Jaoua, and M. Gafert, D-SI to SIRP Converter. (2024). Python. [Online]. Available: https://d-si.ptb.de/#/dsi-converter
[12] ISO, IEC, ‘IEC/ISO SMART - Unleashing the digital power of international standards’. [Online]. Available: https://www.iso.org/smart
[13] S. G. Hackel et al., ‘The Digital Calibration Certificate (DCC) for an End-to-End Digital Quality Infrastructure for Industry 4.0’, Sci., vol. 5(1), p. 11, 2023, doi: 10.3390/sci5010011.
[14] PTB, GEMIMEG Tool powered by OP Layer. (2025). [Online]. Available: https://gemimeg-tool.ptb.de/v2/dcc/create#/
[15] G. Foyer, J. Haller, C. Müller-Schöll, S. Osang, A. Scheibner, and M. Häfner, ‘Instructions on how to use the DCC schema to create a digital calibration certificate for weights (DKD-E-7-2)’. DKD, 2023, doi: 10.7795/550.20220419B.
[16] G. Bartl, D. Heißelmann, K. Janzen, R. Liske, and C. Sander, ‘Instructions on how to use the DCC schema to create a digital calibration certificate for gauge blocks (DKD-E 4-3)’. DKD, 2025, doi: 10.7795/550.20250428.
[17] DKD, ‘Metrology refType Database’. 2023. [Online]. Available: https://digilab.ptb.de/dkd/refType/vocab/
[18] D. Ferreyra, TemaTres. (2025). [Online]. Available: https://vocabularyserver.com/web/
[19] Gartner, ‘Gartner Information Technology Glossary’. [Online]. Available: https://www.gartner.com/en/information-technology/glossary/architecture
[20] Wikipedia, ‘REST (Representational State Transfer)’. [Online]. Available: https://en.wikipedia.org/wiki/REST
[21] SmartBear, Swagger UI. (2025). [Online]. Available: https://swagger.io/tools/swagger-ui/
[22] BIPM, ‘Key Comparison Data Base (KCDB)’. [Online]. Available: https://www.bipm.org/kcdb/
[23] W. El Jaoua, T. Hoffmann, D. Urban, M. Gafert, D. Hutzschenreuter, and C. Brown, Digital Metrological Expert Web Service. (2024). JAVA. [Online]. Available: https://d-si.ptb.de/#/d-comparison
[24] EURAMET e.V., EURAMET Comparison Toolbox. (2025). [Online]. Available: https://www.euramet.org/comparison-toolbox
[25] Broadcom Inc., Spring Boot. [Online]. Available: https://spring.io/projects/spring-boot
[26] D. Sandaruwan, ‘Understanding Spring Boot Architecture Flow’, Medium. [Online]. Available: https://medium.com/@dulanjayasandaruwan1998/understanding-spring-boot-architecture-flow-615d209b95f9
[27] Angular community, Angular. [Online]. Available: https://angular.dev/

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Digital transformation applications in mechanical quantities – hardness measurements

Cihan Kuzu*[1], Alessandro Germak[2], Febo Menelao[3], Moritz Loewit[4], Miha Hiti[5], Andrea Prato[6], Tatiana Apostol[7]

[1] *TÜBİTAK UME, Kocaeli, Türkiye, cihan.kuzu@tubitak.gov.tr*
[2] *INRIM, Torino, Italy, a.germak@inrim.it*
[3] *PTB, Braunschweig, Germany, febo.menelao@ptb.de*
[4] *BEV-PTP, Wien, Austria, moritz.loewit@bev.gv.at*
[5] *ZAG, Ljubljana, Slovenia, miha.hiti@zag.si*
[6] *INRIM, Torino, Italy, a.prato@inrim.it*
[7] *INM, Chişinău, Moldova, tatiana.apostol@inm.gov.md*
** Corresponding author*

***Abstract* – One of the most important and widely used testing method for extracting mechanical properties of material is the hardness test. It is mainly based on realizing a deformation on the material, measuring the geometric dimensions of the deformation and from that calculate the hardness value. Measurements are performed with imaging instruments like optical microscopes, mostly operated manually. However, new developments aim to determine the border of indentation, measure its diameter and diagonal length, save and mark the locations of the measured indents on the surface of the hardness reference block by making use of a fully automated indentation measurement system (IMS). This digitalization approach shifts hardness measurements from manual processes to using pixel-wise image processing and fully automated IMS, leading to increased precision, repeatability and speed and leading the way for further improvements by digital transformation.**

I. INTRODUCTION

Understanding of the critical mechanical properties of materials used in manufactured products, building, and infrastructure projects is essential for safety assurance, production cost control, and the creation of healthier, easier-to-use products. Such understanding relies on accurate, reliable, and consistent reference standards for material properties at the NMI level and the ability to trace measurements from the top-level standards to basic workshop tests through an unbroken chain of traceability.

Hardness testing is one of the most important material testing methods, with Brinell, Vickers and Knoop tests are among the most widely used testing methodologies. Hardness is quantified by measuring the size of an indentation made by a specified, controlled force with an indenter of known geometry. However, the accuracy of such measurements may vary significantly depending on influencing factors such as image magnification, handling by the user, the measuring instrument, and the software used. This leads to differences among NMIs and between NMIs and secondary or testing laboratories.

This difference is largely due to the absence of standard methods and clear specifications for the various indentation measuring systems in use by the many different laboratories. Also, each manufacturer is using his own specifications, settings and software. The need for a solution was highlighted by the 2005 CIPM Brinell pilot study (CCM.H-P3)[1], where large differences between participants' results were seen due to variations in imaging system parameters, e.g., numerical aperture (NA). Such variations become even more pronounced further down the measurement chain, where less advanced equipment is used without stringent control of influencing parameters.

To transcend this, automation has been a promising solution. Not only does it decrease human error and enhance reliability but also improve efficiency, repeatability and allows for the unification of systems throughout the traceability chain. Advances in digital technologies, including computer vision, are enabling new prospects for completely automated hardness measurement systems.

As part of the EPM project 22RPT01 TracInd BVK-H – Traceability for indentation measurements in Brinell-Vickers-Knoop hardness (TracInd-BVK-H) the underlying issues are addressed by both normative measures as well as digitalization and unification of the processes involved in hardness measurements. One of the aims of this project part is to determine the border of indentation, perform its diameter and diagonal length measurement, save the location of the measured indents on the surface of the

hardness reference block to be shared with the users for further measurements. This means currently manually operated measurement steps are replaced by computer vision, algorithm-based evaluation. This includes the fully automatic indentation measurements, the reporting the measurements results together with the indents location with coordinates in a fixed reference coordinate system using pixel-based image recognition resulting faster, more reliable measurements with reduced uncertainty.

Brinell, Vickers and Knoop indentation measurement methods are realized in different ways at the national metrology institutes (NMIs) as well as calibration and testing laboratories. In some applications there are manually measurement application while in some other there are some semi-automatic indentation measurements. In manual measurements the operator determines the border or the corner of indents based on personal preference and perception while in some applications the custom-made software is used for such application.

The possibility of realizing fully automatic indentation measurements, not only one single indentation but *n* indentations including both direction diameter/diagonal lengths, a kind of calibration of hardness reference block as a whole is investigated. For this, certain goals are set for the design and development of such a system.

The basis for the system developed during the TracInd BVK-H project is the semi-automatic IMS that was realized and used in the TÜBİTAK UME Hardness Laboratory[2].

A fully automatic IMS first and foremost require fully automated length measurements in two dimensions, so the indent can be measured without manipulation or any intervention by the operator. For this, the IMS must be capable to record its exact location above the indentation and the travel needed to reach the opposing side.

Next, the IMS should be capable to measure more than one indentation at once, again without operator intervention. For this, the IMS needs to know the exact location of the indents where it should look for them on the hardness reference block. For this, the IMS will be used to mark the locations in the block before their realization in the hardness standard machine. The coordinates of the location of indents will be saved and used in a later time to make the analyses of the indent. With these requirements a fully automated IMS can be realized.

II. CURRENT INDENTATION MEASUREMENT SYSTEMS AND THEIR LIMITATIONS

As mentioned, IMSs at the NMI level are either manual or semi-automatic. In both cases operator input is required to finish the measurements.

While in manual operation the full measurement is done by the operator, in the semi-automatic case automatization enables measurement in one direction of the Brinell, Vickers and Knoop indents, followed by a manual rotation of the indentation by 90º to make the second diameter or diagonal length measurement, necessitating the precise alignment of the diameter or the diagonal length to be measured in the same direction as the first measured one, which is a source of error.

There are few trials to start the measurement of one indentation in two directions and the possibility to measure more than one indentation by specifying the location of *n* indents manually by the operator after their realization, however no such system is currently in operation. Therefore, there are no practical information on how to implement full automation into an IMS.

Automation of measurements is seen as the means to overcome the issues plaguing current systems while also resulting in labour and cost reductions. Digitalization offers new possibilities to further and improve the measurement systems by automatization and computer vision.

For future developments it is extremely important to have an unanimously agreed upon set of parameters to be used by the IMSs in regards to contrast, optical resolution, aperture, luminance, edge definitions and more. In a first step, a fully automated optical measurement system was developed and implemented, greatly improving performance of the hardness measurement while removing operator error and increasing reliability and availability of the system.

The most significant problem during hardness testing, more precisely the measurement of the indentation, is the decision on where the border of the indentation begins.

Due to the test procedure, indentations in hardness tests are not sharp-edged. Instead, indentations merge into the undeformed surface slowly. Pile-up behaviour is observed when material is occasionally pushed out from the indentation, and at times material around the indent is pulled down with it (sink-in behaviour).

Both make the transition from the undeformed surface to the indentation even more complicated to define [3], [4].

Conventional hardness testing methods use optical systems, such as light microscopes, to determine indentation size.

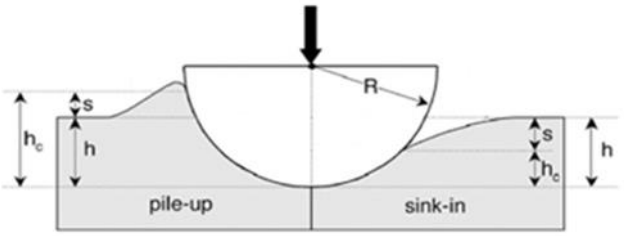


Figure 1: cross section of an indentation profile showing pile-up and sink-in behaviour [5]

The investigation of this behaviour is an important part of the project and its goal of unifying hardness measurement in the whole traceability chain, from NMI level down to the material testing laboratories.

76 indentations were produced and measured and characterized utilizing both conventional optical methods as well as state-of-the art methods like scanning tactile and laser confocal measuring systems to measure the 3D surface areas of indentations.

A 2D profile passing through the largest diameter of the indent is shown in Figure 2, along with the slope (in black) and the rate of change of slope (in red). The two valleys in the curve representing the rate of change suggested in [6], were used to determine the diameter of the indent. This value is treated as the reference value.

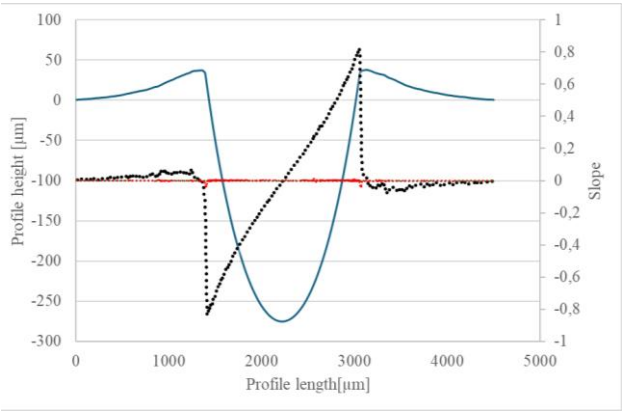


Figure 2: Cross section of the 75,26 HBW 2,5/187,5 indentation profile using the tactile measurement data for the calculation of slope angle and rate of change of the slope angle

Using these results of the investigation of the reference indentations, an analysis of the different measurement methods like scanning tactile and confocal laser versus standard optical measuring systems is conducted. This is highly relevant for the linking of higher-level laboratories and NMIs to the lower level testing laboratories which are almost always using standard optical measurement systems.

In the future with advances in image processing, AI and also a reduction in cost, many of the optical systems will change or be adopted to at least semi-automatic IMS. With this, the use of image processing and edge detection algorithms will increase significantly.

Since the definition of the indention borders will be based on the three-dimensional structure of the indent as shown in Figure 2, a close correlation of measurement data of both the advanced 3D measurement systems and the conventional 2D optical measurement systems is of utmost importance. Otherwise (and partly as still is common practice today), the data will have significant discrepancies and the uncertainty of measurement will increase unproportionable lower down the chain of traceability. This can partly be mitigated by the use of image processing and powerful edge detection algorithms.

Also, the use of advanced image processing systems on low-tech optical systems has a lot of potential for fast implementation and therefore increased speed of hardness measurement improvements. Since it is mostly software, many of the current manual systems can at least be retrofitted to semi-automatic IMS.

III. DEVELOPMENT OF A FULLY AUTOMATED IMS WITH AN INDENTATION MANAGEMENT SYSTEM AND MARKING TOOL

A complete Brinell, Vickers and Knoop scale capable IMS comprises two main components; hardware used for visualisation on the indentation, and a software that is used for realisation of the measurement itself.

As a starting point of development of the fully automatic IMS the currently available automatic indentation measurement systems for Brinell, Vickers and Knoop indentation measurements (either primary reference systems or calibration machines) were reviewed. Different types of design parameters such as electro-mechanical structure and custom-made software, different measurement algorithms and systems including auto-focus, auto-calibration of the system by making use of certified line scales and laser interferometer optic systems were investigated.

Although current systems allow for the (semi-)automated measurements of indentations, they are not connected to the hardness standard machines and require the operator to tell them where to look for the indentation to measure.

The design developed in this project aims to automate as possible in order to completely remove operator influence in both areas of measurement as best possible.

In the scope of the TracInd BVK-H project research [5] this is partially fixed with the introduction of an automatic marking system. The IMS will mark indentation locations on the hardness reference block beforehand, and save its locations. After producing the indentations, the reference block is placed back into the IMS for measurement, with the IMS already knowing where the indentations are located. While the process of indenting the blocks in the hardness standard machines still require an operator, the measurement of the indentations afterwards is speed up significantly. This system is only a first step towards further developments in regards to a fully digitalized measurement process.

As part of the automation, the system was equipped with the capability of marking the locations of indentations that are selected by the operator from the software of the IMS prior to realisation of indentations and later fully automatically measure the size of the realised indentations regardless its number, creating a real automatic indentation measurement capability. For this, revisions were made on

a four axis, motorized stage IMS present at the TÜBİTAK Hardness Laboratory for both mechanical components as well as software.

While hardness testing is considered non-destructive, it still cannot be replicated exactly in the same point as they plastically deform the measuring point. Also, indentations will push material away and therefore influence the immediate material around the indentation, effectively reducing the surface area available for testing. It is therefore necessary to use specially prepared hardness blocks in which the homogeneity of the surface hardness is sufficiently guaranteed so as to be able to highlight significant variations in hardness to be attributed to the variation of the test parameters and not to the variation in hardness of the material.

To deal with this, a mechanical part used to make a circular mark on the surface of the block was designed and adopted to the present system. With this part's help, the operator will place the hardness reference block on the plate of the IMS, and the region where the indentation measurements will be made will be adjusted by the operator and recorded by software developed in the project.

Since the indentation size and therefore the hardness measured is highly dependent on the surface condition of the block there are only limited options in order to mark the blocks. These options exclude any invasive markings like engraving or scratching since it will alter the surface itself, as well as laser engraving or etching since this will influence to hardness of the blocks surface layer. Therefore, non-invasive options needed to be considered. For the ease of use the marking with a waterproof pen was selected.

Different hardness blocks will have different indentation sizes, so the marking tool is designed to be used with different markers in different brush thicknesses in order to be able to mark different hardness reference blocks without abstracting too much area.

The IMS system marks the indentation location automatically (with the help of the four motorised stages). All directional movements of the table are precisely calibrated and recorded by laser interferometers. After the marking process is finished, the indentations will be realised in the traditional way with the hardness standard machines and the block will be placed as it was for marking the indents' places. The IMS performs measurements of the indentation within those circular marks with the help of image processing and edge detection in accordance with the definition of indentation.

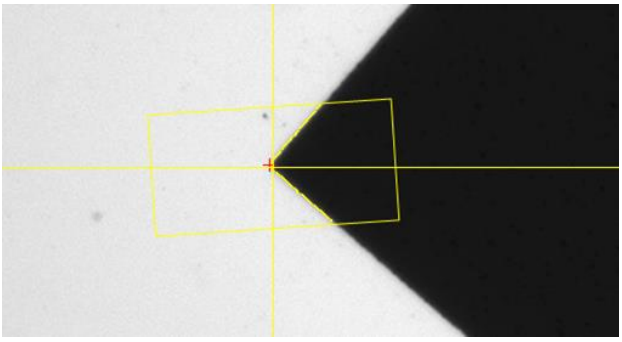


Figure 3: Determination of indentation corner with image processing used in the TÜBİTAK UME IMS

The revised software of the IMS currently available at TÜBİTAK UME Hardness Laboratory was improved with ability of autofocus, edge detection by image processing for circular shaped Brinell indents, determination of the corner of Vickers square and Knoop equilateral-shaped indentations, automatic measurement of one indentation for both diagonal and diameter measurement and extension of this measurement to the entire surface for a serial of *n* number of Brinell, Vickers and Knoop indentations whose coordinates were already recorded by the software during marking for indentations suitable for automated measurements. All these activities were implemented into the IMS available with an existing motorized stage with four-axis (X, Y, Z and rotation).

The improved edge detection algorithm is based on the detection of the gradient between bright and dark sections, where the bright area is the undeformed surface of the hardness reference block reflecting the lighting back up to the lens and photosensor, while the dark area is the deformed surface of the indent, where the light is reflected away from the photosensor.

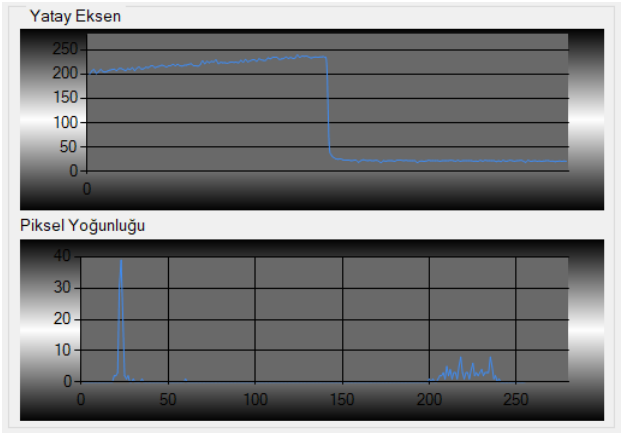


Figure 4: Graphical depiction of the edge detection algorithm: top: the change in brightness along the horizontal axis, bottom: the maximum change in brightness between pixels during the transition

The basic function of the algorithm can be seen in Figure 4. The top graph shows the change in brightness in the horizontal axis in the measurement area (yellow zone shown in Figure 3) with the value 255 being the brightest and 0 being the darkest spot. Where the brightness is dropping significantly, the transition zone of the undeformed surface into the indentation is assumed. Next, the algorithm scans the pixels in the drop zone for the maximum gradient between pixels, in order to determine the maximum rate of change of slope, as suggested in [6] and shown in Figure 2. This allows for the best possible approximation of the real indentation borders without actually using 3D measurement data. With this approach the uncertainty can be reduced to the size/length of the pixels analysed by the algorithm. The pixel-wise uncertainty is mostly given by the resolution of the sensor used (provided there is good enough lighting to saturate the pixels of the sensor used), as well as the magnification of the microscope itself. Insufficient magnification will lead to wider pixels and therefore reduced sharpness in the gradient of brightness, resulting in increased uncertainty of measurement. Also, if there is insufficient lighting, neighbouring pixels may "bleed" into each other, again reducing the sharpness of the image being processed. This highlight yet another problem with the use of automated edge detection algorithms: the algorithm can only work with the picture provided, and will therefore only be as good as the quality of the underlaying image.

This means, the settings selected for the microscope itself in regards to contrast, optical resolution, aperture, luminance, edge definitions and more are very important parts in all measurements. These have a direct influence on the measured length of the indentation and are currently not regulated in any of the relevant international standards and documents such as ISO 6506 [7], ISO 6507 [8], ISO 4545 series [9], etc. Each laboratory uses slightly different settings for their microscopes, so even if all systems would be operating with an algorithm on the same functional basis as described above, there would still be variations in measured hardness due to the incoherent settings on the IMS, especially the microscope part.

Therefore, the TracInd BVK-H project is studding these influences and determining their relations to the hardness value calculated for different surfaces, hardness scales and IMS used. Based on the results a set of unified parameters will be issued in order to be used with the IMS, providing the best possible results.

The fully upgraded system was tested from repeatability and reproducibility points of view and verified by making use of certified line scales. Examples of the function of the new, updated IMS can be seen in Figure 3. After full completion of the system verification and performance tests will be made with other partners' measurement systems and it will be verified from consistency and accuracy points of view. Measurements performed with different systems by the consortium members will be compared with automated measurements of the same indentations made at TÜBİTAK UME and a correlation between the indentation specifications and the accuracy of the automatic measurements and its correlation to the 3D data will be determined. The purpose is to identify the characteristics of indentations that are suitable for automated measurements.

For the possibility of not all indentations being suitable for automatic measurement due to contrast limits, optical resolution, illumination conditions, and indentation quality, the upgraded system will be used to study the effect of these parameters on recognition capabilities. This includes studying the interplay between optical system parameters, material properties, and software processing for autofocus and edge detection. The findings of those tests listed above will contribute to guidelines for manufacturers of future IMS in order to improve and unify measurements by the means of automation even further.

IV. PROJECTED OUTCOMES AND IMPACT ON FUTURE HARDNESS MEASUREMENTS

The implementation of the new IMS is a significant improvement on the old systems in multiple areas, as shown in the tests already performed.

The most obvious improvement is the elimination of user error by fully automating the measurement system. This allows for a lot faster, more reliable and fully reproducible indentation measurements. Human mind and decision making (e.g. on where the border of the indentation measured actually is) is subjective and can be easily influenced by different outside factors and personal preference, while a computer vision system will decide always 100% objectively according to its algorithm. This results in improved measurement uncertainty, increased repeatability and practically fully eliminated a major uncertainty factor: human/operator error.

Furthermore, marking by the IMS used to signal the operator where the indentation is to be made on the block helps with the realisation of the indentations, providing a well visible, consistent marking to follow.

The developed set of parameters as well as findings from the improvement of the TÜBİTAK UME IMS will be implemented directly into the new generation of IMS, ensuring more consistent measurements between the different laboratories.

Future improvements may include also fully automated indentation process by the hardness standard machine being connected to the IMS via a server, both being controlled by a centralized system. This will allow for fully automated processes and may also be used with the digital calibration certificates (DCC) in development at a current time by different research groups and metrology organisations.

At the highest metrology level for calibration of reference hardness blocks at NMIs and calibration certificates needs to include the measurement results currently available in traditional calibration certificates

such as the measured diameter or diagonal values and the calculated hardness values from the geometrical values. The DCCs should also include all relevant technical information regarding the traceability of the hardness machine, i.e. the force generation systems, time measurement system, and the indentation measurement system, and furthermore encompass the information about other relevant parameters such as microscope settings with contrast, optical resolution, aperture, luminance and the information about the edge detection algorithm.

The units of hardness cannot be represented directly in a DCC as a SI units according to D-SI guideline [10], but needs to be defined as a hybrid unit, specifying the test method, test load and the indenter geometry.

V. CONCLUSION

The full digitalisation of the TÜBİTAK UME IMS is a first step towards full connectivity between the IMS and the hardness standard machines, reducing measurement uncertainty and increasing reliability, while providing a basis for further development for the use in all levels of laboratories. Furthermore, important work is done in the area of correlating high-end 3D laser scanning measurement systems with the conventional 2D optical systems in use in the calibration- and testing laboratories around the world. This builds the basis for advanced image processing and edge detection algorithms, which can later be used to be implemented or retrofitted in these current optical systems, therefore marking a significant step towards the unification of hardness measurement and a decrease in measurement uncertainty all along the traceability chain.

This approach of unification by digitalisation will help to reduce the differences between different laboratories and NMIs, improving hardness metrology and measurement at a global scale.

The digital transformation of indentation measurement systems offers significant potential for redefining how laboratories collaborate, report and validate results. The integration of fully-automated systems with cloud-based platforms and DCCs could enable remote verification real-time quality monitoring-features especially valuable for international metrology networks.

ACKNOWLEDGEMENTS

The project 22RPT01 TracInd BVK-H has received funding from the European Partnership on Metrology, co-financed from the European Union's Horizon Europe Research and Innovation Programme and by the Participating States (Funder name: European Partnership on Metrology, Funder ID: 10.13039/100019599, Grant number: 22RPT01 TracInd BVK-H)

REFERENCES

[1] K. Hattori, G. W. Bahng, A. Germak, E. G. Aslanyan, F. Menelao, H. Li, K. Herrmann, L. Brice, Report of Pilot study on Brinell hardness comparison CCM.H-P3, BIPM CCM WG H, 2017 https://www.bipm.org/documents/20126/56729151/CCM-PS-E3_CCM.H-P3_2017.pdf/e886cd1b-7050-c19e-30ff-fdc99e211408?version=1.2&t=1706516358591

[2] C. Kuzu, E. Pelit, İ. Meral, Design and development of hardness indentation measurement system at TÜBİTAK UME, IMEKO 24th TC3, 14th TC5, 6th TC16 and 5th TC22 International Conference, 2022, TÜBİTAK UME, Kocaeli, Turkey

[3] F. B. Kværndrup, C. Engelbrekt, Ö. C. Kücükyildiz, M. A. J. Somers, T. L. Christiansen, G. Winther, G., Area determination with pile-up and sink-in in nanoindentation of oxygen containing titanium., Materials Today Communications, Article 103218, 2022. DOI: https://doi.org/10.1016/j.mtcomm.2022.103218

[4] F. Han, B. Tang, X. Yan, Indentation Pileup Behavior of Ti-6Al-4V Alloy: Experiments and Nonlocal Crystal Plasticity Finite Element Simulations, Metall Mater Trans, A 48, 2051–2061, 2017. DOI: https://doi.org/10.1007/s11661-016-3946-0

[5] C. Kuzu, A. Germak, F. Menelao, A. Štibler, M. Loewit, G. Genta, J. Borovský, G. Maculotti, M. Hiti, A. Prato, E. Pelit, J. Fidelus, T. Apostol, S. Ališić, B. Fakić, P. Reinstädt, F. Turotti, R. Affri, B. Muminovic, A. Alić, Traceability for Indentation Measurements in Brinell-Vickers-Knoop Hardness, Measurement: Sensors 38, 2025, DOI: https://doi.org/10.1016/j.measen.2024.101453

[6] L. Ma, S. Low, J. Song, A method to determine the spherical indentation contact boundary diameter in elastic-plastic materials, Mechanics of Material 69, 2014, DOI: http://dx.doi.org/10.1016/j.mechmat.2013.11.001

[7] International Organization for Standardization, ISO 6506 Metallic materials - Brinell hardness test Parts 1-3, 2014-2018

[8] International Organization for Standardization, ISO 6507 Metallic materials - Vickers hardness test Parts 1-3, 2018-2023

[9] International Organization for Standardization, ISO 6507 Metallic materials - Knoop hardness test Parts 1-3, 2017-2023

[10] D. Hutzschenreuter et all, SmartCom Digital System of Units (D-SI) Guide for the use of the metadata-format used in metrology for the easy-to-use, safe, harmonised and unambiguous digital transfer of metrological data. DOI:10.5281/zenodo.3522631, 2019

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Automating Photometric Measurements Using LabVIEW-Python-Based AI: Enhancing Precision in Luminous Intensity, Responsivity, Illuminance, and Flux Analysis at Saudi Standards, Metrology, and Quality Organization-SASO-KSA

Mohammad D. AlMelfi, Rawan A. AlMutairi, Fahad A. AlMuhlaki, Saad A. Bin Qoud, Rayan A. AlYousefi, I. AlFaleh, N. Qahtani, A. El-Matarawey

NMCC-SASO- Saudi Arabia

Abstract **- This work explores the automation of a photometry laboratory by integrating LabVIEW as the primary control and data acquisition platform, coupled with a Python-based artificial intelligence (AI) algorithm for intelligent analysis and optimization. The system is designed to measure and analyze critical photometric parameters such as luminous intensity, luminous responsivity, illuminance, and luminous flux—all of which are essential for evaluating the performance of light sources like tungsten halogen lamps, LEDs, displays, and optical sensors. By automating traditionally manual processes, the system enhances both the accuracy and efficiency of photometric testing.**

LabVIEW provides an intuitive graphical interface to control instruments, log data, and visualize results in real time. Meanwhile, the Python AI component improves decision-making by learning from historical data, predicting trends, and detecting anomalies or inconsistencies in measurements. The AI model optimizes calibration routines, reduces human error, and enables adaptive testing scenarios based on environmental conditions or device behavior.

Together, LabVIEW and Python form a powerful, flexible platform that brings modern intelligence to photometry labs, making them faster, smarter, and more reliable. This approach demonstrates how combining industrial automation tools with machine learning can revolutionize traditional optical testing environments, paving the way for advanced lighting technologies and robust quality assurance systems.

Keywords: Python, AI, Algorithms, Automation, Photometry, Responsivity, Illuminance, Flux, SASO, NMCC

I. INTRODUCTION

The field of photometry is essential for evaluating the performance of light sources such as tungsten halogen lamps, LEDs, displays, and optical sensors. Key parameters like luminous intensity [1], luminous responsivity [2], illuminance [3], and luminous flux [4] are crucial for characterizing these devices. Traditionally, photometric testing has relied on manual processes, which are not only prone to human error but also time-consuming and often lack adaptability to dynamic testing conditions.

This paper introduces an innovative solution to automate a photometry laboratory by integrating LabVIEW as the primary control and data acquisition platform with a Python-based artificial intelligence (AI) algorithm for advanced analysis and optimization using a powerful integrated library [5-12]. The system automates conventional manual tasks while introducing intelligent features such as predictive analytics, anomaly detection, and adaptive testing scenarios. This combination significantly enhances the accuracy, efficiency, and reliability of photometric measurements, transforming the way photometry labs operate. The SASO-NMCC photometry lab comprises various facilities and equipment that rely on custom software for operation. Recently, the lab faced challenges due to software failures that caused damage to certain light sources. The issue stemmed from the inability of the software to regulate the electric current passing through the lamps, resulting in currents exceeding critical thresholds. To address this problem, the digital transformation team conducted a thorough analysis and proposed replacing the outdated software with a new model. This updated system combines two powerful programming languages: LabVIEW for automation and fractional Proportional Integtration (PI control) algorithm and Python for data analysis, integrated into an interactive AI-driven graphical user interface platform.

II. EQUIPMENT AND SYSYTEM SETUP

Table 1 provides a detailed overview of each piece of equipment utilized in the laboratory, outlining its purpose, functionality, and relevance to photometry and electrical measurements. The information is systematically organized by equipment type, manufacturer, model, serial number, and a provided description for clarity and ease of reference.

Table1. Photometry lab equipment

Equipment	Manufacturer	Model	Serial Number
Digital Multimeter	Agilent Keysight	3458A	MY45050860
Digital Multimeter	Agilent Keysight	3458A	MY45050862
Photocurrent Meter	PRC Krochmann	Radiometer/ Photometer 211	141118-20
Shunt Resistance	Guildline Instruments	9230A-15R	71630
DC Power Supply	Heinzinger electronic GmbH	PTN 250-20	3449 10675
Current Source	Keithley	6220	4056589
Picoammeter	Keysight	B2981A	--

Digital Multimeter A high-precision device designed to measure electrical parameters such as voltage, current, and resistance. It plays a critical role in photometric testing by monitoring the voltage applied to light sources, ensuring stable power delivery during experiments. This stability is essential for accurate photometric calculations, as fluctuations in voltage can impact the performance of light sources. By tracking supply voltage and current, the multimeter provides an indication of the consistency and reliability of the electrical input to photometric devices.

Photocurrent Meter A specialized instrument used to measure the photocurrent generated by Photometer heads, which are photodiodes with $V(\lambda)$ mosaic filter, when exposed to light. Equipped with a thermoelectric cooler, it maintains a constant temperature for the photodiode, ensuring stable and precise measurements. This device is instrumental in calculating responsivity by comparing the incident optical power to the measured photocurrent. It is also used to determine the responsivity of devices under test, making it a vital tool for evaluating detector performance.

Shunt Resistance A precision resistor, typically with a low value (in milliohms) and tight tolerance, used to measure currents by converting them into measurable voltage drops across the resistor. Shunt resistors are commonly employed in current-sensing applications where direct current measurement is impractical. In photometric testing, they are used to measure the current drawn by light sources, providing critical data for analyzing device behavior under various operating conditions.

DC Power Supply A stable and adjustable direct current (DC) power source that supplies power to electronic devices. Known for its high stability and low ripple, this power supply ensures consistent operating conditions during photometric testing. It provides a constant voltage or current output within specified limits, making it ideal for powering light sources such as tungsten halogen lamps and LEDs. This equipment is essential for characterizing the electrical behavior of devices under different power levels and ensuring reliable test results.

Current Source A highly stable and precise instrument used to deliver a controlled amount of current to a device under test. In photometric applications, it drives light sources, such as tungsten halogen lamps and LEDs, at specific current levels to evaluate their performance under controlled conditions. This capability is crucial for testing luminous intensity and luminous flux as a function of current, enabling detailed analysis of device characteristics.

Picoammeter An ultra-sensitive instrument capable of detecting extremely small currents, typically in the picoampere range. It is used to measure dark current or leakage current in photodetectors and optical sensors, which is critical for characterizing device sensitivity and noise performance under low-light conditions. The picoammeter's precision makes it indispensable for advanced research and development in photometry.

Together, these instruments form a comprehensive setup for photometric testing. The Digital Multimeter and Current Source ensure accurate electrical characterization, while the Photocurrent Meter and Picoammeter enable precise optical measurements. The Shunt Resistance and DC Power Supply provide the necessary tools for controlling and stabilizing the testing environment. This combination supports a wide range of applications in photometry, from fundamental measurements to advanced research and development, as illustrated in Figure 1.

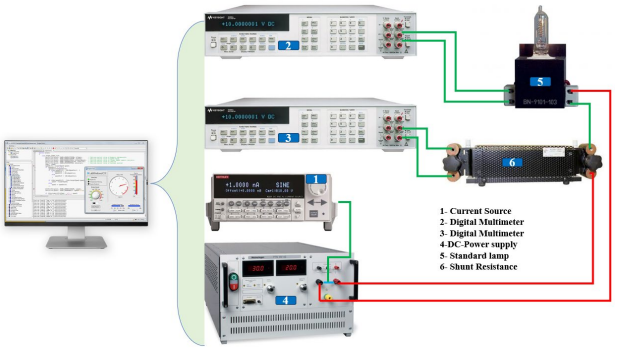


Fig 1. Photometry lab equipment

The diagram 2 and figure 3 illustrate a photometric measurement setup, which is used in laboratory for characterizing light sources and optical devices.

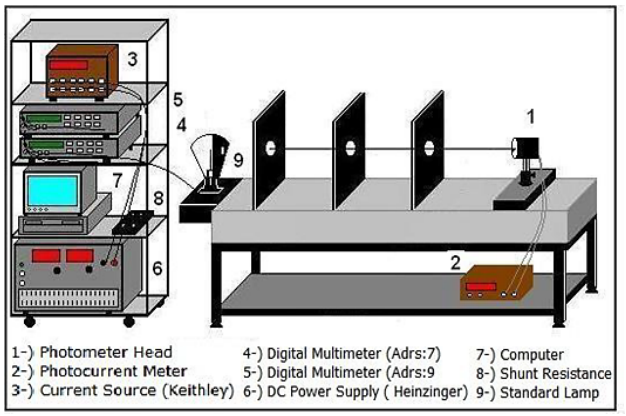


Fig 2. Photometric Measurement setup



Fig 3. Photometric Measurement Equipment

The Photometer Head serves as the sensor component responsible for measuring the intensity or flux of visible light emitted by a source. It captures incoming light and converts it into an electrical signal proportional to the light intensity, which is then connected to the Photocurrent Meter (2) or Picoammeter for further analysis. The Light Source is powered by the DC Power Supply (6) and driven by the Current Source (3) to ensure precise control over its operation. The output of the light source is measured using the Photometer Head (1) and the Photocurrent Meter (2). For calibration purposes, a Standard Lamp (9) is included in the setup. This lamp provides a known light output, enabling accurate calibration of the Photometer Head (1) and ensuring reliable measurement results. At the core of the system, the computer (7) acts as the central hub, controlling the Current Source (3), monitoring real-time measurements, and analyzing collected data. This versatile setup can be adapted for various photometric measurements, including luminous intensity, illuminance, and luminous responsivity of photometers.

III. PROGRAMMING DESIGN

The proposed automation framework is built on two core components: LabVIEW and Python AI. LabVIEW acts as the foundation for instrument control, real-time data acquisition, and visualization, while Python handles advanced data processing through machine learning algorithms. This integration creates a seamless workflow that combines hardware control with intelligent software capabilities.

1. LabVIEW for Instrumentation and Data Acquisition

LabVIEW, developed by National Instruments, offers a graphical programming environment that simplifies the creation of complex measurement systems. In this setup, LabVIEW serves as the central tool for instrument control, interfacing with photometric instruments via standard communication protocols such as GPIB, USB, and Ethernet. It logs measurement data in real time and stores it in structured formats for further analysis. Additionally, its user-friendly graphical interface enables operators to monitor test progress, view live results, and generate reports, as shown in figures 4 and 5. The LabVIEW workflow includes the following key steps:

a) Initialization: All connected instruments are initialized and configured for testing.
b) Data Acquisition: Measurements are taken at predefined intervals and logged into a database.
c) Real-Time Visualization: Results are displayed on a dashboard, allowing operators to monitor the testing process in real time.
d) Exporting Data: Collected data is exported to a CSV file for further analysis by the Python AI component.

This structured approach ensures efficient data collection, real-time monitoring, and seamless integration with the Python-based intelligent processing system.

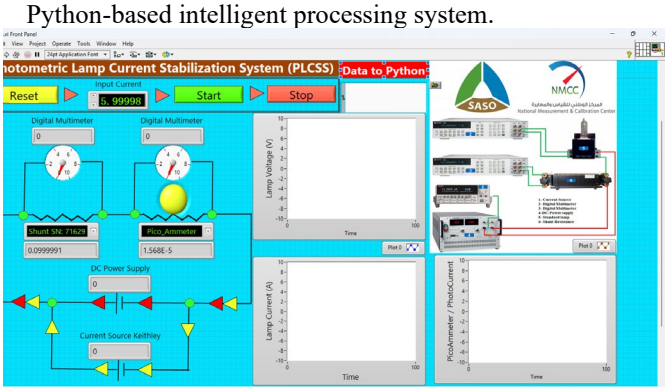


Fig 4. Controlled front panel

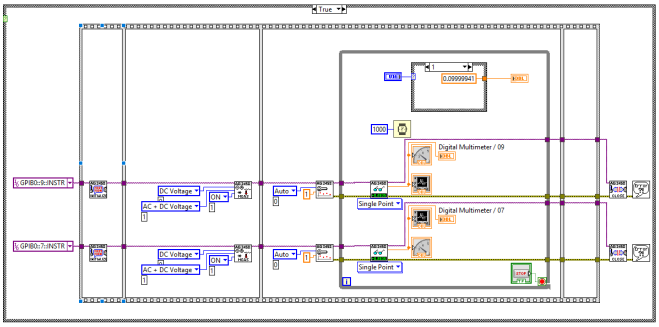


Fig 5. Piece of block diagram related to the automation of multimeters

2. Python AI for Smart Analysis and Optimization

The Python AI component utilizes powerful libraries such as TensorFlow, Scikit-learn, and Pandas to implement advanced machine learning models. These models analyze historical data to predict future trends in photometric parameters, enabling proactive decision-making. Additionally, the AI detects outliers or inconsistencies in measurements, flagging potential issues in the testing process. Adaptive algorithms further enhance the system by optimizing calibration routines and adjusting testing parameters based on environmental conditions or device behavior. Python Workflow includes the following steps:

a) Data Preprocessing: Raw data collected from LabVIEW is cleaned, normalized, and prepared for analysis [13, 14].
b) Model Training: Machine learning models are trained using both live and historical datasets to ensure accurate and reliable predictions.
c) Prediction and Optimization: The trained models dynamically forecast trends in measurement parameters, enabling real-time optimization of testing processes.
d) Hyperparameter Tuning: Parameters generated by the AI are fed back into the LabVIEW system to refine and improve the accuracy of future tests [15, 16].

This workflow ensures that the system continuously learns and adapts, enhancing the precision and efficiency of photometric testing through intelligent analysis and optimization.

The AI component adds a layer of intelligence to the system, enabling it to learn from past data and make informed decisions about future tests related to percentage of current generated by DC power supply and precision current by Keithley. The system adapts to changing conditions, such as variations in ambient light or temperature, ensuring consistent performance across different environments. The modular design of the system allows for easy integration of additional instruments or upgrades to the AI algorithms. While implementing the system, several challenges were encountered. Data synchronization is ensuring seamless communication between LabVIEW and Python, it is required a careful configuration of data exchange protocols [17-22].

IV. MODEL ARCHITECTURE

The flowchart in figure 6 outlines a typical machine learning and data analysis pipeline, the model architecture is described according to series of equations and expressions.

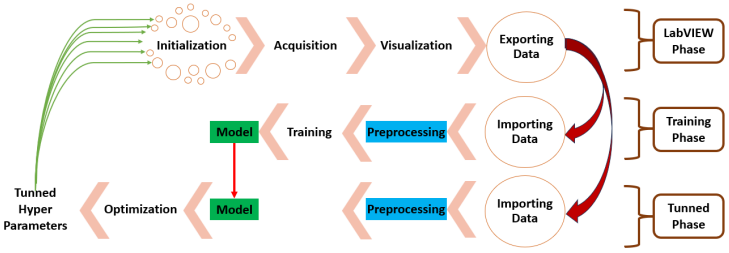


Fig 6. Model architecture and data pipeline

1. *Initialization*, this step involves setting up the environment and defining initial parameters.

$$Initialization\text{: } \mathcal{P} = \{\boldsymbol{\theta_o}, \boldsymbol{\varphi_o}, \boldsymbol{w_o}, \dots\} \quad 1$$

where $\mathcal{P}$ represents the set of initial parameters, and $\boldsymbol{\theta_o}$, $\boldsymbol{\varphi_o}, \boldsymbol{w_o}$ are initial hyperparameters then model parameters.

2. *Acquisition*, data collected from automated equipment.

$$Acquisition\text{: } \mathfrak{D}_{raw} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}, \dots\} \quad 2$$

where $\mathfrak{D}_{raw}$ is the raw dataset consisting of *n* samples.

3. *Exporting Data,* data stored and exported in csv files to processed and analyzed for further use.

$$Exporting\text{: } \mathcal{D}_{exported} = \mathcal{D}_{processed} \quad 3$$

4. *Importing Data (Training Phase),* Data is imported into the training phase for preprocessing and modelling.

$$\text{Importing (Training): } \mathcal{D}_{imported} = \mathcal{D}_{processed} \quad 4$$

5. *Preprocessing*, Raw data is cleaned, normalized, and transformed.

$$\text{Preprocessing: } \mathcal{D}_{processed} = \boldsymbol{P}(\mathcal{D}_{raw}) \quad 5$$

6. *Preprocessing*, model is trained using the preprocesses data.

$$\text{Training: } \boldsymbol{\mathcal{M}} = \boldsymbol{T}(\mathcal{D}_{processed}, \boldsymbol{\theta}) \quad 6$$

where $\boldsymbol{\mathcal{M}}$ is the trained model, $T(\cdot)$ is the training function, and θ are the model parameters.

7. *Optimization*, hyperparameters are tuned to optimize the model's performance.

$$\text{Optimization: } \boldsymbol{\theta^*} = \boldsymbol{O}(\boldsymbol{\mathcal{M}}, \mathcal{D}_{\text{validation}}) \quad 7$$

where θ^* are the optimized hyperparameters, $O(\cdot)$ is the optimization function, and $\mathcal{D}_{\text{validation}}$ is the validation dataset.

8. *Preprocessing (Tuned Phase),* Data is preprocessed again with updated techniques based on insights from the training phase.

$$\text{Preprocessing (Tuned): } \mathcal{D}_{\text{tuned}} = \boldsymbol{P'}(\mathcal{D}_{\text{raw}}) \quad 8$$

where $P'(\cdot)$ is an updated preprocessing function.

9. *Importing Data (Tuned Phase),* tuned data is imported for further processing and deployment.

$$\text{Importing (Tuned): } \mathcal{D}_{\text{imported\_tuned}} = \mathcal{D}_{\text{tuned}} \quad 8$$

10. *Model (Tuned Phase),* The final model is deployed and used for inference.

$$\text{Tuned Model: } \boldsymbol{\mathcal{M}}_{\text{final}} = \boldsymbol{\mathcal{M}}(\boldsymbol{\theta^*}) \quad 9$$

V. RESULTS AND DISCUSSION

Table 2 summarize the average of eight runs measurements for ten lamps, that indicates perfect match with traceable nominal values from calibration certificates.

Table 2. Ten validations of software with Lamp nominal values

Meas. No.	Lamp Serial No.	Lamp Certificate current Input current	Lamp Certificate Voltage	Lamp Measured Current	Lamp Measured Voltage 07	Current Power supply	Precession current source	Shunt voltage	Shunt Resistance	Calculated current	Delta V	Delta I	% supply	% PCS	Stopwatch (Power supply)
1	1009	5.892	30.835	5.892	31.117	5.85	0.03882	0.58919	0.0999941	5.89224764	0.282	0.00024764	99.287169	0.65885947	-
2	1115	5.916	30.939	5.91601	31.18259	5.87	0.03863	0.5915929	0.0999941	5.91627806	0.24359	0.00027806	99.2222799	0.65297388	-
3	1097	5.962	31.291	5.96194	31.60531	5.91	0.0405	0.5961906	0.0999941	5.96225777	0.31431	0.00025777	99.1288071	0.67930908	-
4	A	4		3.99999	30.42339	3.95	0.04112	0.3999972	0.0999941	4.00020801		0.00020801	98.7502469	1.02800257	-
5	B	6		6.00003	62.83999	5.95	0.03919	0.6000021	0.0999941	6.00037502		0.00037502	99.1661708	0.6531634	-
6	C	5		4.99959	45.26392	4.95	0.04345	0.5000256	0.0999941	5.00055103		0.00055103	99.0081187	0.86907126	01:28.6
7	D	7.246	86.08	7.24599	88.30552	7.2	0.03799	0.724591	0.0999941	7.24633753	2.22552	0.00033753	99.3653041	0.52428999	01:28.1
8	1088	5.898	30.548	5.89799	30.51875	5.85	0.03974	0.5897976	0.0999941	5.898324	-0.02925	0.000324	99.186333	0.67378887	01:28.5
9	1076	5.93	30.548	5.93001	30.88055	5.88	0.03774	0.5930004	0.0999941	5.93035389	0.33255	0.00035389	99.1566625	0.63642388	01:28.1
10	1090	5.908	30.705	5.908	30.65335	5.86	0.04023	0.5907978	0.0999941	5.90832659	-0.05165	0.00032659	99.1875423	0.6809411	01:28.2

Both measured and calculated currents show excellent agreement with certificate values; the calculated current is evaluated from the shunt voltage and resistance respectively. Figure 7 shows the relationship between Certificate Current ($I_{Certificate}$) and Measured Current ($I_{Measured}$). A trend line is fitted to the data points indicates a linear relationship between $I_{Certificate}$ and $I_{Measured}$. The slope of the line is approximately 1, suggesting that the measured current is very close to the certificate current. The intercept is very small ($-5\ x\ 10^{-05}$), which is effectively zero for practical purposes. Coefficient of Determination (R^2) is reported as 1, a perfect fit between the trend line and the data points, this suggests that the measured current values are highly correlated with the certificate current values, with no unexplained variance. The small deviations from the trend line suggest high precision in the measurements. The differences between $I_{Certificate}$ $I_{Calculated}$ and $I_{Measured}$ are minimal as shown in figure 8 and 9, there are no significant outliers in the data. All points lie close to the trend line; the scatter plot provides strong evidence that the measured current values are highly consistent with the certified current values.

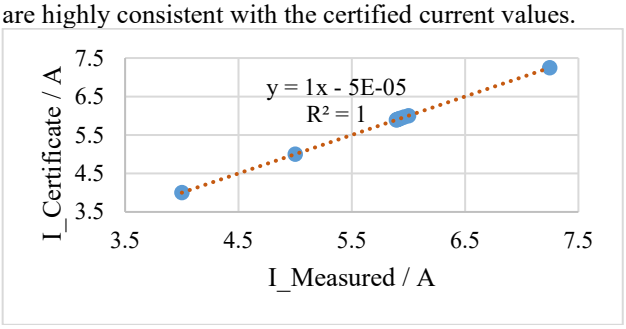


Fig 7. Rrelationship between Certificate Current ($I_{Certificate}$) and Measured Current ($I_{Measured}$)

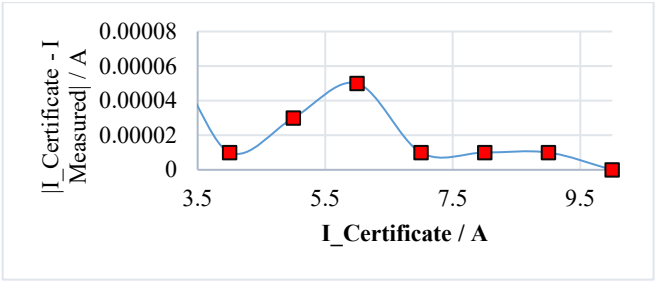


Fig 8. Rrelationship between $|I_{Certificate} - I_{Measured}|$ and ($I_{Certificate}$)

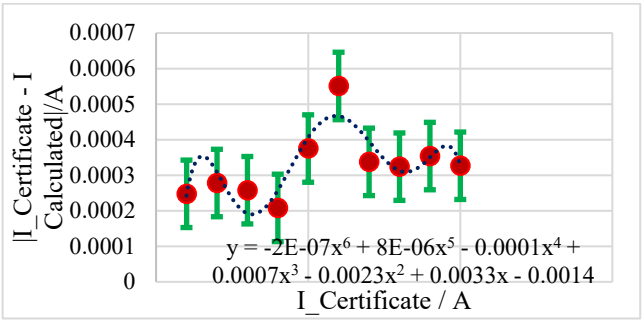


Fig 9. Rrelationship between $|I_{Certificate} - I_{Calculated}|$ and ($I_{Certificate}$)

Figure 10 shows the relationship between % PCS (Percentage of precision Current Source) and % Power supply (Percentage of Power Supply) The equation of the trend line indicates a negative linear relationship between % Power supply and % PCS, As % Power supply increases, % PCS decrease. R^2 value of 0.9247 suggests that approximately 92.47% of the variability in % PCS can be explained by the variability in % Power supply. This indicates a strong correlation between the two variables. The data points are clustered around the trend line, indicating that the linear model fits the data well and there are no significant outliers, as all points follow the general trend. The precise measurements and consistent trend suggest that the relationship is robust and reliable.

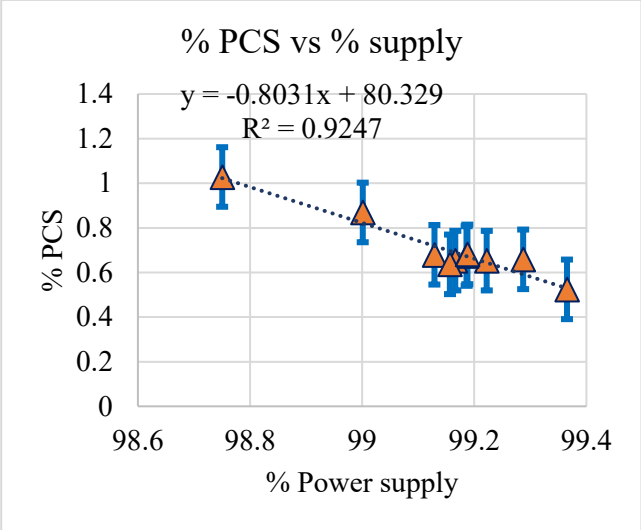


Fig 10. Rrelationship between % Power supply and %PCS

CONCLUSION

The integration of LabVIEW and Python-based AI has successfully transformed the photometry laboratory into a smart, efficient, and reliable testing environment. By automating manual processes and incorporating intelligent analysis capabilities, the system significantly enhances both the accuracy and efficiency of photometric testing. Furthermore, it introduces adaptive and predictive methodologies, enabling more advanced and dynamic testing scenarios. This approach demonstrates the potential of combining industrial automation tools with machine learning to revolutionize traditional optical testing environments, establishing new benchmarks for quality assurance in lighting technologies. Future work will focus on further expanding the AI's capabilities, including the implementation of real-time feedback control systems. Additionally, efforts will explore the application of this framework in other domains, such as radiometry and colorimetry, to broaden its impact and utility across various fields.

REFERENCES

1. The measurement of absolute luminous intensity distributions, CIE Publication, Vol. 70, 1987.
2. The Basis of Physical Photometry, CIE Publication, Vol. 18, 2019.
3. Methods of Characterizing Illuminance Meters and Luminance Meters, CIE Publication, Vol. 69, 1987.
4. The measurement of luminous flux, CIE Publication, Vol. 84, 1989.
5. CASIMER DECUSATIS, Handbook of Applied Photometry, Woodbury, N.Y., AIP Press 1998.
6. RASTELLO M.L., MIRALDI E., PISONI P., J. Appl. Opt. 35 (1996), 4385.
7. Python Software Foundation, *Python*.
8. NumPy Developers, *NumPy*.
9. SciPy Developers, *SciPy*.
10. Pandas Development Team, *Pandas*.
11. Matplotlib Development Team, *Matplotlib*.
12. Seaborn Development Team, *Seaborn*.
13. Python Software Foundation, *Python 3 Reference Manual* . Scotts Valley, CA: CreateSpace Independent Publishing Platform, 2009. ISBN: 978-1-4414-1269-0.
14. T. E. Oliphant, "Python for Scientific Computing," *Computing in Science & Engineering* , vol. 9, no. 3, pp. 10-20, 2007. DOI: 10.1109/MCSE.2007.58.
15. W. McKinney, Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. Sebastopol, CA: O'Reilly Media, 2010.
16. Scikit-learn Developers, *Decision Trees in Scikit-learn*.
17. J. R. Quinlan, "Induction of Decision Trees," *Machine Learning*, vol. 1, no. 1, pp. 81-106, 1986, doi: 10.1007/BF00116251.
18. T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997.
19. C. M. Bishop, *Pattern Recognition and Machine Learning* . New York, NY, USA: Springer, 2006.
20. S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach* , 3rd ed. Upper Saddle River, NJ, USA: Pearson Education, 2010.
21. L. Breiman, "Statistical Modeling: The Two Cultures," *Statistical Science* , vol. 16, no. 3, pp. 199-231, 2001, doi: 10.1214/ss/1009213726.
22. V. N. Vapnik, The Nature of Statistical Learning Theory . New York, NY, USA: Springer, 1995.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Thermometry Machine Learning Model for Digitized Metrological Calibration of Platinum Resistance Thermometer

Oqab. N. Alotaibi[1], Rakan O. Alnefaie[1], Arwa K. Alrushud[1], Fahad A. AlMuhlaki[1], Rayan A. AlYousefi [1], Saad A. Bin Qoud[1], I. AlFaleh[1], N. Qahtani[1], A. El-Matarawey[2]

1. NMCC-SASO- Saudi Arabia
2. National Institute of Standards (NIS-Egypt)

***Abstract* – Temperature measurements rely on various types of thermometers, including but not limited to Platinum Resistance Thermometers (PRTs), thermocouples, and radiation thermometers. Among these, resistance thermometers are considered highly reliable for sensitive temperature measurements. To ensure the accuracy and precision of measurement results, it is essential to consider factors that affect either the temperature value itself (after conversion from ohms) or the uncertainty estimation when using resistance thermometers.**

One critical factor is the interpolation error that arises when converting resistance values to temperature using the ITS-90 equations. Discrepancies in the temperature values obtained through these methods can impact measurement reliability. Therefore, this study aims to develop a robust Python-based algorithm for calibrating PRTs with minimal errors, thereby reducing the impact on measurement uncertainty.

The study will provide an open-source, step-by-step algorithm as part of the global digital transformation trend. This algorithm will serve as a valuable resource for researchers and practitioners seeking to enhance the reliability and accuracy of temperature measurements.

Keyword - Temperature, Resistance, control, Python, fixed point, ITS-90.

I. INTRODUCTION

The National Metrology and Calibration Center (NMCC) at the Saudi Standards, Metrology & Quality Organization (SASO) includes the Thermometry Laboratory, a prominent facility dedicated to the enhancement, advancement, and preservation of national standards in Thermometry under the International Temperature Scale of 1990 (ITS-90) [1]. Additionally, the Thermometry Laboratory provides testing and calibration services to a wide range of industries, including medical, industrial agriculture, pharmaceuticals, and scientific sectors. This underscores the vital role that National Metrology Institutes (NMIs) play in maintaining quality infrastructure. According to the Bureau International des Poids et Mesures (BIPM), "Metrology" refers to the science of measurement. It is essential that measurements of physical properties, such as temperature, are conducted using reliable stabilization systems. Researchers, scientists, equipment developers, and manufacturers have a keen interest in measuring various mediums. In the field of metrology at National Metrology Institutes, the primary focus is on achieving accuracy and precision within a few millikelvins (mK). Industrial Platinum Resistance Thermometers (IPRTs), also known as Pt-100s, utilize platinum as the sensing resistor due to its high sensitivity to temperature, favorable temperature coefficient, ability to withstand high temperatures (up to 962 °C in certain cases), and resistance to chemical corrosion, particularly oxidation. Platinum resistance thermometers are calibrated at specific fixed temperature points within their range, as per ITS-90. To simplify the process for users, calibration outcomes are converted from resistance to temperature. The ITS-90 calculation can be used for unit conversion when calibration is performed at fixed points [2, 3]. In recent years, digital technologies have seen significant advancements, including concepts like big data, cloud computing, artificial intelligence (AI), and machine learning (ML). ML, deep learning, and neural networks have been effectively integrated into the field of metrology [4, 5, 6, 7, 8, 9, 10]. The exponential growth in computing and storage capacities, coupled with faster data exchange speeds and the affordable availability of versatile sensors, has created new opportunities for improving metrology tools, such as Python programming language. Python is one of the most widely used programming languages, serving various purposes, including machine learning, web development, desktop applications, and providing a robust environment for parallel mathematics. Its syntax is straightforward and user-friendly, making it an excellent choice for many applications. Python can be easily adapted for machine learning (ML), artificial intelligence (AI), descriptive and inferential statistics, and advanced mathematics, delivering great results and accuracy even

with limited resources and time. Python offers a wide range of libraries tailored to specific tasks based on the application area. In our case (Advanced Mathematics), we will use libraries that focus on algebra, matrices, mathematics, visualizations, exploratory data analysis (EDA), statistics, data pipelines, and data containers.

II. ITS-90 TEMPERATURE SUB-RANGES

In this study we selected a wide subrange for investigations form -40 °C up to 660 °C that split into two main subranges. The resistance ratio and general deviation function is given by equation 1 and 2 respectively [1, 2, 3, 11, 12, 22].

$$W(T_{90}) = R(T_{90})/R(273.16\,K) \quad (1)$$

$$W(T_{90}) - W_r(T_{90}) = a[W(T_{90}) - 1] + b[W(T_{90}) - 1]^2 + c[W(T_{90}) - 1]^3 + d[W(T_{90}) - W_{Al}]^2 \quad (2)$$

The Mercury Triple Point (-38.8344 °C) to the Gallium Melting Point (29.7646 °C) with c, d = 0

$$W(T_{90}) - W_r(T_{90}) = a[W(T_{90}) - 1] + b[W(T_{90}) - 1]^2 \quad (3)$$

The constants a and b are determined from the calibration of thermometer at mercury triple point and gallium melting point. The Water Triple Point 0 °C to the Aluminum Freezing Point (660.323 °C) with d = 0.

$$W(T_{90}) - W_r(T_{90}) = a[W(T_{90}) - 1] + b[W(T_{90}) - 1]^2 + c[W(T_{90}) - 1]^3 \quad (4)$$

The constants a, b and c are determined from the calibration of thermometer at tin freezing point, zinc freezing point and aluminum freezing point.

III. MACHINE LEARNING MODEL

III. A. Python environment

The Python ecosystem provides a comprehensive suite of libraries for scientific computing, data analysis, and machine learning, often utilized within the interactive environment of Jupyter Lab [13]. At its core, NumPy [14] serves as a fundamental library for numerical computations, offering support for multi-dimensional arrays, matrices, and a wide range of mathematical operations. Built on top of NumPy, Pandas [15] provides high-level data manipulation tools, including Series and DataFrame structures, enabling efficient management and analysis of tabular data. For visualization, Matplotlib [16] allows the creation of complex static plots, while Plotly [17] offers interactive and dynamic visualizations. SciPy [18] extends NumPy's functionality by providing advanced tools for scientific computing, such as optimization, integration, and linear algebra. Seaborn [19], based on Matplotlib, simplifies the creation of visually appealing statistical graphics through a high-level interface. Statsmodels [20] supports statistical modeling, hypothesis testing, and data exploration with extensive result statistics for various estimators. Finally, Scikit-learn [21] stands out as a powerful library for machine learning, offering efficient tools for tasks like classification, regression, clustering, and dimensionality reduction through a consistent API. Together, these libraries form an indispensable toolkit for algorithm development, data analysis, and visualization in Python.

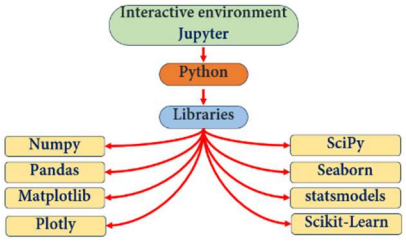


Fig. 1. Interactive environment of Python with an integrated library

III. B. MODEL ARCHITECTURAL.

The workflow for data acquisition, processing, and machine learning model development can be described in a structured manner as shown in figure 2.

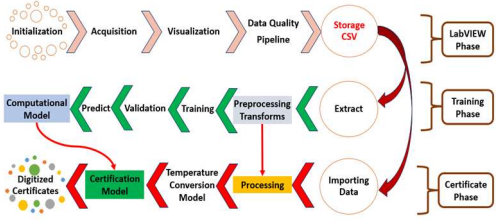


Fig. 2. Model architectural and data pipeline

- *Initialization*: The system initializes with parameters P and configurations C: System← {P, C}, the parameters here are kind of equipment such as ASL thermometry bridge, standard resistor bath with specific configuration as low gain and 1 mA current etc.

- *Acquisition*: Data Frame DF is acquired from SPRT as a time series versus resistance ratio, live standard resistor, oil bath temperature and macroscopic-microscopic feature engineering.

$$DF = \{x_1, x_2, \ldots, x_n\}, x_i \in \mathbb{R}^m \quad 5$$

where n is the number of samples, and m is the dimensionality of each sample.

- *Visualization:* Raw data DF["Ratio"] is visualized using functions $V(DF["Ratio"]): \mathbb{R}^{n\times m}$ For example, scatter plots or histograms are generated to explore distribution.

- *Data Quality Pipeline:* Data quality metrics Q(DF) are computed to ensure integrity

$$Q(D) = \{\mu, \sigma, missing\_values\} \quad 6$$

Where $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2}$ is mean and standard deviation of the trained data respectively. Missing values are flagged and handled for example dropped NA, fill forward or backward using live synchronization.

- *Storage (CSV):* Validated data DF is stored in CSV format, DF→CSV(DF).

- *Extract:* Data is extracted from CSV files for further processing as $DF' = Extract(CSV(DF))$

- *Preprocessing Transforms:* Preprocessing involves transformations $T(DF') = \{Z, E, S\}$ where Z is a zero-mean normalization $z_i = (\frac{X_i - \mu}{\sigma})$, E is Encoding categorical variables and S is scaling features to a range [0,1]

- *Processing:* Feature extraction F(DF′) transforms data into a feature matrix $X = F(D'), X \in \mathbb{R}^{n \times k}$ where k is the number of engineered features.

- *Training:* A machine learning model M is trained on features matrix X and target vector Resistance(R), as $M = Train(X, R; \theta)$, where θ is the model parameters optimized via a loss function L

$$\theta^* = arg\ min_\theta L(M(X;\theta), R) \quad 7$$

- *Validation:* The trained model M is validated using a separate validation set X_{val} and R_{val}, Common metrics include accuracy, mean squared error (MSE), etc.

IV. CASE STUDY

As mentioned before importing the pythonic libraries into the environment (figure 3). for simplicity the libraries were aliased be short name for example python know NumPy by its new name np. The lines of codes were introduced in as a screenshot from the program in order to visualize the different in colors for each piece of codes.

```
import numpy as np
import pandas as pd
import plotly.express as px
import seaborn as sn
from matplotlib import pyplot as plt
import scipy as sp
import statsmodels as sm
from sklearn.linear_model import LinearRegression
```

Fig 3. Loading the required libraries into the python environment

Reading step used Pandas library (pd) that read the csv (comma-separated-values) file, put the data into a container called data frame (df) and display the first 5 rows of the data, Figure 4. The reading step is iterable for all calibration files for other fixed points, water triple point file will be introduced as an example and other files will be the same.

```
# Import the triple point of water data from CSV file
df = pd.read_csv('water_triple_point.csv')
df.head(5)
```

Fig 4. Loading the required libraries into the python environment

The output will be in the form of table called df that includes data about the time of triggering the measurement, index of data value that start in python with zero and finally the value itself. for simplicity in our case, we interesting only in two columns the value and its index number.

Table 1. Reading data from water triple point file

Index	Value/Ω
0	99.392992
1	99.393673
2	99.392815
3	99.393372
4	99.393572

Extracting step are using matplotlib, SciPy and statsmodel libraries for manipulating, visualization, performing normal distribution fitting and finally recalling the median with the required confidence interval after doing the required correction due to the self-heating.

```
#----------Normal Distribution plot----------
plt.figure(figsize=(10,8))
count, bins, ignored = plt.hist(df, 90, density=True,
                                edgecolor="black")
plt.plot(bins, 1/(sigma * np.sqrt(2 * np.pi)) *
               np.exp( - (bins - mu)**2 / (2 * sigma**2)),
         linewidth=2, color='r')
plt.axvline(mu, color = 'red', linestyle = '--',linewidth=2)
plt.axvline(np.quantile(df,0.05), color = 'red',
            linestyle = '--',linewidth=2)
plt.axvline(np.quantile(df,0.95), color = 'red',
            linestyle = '--',linewidth=2)
plt.xlabel('Resistance / Ohm')
plt.ylabel('Density')
plt.rcParams.update({'font.size': 25})
plt.show()
#----------------Median----------------
Rw= np.median(df)
```

Fig 5. Statistical treatment of the df with normal distribution fitting and obtaining the median value.

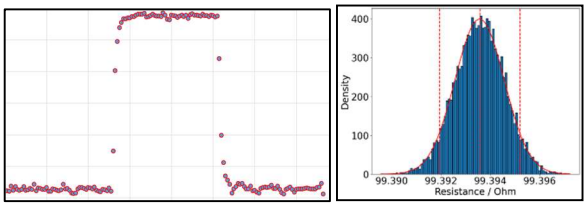


Figure 6. Normal distribution fitting of the data.

The best estimation of the plateau point will be stored in a new variable Rw = 99.39358 Ω.

Starting from instantiating the variables, the value for eac h file will be stored in new name that is represent on its so urce. $R_w, R_{Hg}, R_{Ga}, R_{Sn}, R_{Zn}, R_{Al}$ variables are representin g for the abs-olute resistance of the measurements at fixed points after multiplied by the value of standard resistor Te nsily $R_s = 100.00077$. Now it is the time of resistance ra tio W(T_{90}), by apply equation 1 and stored the variables in a new names $W_{Hg}, W_{Ga}, W_{Sn}, W_{Zn}, W_{Al}$ (Figure 7).

```
# The value of triple point of water
Rw=99.39358;

# The value of triple point of mercury
Rhg=83.90765;

# The value of melting point of gallium
Rga=111.13158;

# The value of freezing point of tin
Rsn=188.10386;

# The value of freezing point of Zinc
Rzn=255.281564;

# The value of freezing point of Alluminium
Ral=335.46533;

# W(T90)
Wsn=Rsn/Rw;
Wzn=Rzn/Rw;
Wal=Ral/Rw;
Whg=Rhg/Rw;
Wga=Rga/Rw;
```

Fig 7. Instantiating the resistance variable at each fixed point.

```
# Chicking the criteria of ITS-90
Wga>=1.11807

True

# Chicking the criteria of ITS-90
Whg>=0.844235

True
```

Figure 8. Checking the criteria of ITS-90

As described in the text of ITS-90, to be considered a suitable platinum resistance thermometer, it is necessary for the thermometer to be constructed using pure and strain-free platinum, and it must meet one of the two conditions outlined as W (29.7646 °C) ≥ 1.118 07 or W (-38.8344 °C) ≥ 0.844 235.

In this case, the thermometer satisfies one condition of the criterion and the algorithm will continue for next step. If the thermometer fails to satisfy both criterions, the thermometer will be introduced for another algorithm based on Callendar Van Dusen equation which will be introduced in another article.

This is the longest part of algorithm and the most computational resources consumable. The advantage of python is its limited consuming operations, python is still work efficiently, fast even with low computers. the temperature step is consisting of several sub-steps, it begins with calculating the conversion constants, evaluating Wr, calculating the temperature, looping for the range of Wr against temperature, calculate the increments in temperature against 0.01 change in the value Wr and finally building table of the specific range. Return the coefficients of the system of equations based on equation (3). Solve equations in two variables a, b as shown in code block figure 9.

```
# Range from HgTP up to GaMP--------------------------------
#(Whg-0.84414211==(a4*(Whg-1))-(b4*(Whg-1)**2)
#(Wga-1.11813889==(a4*(Wga-1))-(b4*(Wga-1)**2)
A = np.array(
    [
        [(Whg-1),(Whg-1)**2],
        [(Wga-1), (Wga-1)**2]
    ],dtype=np.float64
)
b = np.array([(Whg-0.84414211), (Wga-1.11813889)], dtype=np.float64)
y = np.linalg.inv(A).dot(b)
y
```

Figure 9. Input system of equations based on NumPy solve linear algebra function (2 X 2 Equations).

The return constants will be aliased as a_l, b_l for low temperature coefficients, there values are (a_l=-3.54604293e-04, b_l=6.12237423e-05)

Building table of temperature based on the following equation as shown in figure 10 and table 2.

$$\frac{T_{90}}{273.16} = B_o + \sum_{i=1}^{15} B_i \left[\frac{W_r(T_{90})^{\frac{1}{6}} - 0.65}{0.35}\right]^i \qquad 8$$

```
tl=[]
for Wt in [Y / 100 for Y in range(83, 101, 1)]:
    Wr=Wt-(al*(Wt-1))-((bl*(Wt-1)));
    W=((Wr**(1/6))-0.65)/0.35
    t=(((Bo)+(B1*W)+(B2*(W)**2)+(B3*(W)**3)+(B4*(W)**4)
        +(B5*(W)**5)+(B6*(W)**6)+(B7*(W)**7)+(B8*(W)**8)
        +(B9*(W)**9)+(B10*(W)**10)+(B11*(W)**11)+(B12*(W)**12)
        +(B13*(W)**13)+(B14*(W)**14)+(B15*(W)**15))*(273.16))-273.15;
    tl.append(t)
df_TL= pd.DataFrame(tl, columns=['Temperature'])
Y=[Y / 100 for Y in range(83, 101, 1)]
df_Wtl=pd.DataFrame(Y, columns=['W'])

nL=len(df_TL['Temperature'])
incL=[]
for i in range (0, nL-1, 1):
    incr=df_TL['Temperature'][i+1]-df_TL['Temperature'][i]
    incL.append(incr)
df_incL=pd.DataFrame(incL, columns=['Increament'])

df_L=pd.concat([df_Wtl, df_TL, df_incL], axis=1)
df_L.head()
```

Figure 10. Input codes for generating table of low temperature range against resistance ratio and increment.

Table 2. First five rows of the generated data for low range.

Index	W	Temperature	Increament
0	0.83	-42.353109	2.476393
1	0.84	-39.876716	2.478385
2	0.85	-37.398332	2.480364
3	0.86	-34.917968	2.482331
4	0.87	-32.435637	2.484287

Return the coefficients of the system of equations based on equation (4). Solve equations in three variables a, b and c as shown in code block figure 11.

```
# Range from WTP up to ALFP--------------------------------
    # Wsn-1.89279768==(ah*(Wsn-1))-(bh*(Wsn-1)**2)-(ch*(Wsn-1)**3)
    # Wzn-2.56891730==(ah*(Wzn-1))-(bh*(Wzn-1)**2)-(ch*(Wzn-1)**3)
    # Wal-3.37600860==(ah*(Wal-1))-(bh*(Wal-1)**2)-(ch*(Wal-1)**3)
A = np.array(
    [
        [(Wsn-1),(Wsn-1)**2, (Wsn-1)**3],
        [(Wzn-1), (Wzn-1)**2, (Wzn-1)**3],
        [(Wal-1), (Wal-1)**2, (Wal-1)**3]
    ],dtype=np.float64
)
b = np.array([(Wsn-1.89279768), (Wzn-2.56891730),(Wal-3.37600860)],
             dtype=np.float64)
x = np.linalg.inv(A).dot(b)
x
```

Figure 11. Input system of equations based on NumPy solve linear algebra function (3 X 3 Equations).

The return constants will be aliased as a_h, b_h and c_h for lo w temperature coefficients, there values are (a_l= -3.09195 651e-04, b_l= -1.28280350e-05, ch =3.25476196e-06)

Building table of temperature based on the following equation as shown in figure 12 and table 3.

$$T_{90} - 273.15 = D_o + \sum_{i=1}^{9} D_i \left[\frac{W_r(T_{90}) - 2.64}{1.64}\right]^i \qquad 9$$

```
th=[]
for Wt in [X / 100 for X in range(100, 339, 1)]:
    Wr=Wt-(ah*(Wt-1))-(bh*(Wt-1)**2)-(ch*(Wt-1)**3)
    W=(Wr-2.64)/1.64
    t=(Do+(D1*W)+(D2*(W)**2)+(D3*(W)**3)+(D4*(W)**4)+
        (D5*(W)**5)+(D6*(W)**6)+(D7*(W)**7)+(D8*(W)**8)+
        (D9*(W)**9))
    th.append(t)
df_T= pd.DataFrame(th, columns=['Temperature'])
X=[X / 100 for X in range(100, 339, 1)]
df_Wt=pd.DataFrame(X, columns=['W'])

n=len(df_T['Temperature'])
inc=[]
for i in range (0, n-1, 1):
    incr=df_T['Temperature'][i+1]-df_T['Temperature'][i]
    inc.append(incr)
df_inc=pd.DataFrame(inc, columns=['Increament'])

df_H=pd.concat([df_Wt, df_T, df_inc], axis=1)
df_H.head()
```

Figure 12. Input codes for generating table of high temperature range against resistance ratio and increment.

Table 3. First five rows of the generated data for high range.

Index	W	Temperature	Increament
0	1.00	0.010000	2.508940
1	1.01	2.518940	2.510865
2	1.02	5.029805	2.512793
3	1.03	7.542598	2.514723
4	1.04	10.057321	2.516656

Collect step based on Pandas, seaborn, scikit-learn and scipy, the algorithm will merge the two generated tables in one csv file, collect the conversion constants in one csv file. Plot step based on matplotlib and seaborn for good visualizations. The algorithm is plotting the linearity curve, thermometer has a response that is close to good linearity. A common approximation is to draw a "straight line curve," starting from the triple point of mercury up to freezing point of Aluminum. The Y-axis represents to the Temperature / ºC and X-axis represents to the resistance ration Rt/Rs, figure 13.

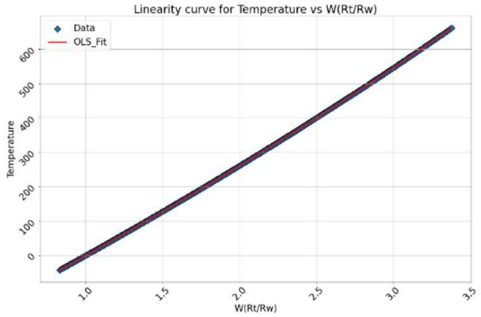


Figure 13. Linearity curve for temperature against W.

In validation step, bulk of codes was created in order to build a function called ***convert_R.*** The function task is to gathering the resistance value in ohm, locate its position in high or low ranges and convert it to temperature / ºC, figure 14. **If** the resistance value less than Rw, the algorithm is going to apply equation 5. **Else**, the the algorithm apply equation 6. The code is called *RPC* for Resistance - Positioning - Conversion code.

The validation mechanism starts with input the actual measured resistance and subtracting from the actual defining temperature in the text of ITS-90, figure 15. the code is called *ARTC* for Actual-Resistance-Temperature-Conversion code. The second step of validation mechanism is building table of discrepancies between T_{90} and T_{Model} as shown in table 4.

```
# Validation of Model
def convert_R(resistance):
    if resistance < Rw:
        Wt=resistance / Rw
        Wr=Wt-(al*(Wt-1))-((bl*(Wt-1))*(np.log(Wt)))
        W=((Wr**(1/6))-0.65)/0.35
        t=(((Bo)+(B1*W)+(B2*(W)**2)+(B3*(W)**3)+(B4*(W)**4)
        +(B5*(W)**5)+(B6*(W)**6)+(B7*(W)**7)+(B8*(W)**8)
        +(B9*(W)**9)+(B10*(W)**10)+(B11*(W)**11)+(B12*(W)**12)
        +(B13*(W)**13)+(B14*(W)**14)+(B15*(W)**15))*(273.16))
        -273.15
    else:
        Wt=resistance / Rw
        Wr=Wt-(ah*(Wt-1))-(bh*(Wt-1)**2)-(ch*(Wt-1)**3)
        W=(Wr-2.64)/1.64
        t=(Do+(D1*W)+(D2*(W)**2)+(D3*(W)**3)+(D4*(W)**4)+
        (D5*(W)**5)+(D6*(W)**6)+(D7*(W)**7)+(D8*(W)**8)+
        (D9*(W)**9))
    return t
```

Figure 14. Algorithm Resistance-Positioning-Conversion code

```
t_w=convert_R(Rw)
print(
    'The corresponding temperature is {} oC'
    .format(t_w)
)
```

```
The corresponding temperature is 0.00999999999990539 oC
```

```
t_w-0.01
```

```
-9.461008365629908e-14
```

Figure 15. Actual-Resistance-Temperature-Conversion code

Table 4. Summary of discrepancies between T_{90} and T_{Model}

T_{90} ºC	T_{Model} ºC	Error (T_{Model} - T_{90}) ºC
-38.8344	-38.83442960353265	2.96035326 x 10^{-05}
0.0100	0.0099999999999053	-9.4610083 x 10^{-14}
29.7646	29.764598216023084	1.78397691 x 10^{-06}
231.9280	231.9280732051055	7.32051055 x 10^{-05}
419.5270	419.5269848357539	-1.5164246 x 10^{-05}
660.3230	660.3230567011886	5.67011886 x 10^{-05}

The propagation of the error was estimated according to fitting the T_{90} against the T_{Model} - T_{90} as shown in figure 16. The interpolating-Extrapolating errors [22] can be estimated from the following equation. $Error\ (predicted) = (3.985e-12)\ T_{90}^3 - (3.913e-09)T_{90}^2 + (9.312e-07)\ T_{90}^1 - (9.835e-07)$

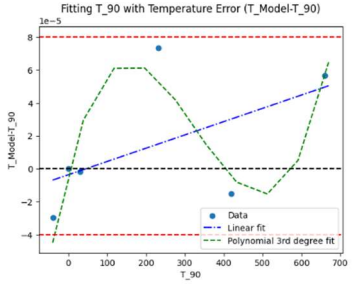


Figure 16. Fitting curve of error propagations against T_{90}

CONCLUSION

Strong, effective and accurate pythonic algorithm has buil t that used in calibration of PRT with limited error impact on measurement uncertainty. The interpolation error that a rises when converting resistance values to temperature usi ng ITS-90 equation has validated within the model.
However, discrepancies in temperature values obtained th at can impact measurement reliability has handled into the uncertainty budget. Maximum difference between mode a nd ITS-90 was found to be 173 µK and minimum one wa s equivalent to 3 µK. Conversion constants from the mod el improve the uncertainty from 0.1 mK to 0.06 mK. The model is statistically significant with p-value = 0.999999

REFERENCES

1. H. Preston-Thomas, "The International Temperature Scale of 1990 (ITS-90)," Metrologia , vol. 27, pp. 3–10, 1990.
2. F. M. Patan and Alper, "The Effect of the Conversion Coefficients of Platinum-Based Resistance Thermometers on the Uncertainty Estimation," International Journal of Thermodynamics, 2023. [Online]. Available: http://doi.org/10.5541/ijot.1220322
3. A. El Matarawy, "Comparison of the realization of water triple point metallic cell through its preparation techniques in new modified adiabatic calorimeter at NIS-Egypt," J. Therm. Anal. Calorim. , 2019. [Online]. Available: https://doi.org/10.1007/s10973-018-7804-8
4. E. C. Jeong and J. Hong Sang, "Machine learning-based virtual metrology on film thickness in amorphous carbon layer deposition process," 2012. [Online]. Available: https://doi.org/10.1016/j.measen.2021.100046
5. S. Rab et al., "Digital Avatar of Metrology," MAPAN, 2023. [Online]. Available: https://doi.org/10.1007/s12647-023-00641-1
6. J. Zhao et al., "A Method for Measuring Tube Metal Temperature of Ethylene Cracking Furnace Tubes Based on Machine Learning and Neural Network," IEEE Access, 2019. [Online]. Available: https://doi.org/10.1109/ACCESS.2019.2950419
7. T. Ren et al., "Machine learning applied to retrieval of temperature and concentration distributions from infrared emission measurements," Applied Energy , 2019. [Online]. Available: https://doi.org/10.1016/j.apenergy.2019.113448
8. M. W et al., "A novel approach to using artificial intelligence in coordinate metrology including nano scale," Measurement , 2023. [Online]. Available: https://doi.org/10.1016/j.measurement.2023.113051
9. S. E et al., "Metrology for the digital age," Measurement: Sensors , 2021. [Online]. Available: https://doi.org/10.1016/j.measen.2021.100232
10. T. D et al., "Uncertainty-aware data pipeline of calibrated MEMS sensors used for machine learning," Measurement: Sensors , 2022. [Online]. Available: https://doi.org/10.1016/j.measen.2022.100376
11. A. El-Matarawy, "New temperature controlling mechanism for thermal metrology regulation at NIS-Egypt," J. Therm. Anal. Calorim. , 2017. [Online]. Available: https://doi.org/10.1007/s10973-017-6340-2
12. A. El Matarawy and Eman El Dien, "Precise temperature controlling algorithm for metrological adiabatic calorimeters based on proportional integration (α) thermal energy," Journal of Thermal Analysis and Calorimetry. [Online]. Available: https://doi.org/10.1007/s10973-021-10806-2
13. Jupyter Documentation, 2023. [Online]. Available: https://docs.jupyter.org/en/latest/
14. NumPy Documentation, 2023. [Online]. Available: https://numpy.org/news/
15. Pandas Documentation, 2023. [Online]. Available: https://pandas.pydata.org/docs/
16. Matplotlib Documentation, 2023. [Online]. Available: https://matplotlib.org/
17. Plotly Documentation, 2023. [Online]. Available: https://plotly.com/
18. SciPy Documentation, 2023. [Online]. Available: https://scipy.org/
19. Seaborn Documentation, 2023. [Online]. Available: https://seaborn.pydata.org/
20. statsmodels Documentation , 2023. [Online]. Available: https://www.statsmodels.org/stable/index.html
21. *scikit-learn Documentation, 2023. [Online]. Available: https://scikit-learn.org/stable/*
22. *P. Rosenkranz, "Uncertainty Propagation for Platinum Resistance Thermometers Calibrated According to ITS-90," Int. J. Thermophys , vol. 32, pp. 106-119, 2011.*

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Sensor Fault Diagnosis Using Spectral Principal Component Analysis and CNN Deep Learning

MOU Jianqiang[1], CUI Shan [1]

*[1]National Metrology Centre, Agency for Science, Technology and Research (A*STAR), Singapore*
E-mail address: mou_jianqiang@a-star.edu.sg, cui_shan@a-star.edu.sg

***Abstract* –A data driven methodology for sensor fault diagnosis in sensor network using principal component analysis (PCA) of coherence spectrum and convolutional neural network (CNN) deep learning is proposed. The methodology was evaluated with the measurement data of a sensor network for ambient relative humidity (RH) monitoring of a chemical laboratory. The results demonstrated accuracy up to 99% for sensor fault diagnosis in the sensor network functioning across a large spectrum of frequencies for environmental monitoring.**

Keywords: Sensor fault diagnosis, sensor network, coherence spectrum, principal component analysis (PCA), convolutional neural network (CNN), deep learning.

I. INTRODUCTION

Sensor fault diagnosis plays an important role for measurement data quality assurance (DQA) in environmental monitoring [1]. In recent years, data-driven algorithms for sensor fault diagnosis and identification have been extensively investigated [2].

In this the paper, a data driven methodology for sensor fault diagnosis in sensor network using principal component analysis (PCA) of coherence spectrum and convolutional neural network (CNN) deep learning traceable to the reference standard was proposed.

The proposed methodology was evaluated with the measurement data of a sensor network for ambient relative humidity (RH) monitoring of a chemical laboratory. The accuracy of the methodology for sensor fault diagnosis in the sensor network was analyzed, and the focus of future work was discussed.

II. METHODOLOGY AND PROCEDURE

The proposed methodology for sensor fault diagnosis uses PCA of coherence spectrum and CNN deep learning and has measurement traceability to the reference standard. Specifically, the methodology is applied to sensor fault diagnosis in sensor network functioning across a large spectrum of frequencies for environmental monitoring.

The methodology is developed by improving our previous work on sensor fault diagnosis in a sensor network for structure health monitoring excited by a single frequency force [3]. It consists of the steps as follows.

2.1 Sensor network configuration

Configure the sensor network as consisting of one pilot sensor x_p and other m sensors x_j *(j = 1, 2, …, m)*. The pilot sensor x_p is defined for disseminating traceability in the sensor network, which is accessible for physical calibration by the reference standard.

2.2 Sensor baseline coherence calculation

At initial stage of the environmental monitoring process, collect time series measurement baseline data under the healthy condition of the sensors. Calculate the coherence spectrum between the pilot sensor and all the other sensors in the sensor network with the baseline time series measurement data. Identify the dominant frequencies ω_k *(k = 1,2,…,l)* of the coherence spectrum.

For time series measurement data $x_p(t)$ and $x_j(t)$ *(j = 1, 2, …, m)* for the pilot x_p and the sensor x_j *(j = 1,2, …, m)* respectively, the coherence spectrum $ch_j(\omega_k)$ between the $x_p(t)$ and each $x_j(t)$ is calculated as:

$$ch_j(\omega_k) = \frac{|S_{pj}(\omega_k)|^2}{S_{pp}(\omega_k)S_{jj}(\omega_k)} \quad (1)$$

Where, the $ch_j(\omega_k)$ is the coherence (magnitude-squared coherence) between the $x_p(t)$ and $x_j(t)$ at the dominant frequencies ω_k *(k = 1,2,…,l)* , the $S_{pj}(\omega_k)$ is the cross-spectral density (CSD), and the $S_{pp}(\omega_k)$ and $S_{jj}(\omega_k)$ are the power spectral density (PSD) respectively [3].

For the indoor ambient environment regulated by an air handling unit (AHU) operated as a linear or piecewise linear fluid dynamic system, the coherence $ch_j(\omega_k)$ between the pilot sensor x_p and the sensor x_j *(j = 1,2,… ,m)* at each dominant frequency ω_k *(k = 1,2,…,l)* will be calculated as close to 1 [3].

When sensor faults occurred in the sensor network, the linearity of the system will be jeopardized and the coherence $ch_j(\omega_k)$ will be degraded.

2.3 Coherence baseline matrix formation

The coherence baseline matrix $C \in R^{n\times(m\times l)}$ is formed by n samples of the coherences for m sensors x_j *(j =*

1,2,…,m) in the sensor network at the *l* dominant frequencies, including the measurement uncertainties.

The type A uncertainty of coherence $ch_j(\omega_k)$ *(j = 1,2,…,m, k =1,2,…,l)* is reflected by the n samples of the coherence spectrum $ch_j(\omega_k)$. The synthetic type B uncertainties added to the matrix $C \in R^{n\times(m\times l)}$ is derived by the Monte Carlo method, following a procedure published in our previous work [4] developed according to the guideline specified in the BIPM JCGM 101:2008 [5].

Normalise the coherence baseline matrix $C \in R^{n\times(m\times l)}$ as the matrix $C_x \in R^{n\times(m\times l)}$, and rescale the coherence for each sensor x_j *(j = 1,2,…,m)* to have zero mean and unity variance.

2.4 Covariance matrix formation and singular value decomposition (SVD)

The covariance matrix C_{cx} of the matrix C_x is formed as:

$$C_{cx} = \frac{1}{n-1} C_x^T C_x \qquad (2)$$

Perform the singular value decomposition to decompose the covariance matrix as [6]:

$$C_{cx} = P \Lambda P^T \qquad (3)$$

and

$$C_{cx} P = P \Lambda \qquad (4)$$

Where, $P = [p_1\, p_2 \ldots p_{(m\times l)}] \in R^{(m\times l)\times(m\times l)}$, $\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_{(m\times l)})$, P^T is the transpose of the P.

The $(\lambda_1, \lambda_2, \ldots, \lambda_{(m\times l)})$ are eigenvalues of the C_{cx}. The columns of matrix P are defined as the principal components (PCs) of the baseline dataset C_x, corresponding to the eigenvalues sorted in descending order.

2.5 Principal components transformation

The relative cumulative contribution rate (*CCR*) represents the selected first *r* principal component (PC) modes contributed to the cumulative variance of the dataset C_x, which is defined as [6]:

$$CCR = \sum_{i=1}^{r} \lambda_i \,/ \sum_{i=1}^{(m\times l)} \lambda_i \qquad (5)$$

Taking the *r* value corresponding to $CCR \geq 95\%$, the first *r* principal components $P_r = (p_1, p_2, \ldots, p_r)$ is selected for principal components transformation of the C_x

Project the dataset C_x into the principal components subspace $P_r = (p_1, p_2, \ldots, p_r)$ to calculate the principal components score matrix T:

$$T = C_x P_r \qquad (6)$$

The matrix C_x is then decomposed as:

$$C_x = T P_r^T + E \qquad (7)$$

$$E = C_x (I - P_r P_r^T) \qquad (8)$$

Where, the TP_r^T is the part of the coherence dataset C_x transformed into the principal components subspace P_r, and E is the PCA residual.

2.6 Principal components projection

Collect time series measurement training data at later stage during the environmental monitoring processes, including the sensor fault introduced.

Following the operations specified at step 2.2 and step 2.3, construct the coherence training matrixes and normalize as $C_{xs} \in R^{s\times(m\times l)}$ for the sensor network. Implement principal components projection using the equation (6), to calculate principal components score matrixes $T_s \in R^{s\times r}$ of the coherence training datasets [3].

2.7 CNN classification by deep learning

Train the convolutional neural network (CNN) classification model by deep learning using the images which are converted from the principal components score matrixes $T_s \in R^{s\times r}$ of the training datasets and categorially labelled with different fault modes.

For new measurement data series, calculate the principal components score matrixes $T_s \in R^{s\times r}$ of the coherence spectrum datasets and convert them into images. Apply the trained CNN model to the images for diagnosis of the sensor fault in the sensor network.

III. MEASUREMENTS AND SPECTRAL PRINCIPAL COMPONENTS ANALYSIS

A sensor network for ambient relative humidity (RH) monitoring of a chemical laboratory is used for evaluation of the proposed methodology and procedure. The ambient condition in the chemical laboratory is regulated by an air handling unit (AHU). The sensor network consists of 7 high accuracy GREYSTONE humidity sensors, marked as *sRHi (i = 1,2…,7)*. Part of the chemical laboratory and the humidity sensor network is displayed in Figure 1. The humidity sensor *sRH1* is configured as the pilot sensor in the sensor network.



Fig. 1. Part of the chemical laboratory and the humidity sensor network.

The time domain measurement data acquisition for the humidity sensor network was carried out from 7.30 PM to 7.30 AM next day over one month under the normal operation of the AHU, with a sampling frequency 0.1Hz, as displayed in Figure 2.

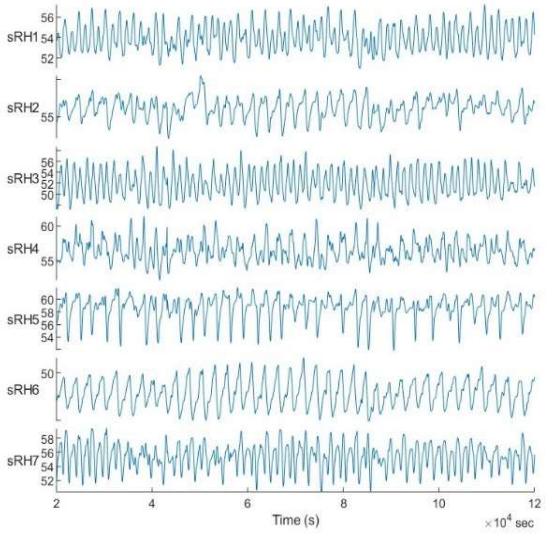


Fig. 2. Time domain measurement data of the RH sensor network.

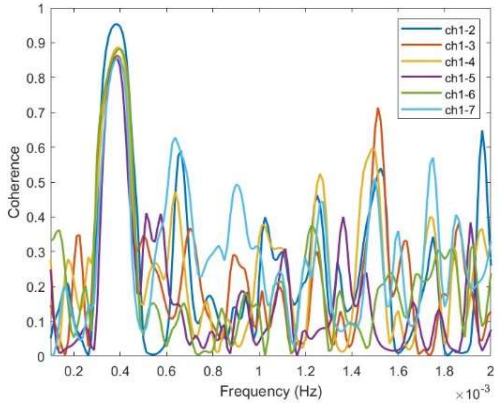


Fig. 3. Coherence spectrum w.r.t the pilot sensor of the baseline measurement data.

The coherence spectrum with respect to (w.r.t) the pilot sensor of the baseline time series measurement data was calculated, as displayed in Figure 3. The dominant frequencies of the coherence spectrum were identified as 13 frequencies in the frequency range from 0.2e-3 Hz to 0.5e-3 Hz, corresponding to the operational frequencies of the air handling unit (AHU). The coherences at the dominant frequencies were calculated and the coherence baseline dataset was constructed, at the initial stage of the testing process without sensor fault in the sensor network. Then the principal components analysis (PCA) on the normalized coherence baseline matrix $C_x \in R^{6000\times(6\times13)}$ for 6000 repeated measurements of the coherence spectrum at the dominant frequencies of the sensor network was performed. As displayed in Figure 4, the first *36* principal components were selected for the PCA model, corresponding to the *CCR* $\geq$*95%.*

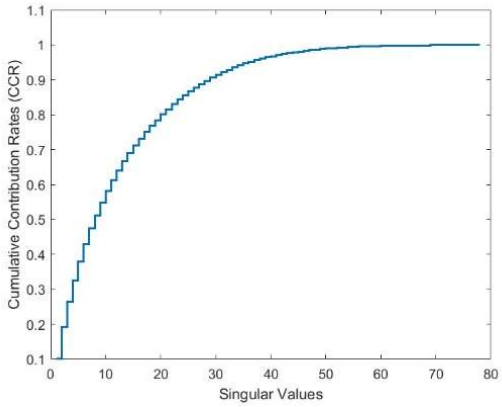
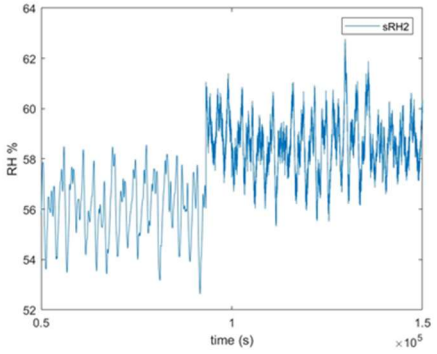
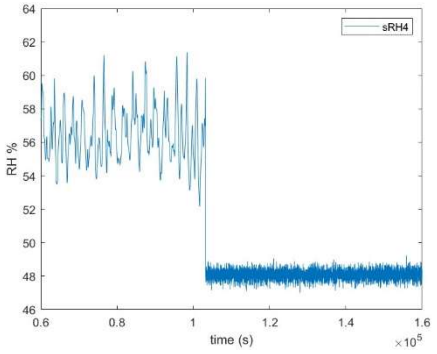


Fig. 4. The CCR contributed to the cumulative variance of the coherence baseline dataset.

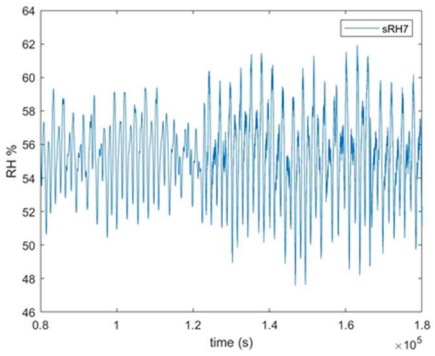
Sensor fault was introduced randomly at the later stage of measurement data acquisition in the evaluation process. Specifically, the sensor fault modes [2, 3] such as senor bias, sensor complete failure and sensor gain were introduced in the time domain measurement data acquired for the sensors sRH2, sRH4 and sRH7 respectively, as illustrated in Figure 5.



a) Sensor sRH2 bias fault.



b) Sensor sRH4 complete failure.



c) Sensor sRH7 gain fault.

Fig. 5. Fault modes of the relative humidity sensors.

Figure 6 displays the principal components scores of the coherence training datasets at the principal component PC1, PC2 and PC3 subspace. It is found that the PCA scores of coherence dataset of the healthy sensors were partially mixed with those for the faulty sensors. To diagnose the sensor fault in the sensor network using the PCA scores of the coherence spectrum at the dominant frequencies, a CNN classification model developed by deep learning is deployed.

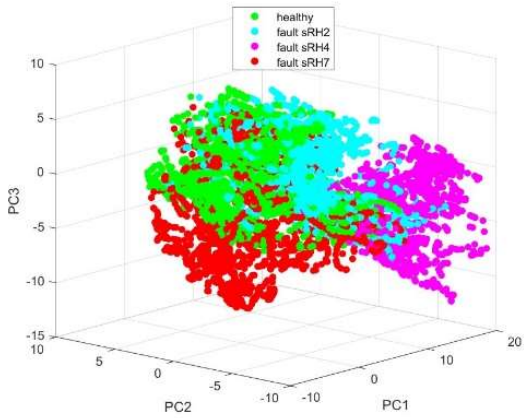


Fig. 6. The PCA scores of coherence training dataset at PC1, PC2 and PC3 subspace.

IV. CNN DEEP LEARNING

A convolutional neural network (CNN) classification model [3] is constructed, as displayed in Figure 7, for sensor fault diagnosis and classification.

The CNN model is trained by deep learning [7] using the images converted from the principal components score matrixes of the coherence spectrum dataset.

Each coherences spectrum dataset is constructed with s = 40 measurements of the coherence spectrum with respect to the pilot sensor, by a time step 120 seconds. The principal components score matrixes $T_s \in R^{40\times 36}$ was then calculated by projecting the normalized coherence spectrum dataset to the principal component subspace truncated at rank r = 36.

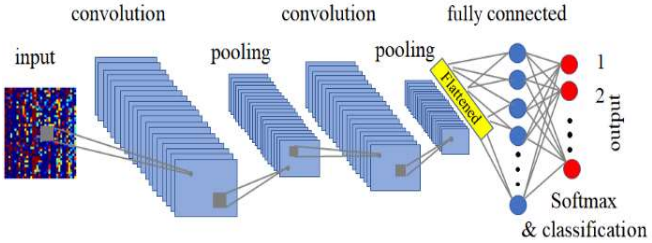
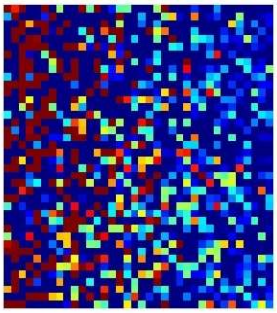
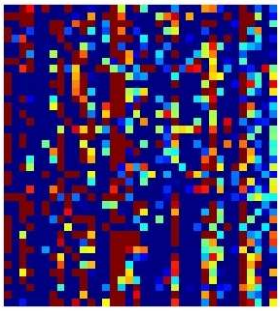


Fig.7. The architecture of the CNN model for sensor fault diagnosis and classification.

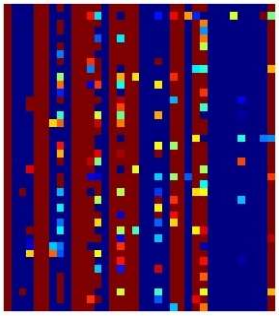
The principal component score matrix $T_s \in R^{40\times 36}$ of the coherence spectrum was converted into a 40×36 pixels image and categorically labelled as input to train the CNN model by deep learning, as displayed in Figure 8.



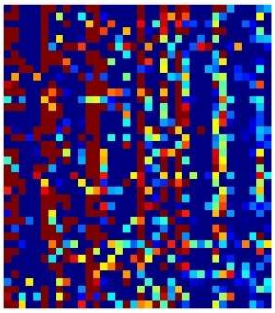
a) Healthy sensor.



b) sRH2 bias fault.



c) sRH4 complete failure.



d) sRH7 gain fault.

Fig. 8. The images converted from the principal components score matrixes of coherence spectrum.

The convolutional neural network (CNN) model was trained by deep learning using the images converted from the principal components score matrixes, which were divided into the training, validating and testing dataset proportionally by 50 %, 10 % and 40 % respectively.

Figure 9 displays the progress of the classification accuracy of the CNN model in the deep learning process using the images of the training dataset and validating dataset. The CNN model was finalized at the classification accuracy of 99.06% in the deep learning process.

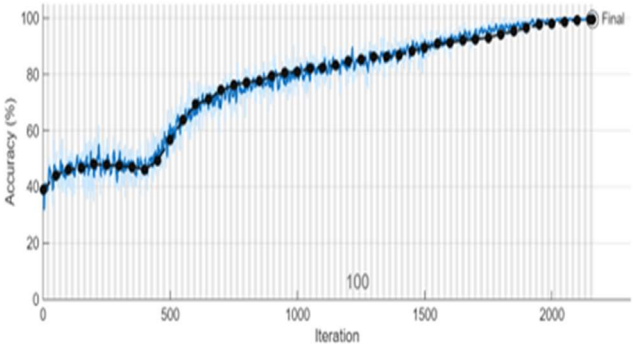


Fig. 9. The progress of the classification accuracy in CNN deep learning

The accuracy of the CNN model was further examined using the images of the testing dataset, by comparing the CNN predicated results with the true class of the sensor status in the sensor network. Table 1 displays the confusion chart of the CNN model trained by deep learning for sensor fault classification examined using the images of the testing dataset, where the classes from 1 to 4 represent the healthy sensor, sRH2 bias fault, sRH4 complete failure and sRH7 gain fault in the sensor network respectively. The results demonstrated that the accuracy up to 99% was achieved for sensor fault diagnosis in the sensor network. However, 2.2% of the healthy sensor (class 1) was misclassified as faulty sensor and confused with the bias fault (class 2) and gain fault (class 4), because of the measurement uncertainties of the coherence spectrum and fluctuation of the AHU operational process. The issue will be further addressed in our future work.

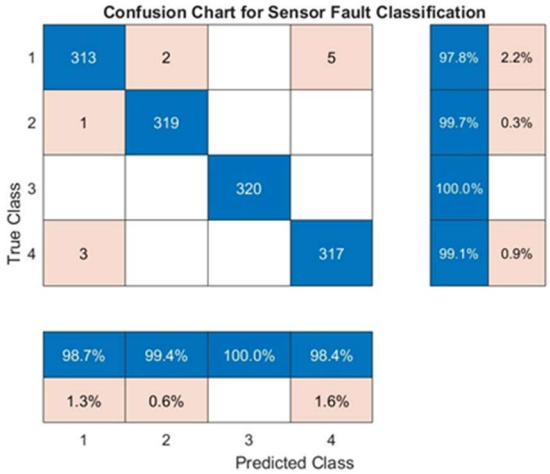


Table 1. Sensor fault classification confusion chart using the CNN model trained by deep learning.

V. CONCLUSIONS

A methodology for sensor fault diagnosis in sensor network using principal component analysis (PCA) of coherence spectrum and convolutional neural network (CNN) deep learning is proposed. The methodology was evaluated with the measurement data of a sensor network for ambient relative humidity (RH) monitoring of a chemical laboratory. The results demonstrated accuracy up to 99% for sensor fault diagnosis in the sensor network. The methodology is applicable for data quality assurance (DQA) in sensor network functioning across a large spectrum of frequencies for environmental monitoring beyond relative humidity, such as for hazardous gas monitoring. Future work will focus on enhancing the accuracy and robustness of the methodology.

ACKNOWLEDGMENTS

This research is supported by A*STAR under its RIE 2025 Urban Solutions and Sustainability domain's Low Carbon Energy Research Phase 2 Programme, Hydrogen and Emerging Technologies Funding Initiative, Directed Hydrogen Programme (Award: U2303D4001). Disclaimer: Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of the A*STAR.

REFERENCES

[1] Y. Liu, N. Tang, "Humidity sensor failure: a problem that should not be neglected", Atmos. Meas. Tech., 7, 3909–3916, 2014.

[2] S. Chauhan, G. Vashishtha and R. Zimroz,

"Analysing Recent Breakthroughs in Fault Diagnosis through Sensor: A Comprehensive Overview", CMES, Vol.141, No.3, pp. 1983-2020, 2024.
[3] J. Mou, S. Cui, C. Cheng, "Sensor fault diagnosis using principal component analysis and convolutional neural network for offshore structural health monitoring", Measurement: Sensors, 28 December 2024, 101465.
[4] R. Chen, S. Cui, D. W. Y. Khoo, B. C. Khoo, "Measurement uncertainty analysis of leak localization in a gas pipeline", Measurement: Sensors, Volume 18, December 2021, 100069.
[5] JCGM101:2008, Evaluation of measurement data — Supplement 1 to the "Guide to the expression of uncertainty in measurement — Propagation of distributions using a Monte Carlo method".
[6] Steven L. Brunton, J. Nathan Kutz, "Data Driven Science & Engineering: Machine Learning, Dynamical Systems, and Control", Cambridge University Press, 2019.
[7] I. Goodfellow, Y. Bengio, and A. Courville (2016), "Deep learning", The MIT press, 2016.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Setup of a distributed sensor network for acquiring environmental data

Alicja Wiora, Józef Wiora

Department of Measurements and Control Systems, Silesian University of Technology,
ul. Akademicka 16, 44-100 Gliwice, Poland
alicja.wiora@polsl.pl; jozef.wiora@polsl.pl

***Abstract* – This study presents the application of an Internet of Things (IoT)-based system for environmental data acquisition in a scientific research setting. The system comprises a network of 12 sensor nodes and a central server. Each node is built around a D1 Mini module, which collects data from an attached sensor and transmits the measurements to the server via web requests. A server-side script processes these requests and stores the data in structured text files. The collected data can be analysed either in real time during the experiment or retrospectively. To ensure durability and reliability in outdoor conditions, all sensor nodes are enclosed in protective housings. This work highlights the practicality, cost-effectiveness, and efficiency of a custom-designed, application-specific IoT measurement system, demonstrating its suitability for rapid deployment in environmental monitoring applications.**

***Keywords* – IoT, ESP8266, particulate matter, CO2**

I. INTRODUCTION

Conducting scientific research on environmental parameters requires collecting data from many sensors, often located in different places. This necessitates the design and construction of a dedicated measurement system. Such a system should be low-cost (depending on available budget), scalable and quick to deploy. This concept has been applied to monitor particulate matter (PM) in the air. The resulting data formed the basis for the analysis described in [1]. The same concept is further continued in a study on monitoring carbon dioxide (CO_2) concentrations.

The Internet of Things (IoT) is a concept in which web technology allows devices to communicate with each other. Smart sensors can additionally perform data preprocessing [2]. This makes the measurement system more advanced and reliable, but also relatively expensive. Another important aspect in wireless sensor networks is the power consumption of the sensor, microcontroller and communication module as well. It is especially important with battery-powered systems [3]. Parallel to the advanced solutions, low-cost sensor networks are being developed. Their low cost drives popularity and contributes data, especially environmental, to society [4].

The ESP8266 microcontroller is currently often applied to collect data from digitally interfaced sensors, like DHT11, for humidity and temperature measurement. It is well known for its affordability for IoT solutions. Its integrated Wi-Fi module allows easy storage of data in the cloud [5]. Also in greenhouse effect monitoring, DHT11 and ESP8266 were used, supported by an MG135 CO_2 concentration sensor [6]. More advanced solutions for monitoring CO_2 and PM utilize the ESP32 microcontroller with FreeRTOS operating system, working with SCD30 infrared sensor and HM3301 scattered light sensor, respectively [7].

The solution used for designing the measurement system must always be adequate to the goals within the set requirements. We observe the increase in complexity, which results in efficiency improvement, cost savings of materials, equipment, and labour, as well as improved safety and quality [8]. It is, in general, true for mass production, especially when produced by teams of experts. Applying advanced solutions for simple and unit tasks unnecessarily increases the project costs, both in equipment and labour parts. When complicated or unpopular approaches are applied, the staff needed to design, run, maintain, and service them is much more expensive. On the other hand, off-the-shelf solutions can be supervised by engineers after short training, significantly reducing the project costs.

The approach of designing the system to allow gathering data for scientific applications differs from those dedicated to mass production. Whereas in the industrial approach, the prototypes are built to increase technological readiness levels and finally create a product ready to sell [9], in academia, the prototype is a target device or system. On the other hand, the academic prototype must still meet the requirements for operating in the target environment, like resistance to rain and dust.

This work focuses on the development and implementation of a sensor network designed for environmental monitoring within a research context. The primary objective is to demonstrate the realisation of a measurement system capable of acquiring environmental data through an IoT approach. A key consideration in the design process was minimising labour and setup time

across mechanical, electronic, and software components. To achieve this, the system was built using widely available, cost-effective components. Popular and low-cost solutions were deliberately selected to ensure affordability, ease of deployment, and replicability.

II. ELEMENTS OF THE NETWORK

The research has been conducted in an area of about 50 m x 50 m covered by Wi-Fi with access to 230 V AC electricity. Installing cables with sockets proved more cost-effective and less time-consuming than designing and implementing battery or solar/wind power systems. Continuous measurements of changes of environmental parameters were required across different locations 24 hours a day, over several months. This necessity led to the development of a system (Fig. 1) with a typical architecture [3], containing nodes with sensors, microcontrollers, Wi-Fi modules and power adapters allowing powering from the 230 V network. The nodes connect to a server where the data is stored. Further data analysis was conducted on a personal computer.

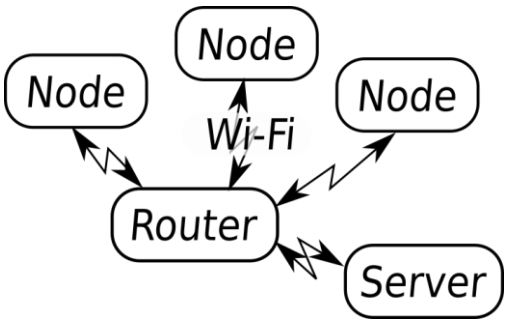


Fig. 1. Schematic overview of the measurement network

A. Nodes

As part of the project, we constructed 12 identical sensor nodes, each following the design shown in Fig. 2. It consists of (1) the sensor (Plantower PMS7003 for PM and Sensirion SCD30 for CO2), (2) a Wemos D1 mini module with ESP8266 microcontroller integrated with Wi-Fi and (3) a 230 V AC to 5 V DC power adapter. All components are enclosed in a 3D-printed PET-G housing.

The choice of sensor type is based not only on objective indicators such as measurement range, accuracy, long-term stability, communication standard, and price, but also on pragmatic ones such as local market availability and experience in use. Both sensors used had a UART serial interface that could be easily connected to a microcontroller. The price for the applied PMS7003 is about 20 EUR, while for SCD30 is 67 EUR on the Polish market.

The variety of microcontrollers is huge, so the choice is

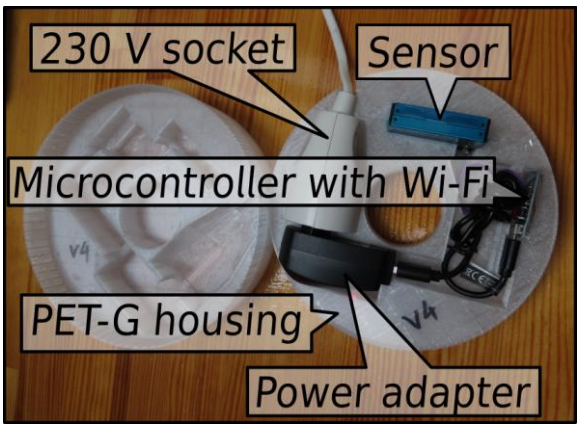


Fig. 2. Internal view of an node for PM measurements.

even more subjective. The D1 Mini module with ESP8266 is a low-cost (ca. 3 EUR) solution that integrates the appropriate computing power for the sensor application with the Wi-Fi communication stack. The possibility of programming using the Arduino IDE was also an advantage.

Owing to the availability of 230 V AC at the research plot, we abandoned a solution with battery power management. The 5 V power adapter selected was among the most affordable available, with a rated current of 1 A. The initial test showed it is suitable for this application. During the few months of experiments, it proved to be the least reliable part—some of these used in the nodes stopped working, terminating data acquisition. However, sourcing low-cost power adapters with reliable long-term performance remains a challenge.

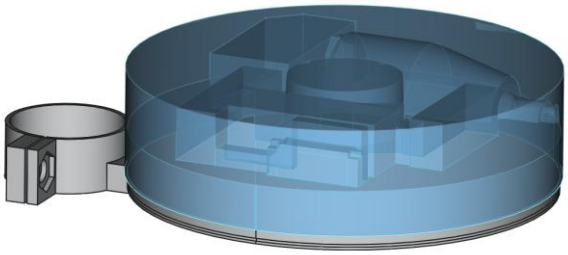


Fig. 3. The housing designed in FreeCAD.

The 3D-printed housing (Fig. 3) was designed in FreeCAD and printed on a Prusa MK3 3D printer using PET-G filament. The housing serves a mechanical construction for all node components. It consists of two parts. The elements can be placed in the niches of the bottom part. The upper part protects against rain and pollutants. An air inlet at the bottom ensures airflow to the sensor. The node can be mounted on a pole using an arm and an appropriate clamp. The housing performed reliably

during tests that took place both in winter at freezing temperatures and in summer in full sunlight.

Prior to deployment, all nodes were co-located in a controlled indoor environment for a few days. During this time, the data were acquired in order to compare their values. After the period, offset and gain errors were established to standardize sensor readings. Due to the unavailability of a dedicated reference sensor for PM, the data were fitted to the data provided by state monitoring stations [1]. However, the primary objective was to observe the spatial and temporal variations rather than to measure the values accurately. For CO2, the reference meter was TESTO 440 with 0554 1111 probe.

B. ESP8266 D1 Mini firmware

The ESP8266 firmware for all nodes is the same. No configuration is required. It uses the web technology approach typical for IoT. It consists of (1) a connection to the local Wi-Fi network, (2) reading the Media Access Control (MAC) address of the integrated Wi-Fi module and (3) an infinite loop. Within the loop, data from the sensor is read via the serial port and stored in the string variable DATA, which contains hexadecimal values encoded in ASCII format. For the PMS7003 sensor, DATA represents the raw data frame directly output by the sensor. In the case of the SCD30 sensor, DATA contains the content of the Modbus response frame, received after issuing the 'Read Measurement' command. Then, the firmware sends a GET request to the server using the following address and parameters:

server/index.php?mac=*MAC*&data=*DATA*

The request is sent immediately after new data is delivered from the sensor. The loop executes without intentional delays, minimizing latency. There is also no preliminary data analysis or correction of the known errors typical for smart sensors, to avoid the need for individual node configuration.

The *server* name or IP address had to be decided when the network was designed. Additionally, the firmware includes a simple web server. Whenever the node is requested, the current sensor readings are provided. This simple solution allowed for checking the sensor state and the operational status of the node. This part of the web server is not mandatory for the correct system operation.

The Arduino IDE environment was used to code and program the microcontrollers. It possesses libraries for the sensors and supports the D1 Mini module.

C. PHP server code

The PHP server is hosted on a Raspberry Pi 3. Only one short file is needed, named index.php. Its code is run if a node makes the request. This PHP script extracts the MAC and DATA from the request, then checks if the DATA contains the correct checksum. If no, such a line is not processed as it is considered unreliable. Otherwise, the hexadecimal numbers placed in the string are converted to decimal equivalents.

After each execution of this script, a text file is opened. The filename is constructed by concatenating the MAC with the current date. This way, each sensor has its data stored in a separate file, keeping individual files manageable in size. In the file, a line is appended to the end. The line starts with the current server date and time. A minor latency between measurement and processing on the server does not affect the analysis due to the low-frequency nature of the signal.

The remainder of the line contains the data obtained from the sensor stored in human-readable decimal format, separated by the space character. The data is stored in a consistent order, matching the format provided by the sensor. All values are natural numbers, and the units of measurement are defined by the sensor manufacturer.

The text files are accessible in real-time facilitating monitoring and debugging during data acquisition. Further processing with mathematical tools like Excel, MATLAB or Python is not complicated as well.

The firmware written for the microcontrollers, together with this PHP script, enables scalability. The number of nodes can be increased without any modifications. Each node has a unique identifier, which is the MAC address of the Wi-Fi module. The identifier serves as the name of the file where the environmental data is stored. The system works well for 12 nodes, making it suitable for the assumed scientific task.

Using a Raspberry Pi as a PHP server seems proved to be an efficient and compact solution for monitoring collected data. It works as a standalone device connected only to the power supply, without a monitor or keyboard. Easy access via SSH or VNC allows us to use the advantages of the Linux operating system. Despite its limited processing power, no impact of VNC operation on data acquisition was observed. Simple Python scripts generate plots visualising the data. File transfer to a Windows computer was done using the SFTP protocol with WinSCP. Application of a desktop computer with a PHP server working under Linux or Windows is also possible but is unlikely to offer significant advantages in this context. Because the server has to work constantly, the power consumption of such a computer is, however, higher. Cloud-based storage is a viable alternative, though it introduces additional cybersecurity aspects and firmware complexity considerations.

D. Router

The router operates with the default Dynamic Host Configuration Protocol (DHCP) settings, which automatically assign IP addresses to connected devices within the network. This default behaviour simplifies the initial setup and is sufficient for general operation. However, manual configuration becomes necessary when

monitoring the operation of individual nodes. In such cases, the user must know the IP addresses of the nodes, which requires reserving IP addresses based on the MAC addresses of Wi-Fi modules. If node monitoring (e.g., via ping) is not performed, no additional configuration is necessary.

E. Python script for data analysis

Data analysis was conducted using Python, specifically the Pandas and Matplotlib libraries, within the Anaconda Navigator environment and Jupyter Notebook. This setup provided a flexible and interactive platform for processing and visualizing the collected data. The Pandas library facilitated efficient data import from structured text files and enabled convenient manipulation of large datasets. Matplotlib was used to generate visualizations, allowing for the extraction of valuable insights and the creation of plots such as those published in [1] or presented in Fig. 4. Offset and gain errors were also corrected during this phase. Performing these corrections at the data analysis stage allowed the same microcontroller firmware to be used across all nodes without requiring individual configuration.

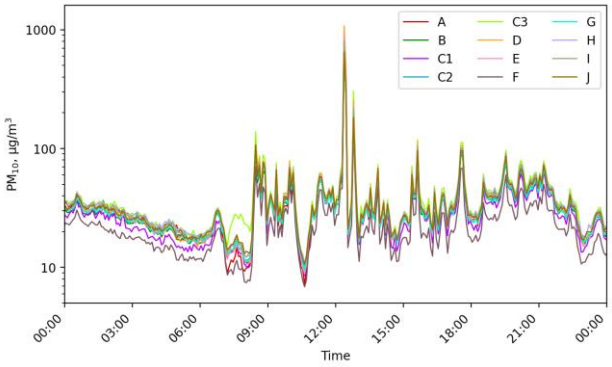


Fig. 4. Example of PM10 changes registed by 12 sensors visualised in Python.

F. Weather monitoring

Additional monitoring of the weather conditions was carried out independently of the main measurement system and was based on a setup adopted from a previous project. For this purpose, a Renkforce WH2600 weather station was used. This station is capable of measuring temperature, pressure, humidity, wind speed and direction, and solar radiation. Data is available through a built-in web interface. A Python script automatically retrieves the data and saves it into a text file every minute. The correlation of weather conditions to PM and CO2 indications is subject to further investigation.

III. CONCLUSION

In contrast to prevailing trends that favour complex and costly solutions, this study presents a straightforward and economical approach to measuring environmental parameters. The measurement network was deployed on a 2 500 m^2 plot with access to electricity and Wi-Fi as part of a research project. By selecting widely available components, we were able to keep the system inexpensive. While the sensor cost could not be reduced, other components were selected for their low cost.

We used an ESP8266 microcontroller to interface the sensor with the network. A Raspberry Pi 3 served as the data server. We implemented communication using HTTP GET requests, transmitting measurement data as URL parameters. This approach allowed us to keep both the microcontroller firmware and the PHP script concise and efficient.

The system operated reliably for several months, initially measuring PM and later acquiring carbon dioxide (CO2) concentrations. During this period, the power adapter was the most unreliable component, as we had not considered its reliability parameters during the design phase. This experience highlights the need to evaluate all system components—not just sensors and processors—when aiming to build robust, long-term monitoring solutions.

ACKNOWLEDGMENTS

This research was funded by EU Funds for Silesia 2021-2027, MONCO2 project No. FESL.10.25- IZ.01-06C9/23-00 “The modern methods of the monitoring of the level and isotopic composition of atmospheric CO2”. (AW); and by the Polish Ministry of Science and Higher Education 02/050/BK-25/0047 (JW).

REFERENCES

[1] A.Wiora, J.Wiora, J.Kasprzyk, “Indication Variability of the Particulate Matter Sensors Dependent on Their Location”, Sensors, vol.24, 2024, article 1683.

[2] S.Eichstädt, A.P.Vedurmudi, M.Gruber, D.Hutzschenreuter, “Fundamental aspects in sensor network metrology” ACTA IMEKO, vol.12, 2023, pp.1–6.

[3] K.Gulati, R.Boddu, D.Kapila, S.Bangare, N.Chandnani, G.Saravanan, “A review paper on wireless sensor network techniques in Internet of Things (IoT)” Materials Today: Proceedings, vol.51, part 1, 2022, pp.161–165.

[4] F.Mao, K.Khamis, S.Krause, J.Clark, D.M.Hannah, “Low-Cost Environmental Sensor Networks: Recent Advances and Future Directions”. Frontiers in Earth Science, vol.7, 2019, article 22.

[5] N.Mitu, V.Vassilev, M.Tabany, “Low Cost, Easy-to-Use, IoT and Cloud-Based Real-Time Environment Monitoring System Using ESP8266 Microcontroller”, International Journal of Internet of Things and Web Services, vol.6, 2021, pp.30–44.

[6] Z.Wan; Y.Song; Z.Cao, “Environment Dynamic

Monitoring and Remote Control of Greenhouse with ESP8266 NodeMCU", IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), 2019, pp.377-382.
[7] D.Pineda-Tobón, A.Espinosa-Bedoya, J.Branch-Bedoya, "Aquality32: A low-cost, open-source air quality monitoring device leveraging the ESP32 and google platform", HardwareX, vol.20, 2024, article e00607.
[8] L.Zhang, Y.Li, Y.Pan, L.Ding, „Advanced informatic technologies for intelligent construction: A review", Engineering Applications of Artiɔcial Intelligence vol. 137, 2024, article 109104.
[9] S.Yfanti, N.Sakkas, „Technology Readiness Levels (TRLs) in the Era of Co-Creation", Applied System Innovation, vol.7, 2024, article 32.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Metrology-Driven Standardization of Sensor Networks in Mining: A RAMI 4.0 Approach to Sustainable and Efficient Ventilation Systems

Armin Shirbazo[1], Binghao Li[2], Seher Ata[3], Hamed Lamei Ramandi[4], Serkan Saydam[5]

[1] MERE, UNSW, Sydney, Australia, a.shirbazo@unsw.edu.au
[2]MERE, UNSW, Sydney, Australia, binghao.li@unsw.edu.au
[3]MERE, UNSW, Sydney, Australia, s.ata@unsw.edu.au
[4] MERE, UNSW, Sydney, Australia, h.lameiramandi@unsw.edu.au
[5] MERE, UNSW, Sydney, Australia, s.saydam@unsw.edu.au

Abstract— The mining sector is undergoing a profound shift as Industry 4.0 technologies—IoT, AI, robotics—reshape operational planning and execution. This study explores the integration of Mine IoT to modernize mining practices, emphasizing metrology-driven advancements in real-time monitoring, predictive maintenance, and autonomous systems. Accurate measurement and standardized data are central to improving efficiency, enhancing safety, and advancing sustainability goals. A smart ventilation case study illustrates how RAMI 4.0, combined with digital metrology, enhances interoperability, enables seamless integration, and supports scalable, adaptable systems for improved energy use, safety, and long-term resilience. Metrological traceability is embedded throughout system layers to support interoperability and long-term performance. RAMI 4.0's structured framework ensures traceable data from calibrated sensors and uncertainty-aware analytics, aiding reliable decision-making and regulatory compliance. The study also highlights the role of standardization in facilitating communication across devices, platforms, and vendors. Achieving these outcomes requires strategic planning, skilled personnel, and cross-sector collaboration.

Keywords— *RAMI 4.0, Mine IoT, Sustainability, Metrology, Standardization*

I. Introduction

The mining industry, historically defined by high resource demands and operational risks, is undergoing a major transformation through Industry 4.0 technologies. Shifting from manual processes and isolated systems, mining operations are increasingly shaped by cyber-physical systems, IoT, AI, robotics, and data analytics. These technologies not only automate processes but also enable predictive maintenance, autonomous functions, and real-time decision-making. Machine learning and intelligent sensors forecast equipment failures, while autonomous haulage systems reduce human exposure to hazardous environments. These innovations drive measurable improvements in efficiency, safety, cost management, and environmental impact reduction. Realizing their full potential, however, depends on more than technological adoption. Seamless integration, reliable data flow, and interoperability across platforms require standardized frameworks and communication protocols. Standardization supports scalability, regulatory compliance, and system resilience, while accurate measurement underpins smart system performance. Metrological traceability ensures that data from calibrated sensors and uncertainty-aware analytics remains reliable, interoperable, and actionable. Traceable digital records enhance decision-making throughout all operational layers. This study explores the application of Industry 4.0 technologies across the mining value chain, emphasizing the foundational roles of standardization and metrology. It presents a framework to help stakeholders implement scalable, transparent, and sustainable mining systems aligned with international best practices.

Ventilation systems in underground mining operations are critical for maintaining air quality, managing heat, and ensuring worker safety. However, traditional systems often lack adaptability and energy efficiency, leading to excessive operational costs and environmental impact [1]. An approach that leverages digital integration—such as the RAMI 4.0 framework—offers potential for optimising ventilation control, improving responsiveness, and supporting sustainability goals [2]. Improving mine ventilation efficiency can greatly cut energy consumption, lower greenhouse gas emissions, and reduce operational costs. Research highlights the importance of measures such as better airflow management, the use of energy-efficient motors, and system-wide upgrades [3]. Incorporating real-time monitoring and intelligent control systems allows ventilation to adjust dynamically, boosting energy savings while also enhancing worker safety [4].

Underground mine ventilation significantly impacts energy consumption and is critical for ensuring safety and stable working conditions. However, current systems are

often rigid, resource-intensive, and lack real-time adaptability. This paper addresses these challenges through a RAMI 4.0-based framework that integrates standardisation, interoperability, and digital metrology principles. Using a case study on smart underground ventilation, it demonstrates how structured sensor networks and verifiable measurements can support intelligent control, reduce energy use, and improve air quality. The approach also offers a scalable model applicable to other safety-critical mining systems.

II. Technological Advancements in Mining

Section II examines the technological advancements in reshaping the mining industry, focusing on innovations that improve efficiency, safety, and sustainability. It highlights the integration of Industry 4.0 technologies, such as IoT, AI, predictive maintenance, and automation, and their impact on modernizing mining operations.

A. IoT and Real-Time Monitoring

IoT devices are revolutionizing mining operations by enabling real-time monitoring of environmental conditions—such as temperature, pressure, gas levels—and equipment health, facilitating predictive maintenance, enhancing safety protocols, and optimizing operational efficiency [5]. Sensors, thermal cameras, and GPS modules embedded in assets send structured data to centralized platforms for real-time analysis, enabling early fault detection, reduced downtime, and improved safety. Personnel tracking and automated alerts further protect workers in hazardous areas. These capabilities support predictive maintenance, automation, and digital twins, advancing efficiency and sustainability [6], [7]. Metrological principles are essential to ensure the accuracy, traceability, and interoperability of sensor data. In smart ventilation systems, calibrated sensors provide time-stamped, uncertainty-aware measurements, supporting consistent environmental monitoring and regulatory compliance.

B. Predictive Maintenance and AI

Artificial intelligence (AI) combined with sensor networks is transforming maintenance practices in mining. Instead of relying on scheduled or reactive maintenance, AI algorithms use real-time and historical sensor data to forecast equipment failures, enabling timely interventions. This predictive approach reduces downtime, cuts maintenance costs, and extends equipment life, while also conserving energy and supporting sustainability goals [8]. Enhanced performance and optimized resource use make predictive maintenance a cornerstone of modern mining operations. Metrological validation supports this transformation by ensuring predictive models draw on sensor data with known accuracy, uncertainty, and traceability for reliable, evidence-based decision-making. This integration strengthens the credibility and impact of AI-driven forecasts in safety-critical environments.

C. Automation and Robotics

Automation technologies—including autonomous haulage systems (AHS), robotic drilling rigs, tele-remote loading systems, autonomous blasthole drills, automated ore sorting equipment, and drone-based inspection platforms—are transforming mining operations. These systems reduce human exposure to toxic gases and extreme conditions, significantly improving safety. Robotic drills increase precision, minimize material waste, and enhance resource efficiency. Operating continuously, automated systems elevate productivity and ore recovery while reducing environmental impact. AI-driven analytics help adapt operations to varying ore qualities, ensuring consistent performance. Fully autonomous mines are emerging as the industry's next milestone, promising enhanced efficiency, safety, and sustainability [9], [10]. Embedding metrological safeguards into these systems—such as remote calibration, traceable measurements, and uncertainty-aware control logic—ensures that automation operates on reliable, verifiable data, supporting safety-critical decisions and long-term system integrity.

III. Standardization and Metrology for Sustainable Digital Mining

The shift toward Industry 4.0 technologies—IoT, AI, robotics—is reshaping mining. Realizing their full potential requires more than just digital tools. Standardization and metrological integration are critical for ensuring interoperability, data quality, and responsible, scalable operations. Together, they support efficiency, traceability, and ethical practices across the mining value chain.

A. Interoperability and Metrology for Scalable, Reliable Mining Systems

Digital transformation depends on systems speaking the same language. Standardized data formats and communication protocols enable seamless interaction among sensors, control systems, and enterprise platforms. This consistency breaks down operational silos, enhances automation, and supports informed, cross-functional decision-making. Standardization also allows successful technologies to scale across sites, simplifying compliance and accelerating innovation. Vendors and partners benefit too, as aligned systems reduce integration costs and friction. A connected digital ecosystem emerges—resilient, agile, and optimized for sustainable performance. Accurate data drives every smart mining process. IoT-enabled sensors monitor conditions like airflow, gas levels, and temperature. To ensure this data is usable, it must follow digital standards and include traceable metadata. Digital calibration certificates and consistent units improve auditability and system trust. Underground environments introduce uncertainty measurement. The conceptual system anticipates this challenge through features like uncertainty tagging and quality-aware processing. These design choices contribute to safer, more reliable automation.

B. Lifecycle Reliability, Sustainability, and Traceability in Smart Mining

Reliable operations require consistent calibration. The system framework includes regular calibration routines—manual or automated, conducted on-site or remotely—with records stored digitally. Embedding calibration into system design improves transparency and ensures alignment with evolving digital testing standards. AI and real-time analytics enhance sustainability. Smart ventilation systems adjust airflow based on sensor input, cutting energy use. Predictive models help recover more resources while reducing waste. Load forecasting and smart grid integration support cleaner, more efficient energy use. These outcomes depend on reliable, interoperable systems. Validated data and shared protocols underpin every algorithm and decision, linking sustainability to system integrity. IoT and blockchain tools bring transparency to sourcing and operations. Material origins, carbon emissions, and water usage are logged and verified in real time. These tools help companies meet Environmental, Social, and Governance (ESG) goals, counter illegal practices, and build trust with regulators and stakeholders. Traceable systems also unlock access to certifications and markets where ethical sourcing is non-negotiable.

C. RAMI 4.0 as an Integration Framework

The Reference Architecture Model for Industry 4.0 (RAMI 4.0), depicted in Fig. 1, offers a powerful framework for structuring standardization and metrological integration. In complex mining environments, RAMI 4.0 organizes system components across hierarchy levels and IT layers, enabling interoperability from sensors to enterprise software [11]. Hierarchy levels represent the structure of manufacturing systems—from field devices and control units to enterprise-wide coordination and inter-company collaboration. The IT layers define how data and processes are managed across this structure, covering everything from physical assets and communication to information processing and business operations. Metrological elements—such as sensor calibration status, measurement uncertainty, and traceability metadata—can be embedded across RAMI 4.0 layers. At the asset level, devices store calibration data; at the information level, measured values are digitally represented; and at the functional layer, validated inputs guide automated decisions. This structure supports scalable, transparent systems that align with both industrial and regulatory standards.

In underground mining, reliable ventilation is essential—not just for maintaining air quality, but for protecting lives. Given the harsh and dynamic nature of these environments, any disruption to the ventilation system can have serious consequences. The RAMI 4.0 framework proposed in this work accounts for such risks by supporting redundancies across the Communication, Integration, and Information layers. Built-in backup systems, alternative data routes, and automatic failover responses help ensure that critical ventilation functions remain active, even if part of the system encounters a fault. This kind of resilience is key to enabling safer, more efficient, and digitally integrated ventilation management in modern mining operations.

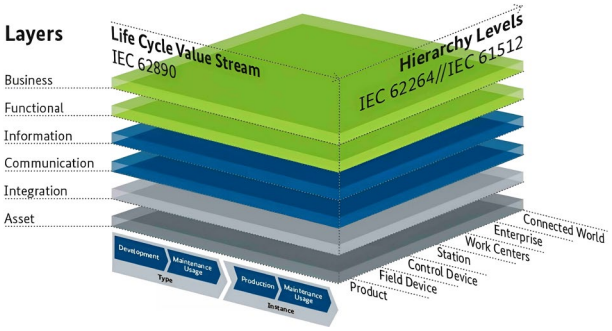


Fig. 1. RAMI 4.0 Layout [12].

IV. Strategic Implementation and Case Study

Successful adoption of Industry 4.0 in mining demands more than technical readiness. Organizational alignment, strategic planning, and workforce development are equally critical. The value of digital transformation stems not only from advanced technologies but from the ability to adapt structures and processes to unlock their full potential.

A. Background and System Description

In underground mining, ventilation is critical for both worker safety and regulatory compliance. Traditional systems, often reliant on manual or semi-automated controls, struggle to adapt to shifting underground conditions, leading to inefficiencies and safety concerns. Smart ventilation systems present a more responsive and energy-efficient alternative. Using sensors, actuators, and programmable logic controllers (PLCs), they monitor airflow, gas levels, temperature, and humidity in real time. Variable frequency drives (VFDs) and motorized dampers dynamically adjust airflow, while Supervisory Control and Data Acquisition (SCADA) platforms provide centralized oversight, enabling rapid responses and reduced downtime. RAMI 4.0 offers a structured approach to standardizing these systems across all layers—from field devices to enterprise software—promoting interoperability and smooth integration with legacy infrastructure. Core system components include:

- Sensors that collect continuous environmental data across the mine.
- Actuators and VFD-equipped fans that regulate airflow based on real-time input.
- PLCs that process data and execute control logic for adaptive ventilation.
- A central monitoring system that visualizes performance, supports control decisions, and activates alarms when hazards arise.

This integrated approach enhances safety, efficiency, and system resilience, making it a foundational element of smarter underground mining operations. The system design anticipates metrology-related challenges—such as sensor drift—which could be managed through embedded calibration protocols and uncertainty-aware analytics, structured via the RAMI 4.0 framework.

B. Setting Clear Objectives

Well-defined goals aligned with business strategy lay the groundwork for effective digital initiatives. Whether targeting cost reduction, safety, efficiency, or sustainability, clear objectives steer investments and reduce fragmentation. Success hinges on measurable KPIs, phased implementation, and iterative learning through pilot projects, which help refine strategies before broader deployment [13], [14]. Automation and AI reshape workforce demands, requiring new skills and adaptive mindsets. Employees must gain data literacy, technological proficiency, and the ability to collaborate across functions [15]. Training supports this evolution, while transparent leadership helps manage concerns about job displacement. Communicating the benefits—such as improved safety, less manual labor, and smarter decision-making—builds trust. Involving staff at all levels and fostering a culture of innovation smooths the transition and builds engagement.

C. Case Study: Smart Underground Ventilation System

In deep-level mining operations, effective ventilation is critical for both safety and operational efficiency. Fluctuating air quality, the accumulation of hazardous gases like methane and CO_2, along with dust and heat, create significant risks. These conditions make consistent and clean airflow essential for workers' safety and compliance with environmental regulations. A smart ventilation system addresses these challenges through a network of sensors that monitor airflow, gas concentrations, temperature, and humidity. PLCs process data to regulate fan speeds and damper positions in real time, ensuring optimal air quality. Large fans with VFDs and motorized dampers precisely control airflow across different mine zones. Local control stations, equipped with human-machine interface (HMI) terminals, enable operators to monitor and adjust the system on-site. Simultaneously, a surface-based SCADA system oversees the entire operation, providing a visual representation of ventilation patterns and enabling both automated and manual control.

The system adapts to varying conditions through different operating modes. During normal operations, it balances airflow and energy efficiency. After blasting, ventilation is increased to clear harmful gases, and in emergencies, the system maximizes output while triggering alerts for evacuation. Additionally, integration with enterprise resource planning (ERP) enables energy tracking, predictive maintenance, and compliance reporting. This smart ventilation setup demonstrates how automation and digital integration enhance safety, efficiency, and environmental responsibility in mining operations. In the proposed ventilation system, sensors are expected to produce measurements that adhere to traceability principles through digital calibration records and standardized data structures. This framework aligns with RAMI 4.0, distributing metrological functions across system layers—from calibration metadata at the asset level to validated input at the functional layer—supporting transparency, reliability, and future standardization.

D. Challenges and Solutions in Underground Ventilation: Leveraging Smart Systems and RAMI 4.0

Effective underground ventilation is critical for worker safety and regulatory compliance. Conditions such as fluctuating air quality, gas buildup, and heat present ongoing challenges. Manual or semi-automated systems often lack the adaptability and integration to manage these risks. Smart ventilation technologies—leveraging sensors, actuators, VFDs, and real-time control—offer more responsive and efficient solutions. Edge-cloud collaboration and feedback loops enhance reaction times and energy use. RAMI 4.0 provides a structured framework for integrating components across all system layers, improving performance in complex underground environments.

V. Discussion and Results

To provide a clearer understanding of how the smart underground ventilation system integrates within the RAMI 4.0 framework, this subsection offers a structured mapping of its components across both the architecture's hierarchy levels and IT layers.

A. RAMI 4.0 Mapping

The alignment of physical devices, control systems, and enterprise tools with the dimensions of RAMI 4.0 offers a structured view of system integration, interoperability, and performance enhancement. Table 1 maps components across hierarchy and IT layers. It is demonstrating practical standardization in a complex mining context. This mapping also shows how metrological principles can be embedded, with calibration metadata handled at the asset layer and uncertainty-aware decision logic integrated at higher layers.

Table 1. The mapping across hierarchy and IT layers

IT Layer	Hierarchy Level	Mapped Component
Asset	Field Device	Sensors with digital calibration and uncertainty metadata
Integration	Control Device	PLCs processing validated, traceable input

Communication	Station	OPC-UA/MQTT protocols transmitting traceable sensor data
Information	Work Centers	SCADA systems aggregating quality-tagged data
Functional	Enterprise	Decision logic incorporating uncertainty and quality status
Business	Connected World	ERP and reporting tools reflecting metrological compliance and traceability

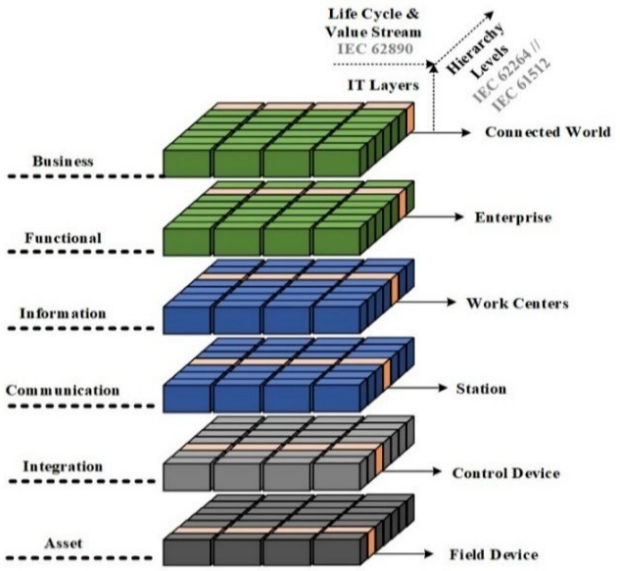


Fig. 2. RAMI 4.0 Mapping

Fig. 2 visualizes how smart ventilation systems align with the RAMI 4.0 framework, with metrological traceability embedded throughout the system lifecycle. From calibrated field sensors to enterprise-level platforms, validated measurements and uncertainty data flow across all layers and hierarchy levels. This structured integration supports real-time, quality-aware decision-making, enhances interoperability, and ensures scalable, reliable, and compliant mining operations.

B. Implementation Benefits

Introducing RAMI 4.0 into underground smart ventilation systems offers tangible improvements in operational efficiency and safety. One key advantage is enhanced interoperability, allowing systems and devices from different vendors to work together seamlessly—a crucial benefit in mining environments with diverse equipment suppliers. Real-time adjustments enable the system to quickly respond to hazards, such as gas buildup or excessive heat, ensuring safe conditions and optimizing energy use. Detailed performance data supports strategic maintenance by helping operators anticipate issues and reduce unplanned downtime. Integrating ventilation operations with higher-level business systems also supports decision-makers, providing accurate, up-to-date information. The system's modularity and scalability make it adaptable, supporting long-term value through easier upgrades as the mine evolves or as new technologies emerge. The proposed integration of metrological validation is expected to improve output consistency and reliability, thereby enhancing safety and efficiency.

C. Challenges and Limitations

Despite these benefits, practical challenges remain. Many mines still rely on legacy equipment that may not integrate easily with modern digital systems, making retrofitting or replacement complex and costly. Data standardization is another obstacle, as older systems often use non-standard formats, hindering integration and limiting interoperability. Additionally, advanced ventilation systems require a different skill set for maintenance. Technicians and engineers need expertise in digital controls and automation, alongside traditional mechanical knowledge. Workforce training is crucial to ensure reliable system operation and minimize inefficiencies caused by technical skill gaps. Ensuring metrological traceability in harsh mining environments presents clear challenges, but future systems may address these through adaptive calibration strategies and standardized digital validation protocols.

VI. CONCLUSION

The adoption of Industry 4.0 technologies represents a transformative shift in mining, enabling smarter, more efficient, and sustainable operations. Through IoT, AI, robotics, and advanced analytics, companies can optimize performance, support real-time decision-making, and reduce environmental impact. Structured frameworks like RAMI 4.0 strengthen this transformation by guiding the integration of technologies across physical, functional, and enterprise layers. The smart underground ventilation case highlights how standardization enhances system responsiveness, interoperability, and long-term scalability—benefits that extend to other critical systems such as dewatering, safety, and environmental monitoring. Realizing the full potential of these innovations requires a robust implementation strategy, supported by workforce development, ethical practices, and a commitment to standardization. Accurate, traceable, and standardized measurement underpins every aspect of digital mining. Integrating digital metrology principles—such as calibration traceability, quantified uncertainty, and sensor verification—into smart systems ensures data reliability, supports safety-critical automation, and facilitates compliance. RAMI 4.0 offers a structured approach for embedding these principles across all system layers, making metrology a cornerstone of future-ready, trustworthy mining operations.

REFERENCES

[1] M. Bascompta, A. M. Castañón, L. Sanmiquel, and J. Oliva, "A GIS-based approach: Influence of the ventilation layout to the environmental conditions in an underground mine," *J Environ Manage*, vol. 182, pp. 525–530, Nov. 2016,

[2] Y. Li, S. Shao, and F. Zhang, "An Analysis of the Multi-Criteria Decision-Making Problem for Distributed Energy Systems," *Energies (Basel)*, vol. 11, no. 9, p. 2453, Sep. 2018,

[3] N. Demirel, "Energy-Efficient Mine Ventilation Practices," 2018, pp. 287–299.

[4] M. Shriwas and C. Pritchard, "Ventilation Monitoring and Control in Mines," *Min Metall Explor*, vol. 37, no. 4, pp. 1015–1021, Aug. 2020,

[5] H. Zhang, B. Li, M. Karimi, S. Saydam, and M. Hassan, "Recent Advancements in IoT Implementation for Environmental, Safety, and Production Monitoring in Underground Mines," *IEEE Internet Things J*, vol. 10, no. 16, pp. 14507–14526, Aug. 2023,

[6] A. Strongrich and A. Alexeenko, "Wireless sensor networks for pharmaceutical lyophilization: Quantification of local gas pressure and temperature in primary drying," *European Journal of Pharmaceutics and Biopharmaceutics*, vol. 169, pp. 52–63, Dec. 2021,

[7] R. Klis and E. N. Chatzi, "Vibration monitoring via spectro-temporal compressive sensing for wireless sensor networks," *Structure and Infrastructure Engineering*, vol. 13, no. 1, pp. 195–209, Jan. 2017,

[8] P. Aqueveque, L. Radrigan, F. Pastene, A. S. Morales, and E. Guerra, "Data-Driven Condition Monitoring of Mining Mobile Machinery in Non-Stationary Operations Using Wireless Accelerometer Sensor Modules," *IEEE Access*, vol. 9, pp. 17365–17381, 2021,

[9] A. Shukla and H. Karki, "Application of robotics in offshore oil and gas industry— A review Part II," *Rob Auton Syst*, vol. 75, pp. 508–524, Jan. 2016,

[10] M. Whatley, A. Ayanambakkam, D. Patterson, B. Farrow, L. Feng, and P. Hewitt, "Machine and Process Automation are Improving Personnel Safety and Drilling Performance," in *SPE Annual Technical Conference and Exhibition*, SPE, Sep. 2022.

[11] A. Shirbazo, B. Li, S. Ata, H. L. Ramandi, and S. Saydam, "A Guideline for the Standardisation of Smart Manufacturing and the Role of RAMI 4.0 in Digitising the Industrial Sector," *IEEE Internet Things J*, pp. 1–1, 2025,

[12] K. Schweichhart, "Reference Architectural Model Industrie 4.0 (RAMI 4.0). An Introduction," 2016.

[13] M. Ryghaug, B. T. Haugland, R. A. Søraa, and T. M. Skjølsvold, "Testing Emergent Technologies in the Arctic," *Science & Technology Studies*, Jan. 2022,

[14] M. Samis and J. Steen, "Financial evaluation of mining innovation pilot projects and the value of information," *Resources Policy*, vol. 69, p. 101848, Dec. 2020,

[15] A. Young and P. Rogers, "A Review of Digital Transformation in Mining," *Min Metall Explor*, vol. 36, no. 4, pp. 683–699, Aug. 2019,

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Infrastructure requirements for metrological distributed sensor networks

Martin Koval[1], Gertjan Kok[2], Maximilian Gruber[3], Shahin Tabandeh[4], Martin Staněk[1]

[1] Czech Metrology Institute, Okružní 31, 63800 Brno, Czech Republic, martin.koval@cmi.gov.cz, martin.stanek@cmi.gov.cz
[2] Van Swinden Laboratorium (VSL), Thijsseweg 11, 2629 JA, the Netherlands, gkok@vsl.nl
[3] Physikalisch-Technische Bundesanstalt (PTB), Abbestraße 2-12, D-10587, Berlin, Germany, maximilian.gruber@ptb.de
[4] VTT MIKES, Tekniikantie 1, 02150 Espoo, Finland, shahin.tabandeh@vtt.fi

***Abstract* – Distributed sensor networks (DSNs) are increasingly being deployed in various systems, indicating a more active implementation of digitalisation in the field of metrology. DSNs bring a wide range of benefits in the management of processes, where we are now looking not only at real-time monitoring, but also at advanced process optimization based on efficiently acquired data, support in the creation of digital twins, the prediction of future states such as calibration or service maintenance, the usage of artificial intelligence, and much more. For DSNs to operate efficiently and reliably, it is essential to properly establish the network infrastructure and identify the associated requirements at the design phase, including metrological aspects. This paper proposes a structured set of infrastructure requirements and metrological design guidelines that enable long-term reliability, traceability, and data quality in DSNs, and discusses practical approaches to sensor architecture, network topology, calibration strategies, and QA/QC (Quality Assurance/ Quality Control) planning tailored to metrological applications.**

I. INTRODUCTION

Distributed Sensor Network (DSN) is a modern concept where individual sensors are distributed in space and interconnected to efficiently collect, transmit, and process data (measurement data and metadata). This concept is becoming increasingly important in the context of digitalization, particularly in the field of metrology, where it enables more efficient real-time monitoring and control of processes. The deployment of DSNs opens up new possibilities not only for tracking measurement data, but also for data processing, process optimization, digital twins, predictive analytics, and integration of artificial intelligence [1, 2]. DSNs can enable the creation of dynamic models that can respond to current conditions, predict the development of monitored parameters, and positively influence system control. The concept of digital twins is one example where DSNs play a key role. Data collected from distributed sensors allow for constantly updating digital models, improving control and offering advanced opportunities for optimization [1].

DSNs can be found in various applications from Industry 4.0, where they support automation and intelligent control of production lines, through the energy sector (monitoring of distribution systems, smart grids), healthcare (medical devices complex system or wearable devices), to smart cities and environmental monitoring (e.g., air, water, and soil quality monitoring) [3]. In each of these areas, DSNs contribute to improved efficiency, reliability, and sustainability. However, the effective deployment and operation of a DSN is not guaranteed without a properly designed network architecture and clearly defined infrastructure requirements, otherwise, key aspects such as DSN lifecycle and adaptability can be significantly compromised. Factors such as network topology, communication cybersecurity, and scalability as well as metrological considerations with respect to sensor uncertainty, initial calibration and re-calibration strategy greatly influence not only real-time performance, but also the long-term benefits of DSN [3, 4]. The aim of this paper is to analyse key aspects of DSN infrastructure design, highlight the influence of the DSN topology on its performance, and define fundamental requirements necessary for the long-term sustainability and effective operation of a metrological DSN.

II. TERMINOLOGY AND DEFINITIONS

To properly understand the specifics of DSNs, it is necessary to define basic important concepts and the relationships between them. This section only includes terms that are necessary for a better understanding of the context of the article.

DSN architecture: The architecture of a DSN is the overall design, how infrastructure components interact. This is done by designing the logical as well as the physical relationships between these components, specifying access methods and choosing communication protocols.

DSN topology: The topology of a DSN is the graph that represents how measuring devices and computation nodes are linked, either physically or logically.

DSN infrastructure: The infrastructure of a DSN covers the components that are required to develop, operate, monitor, control and support the DSN. These are mainly physical components (i.e., measuring devices, compute nodes, power supplies, networking hardware), but also the required software and network applications.

Node: A node in DSN is a basic functional unit capable of performing measurements, aggregation data, processing data, and communicating with other nodes or a central system.[13]

Sensor: For the purposes of this article, a sensor is considered a physical device that detects, measures, and responds to a specific physical quantity and converts it into a usable signal, typically electrical or digital. In the context of DSNs, a sensor can also be an intelligent sensing unit. This includes additional functions such as signal preprocessing, generation of metadata, bidirectional communication, or embedded firmware, among others.

III. INFLUENCE OF NETWORK TOPOLOGY ON DSN INFRASTRUCTURE

The topology of a DSN describes the way in which individual sensors or sensor networks are interconnected. Standard topologies such as star, tree, mesh, and hybrid structures are well-known in the field, but their impact on the performance and reliability of a DSN is sometimes underestimated. In simple topologies like the star, sensors are directly connected to a central gateway. In tree-based structures, sensors are connected hierarchically through intermediate nodes. Mesh topologies allow nodes to communicate with multiple neighbouring nodes, creating redundant paths and enabling decentralized communication. These configurations help to determine the robustness of a DSN, to visualize data flows, how faults are detected, and many other important information [1,2,6]. The choice of topology directly affects the number of required sensors, the presence of critical points in the network (e.g., huge density of sensors, single points of failure), and the network's ability to detect anomalies either in the monitored process or in the sensor (or sensor nodes) themselves. An effective design of topology enables early identification of issues, whether they are from the object of measuring process, sensor degradation/failure, or communication faults. But an inappropriately chosen topology can lead to blind spots, higher failure rates, or excessive load on specific network segments. Moreover, network topology strongly influences scalability, energy efficiency (especially in wireless systems), latency in data transmission, and overall infrastructure maintainability [1,5]. For these reasons, network topology must be taken as an important design element for the design of a DSN. It determines how well the DSN can perform throughout its lifecycle and how resilient it is to both internal and external influences [2,3,6]. Examples of topologies and their impact on the infrastructure of DSN are the following:

Star topology: In this architecture, all sensors are directly connected to a single central point, typically a gateway or processing unit. This setup allows for simple deployment, efficient communication control, and low data transmission latency. The disadvantage is that the failure of the central unit results in the disconnection/failure of the whole DSN. Star topologies are suitable for smaller networks with a limited number of sensors and a simple physical layout.

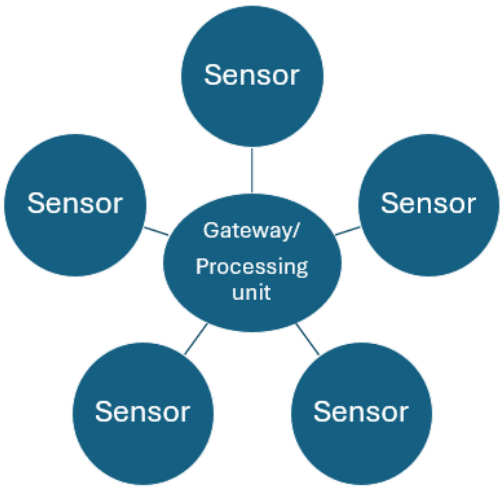


Fig. 1. Example of star topology.

Tree topology: The tree structure has hierarchical levels in which sensors communicate through intermediary nodes (in this case these nodes work as aggregators of data). It is commonly used in applications where the network is logically divided into specific segments. The tree topology facilitates network scalability, but it lacks redundancy and may suffer from increased latency in case of more or deeper branches, which can result in delays in receiving the measuring data or cause faults.

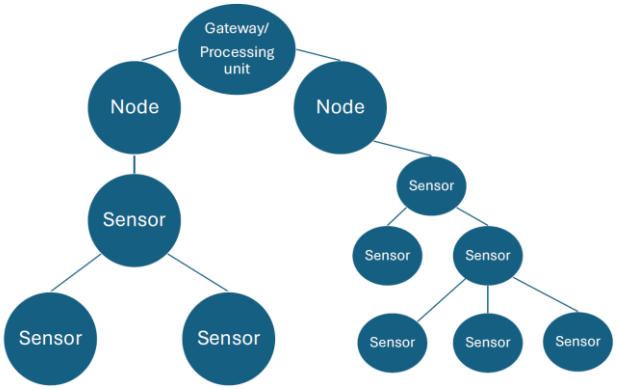


Fig. 2. Example of tree topology.

Mesh topology: In a mesh topology, each sensor can communicate with multiple neighbouring sensors or nodes, forming redundant data transmission paths. This increases the network's fault tolerance and allows for dynamic rerouting in case of sensor failures or environmental changes. However, mesh networks typically require more energy and more complex routing protocols. They are well-suited for applications that demand high reliability or autonomous operation [3].

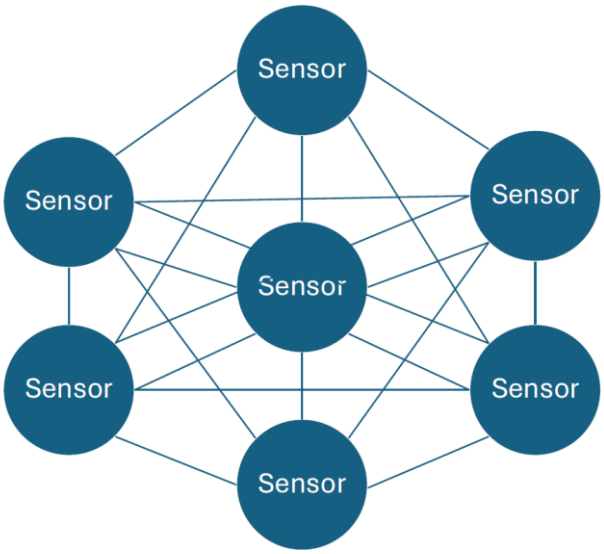


Fig. 3. Example of mesh topology.

Hybrid topologies: In practice, hybrid approaches are most often used, combining the benefits of each topology. For example, tree-mesh structures can balance reliability with energy efficiency while maintaining moderate deployment complexity. Such configurations offer flexibility and adaptability to the specific requirements of the monitored processes where DSN is applied.

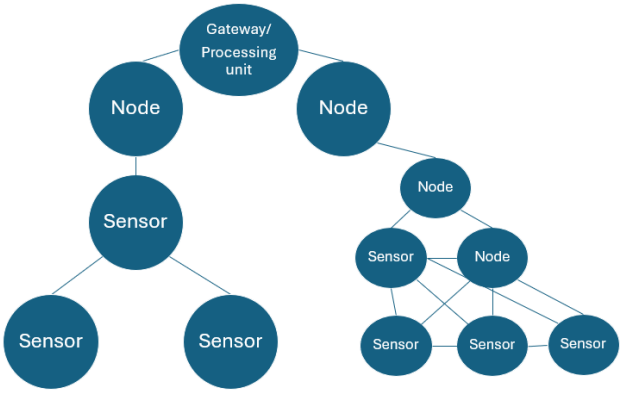


Fig. 4. Example of hybrid topology.

IV. METROLOGICAL ASPECTS OF DSN DESIGN

When designing a DSN with a metrological purpose, it is essential to consider all metrological aspects. The first step is to specify the exact purpose of the network or, in metrological terms, to define the overall DSN measurand, i.e., the quantity of interest that serves as the final output of the DSN. For example, it is important to determine whether the DSN measurand represents a global space- and time-averaged value (e.g., temperature or pollution) or if precise space- and time-resolved data are also of key interest. Additionally, the required measurement uncertainty, along with the spatial and temporal resolution for the measurand, will influence the DSN design and sensor selection. In many cases, the individual sensors in the DSN will measure the same metrological quantity as the overall DSN measurand. However, there are instances where this is not the case. For example, the measurand of the DSN might be the average refractive index over a certain distance, while the individual sensors could be measuring temperature, relative humidity, and CO_2 concentration [7]. When the overall DSN measurand, the type of individual sensors, and the mathematical relationship between sensor values and the DSN measurand are known, it is possible to specify the required uncertainty of the individual sensors. However, it should be noted that correlations in the measurement errors of the sensors need to be carefully considered [8]. Incorrectly assuming that all sensor errors are independent can lead to calculating overly optimistic uncertainties for the DSN measurand. In any case, the combination of multiple data streams requires a common index, usually achieved through synchronized clocks. It is important to evaluate the effects of the implemented time synchronization protocol on the data fusion task at hand, as timestamp uncertainties (e.g., NTP [11]) can be a major part in the uncertainty budget [12].

Another relevant metrological aspect is the expected in-situ behaviour of the sensors, i.e., deployed in the field, the sensors may behave differently than under laboratory conditions. Furthermore, the expected

maximum allowed drift over time should be known. More generally, an initial calibration strategy as well as a metrological QA/QC plan should be designed, where the QA/QC strategy should include a re-calibration plan. The QA/QC-plan may require the inclusion of some high-quality reference sensors and/ or including some redundant sensors in the DSN. As QA/QC can be a major concern for the operation of a DSN over its lifetime, this aspect should be carefully considered in the planning and design phase.

The outputs of a truly metrological DSN should be metrologically traceable, which implies that the individual sensors must also be metrologically traceable and possess a known measurement uncertainty level. If the outputs of a DSN are intended for use in a regulatory context, this can be a significant concern. If the DSN is meant to provide only some qualitative information, the requirements for metrological traceability may be somewhat relaxed; however, an appropriate QA/QC plan is still necessary to ensure that the end-user is not misled by the DSN output and does not base their conclusions incorrectly on, for example, drifting sensors.

A growing body of work shows that several *in situ* calibration strategies, e.g., co-calibration with the neighbouring nodes, consensus calibration, only become viable when the DSN provides sufficient metrological redundancy, i.e. multiple independent observations of the same or strongly correlated measurands within the relevant spatio-temporal window [9] These algorithms exploit the redundant information to solve simultaneously for individual sensor errors and for the measurand field itself; consequently, the expanded uncertainty of the calibrated sensors scales with the degree of redundancy that is preserved [10]. Hence, during the design phase, it is essential to specify a node density that not only meets the desired spatial resolution but also provides the level of redundancy demanded by the chosen calibration strategy and its target calibration uncertainty. The optimum density depends on both the calibration algorithm and the physics of the measurand, as slowly varying or highly correlated fields tolerate larger inter-sensor spacing than fields with sharp and highly localised spatial gradients or rapid temporal dynamics.

V. DSN INFRASTRUCTURE REQUIREMENTS

The effective operation of a DSN depends on a carefully designed infrastructure that takes into account specific hardware, software, and communication requirements. These requirements should be defined during the design phase. The fundamental requirements that need to be considered for any DNS include:

Estimate the numbers or range of used sensors: The number and placement of sensors in a DSN should be determined based on specific targeted processes and take into account measurement accuracy, coverage of the monitored area, or redundancy requirements. ISO/IEC 30141 also recommends considering the scalability of the system so that additional sensors can be added in the future without major changes to the infrastructure.

Types of data collected: Defining the types of data that the sensors will collect is crucial for selecting appropriate sensors and communication protocols. The data should include measured data but may also include derived values obtained by processing the raw data and its measurement uncertainty. In addition to the measured data itself, it is important to collect metadata such as time stamps, sensor identifiers, network topology status, calibration status and other operational parameters depending on the monitored process. This information is essential for proper data interpretation and system maintenance.

Time intervals: It is necessary to determine how often the sensors will collect and send data. These intervals should be defined based on the dynamics of the monitored process and the requirements for data timeliness.

Expected ranges of data values: For each sensor, the data type should be identified and the expected minimum and maximum values defined. This information is important for anomaly detection and sensor calibration.

Distributed computing power: The decision of where data processing will take place (in sensors, nodes, or a central gateway) influences the infrastructure design. Edge computing can reduce latency and network load, while centralized processing can be advantageous for more complex analytics.

Security and Cybersecurity Requirements: DSN security should include data protection during measurement data generation, transmission and storage, as well as device and user authentication, as well as protection against unauthorized access and audit trails in case of operator intervention and all relevant incidents. The ISO/IEC 30141 standard defines requirements for confidentiality, integrity and availability of data, as well as for the protection of personal data.

Network Topology Information: The network topology affects the reliability, latency, and energy efficiency of the DSN. It is necessary to define what type of topology will be used and ensure that the infrastructure is designed with the requirements of the specific process being monitored in mind.

Interoperability: The DSN should be able to work with different devices and services, with an emphasis on standardized interfaces and protocols.

Scalability and flexibility: The architecture should allow for easy expansion of the DSN and adaptation to changing

requirements.

VI. LIFECYCLE OF A DISTRIBUTED SENSOR NETWORK

DSNs are systems that are mostly considered for long-term deployment, requiring a systematic approach to their planning, operation and management throughout their entire life cycle.

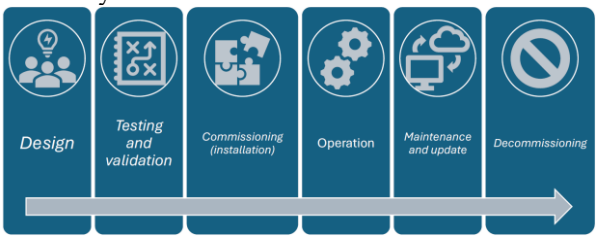


Fig. 5. Lifecycle of DSN

Design phase: This phase sets the network goals, defines the architecture, infrastructure, sensor types, topology, communication, data processing, and security requirements, and considers every known aspect. It also models the operating environment, the expected network load, and plans for scaling.

Testing and validation phase: This phase is essential for the proper functioning of the DSN. Testing includes verification of sensor functionality, compatibility between devices, communication stability and failure tolerance. Simulations of various boundary situations, such as failures, limit values, service (sensor calibration/repair), etc.

Commissioning (installation) phase: physical installation of sensors for monitoring the given process, configuration of nodes and their integration into higher-level systems (e.g., SCADA), or digital twins or cloud platforms. In this phase, communication parameters, security layers are set.

Operational phase: DSN fulfils its main task – data collection and transmission, or local processing. At the same time, continuous monitoring of data integrity, network performance, etc. must be ensured. Early detection of failures, deviations in measurements, or degraded communication quality is also critical.

Maintenance and update phase: This phase includes regular verification of the functionality of all DSN elements, updating of security elements, such as cybersecurity elements. Calibration/service of DSN components, etc. The maintenance interval is set depending on the monitored process, sensor types, and also on the development of cybersecurity (being state of the art in terms of possible attacks).

DSN decommissioning (last phase): includes dismantling components, ensuring secure data handling (archiving, anonymization or total data deletion). This phase should be planned during the design, especially for large-scale or critical applications.

VII. CONCLUSION

Distributed sensor networks are part of the digitalization of metrology. Their ability to continuously monitor processes, process data in real time, and provide data for decision-making, optimization, and prediction makes them a key tool in the field of metrology and its impact on industry, healthcare, and environmental monitoring. However, the successful deployment of DSNs is not only a technical challenge, but also requires a multidisciplinary approach to the design of their infrastructure. As we have shown, it is necessary to take into account the network topology, the number and type of sensors, the method of communication, the location of computing power, and even security requirements already at the design stage. In addition, metrological concerns like the calibration strategy, QA/QC procedures, required uncertainty of the overall DSN measurand and, based on the measurement model and expected sensor uncertainties, the required uncertainty of the sensors should be taken into account. It is equally important to consider the DSN life cycle.

Standards such as ISO/IEC 30141 provide a useful framework and recommendations that can help system designers to create safe and long-term sustainable solutions. It can be stated that DSN have the potential to fully develop and their further development will be increasingly linked to the concepts of digital twins, artificial intelligence and autonomous systems, which places even higher demands on their design and management.

FUNDING STATEMENT

The project (22DIT02 FunSNM) has received funding from the European Partnership on Metrology, co-financed from the European Union's Horizon Europe Research and Innovation Programme and by the Participating States.

REFERENCES

[1] M. Lanza et al., "Towards a Distributed Digital Twin Framework for Predictive Maintenance in Industrial Internet of Things (IIoT)," *Sensors*, vol. 24, no. 8, p. 2663, 2024. [Online]. https://www.mdpi.com/1424-8220/24/8/2663

[2] C. Zhao, Y. Wang, L. Xie, and S. Zhang, "Overview of predictive maintenance based on digital twin technology," *Heliyon*, vol. 9, no. 8, e18385, 2023. [Online].

https://www.sciencedirect.com/science/article/pii/S2405844023017413

[3] IEEE Instrumentation and Measurement Society, *IEEE Std 1451.1-1999: Standard for a Smart Transducer Interface for Sensors and Actuators*. IEEE, 1999. [Online]. https://www.nist.gov/document/information-14511588-v36pdf

[4] International Organization for Standardization (ISO), *ISO/IEC 30141:2018 - Internet of Things (IoT) – Reference architecture*. ISO, 2018. [Online]. https://www.w3.org/WoT/IG/wiki/images/9/9a/10N0536_CD_text_of_ISO_IEC_30141.pdf

[5] S. Tabandeh *et al.*, Sensor network metrology: Current state and future directions, Measurement: Sensors, (2025), 101798., DOI:https://doi.org/10.1016/j.measen.2024.101798

[6] M. Koval, J. Tesař, M .Havlíček, General Sensor Network application approach, Acta IMEKO, Vol.12 No.1(2023).DOI:https://doi.org/10.21014/actaimeko.v12i1.1409

[7] G. Kok, F. Gugole, A . Seymourk, R. Koops, Improved uncertainty evaluation for a long distance measurement by means of a temperature sensor network, Acta IMEKO, Vol.12 No.1(2023).DOI: https://doi.org/10.21014/actaimeko.v12i1.1411

[8] G. Kok, M.V. Dijk, P. Harris, A. Vedurmudi, Modelling and determining correlations in sensor networks, Measurement: Sensors, 2025, 101793, ISSN 2665-9174, https://doi.org/10.1016/j.measen.2024.101793.

[9] Vedurmudi, A. P., et al. "Automation in sensor network metrology: An overview of methods and their implementations." Measurement: Sensors (2025): 101799.

[10] Harris, P., et al. "Measurement Uncertainty Evaluation for Sensor Network Metrology." Metrology 5.1 (2025): 3.

[11] D. L. Mills, J. Martin, J. Burbank, and W. Kasch, "Network Time Protocol Version 4: Protocol and Algorithms Specification, " RFC 5905, IETF, Jun. 2010.https://datatracker.ietf.org/doc/html/rfc5905-

[12] Dorst, T., Robin, Y., Eichstädt, S., Schütze, A., and Schneider, T.: Influence of synchronization within a sensor network on machine learning results, J. Sens. Sens. Syst., 10, 233–245, https://doi.org/10.5194/jsss-10-233-2021, 2021.

[13] S. Mahlknecht, J. Glaser, T. Herndl, Pawis: Towards a Power Aware System Architecture for a SoC/SiP Wireless Sensor and Actor Node Implementation, Editor(s): Miguel León Chávez, Fieldbus Systems and Their Applications 2005, Elsevier, 2006, Pages 129-134, ISBN 9780080453644, https://doi.org/10.1016/B978-008045364-4/50058-7.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Single-Device Integration of Legal Metrology and Third-Party Software via Virtualization

Nicola Zingirian[1], Marco Profeti[2], Federico Botti[3]

[1] *Dept. of Information Engineering, University of Padova, Italy, nicola.zingirian@unipd.it*
[2] *Sampi S.p.A, Altopascio (LU), Italy, mprofeti@idexcorp.com*
[3] *Click & Find s.r.l, Genova (GE), Italy, f.botti@clickandfind.it*

***Abstract* – Integrating process-support functionalities within legal for trade-certified metering heads is a significant challenge in the digital transformation of industrial metering. In this context, we present a virtualized software architecture successfully implemented on the TEX® electronic metering head by IDEX/Sampi S.p.A., a commercially available, legally certified metrological device. The resulting system includes a host machine running legal metrology software and a virtualized guest machine managing auxiliary functions, which may be developed by third-party integrators. The guest handles the display, keypad, and I/O ports when no measurement is in progress. During measurement, the legal metrology software takes preemptive control to ensure legally relevant pulse counting, visualization of metrological data, printouts via sealed printers, and data storage in a database with read-only access for the guest system. This approach enables software extensibility without requiring frequent re-certifications while maintaining regulatory compliance, data integrity, and system flexibility. The proposed solution advances modern industrial metering, as demonstrated in a real-world use case.**

Keywords – Industry 4.0, Measurement Data Protection

I. INTRODUCTION

In modern industrial processes, the precise measurement and control of volumes and weights are essential to ensuring efficiency, safety, and regulatory compliance. Sectors such as petrochemical, food and beverage, and pharmaceuticals rely on metering devices to deliver accurate and legal for trade measurements. However, rising demands for automation, digital integration, and cloud connectivity introduce significant challenges, particularly in integrating metrological systems with enterprise resource planning (ERP) platforms [1]. Achieving this integration while remaining compliant with metrological regulations calls for innovative software and hardware architectures [2].

Metrology software that is certified as legal for trade—hereinafter simply referred to as "certified"— must guarantee that measurement data remain unaltered and protected from unauthorized changes. At the same time, industrial operators increasingly require flexible integration with automation systems, control platforms, and quality assurance tools to record, process, and analyze data automatically. This boosts efficiency and reduces errors common in manual reporting [3],[4]. It also enables automation of procedures like loading, unloading, and digital authorization, improving real-time inventory tracking and overall process optimization [5]. Security requirements further complicate integration. As industrial devices grow more connected and cloud-based solutions become widespread, robust cybersecurity measures are needed to safeguard measurement data integrity and compliance. Software must enable secure interactions between control systems and external applications without undermining metrological standards [6],[7].

This paper addresses these challenges by presenting an innovative hybrid architecture that combines a certified metrology system with a virtualized environment for process-support applications. Virtualization allows non-certified auxiliary software to share the same hardware platform, ensuring compliance with regulatory and technical standards while enabling automation and digital integration. Key benefits include software segregation and data protection within a single device. Unlike traditional systems requiring separate hardware for compliance, often rigid and costly, this solution uses virtualization to logically isolate metrological and non-metrological functions, offering greater scalability and flexibility. Support applications can thus be integrated without additional hardware, overcoming major limitations of conventional approaches and enhancing operational efficiency, adaptability, and security in line with Industry 4.0 goals.

The remainder of this paper reviews the state of the art (Sect. II), defines the design goals (Sect. III), describes the implementation (Sect. IV), and presents a real-world use case (Sect. IV), followed by discussion and conclusions (Sect. V).

II. STATE OF THE ART

Modern metrological systems increasingly adopt digital architectures that allow integration with external applications. Some of the most relevant trends include:

- *Embedded Software in Certified Devices:* Many metrological instruments now include embedded software functions capable of executing certified measurement operations while providing controlled data access to external applications [8].
- *Cloud Connectivity and IoT Integration*: Industrial metrology is progressively adopting cloud-based architecture, allowing measurement data to be securely transmitted to cloud platforms for analysis, storage, and remote monitoring [9].
- *Edge Computing for Metrology:* Edge processing enables real-time data analysis and decision-making at the device level, reducing latency and enhancing system autonomy [10].

The virtualization approach presented in this paper offers an effective alternative for integrating ancillary functionalities, while maintaining compliance with key software certification requirements, according to the international guidelines issued by organizations such as OIML and WELMEC, taking care of:

- *Separation of Certified and Non-Certified Software*: Regulations mandate that certified metrology software must operate independently from any auxiliary software to prevent unauthorized influence on measurements [11][12].
- *Data Integrity and Security*: Measurement results must be securely stored, tamper-proof, and only accessible through controlled interfaces that do not allow modifications [13].
- *Software Update and Certification Compliance*: Any software updates affecting the measurement process must undergo a new certification process to ensure compliance with legal metrology requirements [13][14].
- *Interfaces and Interoperability*: Regulatory frameworks specify how metrological devices can expose data to external applications, ensuring that data integrity is maintained while allowing industrial interoperability [13][15].

III. DESIGN GOALS

Table I summarizes the key design goals of the virtualized metrology system, highlighting its architecture, security mechanisms, and operational safeguards. At its core, the system should support a multi-layered structure responsible for running the certified metrology software, and a virtualized execution environment, which safely hosts third-party applications without interfering with legally compliant operations. Metrological task preemption should ensure that certified functions always take precedence over auxiliary processes, with hardware resource locking preventing unauthorized access during measurements. To maintain security and regulatory compliance, the system should enforce certified firmware execution, using checksum validation, OTP authentication, and MD5 hashing.

Table 1. Design Goals

Design Goal	Description
Multi-Layered System Architecture	Host Layer, Virtualized Guest Environment, and Metrological Software & Data Handling.
Metrological Host Layer	Runs certified metrological software, manages low-level hardware, and ensures isolation of RS232, CAN, RS485, keyboard, display, and printer. Implements checksum validation and OTP authentication.
Virtualized Execution Environment	Provides an independent execution space for non-certified applications, ensuring fast boot and supporting UI, logging, and automation tasks.
Metrological Data Processing	Handles pulse counting, digital flow measurement, and secure read-only storage. Communicates via RS232 interface to metrological printer and exposes structured API.
Secure External Integration	Facilitates safe communication and execution of third-party software without compromising metrological processes.
Access-Controlled Display Management	Requires third-party applications to provide Application ID, Security Key, and Shared Memory ID for accessing shared display memory.
Secure Connectivity & Authentication	Supports LAN, Wi-Fi, and USB connectivity with strict access control policies.
Controlled Third-Party Application Execution	Regulates execution of auxiliary applications with CPU and memory allocation policies to ensure no interference with metrology.
Metrological Task Preemption	Ensures that metrological operations always take priority over auxiliary software.
Hardware Resource Locking	During active metrology operations, critical hardware is locked, suspending all auxiliary processes until completion.
State-Based Process Management	Defines operational states: Startup & Initialization, Idle Mode, Measurement Mode, Post-Measurement Processing.
Regulatory Compliance Enforcement	Security mechanisms to ensure firmware validation and legally compliant execution policies.
Read-Only Data Protection	Stores metrology data in a read-only database to prevent modification, with all interactions logged and audited.

Additionally, secure software update mechanisms guarantee that all updates follow encrypted and signed protocols, mitigating risks of unauthorized modifications. The combination of strict access control policies, secure connectivity, and state-based process management makes this architecture scalable, compliant, and future-ready for industrial metrology applications.

IV. IMPLEMENTATION

The diagram in Fig. 1 illustrates how virtualized architecture has been implemented on the TEX electronic metering head, through a host-guest virtualization model.

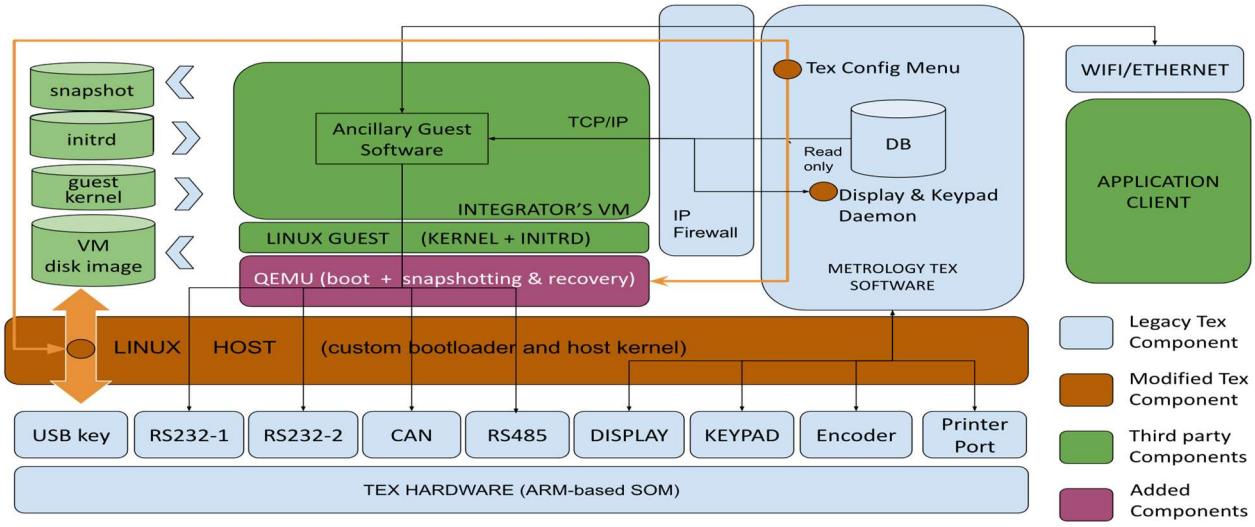


Fig.-1. System modules

A. Host Layer

The Host Layer runs a customized Linux-based operating system, modified to support efficient host-guest virtualization on ARM CPU. Some of the device drivers may be selectively shared with the Guest Layer through virtualized channels or controlled passthrough mechanisms, allowing the guest environment to access specific resources under tightly regulated conditions.

B. Virtualization Layer

The Virtualization Layer utilizes QEMU (Quick Emulator) [16], an open-source hypervisor capable of hardware emulation and Virtual Machine (VM) management. QEMU also manages VM snapshots, enabling quick, safe and reproducible initialization of third-party environments. Leveraging QEMU with KVM (Kernel-based Virtual Machine) [17] ported to the ARM CPU, this layer efficiently isolates the guest virtual machine, managing hardware resources and ensuring effective virtualization.

C. Guest Layer

The Guest Virtual Machine operates as an isolated environment within the QEMU/KVM infrastructure. It relies on a lightweight custom kernel, designed for rapid initialization and efficient resource management. The guest VM specifically handles non-metrological software applications, providing essential support services without affecting certified metrological operations. It has read-only access to metrological data, ensuring compliance with regulatory standards. During system installation, a USB key is used to load the third-party guest kernel and *initrd* (initialization RAM disk) through a configuration procedure integrated within the metrological software, for the secure and consistent guest VM bootstrapping.

D. Certified Metrological software

The certified metrological software runs directly on the Host Layer and relies on a secure database specifically customized for storing metrological data. To ensure data integrity, the Guest VM is granted read-only access, preventing any unauthorized modifications. Compliance with metrological regulations is further reinforced by restricting access to critical metrically relevant interfaces, such as the dedicated I/O port for the measuring encoder and the serial port for the fiscal printer.

E. Ancillary Guest Software

Non-certified applications, possibly developed by third-party integrators, run as Ancillary Guest Software exclusively within the VM, isolated from certified operations to ensure regulatory compliance, operational security, and data integrity. The software interacts with the TEX display and keypad through an internal LAN-based communication channel. This communication is handled by a dedicated daemon on the host system, which mediates access to the display and keypad. When permitted, the guest can read keypad input and write to the display without compromising metrological compliance.

F. Application Clients via Wi-Fi

The Application Client operates externally to the TEX system and connects Wi-Fi (or Ethernet) to interact with internal software components. Communication can be established through secure RESTful web services exposed by the Guest VM.

V. USE CASE: INTO-PLANE DELIVERY

The Dyna-Bus aviation refueling system, developed by Click & Find integrator, is the first real application based on the virtualized development environment within the TEX electronic metering head. This system demonstrates how third-party software can leverage virtualization to

extend the functionality of metrological systems without compromising regulatory compliance.

A. Segregation of Metrological and Non-Metrological Processes

Virtualization architecture ensures strict separation between certified metrological functions and auxiliary process control applications. In the Dyna-Bus system, the metrological software exclusively manages:

- legal-for-trade fuel measurement and dispensing operations,
- pulse counting and calibration procedures
- data logging and communication with metrological storage

Meanwhile, the virtual machine (VM) executes all non-certified functions, including:

- supervision and control of interlock sensors (e.g., valves, hatches, handrails, overfill, etc.),
- automated pre-dispensing checks and actuation
- real-time visualization and user interface

This ensures that auxiliary software enhances automation without affecting legal metrology compliance.

B. Preemptive Hardware Control and Process Automation

The virtualization layer enables dynamic hardware resource allocation, ensuring that the metrological system always has priority access. The Dyna-Bus application benefits from this by:

- Taking over the user's interface during measurement execution
- Allowing automated actuation of fueling system components without interfering with legal metrology functions
- Implementing a multi-step interlock verification process, ensuring that all safety and compliance conditions are met before and during refueling

This preemptive control mechanism prevents unauthorized software interactions while allowing seamless process automation.



Fig. 2 – Application Screenshot: The yellow top bar, managed by the metrology software, indicates that the displayed data are not valid for legal metrology purposes.

C. Secure Data Handling and Interoperability

To prevent unauthorized modifications, the virtualized environment enforces strict data protection policies:

- Metrological data remains locked within the host environment, ensuring compliance
- Dyna-Bus applications access operational data in read-only mode, preventing tampering
- Remote diagnostics and configuration updates can be performed without altering legally certified measurement records

This ensures that real-time monitoring and process control functionalities can be integrated securely and efficiently.

D. Remote Supervision and Cloud Integration

The Dyna-Bus system also utilizes the virtualization architecture to support:

- Cloud-based telemetry and remote diagnostics via the Click & Find telecontrol platform
- Firmware updates and system adjustments over secure communication channels
- Data-driven decision-making through real-time fuel level monitoring and predictive maintenance

By running these functionalities within the virtualized guest environment, the system maintains compliance, security, and extensibility while minimizing hardware costs.

E. Performance

To ensure optimal resource isolation and compliance with the certified metrological execution environment, the virtualized architecture was deployed on a dual-core ARM platform, explicitly assigning a single core as a resource limit to the guest virtual machine. This constraint was introduced to guarantee that the metrology host layer always retains exclusive access to not less than one core, avoiding contention during legally relevant operations. Despite this limitation, both the certified metrological software and the auxiliary guest applications performed flawlessly under real-world workloads. The virtual machine boot time, often a critical factor in embedded deployments, was minimized by leveraging QEMU snapshot mechanisms. This approach enabled near-instantaneous startup of the guest environment without introducing latency or delays during system initialization.

VI. DISCUSSION AND CONCLUSIONS

The implementation of the proposed architecture required addressing several non-trivial technical challenges, notably the adaptation of the QEMU/KVM virtualization stack to operate on ARM-based metering hardware. This involved low-level engineering tasks such as kernel modification, driver integration, and the development of secure guest initialization procedures—none of which are provided by standard toolchains. These efforts went beyond standard hypervisor deployment and

were necessary to meet the stringent operational and regulatory requirements of legal-for-trade systems.

The results demonstrate that this technical investment is fully justified by the benefits achieved in the target domain. The novel system-level approach to separating certified metrological functions from non-certified auxiliary processes, supports automated hardware resource arbitration and allows certified software to preemptively control critical interfaces during measurements. This reduces the risk of interference and enhances robustness compared to manual access control in non-virtualized systems. The modular separation also simplifies maintenance: updates to auxiliary components can be performed within the guest environment without re-certification of the metrological core. Table 2 summarizes the advantages of this virtualization-based approach, including improved software isolation, data integrity, hardware control, and regulatory compliance. These characteristics make the solution suitable for scalable and compliant industrial metrology systems.

Table 2. Virtualization added value

Aspect	Traditional Approach	Virtualized Architecture
Software Segregation	Dedicated hardware partitions or separate physical devices.	Hypervisor-based VM isolation for non-metrological software.
Data Protection	Read-only storage with OS-level access control.	Read-only database accessible only by the host.
Hardware Access Control	Manual restrictions; API-based access for non-certified software.	Preemptive hardware control by metrology software.
Security Enforcement	Firmware-level restrictions and predefined OS policies.	Hypervisor isolation, VM sandboxing, and controlled data exchange.
Flexibility for Third-Party Applications	Limited: separate execution environments or dedicated devices needed.	High; third-party apps run in VM without interfering with metrology processes.
Regulatory Compliance	Physical isolation and certified firmware; difficult modifications.	Host software remains untouched; VM enables software extensions.
Physical Protection of Wired Commun.	Shielded cables, dedicated wired links, tamper protection.	Virtualization-based isolation and cryptographic protocols.

REFERENCES

[1] C. R. H. Barbosa, M. C. Sousa, M. F. L. Almeida, and R. F. Calili, "Smart manufacturing and digitalization of metrology: a systematic literature review and a research agenda," Sensors, vol. 22, no. 7, 2022.

[2] J. Sousa, J. P. Mendonça, and J. Machado, "A generic interface and a framework designed for industrial metrology integration for the Internet of Things," Computers in Industry, vol. 142, 2022.

[3] P. A. Dreyfus, F. Psarommatis, and G. May, "Virtual metrology as an approach for product quality estimation in Industry 4.0: a systematic review and integrative conceptual framework," International Journal of Production Research, vol. 60, no. 21, pp. 6620-6642, 2022.

[4] P. Tambare, C. Meshram, C. C. Lee, R. J. Ramteke, and Y. T. Chang, "Performance measurement system and quality management in data-driven Industry 4.0: A review," Sensors, vol. 21, no. 20, 2021.

[5] A. Varshney, N. Garg, K. S. Nagla, et al. "Challenges in sensors echnology for industry 4.0 for futuristic metrological applications," Mapan, vol. 36, no. 2, pp. 215-226, 2021.

[6] S. Catalucci, A. Thompson, and S. Piano, "Optical metrology for digital manufacturing: a review," The International Journal of Advanced Manufacturing Technology, vol. 120, pp. 597-616, 2022.

[7] E. V. Dudukalov, V. D. Munister, A. L. Zolkin, et al., "The use of artificial intelligence and information technology for measurements in mechanical engineering and in process automation systems in Industry 4.0," Journal of Physics: Conference Series, vol. 1889, no. 3, 2021.

[8] L. De Santis, "Blockchain: the distributed paradigm for secure metrology digitalization," Elea, University of Salerno, 2022.

[9] K. Soni, N. Kumar, A. S. Nair, P. Chourey, and N. J. Singh, "Artificial Intelligence: Implementation and obstacles in industry 4.0," Handbook of Metrology and Applications, 2021.

[10] E. G. Machado, R. T. R. Gomes, E. V. A. Martins, and I. L. Vieira, "Blockchain network to conformity assessment bodies," Metrologia 2023.

[11] International Organization of Legal Metrology (OIML). (2019). General requirements for software-controlled measuring instruments (OIML D 31:2019).

[12] Physikalisch-Tech Bundesanstalt, “Requirements for software separation in measuring instruments and ancillary equipment in legal metrology”, 2020, PTB-8.51-MB07-SWtrennung-EN-V04

[13] European Cooperation in Legal Metrology e.V., WELMEC Guide 7.2:2021, 2021.

[14] B.Wichmann, G. Parkin, R. Barker, “Software Support for Metrology Best Practice Guide No. 1: Validation of Software in Measurement Systems”, NPL Report DEM-ES 014, 2007.

[15] J. Horst, S. McSpadden, “A roadmap for metrology interoperability.” NIST Interagency/Internal Report (NISTIR) – 7381

[16] F. Bellard, "QEMU, a fast and portable dynamic translator," in USENIX Annual Technical Conference, FREENIX Track, 2005.

[17] Y. Kamay, D. Laor, U. Lublin, and A. Liguori, "KVM: the Linux virtual machine monitor," in Proc. Linux Symposium, vol. 1, pp. 225–230, 2007.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Supporting Medicines Manufacturing through Semantic Technologies

Nina Perić[1], Moulham Alsuleman[2], João Gregório[2], Paul Duncan[2], Michael Chrubasik[2]
[1] *Informatics, Data Science Department, National Physical Laboratory, Teddington, TW11 0LW, London, United Kingdom, nina.peric@npl.co.uk*
[2] *Informatics, Data Science Department, National Physical Laboratory, Glasgow G1 1RD, United Kingdom, moulham.alsuleman@npl.co.uk, joao.gregorio@npl.co.uk, paul.duncan@npl.co.uk, michael.chrubasik@npl.co.uk*

***Abstract* – Pharmaceutical manufacturing involves complex, highly regulated processes that generate large volumes of heterogenous data. While highly valuable, this data is often siloed across systems and not easily interoperable, limiting its useability for provenance, advanced analytics and AI as well as automated decision-making processes. This paper presents a semantic technology driven use case to address these issues through the integration of a pharma-based digital architecture with established domain ontologies to support intuitive, semantically rich queries over manufacturing data. The work presented lays the structural groundwork to integrate sensor and event-based observations, linking them to higher-level domain structures in the future using ontologies such as SSN and the Industrial Ontologies Foundry ontology stack, a key requirement for a fully digitalised, interoperable and linked industrial workflow promoting industry 4.0 principles.**

I. INTRODUCTION

The transition to Industry 4.0 and then Industry 5.0 is largely dependent on the digitalisation of manufacturing processes. Rather than seeing data as a simple byproduct, it is recognised as an essential element of more sustainable, automated, and sophisticated manufacturing processes. This transformation requires that all process data, including basic measurement information, should be contextualised, interoperable, and accessible across large and complex digital architectures (US Food and Drug Administration and others). Still, fragmentation of data and siloing persist in many industries, including pharmaceutical manufacturing, where it is often locked within isolated and proprietary systems or structured inconsistently across abstraction layers (International Society of Automation) (International Society of Automation) [4] [5]. This can prevent the implementation of modern analytics or AI-based workflows, as well as reasoning and traceability across systems intra and inter organisation, meaning that the intended efficiency and productivity gains of digitalisation are not being realised.

The National Physical Laboratory (NPL) has developed a case study in collaboration with the Medicines Manufacturing Innovation Centre (MMIC) operated by the Centre for Process Innovation (CPI). This case study focusses on how semantic technologies, specifically ontologies and knowledge graphs, can bridge the gap between business-level Conceptual Data Models (CDMs) and their technical implementation in Logical Data Models (LDMs) [5] [6], linking data and making knowledge more accessible using domain-specific language. This creates a foundation for data interoperability resulting in democratised data access. While the case study does not initially include specific measurement data, the architecture is deliberately built on modular, ontology-based standards (Basic Formal Ontology (BFO) [7], Industrial Ontologies Foundry Core Ontology (IOF-Core) [6], Supply Chain Reference Ontology (SCRO) [8] [9], Biopharmaceutical manufacturing ontology (NIIMBL) [10]) for enabling future integration of sensor-based [11] or quantitative measurement data [12] [13]. By implementing such a semantic layer, architectures enable the future integration of traceable [14], measurement-driven processes, for more robust analytics, compliance, and automation in line with metrological traceability.

Mapping CDMs to domain model ontologies also enables business professionals and domain experts to query LDMs using familiar language. This improves accessibility for non-technical stakeholders and enhances interoperability between various data sources and domains. The simplified process of linking and combining diverse datasets can potentially reveal new insights, significantly enhancing data management capabilities. This system allows for more efficient use of pharmaceutical manufacturing data across the MMIC and its partners, paving the way for more robust analytics, collaboration, compliance, and automation.

II. METHOD

This work aims to enhance data interoperability in pharmaceutical manufacturing, using a tablet manufacturing process as a case study, focusing on supply chains and internal material movements. By mapping terminology from conceptual data models to ontology

Table 1. Example of competency questions (CQs) and their expected answers.

ID	Competency	Competency Questions	Purpose of the question
1	Material Flow	For a process, which materials are used and in which quantities? Where are the final locations of the materials?	If two or more processes require the same material, can that material be stored in more than one location?
2	Scheduling	For all processes, what are the start and end times? What are the events associated with each process?	Processes may have multiple events or just one. How does this affect the start and end times of the process?
3	Material Availability	For all processes, which materials are required (specifically, which material lots)? What are the storage locations of specific material lots?	If a material is required in two processes, how do we check its availability? How do we monitor material availability if it is stored in multiple locations?

classes, the semantic layer acts as a translator between domain experts and technical infrastructure. This allows non-technical users to frame queries using familiar business terms without needing to understand the underlying data scheme. For example, a process engineer can retrieve information about material lots or job steps using domain-relevant language.

Example competency questions and their answers can be found in Table 1. Example of competency questions (CQs) and their expected answers. Competency questions were chosen based on discussions with domain experts, focusing on key tasks such as material tracking and scheduling. These discussions reflected real information needs and helped guide ontology development. Each question was tested using SPARQL queries based on ontology terms, to check that relevant and accurate answers could be retrieved—validating both the ontology structure and its practical value. To ensure the accuracy and reliability of the ontology, we generated a test dataset with predefined and controlled scenarios. This allowed us to validate the outputs of SPARQL queries against known expectations, thereby confirming the ontology's structure and its practical utility. The result was a clear list of connections that allowed logical data fields to be understood as formal ontology elements, making it easier to share information and ask questions in domain-specific language.

A. Ontology Review and Selection

A detailed examination of openly available and specialised ontologies was completed to find frameworks that effectively represent entities and relationships in structured data. The criteria for choosing ontologies included

- their importance to manufacturing (generally and pharmaceutical more specifically), logistics, and quality testing
- how well they fit with existing Conceptual and Logical Data Models
- their ability to support inferencing and SPARQL-based semantic querying
- their agreement with well-known upper ontologies—high-level, abstract frameworks that provide general concepts and relationships applicable across various domains, serving as a foundation for more specific domain ontologies to ensure interoperability and semantic consistency—such as BFO.

The selected ontologies, IOF-Core, SCRO, and NIIMBL ontology, help follow industry standards and improve how different systems understand each other.

B. Ontology Mapping

The mapping process involved aligning elements from the LDM's with classes and properties defined in the selected ontologies, exemplified in Tab. 2. This task required a detailed examination of schema definitions and ontology structures to ensure semantic consistency and coverage. System Architecture

To help combine ontologies with MMIC's systems, a thorough study of the existing underlying digital architecture was completed. This architecture comprises two layers: conceptual and physical. The physical layer provides technical specifications of data classes, attributes, relations, and constraints, which are essential for database design and deployment. Key ideas of the architecture were recognised to help connect the ontology layer, which acts as a bridge between the conceptual and physical layers.

The system architecture consists of three main components:

- *PostgreSQL Relational Database:* Hosted on Azure to store MMIC's structured data aligned with the LDM.
- *GraphDB Triplestore:* Hosted on NPL's servers to enable SPARQL-based semantic querying.
- *Data Connector:* Uses the Ontop virtualisation engine and R2RML mappings to convert relational data into RDF triples that follow the structure of the ontology.

The Ontotop virtualisation engine is a system that exposes relational database content as virtual knowledge graphs, while R2RML mappings are a W3C standard language used to define customised mappings from relational databases to RDF datasets. This architecture enables querying over semantically enriched data without duplication, preserving alignment with the source data while allowing flexible, ontology-driven exploration.

C. Data Interfacing

Data interfacing connects systems to facilitate

Tab. 2. Mapping table. Relationship between SQL (structured) classes and corresponding ontology classes used in knowledge graph development. PK (Primary Key) depicts a unique, non-null identifier for each record in a database class.

SQL Table	Ontology Class	Superclass	PK Column	Identifier Property
process. job	:Job	iof-core: PlannedProcess	jobid	:jobIdentifier
process . isa95joborder	:JobOrder	iof-core: DirectiveInformationContentEntity	joborderid	:jobOrderIdentifier
process. isa95jobresponse	:JobResponse	iof-core: DescriptiveInformationContentEntity	jobresponseid	:jobResponseIdentifier
process. materiallot	:MaterialLot	scro: MaterialEntity	materiallotid	:designatedByLotNumber

information exchange. The semantic data interface removes the need for manual porting and streamlines the conversion from relational databases into a semantically connected fabric. Two options were considered for the development of the data connector:

- *Virtualised Instantiation:* Using R2RML to dynamically pull data from a relational database and map it into a graph within a triple store.
- *Python-based ETL Process:* Hardcoding an ETL processing and mapping layer through Python, providing materialised triples for loading into a triple store.

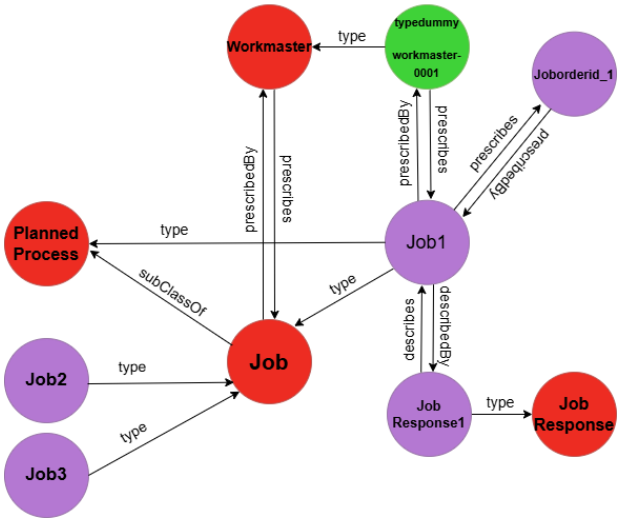


Figure 1. Knowledge graph visualisation illustrating a manufacutring process. Red nodes represent classes, green nodes represent individual instances and purple nodes denote instance-level data. data. Solid arrows represent explicit relationships such as type, subClassOf, prescribes and describes.

R2RML, along with RML, provides a strong and adaptable system that supports both relational and different types of data sources (RML), making sure that mappings are consistent and simple to access. An Extract-Load-Transform (ELT) approach was not considered due to factors like the limited scope of the project, the nature of the available data sources as well as available infrastructure.

III. RESULTS

To enable interoperability of structured data, key entities and attributes were mapped to an LDM using ontology classes. The process of mapping the structured data to an ontology involved several key steps:

- *Identifying Relevant Entities and Attributes:* Key components within the LDM were identified.
- *Aligning with Ontology Classes:* These elements were aligned with ontology classes to ensure semantic compatibility. An application ontology was generated as a subclass of the domain ontology, allowing for different levels of abstraction and future growth.
- *Validating Mappings:* Mappings were validated to confirm that the structured representation accurately reflects domain requirements.

This customised ontology, with relevant entities and attributes, extended standard classes to meet domain-specific requirements. It takes organised data from the relational database and represents by using ontologically enriched RDF triples. This graph structure allows for consistent representation of entities such as materials, manufacturing steps, and material lots. The data was instantiated using this ontology and SPARQL queries were ran to test the system's capabilities. These queries were developed based on the competency questions to validate the system's functionality and demonstrate its capability to answer stakeholder-driven questions.

Evaluation was conducted by translating each competency question into one or more SPARQL queries a validating the returned results against known records from the LDM. For instance, to test material availability (CQ3), queries checked for the correct retrieval of lot identifiers and their associated storage locations. Success was defined as accurate retrieval of all expected links between materials, locations, and process requirements. Partial

failures highlighted ontology gaps, which are discussed in Section IV.

The result of this methodology is a knowledge graph that integrates data from a relational database as RDF triples, semantically enriching information on materials, processes, and production steps. A snippet of this graph is shown in **Error! Reference source not found.**. This structure supports consistent, machine-operable links across the manufacturing lifecycle.

IV. DISCUSSION

A. Data Integration and Interoperability

This work demonstrates how semantic technologies can address key data integration and interoperability challenges in pharmaceutical manufacturing. By aligning structured data with a domain-specific ontology, we enabled richer, more contextualised representations of manufacturing processes. The knowledge graph helps access and interpret data more consistently as well as offers a flexible way to add new features in the future, such as connecting real-time process data or including external standards [4].

B. Stakeholder Engagement and Accessibility

One of the main benefits of the knowledge graph is its ability to enable non-technical users to explore data using domain-specific language and concepts. For example, a process engineer can issue a query like "which material lots are stored at Facility A?" without needing to know SQL schema. The semantic layer acts as a translator between user-friendly terms in the conceptual model and the underlying RDF structure. This significantly reduces the learning curve for domain users and encourages wider adoption across teams.

C. Scalability and Applicability

While the current demonstrator is specific to pharmaceutical manufacturing (tablet production), the approach is broadly applicable across other manufacturing contexts. The scalable framework provided by the knowledge graph allows for future extensions, including the integration of real-time process data and external standards. Ongoing work includes extending the ontology, improving automation of data mapping, and evaluating impact in operational settings.

D. Automation and LDM Alignment

Several challenges were encountered during the automation and alignment of the LDM with the ontology. The physical data model was auto-generated from an enterprise system to be regenerated on NPL servers, but due to using different database systems, the auto-generated PostgreSQL did not consider dependencies in the order of generation, and some constraints and foreign keys were missing. Manual adjustments were made to generate the model on the NPL server, but differences between the LDM and the implemented model remain. Future work will focus on aligning both models and generating synthetic data to cover all tables and test cases.

E. Ontology Maturity and Scope Limitations

While IOF-Core, NIIMBL, and SCRO provided a robust starting point, these ontologies required adaptation to reflect the realities of pharmaceutical manufacturing. Several domain-specific concepts were not supported, limiting immediate coverage of some user requirements. This challenge underscores the importance of treating semantic modeling as an iterative process, continuously refining domain ontologies as the data landscape and use cases evolve.

To ensure semantic consistency across data and knowledge layers, we introduced the ontology classes JOB and JOB STEP as bridging constructs between the architecture (LDM, CDM) and the ontology layer. While the LDM and CDM provided structural definitions of operations, they lacked explicit semantic representation of process hierarchies and temporal dependencies. By modeling JOB as a subclass of `PlannedProcess` and JOB STEP as a granular temporal part thereof, we were able to represent domain-specific workflows such as granulation and blending more precisely. This design also supports alignment with definitions found in the ISA-95 and MMIC data models, where a job is commonly understood as a unit of scheduled work composed of ordered steps. Mapping this to the ontology enabled more meaningful query results and improved traceability across digital models.

F. Alignment Between LDMs and Ontologies

Mapping technical attributes from specific logical data models to high-level ontology concepts presented difficulties. Not all terms aligned clearly, and some attributes had to be reinterpreted or abstracted during the mapping process. Close collaboration between domain experts and data scientists was essential to maintaining semantic integrity without distorting the original meaning of the data. However, due to time constraints, not all modelling was performed according to BFO and IOF standard patterns, particularly in the representation of temporal aspects.

G. Virtualised Queries

Using R2RML virtualisation to avoid data duplication was a practical choice, but it introduced trade-offs. Virtualisation supports live-access data but can introduce lag and increased query time, reducing the speed and availability of data through advanced inferencing. Some advanced SPARQL query features are not yet implemented, such as asking whether an entity "exists" in the virtualised graph. Future improvements will focus on optimising query performance and queryability.

V. CONCLUSION

This project has demonstrated the value and feasibility of integrating semantic technologies within the MMIC digital architecture and, more generalised, pharmaceutical manufacturing. By developing a domain-specific ontology and applying it to real-world manufacturing data, we created a functional knowledge graph that supports more informed decision-making. The approach is flexible, extensible, and offers a foundation for broader digital transformation in the sector. The lessons learned from this pilot will inform the next generation of data-driven tools in medicine manufacturing. With continued collaboration, semantic technologies have the potential to drive efficiency, interoperability, insight generation, and digital maturity across the pharmaceutical ecosystem.

VI. FUTURE WORK

Future work will expand the use case to implement actual sensor and process events, demonstrating full-scale implementation. This includes modelling units [12] and sensors [11] with appropriate ontologies to support provenance and traceability [14]. Expanding the ontology to include testing materials and incorporating quality assurance, formulation processes, deviation handling, regulatory documentation, and batch-specific testing protocols will deepen the semantic fidelity of the graph.

Improving user access with graphical query builders, dashboards, and natural language interfaces will demonstrate the value of natural language to SPARQL to knowledge graph for data democratisation. Building tutorials and methods for ontology mapping and integration across the architecture will facilitate adoption.

Improvements in graph query performance are needed for production-level deployment, including materialising selected RDF graphs rather than relying on R2RML or applying query rewriting and indexing strategies within GraphDB. A more scalable and enterprise-ready implementation of this architecture could be applied to numerous other pharmaceutical manufacturing domains, such as environmental monitoring and compliance, production scheduling and capacity planning, as well as formulation management and change control.

ACKNOWLEDGMENTS

This work was funded by the Department for Science, Innovation and Technology through the UK's National Measurement System. The authors would further like to thank our collaboration partners at the Medicines Manufacturing Innovation Centre consisting of our partners at CPI, University of Strathclyde, UK Research & Innovation, Scottish Enterprise and founding industry partners, AstraZeneca and GSK as well as the Industrial Ontologies Foundry, specifically members of the Supply Chain Reference Ontology working group for the discussions and suggestions.

REFERENCES

[1] US Food and Drug Administration, "Guidance for Industry: Process Validation: General Principles and Practices," Center for Biologics Evaluation and Research. US Department of Health and Human Services, Rockville, MD, Jan. 2011.

[2] International Society of Automation, "ISA95, Enterprise-Control System Integration," 2000. [Online]. Available: https://www.isa.org/standards-and-publications/isa-standards/isa-standards-committees/isa95. Accessed: 2 May 2025.

[3] International Society of Automation, "ISA88: Batch Control," 2010. [Online]. Available: https://www.isa.org/standards-and-publications/isa-standards/isa-standards-committees/isa88. Accessed: 2 May 2025.

[4] Y. Chen, C. Sampat, Y-S. Huang, S. Ganesh, R. Singh, R. Ramachandran, G. V. Reklaitis, M. Ierapetritou, "An integrated data management and informatics framework for continuous drug product manufacturing processes: A case study on two pilot plants," International Journal of Pharmaceutics, vol. 642, 2023, pp. 123086. doi:10.1016/j.ijpharm.2023.123086

[5] H. Cao, S. Mushnoori, B. Higgins, C. Kollipara, A. Fermier, D. Hausner, S. Jha, R. Singh, M. Ierapetritou, R. Ramachandran, "A Systematic Framework for Data Management and Integration in a Continuous Pharmaceutical Manufacturing Processing Line," Processes, vol. 6, No. 5, 2018, pp. 53. doi:10.3390/pr6050053

[6] B. Kulvatunyou, M. Drobnjakovic, F. Ameri, C. Will, B. Smith, "The Industrial Ontologies Foundry (IOF) Core Ontology," in Formal Ontologies Meet Industry (FOMI) 2022, Tarbes, FR, 2022. doi:10.1007/978-3-031-13889-8_10

[7] R. Arp, B. Smith, A. D. Spear, "Building Ontologies with Basic Formal Ontology," MIT Press, Cambridge, MA, USA, 2015.

[8] Semantic Computing Lab Arizona State University, "Supply Chain Reference Ontology (SCRO)," Nov. 18, 2022. [Online]. Available: https://spec.industrialontologies.org/iof/ontology/supplychain/SupplyChainReferenceOntology/?version=release/202301. Accessed: 2 May 2025.

[9] F. Ameri, E. K. Wallace, B. Kulvatanyou, "Towards a Reference Ontology for Supply Chain Management," in I-ESA Workshops, 2020. doi:10.1007/978-3-030-68462-4_2

[10] The National Institute for Innovation in Manufacturing Biopharmaceuticals, "Open-sourced Biopharmaceutical Manufacturing Ontology," 2024. [Online]. Available: https://www.niimbl.org/projects/opensourced-

biopharmaceutical-manufacturing-ontology/
[11] M. Compton, P. Barnaghi, L. Bermudez, R. García-Castro, O. Corcho, S. Cox, J. Graybeal, M. Hauswirth, C. Henson, A. Herzog, V. Huang, K. Janowicz, W. Kelsey, D. Le Phuoc, L. Lefort, M. Leggieri, H. Neuhaus, A. Nikolov, K. Page, A. Passant, A. Sheth, and K. Taylor, “The SSN Ontology of the W3C Semantic Sensor Network Incubator Group,” Web Semantics: Science, Services and Agents on the World Wide Web, vol. 17, pp. 25–32, Dec. 2012.
[12] FAIRsharing.org: QUDT; Quantities, Units, Dimensions and Types, DOI: 10.25504/FAIRsharing.d3pqw7, Accessed:22 May 2025.
[13] R. Albertoni, D. Browning, S. Cox, A. N. Gonzalez-Beltran, A. Perego, P. Winstanley, "The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake," Data Intelligence, vol. 6, no. 2, 2024, pp. 457–487. doi:10.1162/dint_a_00241
[14] L. Moreau, P. Groth, J. Cheney, T. Lebo, S. Miles, "The rationale of PROV," *Journal of Web Semantics*, vol. 35, Part 4, 2015, pp. 235-257. doi:10.1016/j.websem.2015.04.001. Available: https://www.w3.org/TR/prov-o/
[15] J. Gregório, M. Alsuleman, M. Chrubasik, P. Duncan, G. Bisland, "A competency question driven approach to conceptual data model design for digital verification and validation," in Advanced Mathematical and Computational Tools in Metrology and Testing XIII, 2025, pp. 71-82.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Designing a Fatigue Monitoring Sensor System with Industry 5.0 principles

Cai Qingwei[1], Francesco Pilati[2], Francesca Calabrese[3], Matteo Zendri[4]

[1]*University of Trento, qingwei.cai@unitn.it*
[2]*University of Trento, francesco.pilati@unitn.it*
[3]*University of Trento, francesca.calabrese@unitn.it*
[4]*University of Trento, matteo.zendri@unitn.it*

***Abstract* – In the Industry 4.0 era, the focus has been on replacing human operators with industrial robots to reduce labors costs. However, it has led to various social and environmental challenges. Industry 5.0 aims to enhance human-robot collaboration (HRC) in intelligent manufacturing environments, promoting both efficiency and flexibility while prioritizing human operators' health and well-being. This study reconfigures a new HRC systems, optimizes layout, and introduces a new Internet of Things (IoT) structure to monitor human-centric manufacturing processes. By utilizing medical-grade sensors for real-time collection of physiological data, the system ensures privacy compliance. The collected data are input into a artificial intelligence algorithm to achieve two main objectives: evaluating operator's health levels and identifying the most efficient movement patterns during the manufacturing process. This research has significant implications for enhancing operator health and well-being in industrial settings.**

Keywords: Industry 5.0, Wearable Sensor, Human well-being, Internet of Thing.

I. INTRODUCTION

A. *Research Background*

The fourth industrial revolution (Industry 4.0) has greatly transformed manufacturing systems through the integration of cyber-physical systems(CPS), Internet of Things (IoT), and advanced automation. By enabling smart factories, real-time data exchange, and decentralized decision-making, Industry 4.0 has significantly enhanced productivity, flexibility, and operational efficiency across industries. However, as the complexity of manufacturing systems increases and societal expectations evolve, the extreme pursuit of production efficiency by Industry 4.0 has promoted the digital transformation of the manufacturing industry, but has also caused a profound humanistic crisis. The emphasis on automation and technological optimization, while economically beneficial, often overlooks critical aspects such as human well-being, environmental sustainability, and the need for individualized products. The technology-first development model alienates workers into accessories of the production system.

Current research evidence shown that the industrial model of Industry 4.0 has exacerbated social inequality and structural unemployment[7][8][14]. Employment opportunities have been greatly reduced[9], and changes in the labor market have led to a shortage of skilled personnel[10]. New skills and training requirements are needed[12][13], and in the 2030 year disappearance of 800 million jobs is expected to occur[11].

These challenges and crisis underscore a systemic dilemma-whether the single-minded pursuit of hyper-efficiency can be justified without accounting for broader human and ecological considerations. This systemic dilemma directly gave rise to the human centric turn of Industry 5.0. With the continuous growth in industrial production demands, ensuring the well-being of human operators and environmental sustainability has become imperative. Unlike its predecessor, places the human operator at the core of manufacturing systems, promoting collaboration between humans and intelligent machines, where technological advancement serves human needs rather than displacing human agency [1].

B. *Research Objective*

This study aims to develop an innovative digital framework employing wearable devices, real-time data monitoring, and artificial intelligence to enhance operator safety and human well-being in manufacturing environments. The primary objectives include real-time monitoring of physiological states (e.g., fatigue, stress, musculoskeletal risks), providing immediate alerts and corrective measures, and utilizing AI algorithms to predict potential hazards-combining short-term interventions with long-term optimizations (e.g., workstation redesign, task allocation adjustments) to mitigate occupational risks. Additionally, the research focuses on designing a low-power, modular, and easily deployable IoT system, ensuring wearability and durability while maintaining energy efficiency and wireless connectivity. By validating data accuracy and testing system reliability in real-world manufacturing settings, the

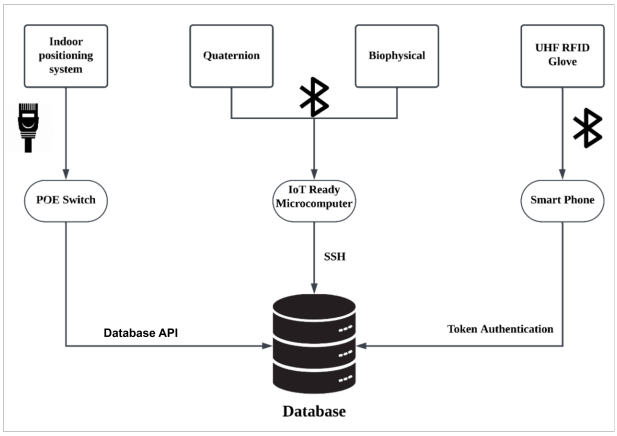


Fig. 1. Fusion Data Acquisition Architecture

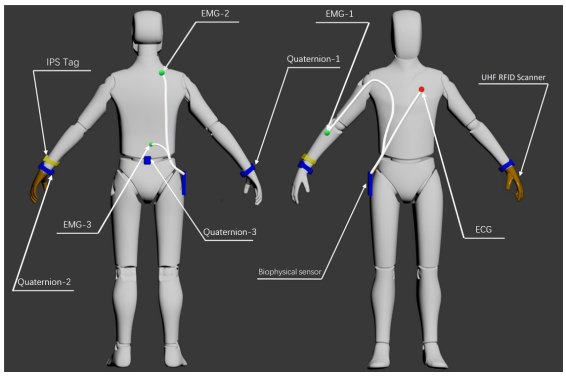


Fig. 2. Deployment of a Sensor Fusion System

ultimate goal is to establish a scalable, adaptable digital safety-enhancing solution that improves productivity and safeguards operator well-being.

II. METHODOLOGY

A. Data-Collection Architecture

Our study proposes an IoT-based multi-modal data acquisition architecture that integrates indoor positioning, inertial measurement, biophysical signals, and ultra-high frequency radio frequency identification (UHF-RFID) technology. The system architecture incorporates multiple data sources, which transmit data to IoT ready Microcomputer via database API, Bluetooth, and secure communication protocols (SSH, Token Authentication) being stored in a database for further analysis and applications. The overall architecture is shown in Fig 1.

Future research work will focus on optimize data fusion algorithms and enhancing system adaptability in complex environments.

B. Wearable Sensor Architecture

In this section, we show the sensor deployment locations, including the rationale their placement. Special attention will be given to the interplay between sensor positioning and physiological motion patterns, ensuring that the collected data accurately represents the intended actions, and ensuring ergonomic feasibility. See Fig 2 for sensor location deployment. The details of the sensors we deployed are shown in the following Table 1.

Table 1. Sensor Location & Function

Sensor	Position	Detection
IPS	Forearm	Movement
Quaternion-1	Dominant Forearm	Activity
Quaternion-2	Non-dominant Forearm	Activity
Quaternion-3	Waist	Posture
EMG1	Dominant-Forearm	Strength
EMG2	Shoulder	Strength
EMG3	Waist	Strength
ECG	Chest	Heart Rate
UHF-RFID Scanner	Dominant-hand	Tool usage

C. Fatigue level declaration

In the fatigue declaration part, we did not use the traditional Borg-Scale [18] [17] [19] or self-report questionnaire. We innovatively used a new method. An innovative approach was employed to assess operator fatigue by utilizing a UHF-RFID glove scanner for real-time fatigue vote declaration. During work periods, operators briefly interacted with different fatigue-level tags using the glove scanner to record their perceived fatigue. The fatigue vote was categorized into five levels, ranging from 1 (slightly tired) to 5 (very tired), different fatigue level have different Electronic Product Code (EPC) value.

During our data collection process, the workers were continuously engaged in the entire spectrum of assembly work at their workstations without interruption. There were no instances where a worker alternated between being fatigued for five minutes and then not fatigued for the next five minutes. Given this sustained activity, we did not set any declarations of "not tired" or "energetic" levels, as the level of fatigue was considered to accumulate progressively over time-only diminishing when the workers took breaks.

Unlike traditional methods such as questionnaires or the Borg scale, which rely on retrospective self-reporting and introduce lag in data collection, this fatigue declaration system enables fatigue simultaneously logging without requiring manual input or stop working. Fatigue declaration system display diagram is shown below Fig 3.

D. Privacy friendly Consideration

Due to this Multi-sensor database storing highly privacy personal data, such as medical/health-related, positional information, etc. The data collection system adopts



Fig. 3. Mock-up of a fatigue declaration system (AI-generated)

an anonymized approach, retaining only identifiers (e.g., 1, 2, 3...) corresponding to the data to ensure that personal identity information cannot be traced. To safeguard the network security of the data collection system, we utilize nProbe to monitor and analyze the upstream and downstream communication traffic related to the database, identifying abnormal behaviors. nProbe can output NetFlow/IPFIX data and, when integrated with visualization platforms such as ntopng, enables continuous tracking and analysis of the data transmission process.

III. CONTRIBUTION

The contributions of this research are outlined here, with a focus on the innovative aspects explored in detail.

1. We utilize a privacy-friendly system of wearable sensors; unlike traditional computer vision technology, we use wearable sensors to monitor the operator's bio-signals, movement patterns, and positional change data. This approach avoids the privacy issues that may arise from deploying computer vision systems because there is no need to capture or track the operator's facial or body images. Operators can be monitored without worrying about personal privacy leaks, which improves the overall acceptability and reliability of the system.

2. Innovatively created a joint fatigue database of biosignals and behavioral patterns, by fusing electromyography (EMG), electrocardiography (ECG) data, real time locating system (RTLS) data, and the quaternion data from the operator's real-time movement patterns (from three Quaternion wireless sensors), we successfully created a multi-modal database. Through utilize this database, machine learning algorithm or deep learning algorithm can not only analyse the changes in physiological activities of human operators during work, but also identify fatigue manifestations under different movement patterns, thereby improving the accuracy of fatigue assessment.

3. Innovatively utilizing a UHF-RFID glove scanner, we can track the specific tools used by operators in real time and correlate these data with physiological metrics for analysis. Prior to deploying the UHF-RFID gloves, we labelled all tools with passive tags that operators might use during their work processes. The subjects do not need to deliberately interact with the tools to accurately identify the passive tags during their tasks. And can collect the strength of RSSI value. Received signal strength indicator (RSSI) is an indication of the power level being received by the receiving radio after the antenna and possible cable loss. When an RSSI value is represented in a negative form (e.g. -100), the closer the value is to 0, the stronger the received signal has been. This approach helps in understanding fatigue variations under different tool usage conditions, thereby enabling optimization of tools usage and workflows to mitigate fatigue-related risks.

4. By deploying a real-time positioning system (RTLS), the spatial position changes and movement behaviors of operators can be monitored in real time, and the position change data can be combined with the bio-signal data. By combining the different tasks that need to be completed at different workstations on the actual assembly line, different task from different workstation and the impact of operation scenarios on fatigue levels can be better captured, thereby improving the accuracy of fatigue prediction.

5. An innovative approach was employed to assess operator fatigue by utilizing a UHF-RFID glove scanner for real-time fatigue vote declaration. During work periods, operators briefly interacted with different fatigue-level tags using the glove scanner to record their perceived fatigue. The fatigue vote was categorized into five levels, ranging from 1 (slightly tired) to 5 (very tired). Unlike traditional methods such as questionnaires or the Borg scale, which rely on retrospective self-reporting and introduce lag in data collection, this system enables fatigue simultaneously logging without requiring manual input.

IV. RESULTS

A. *Multi-sensor Dataset*

This study utilize InfluxDB as data storage, a database designed specifically for time series data, as a time series database (TSDB) to efficiently store, retrieve, and extract time series data generated during the operation of the fusion sensor system. InfluxDB utilize in house built data engine, the Time Structure Merge Tree (TSM Tree), millisecond-range query performance is achievable even for datasets exceeding one million records, it has massive scalability, allows an applications to scale easily to support millions of IoT devices, and perform real-time analysis [16]. In terms of API extension, by utilize Token-based authentication was implemented to establish a secure connection between Grafana Dashboard and InfluxDB 2.0, facilitating real-time monitoring of multi-sensor data streams. Fig 4 shows the Grafana dashboard connected to InfluxDB.

During the fused sensors' data collection process,

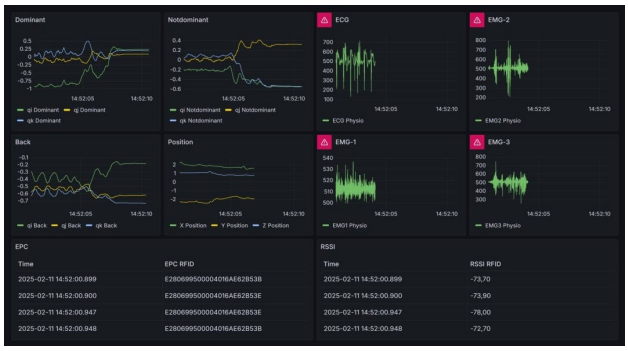


Fig. 4. Multi-sensor Grafana Dashboard

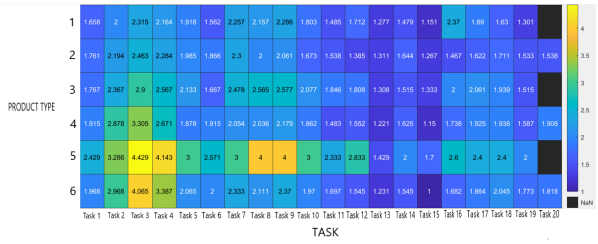


Fig. 5. Heatmap of Product Types and Tasks

Grafana receives information from InfluxDB and display the following data:

Data	Unit
Quaternion data from dominant hand.	dimensionless
Quaternion data from not-dominant hand.	dimensionless
Quaternion data from waist.	dimensionless
ECG signal.	mV
EMG-1 (Forearm) signal.	μV
EMG-2 (Shoulder) signal.	μV
EMG-3 (Waist) signal.	μV
EPC value with acquisition timestamp.	unique identifier
RSSI value with acquisition timestamp.	dBm

B. Limitation

The current study has several limitations. First, due to constraints in the industrial deployment site, our participant consisted solely of male operators, lacking data from female operators. Future research will prioritize addressing this gender imbalance. Additionally, hardware-related issues with the UHF-RFID glove scanner had noise in the collected data, complicating subsequent data analysis. Furthermore, occasional positional drift in the indoor positioning system (IPS), caused by uncontrollable factors in the industrial environment, was observed. Future research will address these limitations to validate the framework's broader applicability.

V. CONCLUSIONS

In this study, we innovatively introduced a UHF-RFID glove scanner for tool identification, which to our knowledge has not been previously deployed in real-world industrial settings. By integrating heterogeneous sensors, we successfully developed a novel data fusion framework incorporating spatial positioning data, IMU-derived quaternion streams, physiological sensor inputs ,UHF-RFID tool-detect signals creating the first known multimodal from real industrial activity database of its kind. Demonstrating the feasibility of this framework for real-time, multi-modal data transmission in industrial environments. This achievement lays a solid foundation for subsequent research.

About sensitivity analysis perspective of this research, we involved a statistical analysis focusing on the mean fatigue level across different tasks and product types. As shown in Fig. 5, the heat-map presents the `Mean Fatigue Level` for each activity-product pair, reflecting how different industrial activities impact worker fatigue depending on their operation type.

To evaluate the robustness and stability of the data analysis under varying conditions, we analysis how fatigue levels varied across product type for the same task, and across tasks for the same product type. These variations help to identify the sensitivity of certain activity-product combinations to fatigue-related factors, such as heart effort, muscle load, posture effort, and movement frequency.

Future research will focus on fusion data analysis, data pre-processing, by using artificial intelligence algorithm make prediction on fatigue level of human operator in real industry environment.

ACKNOWLEDGEMENTS

This research has been funded by the Italian Ministry for University and Research (MUR) through the "Departments of Excellence 2023-27" program (L.232/2016) awarded to the Department of Industrial Engineering.

REFERENCES

[1] European Commission, Directorate-General for Research and Innovation, M. Breque, L. De Nul, and A. Petridis. – *Towards a sustainable, human-centric and resilient European industry*. Publications Office of the European Union, 2021. DOI: 10.2777/308407.

[2] International Organization for Standardization. *Robots and robotic devices – Collaborative robots*. ISO/TS 15066:2016, 2016. Geneva, Switzerland: International Organization for Standardization (ISO). Technical Specification.

[3] S. Chand, H. Zheng, and Y. Lu. A vision-enabled fatigue-sensitive human digital twin towards human-centric human-robot collaboration. *Journal of Manufacturing Systems*, 77:432–445, 2024. ISSN 0278-

6125.

[4] C. Msigwa, D. Bernard, and J. Yun. IoT-driven wearable devices enhancing healthcare: ECG classification with cluster-based GAN and meta-features. *Internet of Things*, 28:101405, 2024. ISSN 2542-6605.

[5] M. L. Chelly and N. Samama. New techniques for indoor positioning, combining deterministic and estimation methods. In *Proceedings ENC-GNSS 2009: European Navigation Conference - Global Navigation Satellite Systems*, pages 1–12, May 2009. Naples, Italy.

[6] K. Pouliakas and N. Becuwe. AI and the EU skilling challenge: First insights from Cedefop's AI skills survey. Presented at the 16th Cedefop Brussels Seminar, June 24, 2024. Cedefop.

[7] R. Kurt. Industry 4.0 in terms of industrial relations and its impacts on labour life. *Procedia Computer Science*, 158:590–601, 2019. 3rd WORLD CONFERENCE ON TECHNOLOGY, INNOVATION AND ENTREPRENEURSHIP. ISSN 1877-0509. DOI:

[8] C. B. Frey and M. A. Osborne. The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114:254–280, 2017. Elsevier.

[9] L. Nedelkoska and G. Quintini. *Automation, skills use and training*, volume 202. OECD Publishing Paris, 2018.

[10] J. Bokrantz, A. Skoogh, C. Berlin, and J. Stahre. Maintenance in digitalised manufacturing: Delphi-based scenarios for 2030. *International Journal of Production Economics*, 191:154–169, 2017. Elsevier.

[11] J. Manyika, S. Lund, M. Chui, J. Bughin, J. Woetzel, P. Batra, R. Ko, and S. Sanghvi. Jobs lost, jobs gained: Workforce transitions in a time of automation. *McKinsey Global Institute*, 150(1):1–148, 2017.

[12] J. Jager, O. Schollhammer, M. Lickefett, and T. Bauernhansl. Advanced complexity management strategic recommendations of handling the "Industrie 4.0" complexity for small and medium enterprises. *Procedia CIRP*, 57:116–121, 2016. Elsevier.

[13] Y. Krykavskyy, O. Pokhylchenko, and N. Hayvanovych. Supply chain development drivers in industry 4.0 in Ukrainian enterprises. *Oeconomia Copernicana*, 10(2):273–290, 2019. Instytut BadaÅ Gospodarczych.

[14] A. Kuzior. Technological unemployment in the perspective of industry 4.0. *Virtual Economics*, 5:7–23, April 2022. DOI: 10.34021/ve.2022.05.01(1).

[15] P. Mohapatra, V. Aravind, M. Bisram, Y.-J. Lee, H. Jeong, K. Jinkins, R. Gardner, J. Streamer, B. Bowers, L. Cavuoto, A. Banks, S. Xu, J. Rogers, J. Cao, Q. Zhu, and P. Guo. Wearable network for multilevel physical fatigue prediction in manufacturing workers. *PNAS Nexus*, 3(10):pgae421, October 2024. ISSN 2752-6542. DOI:

[16] S. N. Z. Naqvi, S. Yfantidou, and E. ZimÃ¡nyi. Time series databases and influxdb. *Studienarbeit, UniversitÃ© Libre de Bruxelles*, 12:1–44, 2017.

[17] Borg, Gunnar A. Psychophysical bases of perceived exertion. *Medicine and Science in Sports and Exercise*, 14(5):377–381, 1982.

[18] Borg, G. Perceived exertion as an indicator of somatic stress. *Scandinavian Journal of Rehabilitation Medicine*, 1970.

[19] Dawes, H. N., Barker, K. L., Cockburn, J., Roach, N., Scott, O., and Wade, D. Borgâs rating of perceived exertion scales: do the verbal anchors mean the same for different clinical groups? *Archives of Physical Medicine and Rehabilitation*, 86(5):912–916, 2005.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Preliminary Experimental Assessment of an IoT-Based Fatigue Monitoring System for Industrial Operators

Enrico Picariello[1], Francesco Picariello[2], Ioan Tudosa[1]

[1]*Dept. of Engineering, University of Sannio, Benevento, Italy , {epicariello, itudosa}@unisannio.it*
[2]*Dept. of Engineering and Science, Mercatorum University, Rome, Italy , francesco.picariello@unimercatorum.it*

***Abstract* – Operators' well-being is essential to implement the Industry 5.0 framework. To this end, this paper presents the first experimental results of a wearable IoT system for monitoring the muscle fatigue of operators on assembly lines. The system is completely modular and composed of 4 subsystems: i) a system for the acquisition of physiological signals, an IMU-based system for the acquisition of inertial data, an RFID glove for tag recognition, and an Indoor Positioning System for the evaluation of operators' movement. For each task, operators can express a fatigue rating from 1 to 5, and a bagged decision tree classifier was used for the classification of muscle fatigue. From the results obtained, it is possible to note that the model can predict muscle fatigue with an accuracy higher than 90%.**

I. INTRODUCTION

In recent years, we have been experiencing a new industrial revolution that will lead to what the experts call the Industry 5.0 paradigm. It is build on the foundation laid by Industry 4.0, which was focused on automation, data exchange, and smart technologies, such as the Internet of Things (IoT), artificial intelligence (AI), and cyber-physical systems, but with the the aim of bringing the human element back to the center of industrial processes [1]. At its core, Industry 5.0 emphasizes collaboration between humans and machines. This paradigm shift aims to combine the efficiency and precision of automation with human creativity, critical thinking, and emotional intelligence, where the human workers work alongside advanced technologies, rather than replacing them [2]. The implementation of Industry 5.0 necessitates several key components: (i) the development of a human-centric system, (ii) the integration of collaborative robots (cobots), (iii) the adoption of flexible production processes that emphasize customization and adaptability, (iv) the deployment of systems designed to enhance operator safety and well-being, and (v) the incorporation of sustainable and environmentally conscious practices. For example, in [3], the authors proposed a wireless sensor network designed for the safety and health monitoring of workers. The system comprises two categories of sensors: (i) a safety node, which monitors environmental parameters such as temperature, relative humidity, ultraviolet radiation, and carbon dioxide concentration; and (ii) a wearable health node that measures physiological indicators including body temperature and heart rate (HR) through PhotoPlethysmoGraphy (PPG). Similarly, the study presented in [4] introduces a monocular vision-based system for recognizing worker postures during construction activities. These studies primarily focus on monitoring either environmental and physiological parameters or ergonomic aspects. However, the physiological parameters monitored, specifically HR and body temperature, are insufficient for estimating an operator's physical fatigue, which is typically assessed using surface electromyography (sEMG) [5]. Fatigue estimation is of particular importance, as it can be used to evaluate risk indices (among which the most used are the EAWS indices [6]), which are fundamental for the physical well-being of operators. Thanks to them, it is possible to identify critical issues in the manufacturing process, which can then be customized to the physical needs of the operator. Moreover, the use of a monocular camera is inherently constrained by its limited field of view, rendering it unsuitable for dynamic assembly tasks in confined or obstructed workspaces where operator mobility is required. To achieve these objectives, one promising approach that has attracted growing attention in both scientific and industrial domains in recent years involves wearable IoT devices [7, 8] used to evaluate fatigue, stress, ergonomic index [5] and postures[7]. In this paper, a system for the automatic classification of operators' fatigue on an assembly line has been developed. Wearable sensors worn by operators during a normal pump assembly shift have been used. In particular, the ECG signal and 3 sEMG signals are acquired in correspondence with the forearm, trapezius, and erector spinae muscles, a BITalino acquisition board, placed on the operator's belt. 3 IMUs with bands are also used to monitor the quaternions of the dominant hand, non-dominant hand and the back. The fatigue declaration is performed after each task, through an RFID glove, with the possibility of giving a qualitative fatigue rating ranging from 1

to 5 (1 being almost no fatigue and 5 being exhausted). Time and frequency domain features are calculated on the acquired signals and used to train a bagged tree ML algorithm for the automatic fatigue assessment. The rest of the paper is structured as follows. In Section 2, a brief review of the state-of-the-art on automatic fatigue classification is presented. In Section 3, a description of the proposed monitoring system is reported. A preliminary experimental campaign is presented in Section 4. Section 5 draws the main conclusions and several future work directions.

II. RELATED WORKS

In order to assess the well-being of operators, is essential to study how the state of the art deals with fatigue classification. This classification is very important since it is directly correlated to risk indices and to the well-being of operators. The customization proposed by Industry 5.0 should be based on these indices and, therefore, it depends on the perceived fatigue in the various tasks that the operators have to perform. In [9], the authors investigate how a commercially available wearable low-cost sensor and two machine learning algorithms can be applied to measure and evaluate heart rate, heart rate variability and respiration rate to establish a relationship with workload. The sensors used consist of a sensor shirt, a sensor belt and a multifunctional sensor module, and 6 male healthy subjects were used in the experimental evaluation. Furthermore, the RR is used but exclusively for the training of the two chosen Machine Learning models: Random Forest and K-Nearest Neighbors. The task recognition accuracy proposed are very high, around 90-95%. Even if these results are performed on few tasks, and not in an industrial environment, they highlight the possibility of using WHD for task recognition and the correlation between physical stress and fatigue.

In [10], the authors attempted to characterize the nonlinear variations in sEMG measurements and to detect fatigue for the associated activities. The experimental evaluation was conducted on 52 healthy individuals performed repetitive contractions of the biceps muscle and the sEMG signal was recorded. The signals are then codified into a binary sequence and symbolic transition networks were created to model the dynamics of the sEMG signals. From these networks, metrics such as symbolic entropy, network entropy, uniformity, and effective degrees (minimum and maximum) were derived and fed into a classifier (k-Nearest Neighbor or Naive Bayes) achieving a maximum accuracy of 90 %. The study demonstrates that analyzing sEMG signals is effective in detecting muscle fatigue since a significant decrease in signal complexity was observed as muscle fatigue progressed. In [11] the authors performed extensive research to monitor and predict physical fatigue among manufacturing workers using wearable technology and machine learning. The main objective was to develop a system that can continuously monitor and predict varying levels of physical fatigue in real-time, moving beyond traditional binary fatigue assessments. Data were collected from 43 individuals performing two manufacturing tasks: composite sheet layup and wire harnessing and six wearable sensors were strategically placed on the upper body to capture vital signs (heart rate, heart rate variability, skin temperature) and movement data (via inertial motion units). The fatigue level was self-reported by the participants and recorded alongside sensor data during task performance. To perform the automatic classification, a gradient-boosted tree-based regression model with a custom asymmetric loss function was developed to predict continuous fatigue levels, emphasizing accurate predictions of higher fatigue states. The system effectively predicted multilevel fatigue states in real-time, demonstrating its potential applicability in actual manufacturing settings. From the analysis of the state of the art, it is therefore possible to deduce that using features extracted from physiological signals and inertial signals, it is possible to use ML techniques for the automatic classification of fatigue, with high accuracy. Unfortunately, we believe that there are some critical points that are not analyzed in the literature. In all the articles analyzed, regardless of the accuracy of the fatigue classification, the tasks that the subjects perform are typically very simple, related to repetitive movements and muscle activations. Furthermore, even when more complex tasks are considered, they are too few, and do not generalize an industrial assembly process, where each operator finds himself performing even 20 or 30 different tasks. Some of these tasks are also performed only a few times throughout the assembly line. In addition, the industrial environment presents numerous sources of disturbance, not analyzed in laboratory analysis or in more controlled environments. In this paper, the proposed system tries to tackle the previous shortcomings of the state of the art on fatigue classification. The system is deployed in an actual industrial environment, where it is used to monitor the task performed by 3 operators on a pump assembly line. The system is completely modular and user-friendly in fact, the operators were instructed on how to start and stop the acquisition, with a very intuitive GUI. The system was extensively used to monitor the activities of the operators for 1 month, but this preliminary evaluation will be focused on a reduced dataset, in order to validate the whole system.

III. PROPOSED IOT SYSTEM FOR MONITORING OPERATORS

To collect data and create a large, reliable dataset suitable for training various ML/DL algorithms to classify fatigue levels, a modular, easy-to-install, and user-friendly IoT system was designed and implemented. The general architecture of the proposed IoT system is shown in Fig. 1.

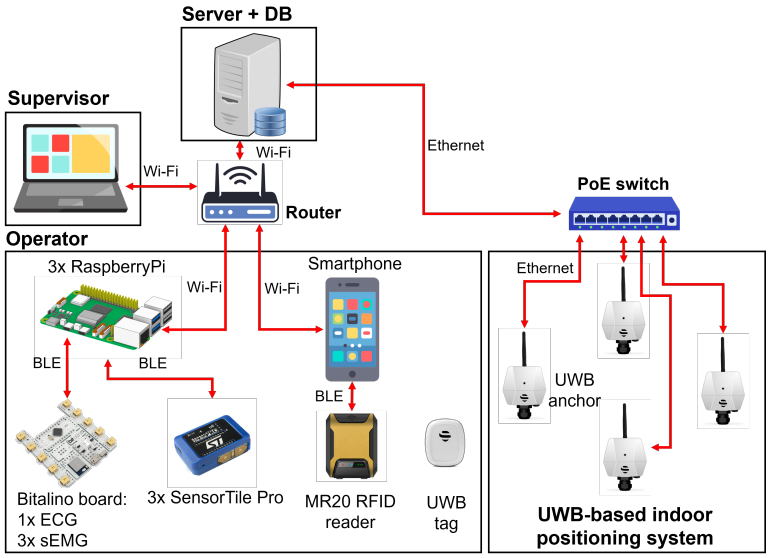


Fig. 1. General architecture of the proposed IoT-based system

Each monitored operator wears four types of sensors: (i) a BITalino board [12] for acquiring a precordial ECG signal and three surface EMG (sEMG) signals from the forearm, trapezius, and erector spinae muscles, (ii) three SensorTile Pro inertial measurement units (IMUs) by STMicroelectronics [13] for capturing orientation data, (iii) an MR20 RFID glove reader [14] for identifying passive tags attached to tools and components; and (iv) a UWB tag for indoor positioning measurements. The BITalino board and the IMU on the dominant arm transmit data via Bluetooth Low Energy (BLE) to a Raspberry Pi: the BITalino streams data at 1 kHz per channel, while the IMU streams quaternion data at 100 Hz. In particular, the accuracy of the pitch angle measured by the IMUs has been experimentally evaluated in [15], where uncertainties are lower than 0.2°. Two additional Raspberry Pis are used to collect data at 100 Hz from the non-dominant arm and back IMUs via BLE. This three-device setup was adopted after extensive testing proved it to be the most stable configuration over time. The RFID glove streams the detected tag's EPC value and corresponding RSSI to an Android-based smartphone. All Raspberry Pis are connected to a Wi-Fi router and stream the collected data to a central server with a database. The smartphone application also transmits RFID data to the same server. For indoor positioning, four Sewio UWB anchors [16] are connected to the server via Ethernet through a PoE switch. Each operator wears a UWB tag, and its position is estimated at 10 Hz using Sewio's cloud-based software. On the server, an InfluxDB time-series database is installed. For each operator, a record is maintained with fields corresponding to: (i) ECG and three sEMG signals, (ii) quaternion data from the dominant arm, non-dominant arm, and back, (iii) EPC tag and RSSI from the RFID glove, and (iv) position data from the UWB system. The server also hosts Grafana, an open-source data visualization and monitoring platform. It provides a simple and intuitive interface that allows one or more supervisors to visualize the collected data in real time, with updates every 5 seconds on any laptop connected to the Wi-Fi network. To collect subjective feedback on the perceived fatigue level, each workstation is equipped with five RFID tags numbered from 1 to 5. These tags allow the operator to send the corresponding fatigue level label to the database via the RFID glove reader.

IV. EXPERIMENTAL CAMPAIGN AND RESULTS

A. *Experimental campaign*

An extensive experimental campaign was carried out, where the system was installed at EBARA Pumps Europe S.p.A for one month. The system was installed at the pump assembly stations, to monitor the activities of 3 operators during the two daily shifts. The data of the operators who agreed to participate in the experimental campaign were acquired in compliance with privacy regulations, and anonymously, with written consent from the operators. The operators wore the system before starting their work shift. The RFID glove is worn on the right hand, the EMG signals are acquired via electrodes placed on elastic bands, on the back, arm and forearm while the ECG signal is acquired using one of the precordial leads. The BITalino is hooked to the operator's belt. The IMUs were inserted in boxes equipped with an elastic band (like a bracelet) and were placed on both hands, while the one for the back was connected with Velcro strip to a band placed around the chest, and are directly able to acquire quaternions. The UWB tags can instead be worn like a watch. To wear the

entire system the operators take 2-3 minutes on average. When the wearing of the system is finished, the operator uses the GUI to start the communication between the sensors, the raspberries and the server. The operator uses the glove to scan a tag placed near the server to signify the beginning of his shift. On each of the stations, RFID tags have been placed on the tools that the operator uses, in addition to 5 tags to express the fatigue. After each task, the operator votes, scanning one of the 5 fatigue tags based on the fatigue vote he wants to assign to that task. The data are sent to the InfluxDB server and then downloaded for processing. The system has been in operation for about a month, but in this paper a preliminary evaluation with a reduced dataset will be included. In particular, the data regarding the assembly of 5 pumps by the 3 operators on the line will be analyzed.

B. *Features Selection*

In order to classify the fatigue of the operators during the assembly tasks, the signals acquired by the IoT wearable system were processed to obtain features used by the Machine Learning model. Instead of using the raw EMG data, the latter is processed to identify moments of muscle activity. At the beginning, the sEMG signal is band-pass filtered with 4-th order Butterworth filter working at 30 Hz to 300 Hz. To evaluate muscle activity, the Teager–Kaiser Energy Operator (TKEO) is then calculated as follows :

$$TKEO = x(n)^2 - x(n-1) \cdot x(n+1) \qquad (1)$$

where, $x(n)$ is the n-th filtered EMG sample. A wavelet decomposition at three levels with a db2 mother wavelet is performed on the TKEO. The signal $\eta(n)$ is calculated from the three coefficients and the residual as follows:

$$\eta(n) = \max_a CWT(a, n) \qquad (2)$$

where a is the scale parameter and $CWT(a, n)$ are the estimated coefficients, at each scale factor, and the residual. In this way, $\eta(n)$ contains the highest $CWT(a, n)$ coefficients for each sample, thus increasing the muscle activation detectability. The $\eta(n)$, ECG and quaternions, were divided into 1-second segments, on which the following features were evaluated: Clearance Factor, Crest Factor, Impulse Factor, Kurtosis, Mean, Peak Value, Root Mean Square (RMS), Signal-to-noise Distortion Ratio (SINAD), Signal-to-noise Ratio (SNR), Shape Factor, Skewness, Standard Deviation, Total Harmonic Distortion (THD), Autocorrelation of errors at lag 1 (ACF1), Interquartile Range (IQR), Maximum, Minimum, Median, Quartile 1 and 3 and Maximum Allowable Variation (MAV). Furthermore, on the $\eta(n)$ signal, a threshold is applied to obtain the onset signal (i.e., a signal assuming one value when the threshold is exceeded and zero otherwise), which is also been used as an additional feature for the classifier. The chosen classifier is a bagged ensemble tree with 30 classifiers.

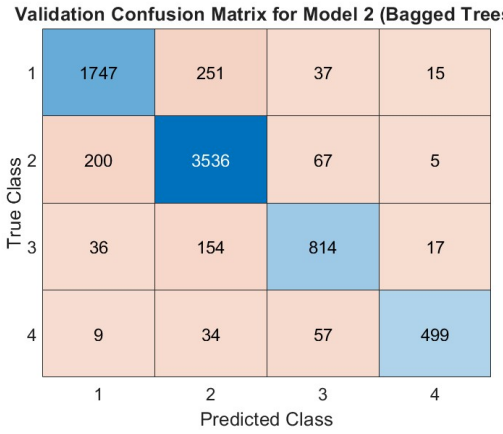


Fig. 2. Validation Confusion Matrix

Fatigue Vote	Accuracy	Precision	Recall	F1-Score
1	92.67	87.70	85.22	86.44
2	90.49	88.96	92.86	90.86
3	95.08	83.49	79.73	81.56
4	98.17	93.10	83.31	87.93

Table 1. Obtained metrics for the Validation Phase

V. RESULTS

To train the model, the features of the signals acquired during the assembly of the 5 pumps were divided following an 80-20 split (80 for training and 5-fold validation, and 20 for the test). In particular, the validation/test data split was performed by examining the fatigue scores for each task and then using a Random Number Generator function to mix the data. This was necessary because it ensures that all tasks are present in both training and testing. For the validation of the classifier, the most common metrics used in Machine Learning applications have been calculated for every fatigue vote: Accuracy, Precision, Recall and F1-Score. The validation Confusion Matrix is reported in Figure 2, the validation metrics are reported in Table 1. As can be seen from the confusion matrix and the tables, in the available ground truth, there is no fatigue rating equal to 5. This does not affect the purpose of this work, as these results are related to a first experimental validation. Nevertheless, for the correct generalization of the system, it is necessary to use an expanded dataset, which contains all 5 fatigue declarations. From the analysis of the metrics, it can be noted that the model exhibits an accuracy greater than 90%. From these values, it is therefore possible to declare that this first experimental validation of the model had a positive outcome, and the model is able to effectively classify the fatigue of operators during complex assembly operations, in an industrial environment.

VI. CONCLUSION AND FUTURE WORK

In this paper, the first experimental results of a wearable IoT system for the classification of fatigue of operators on assembly lines have been presented. The system has been designed to be completely modular, and is composed of (i) aBITalino board for acquiring a precordial ECG signal and three surface EMG (sEMG) signals from the forearm, trapezius, and erector spinae muscles, (ii) three SensorTile Pro inertial measurement units (IMUs) by STMicroelectronics for capturing orientation data, (iii) an MR20 RFID glove reader for identifying passive tags attached to tools and components; and (iv) a UWB tag for indoor positioning measurements. The system was installed and used by operators during the assembly of industrial pumps for one month, and the data from 5 of these assemblies were used for preliminary validation testing of the fatigue classification system. For each task performed, operators assigned a fatigue score, and a set of features in both the time and frequency domains were used to train a bagged ensemble decision tree for classification. From the results obtained it is possible to verify that the system correctly classifies the fatigue of the operators on the different tasks with accuracies greater than 90%. Future works will be focused on expanding the dataset, using the full month-long available dataset for the training of the model, to improve generalization and to evaluate the performance of the classification on more data, and to implement the classification system to be able to work in real-time, while the operators perform the assembly operations.

ACKNOWLEDGMENT

Funded by the European Union under NextGenerationEU, PRIN 2022 Prot. 2022ZTMNPC.

REFERENCES

[1] E. Commission, D.-G. for Research, Innovation, and J. Muller, *Enabling Technologies for Industry 5.0 : results of a workshop with Europe's technology leaders*. Publications Office, 2020.

[2] F. Pilati, A. Sbaragli, M. Nardello, L. Santoro, D. Fontanelli, and D. Brunelli, "Indoor positioning systems to prevent the covid19 transmission in manufacturing environments," *Procedia Cirp*, vol. 107, pp. 1588–1593, 2022.

[3] F. Wu, T. Wu, and M. R. Yuce, "Design and Implementation of a Wearable Sensor Network System for IoT-Connected Safety and Health Applications," in *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*, pp. 87–90, 2019.

[4] Y. Yu, H. Li, J. Cao, and X. Luo, "Three-Dimensional Working Pose Estimation in Industrial Scenarios With Monocular Camera," *IEEE Internet of Things Journal*, vol. 8, no. 3, pp. 1740–1748, 2021.

[5] L. De Vito, E. Picariello, F. Picariello, I. Tudosa, A. Sbaragli, G. P. R. Papini, and F. Pilati, "Measurement System for Operator 5.0: a Learning Fatigue Recognition based on sEMG Signals," in *2023 IEEE Int. Symposium on Medical Measurements and Applications (MeMeA)*, pp. 1–6, 2023.

[6] K. Schaub, G. Caragnano, B. Britzke, and R. Bruder, "The european assembly worksheet," *Theoretical Issues in Ergonomics Science*, vol. 14, pp. 1–23, 01 2012.

[7] F. Pilati, A. Sbaragli, F. Tomelleri, E. Picariello, F. Picariello, I. Tudosa, and M. Nardello, "Operator 5.0: Enhancing the physical resilience of workers in assembly lines," in *2023 IEEE Int. Workshop on Metrology for Industry 4.0 & IoT (MetroInd4.0&IoT)*, pp. 177–182, 2023.

[8] J. P and E. Ruby, "A review on wearable device technology for healthcare industry applications," in *2023 2nd International Conference on Ambient Intelligence in Health Care (ICAIHC)*, pp. 1–5, 2023.

[9] M. Brillinger, S. Manfredi, D. Leder, M. Bloder, M. Jäger, K. Diwold, A. Kajmaković, M. Haslgrübler, R. Pichler, M. Brunner, S. Mehr, and V. Malisa, "Physiological workload assessment for highly flexible fine-motory assembly tasks using machine learning," *Computers & Industrial Engineering*, vol. 188, p. 109859, 02 2024.

[10] N. Makaram, P. A. Karthick, and R. Swaminathan, "Analysis of dynamics of emg signal variations in fatiguing contractions of muscles using transition network approach," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–8, 2021.

[11] P. Mohapatra, V. Aravind, M. Bisram, Y.-J. Lee, H. Jeong, K. Jinkins, R. Gardner, J. Streamer, B. Bowers, L. Cavuoto, A. Banks, S. Xu, J. Rogers, J. Cao, Q. Zhu, and P. Guo, "Wearable network for multilevel physical fatigue prediction in manufacturing workers," *PNAS Nexus*, vol. 3, p. pgae421, 10 2024.

[12] D. Batista, H. Silva, and A. Fred, "Experimental characterization and analysis of the bitalino platforms against a reference device," in *2017 39th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2418–2421, 2017.

[13] "Steval-mksbox1v1 by st-microelectronics," 2016. https://www.st.com [Accessed: 23/05/2025].

[14] "Steval-mksbox1v1 by st-microelectronics," 2025. https://www.chainwayeurope.eu/eshop-mr20-wearable-bt-rfid-reader.html [Accessed: 23/05/2025].

[15] L. De Vito, E. Picariello, F. Picariello, S. Rapuano, I. Tudosa, A. Sbaragli, and F. Pilati, "Iot-based system for monitoring the well-being of industrial operators through wearable devices," in *2024 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 1–6, 2024.

[16] "Sewio real-time location system (rtls)," 2025. https://www.sewio.net/ [Accessed: 23/05/2025].

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Soil Digitalization Using Micro-Sensors

Federico Fina[1], Massimo Piotto[2], Simone Contardi[3], Fabio Leccese[1]

[1]*Dipartimento di Scienze, Università degli Studi Roma Tre, Rome (Italy), federico.fina@uniroma3.it*
[2] *Dipartimento di Ingegneria dell'Informazione, Università di Pisa, Pisa (Italy), massimo.piotto@unipi.it*
[3] *Sensichips.srl, Aprilia (LT, Italy), simone.contardi@sensichips.com*
[1]*Dipartimento di Scienze, Università degli Studi Roma Tre, Rome (Italy), leccese@uniroma3.it*

***Abstract* – In this paper, we propose a method of data collection from different sensors through a software developed by Sensichips srl, SLM-Studio, which can then be compared with MLP machine learning models trained through training data available on the Sensichips website. The measurements performed by the SCW water sensors, the SCA air sensors and the SCP multispectral sensor can be applied in different fields of precision agriculture ranging from irrigation water monitoring to the health status of plants and soils up to the monitoring of the chemical-physical conditions of the air.**

***Keywords* - microsensors, machine learning, colorimetry, digitalization**

I. INTRODUCTION

Precision agriculture [1] aims to develop techniques and technologies to be used in the agri-food sector to reduce both wastes associated with resource consumption and to develop methods to reduce the impact of crops on the environment and to obtain high quality products that can improve the quality/price ratio. To do that, it is important that cultivated fields have a series of tools to monitor the health of a plant, the ripening of a fruit and the possible onset of problems related to the presence of fungi, the health of the soil as well as its chemical-physical characteristics. In recent years, materials science and various engineering sciences have introduced various sensors into the agricultural market that can be installed in strategic points of the agricultural field or mounted on special drones [1]. The measurements taken by these sensors can then be transmitted via Wi-Fi [2], ZigBee [3-4] or Bluetooth [5] systems to a sort of central system that is consulted by the farmer, who can decide whether to intervene to reduce the possible onset of problems that can damage the quality and the harvest itself. In the future, these data could be an integral part of a single dataset in which it is possible to consult the possible problems of a plant or a vegetable as the chemical-physical conditions and the type of soil, the meteorological conditions and the state of health of the plant itself vary.

A soil is the most superficial layer of the earth's crust and extends from the surface, where there is almost exclusively humus and organic matter, to a layer of rock, the bedrock, which has a low degree of fragmentation. Each layer of the soil is called a horizon and at the level of nomenclature is named by specific letters as shown in Fig. 1.

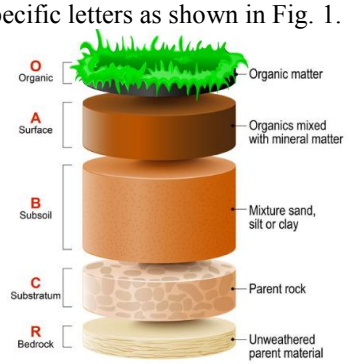


Fig. 1Geological structure of a soil (Soil is the thin layer of material covering the earth's surface)

In nature there are different types of soil that are classified based on the granulometry through sieves that evaluate the percentage of sand, silt and clay present in each soil sample (Fig. 2). Each soil is also associated with a specific color that is associated both with the content of the grains present but above all based on their mineralogical content.

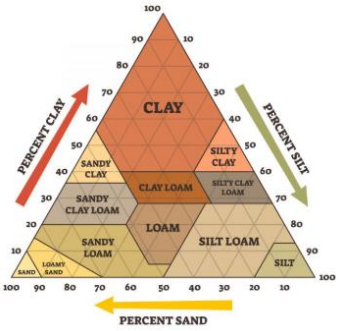


Fig. 2. Soil types classified according to the percentage content of clay, sand and silt (Soil is the thin layer of material covering the earth's surface).

In agricultural fields there are different types of sensors that incorporate different scientific technical knowledge such as geophysics, chemistry up to electronic and telecommunications engineering.

Geophysical sensors are mainly applied for monitoring the first horizons of the soil. At a theoretical level, geophysical sensors analyze two types of waves: electromagnetic waves (in the case of a GPR [6] or a three-wire probe connected to a VNA [7]) or an acoustic wave [8] (in the case of a seismometer). Seismometers, for example, in addition to having an interesting application for the study of site effects in different types of agricultural land [9-10], can also have interesting applications for soil moisture analysis as well as in the case of a permanent seismic station allow other research institutions to perform geophysical analyses such as seismic tomography [11]. Instead, with the use of GPR from the interaction of the waves with the dielectric properties of water, from the analysis of the reflected signal it is possible to analyze the water content of the soil and the soil horizon in which it is located [12-13]

Chemical sensors instead work mainly in monitoring the chemical-physical characteristics of agricultural soil. Temperature [15-16], pH [17] and humidity [18] sensors and conductivity meters [19] are available on the market as gas detectors [14]. The parameters just mentioned can be applied both in monitoring the water with which the fields are irrigated, and to monitor the horizon A of a soil and to monitor the atmospheric environment in which the crop is located. In recent years, the microelectronics industry is developing smart microsensors that can be applied to the world of precision agriculture and through machine learning techniques it is possible to reveal chemical anomalies of the soil or of the leaves themselves even before the effects are perceptible by the human eye [20-21].

The optical sensors available on the market exploit different forms of material radiation interaction such as reflectance [22], transmission [23] and fluorescence [24] of a given substance. This technique is applied through different portable spectrophotometers that work in different ranges of the electromagnetic spectrum: from UV, to VIS [24] up to NIR [25]. The different spectroscopic techniques help us to discriminate for example the presence of microfungi substances on the surface of a leaf as well as the health status of the plant [26]. Fluorescence, for example, is a technique that allows us to observe the water content of a leaf and therefore can provide interesting data on its hydration [27]. Another important type of optical sensors available on the market are colorimeters that can be applied in different fields such as: monitoring the health status of a leaf [28-29] or the color of a soil [30] which is also a diagnostic element of a soil.

II. METHODS

In this context, the company Sensichips.srl (Sensors - Sensichips - learning MICROSENSORS), has developed several microsensors that can be applied in different agricultural fields by providing a data set that can be found online on the site. These sensors can be applied for the monitoring of water and air pollutants that can be applied in different fields, including precision agriculture as well as for monitoring the colorimetry of a colored surface such as a leaf or soil.

A. SCA

SCA, Fig. 3, is a multi-sensor microsystem (MSM) for monitoring the presence of toxic chemicals (TICs), pollutants, volatile organic compounds (VOCs) and flammable gases in the air.

The SCA integrates temperature, relative humidity, ionic compounds, and metal oxide hotplate sensors. It uses temperature and voltage cycling combined with an electrical impedance spectrometer (EIS) to achieve better gas discrimination. The alternating current (AC) reading is less sensitive to drift, and the EIS allows the sensor's R/C equivalent circuit to be derived to decouple drifting components from gas response components, supporting drift mitigation algorithms. With its versatile analytical tools and integrated interdigitated electrodes, the SCA is also an excellent experimental board for novel sensing materials. Multiple SCAs can be installed on long cables for continuous monitoring of large areas.



Fig. 3. SCA sensor for air quality monitoring

In an agricultural land, the SCA can be installed in some strategic points in order to monitor the air quality and possibly the presence of pollutants harmful to the health of the plant. This sensor will be installed on special poles where it is possible to install a thermometer to monitor the air temperature, an anemometer and hygrometer to analyze the humidity currents and IR cameras facing the agricultural field that allow to monitor the surface temperature of the plant leaves and therefore its health status.

B. SCW

SCW, Fig.4, is a multi-sensor microsystem (MSM) for monitoring the presence of toxic chemicals (TIC), pollutants, hydrocarbons and organics in water.

At the heart of SCW is SENSIPLUS, Sensichips' microsensor platform that can interrogate on-chip and off-chip sensors with its versatile and accurate Electrical Impedance Spectrometer (EIS) and potentiostat. EIS analyses exploit the RedOx dynamics of catalytic noble metals to aid in chemical discrimination and measurement of conductivity and permittivity spectra. The on-chip

potentiostat is used for a variety of different voltammetric or amperometric measurements and for real-time pollutant discrimination.

By flowing the electrodes with overvoltage, the device prevents or mitigates the formation of biofilm. SCW is a complete and multiparametric microsystem for water analysis.

Thanks to its versatile analytical instruments and the availability of integrated interdigital catalytic electrodes, SCW is also an excellent experimental microsystem for discriminative measurement techniques.



Fig. 4. SCW sensor for irrigation water quality monitoring

This sensor can be applied in different agricultural fields such as monitoring irrigation water and wastewater. In particular, in an agricultural land one can think of installing several SCWs near the irrigation system monitoring pH, temperature and contaminants and the same reasoning can be applied for wastewater. Furthermore, one could think of installing this sensor in different strategic points of the agricultural soil to monitor the nutrients that are dissolved from the soil in the irrigation water near the tree trunks.

C. SCP

The SCP multispectral sensor, Fig. 5, features 12 LEDs covering most of the VIS to NIR spectrum, Fig. 6, and a SFH-2201 photodiode as a detector. The sensor operation is based on the emission of a discrete frequency modulated signal by each single LED on the SCP that is reflected by a colored target such as a leaf or a soil sample and then analyzed by a photodiode that converts it into a current signal. Each LED sensor will be associated with two currents analyzed by the photodiode, one IN-PHASE and one 90° phase-shifted QUADRATURE.

Any LED can be used for light emission and any other LED or photodiode can be used for measurement. Light emission and measurement can be performed with a lock-in amplifier modulation at a user-programmable frequency, both in-phase and quadrature components can be acquired. The use of the lock-in technique helps to reduce ambient light interference on the measurement. This sensor can be thought of as an integrated part of a drone that can ideally fly at a height of about 2.0 m from the ground monitoring the colorimetry of the latter, this data can provide information on the type of soil and any water shortage.

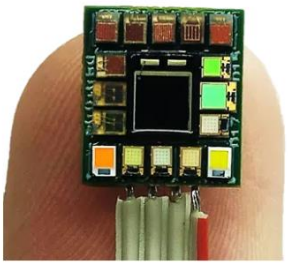


Fig 5. SCP multispectral sensor for monitoring plant and soil colorimetry

Furthermore, the spectral analyses from the drone can also be applied to the analysis of leaves from which the possible onset of diseases as well as a lack of nutrients and water can be analyzed.

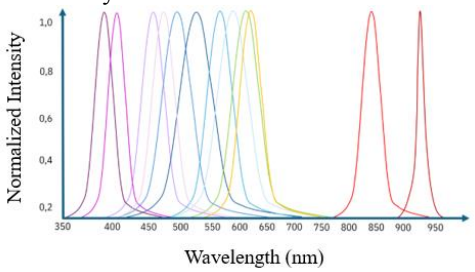


Fig. 6. Emission spectrum of the 12 LEDs present on the SCP multispectral sensor. As can be observed, they cover a large part of the VIS (10 LEDs) and a part of the NIR between 800 nm and 950 nm.

III. MEASUREMENTS

The measurements were performed using the software developed by Sensichips, SLM-Studio, which allows data to be collected from different sensors in a single device. This communicates with the various sensors via a Wi-Fi or ZigBee system that collects the measurements from the various sensors and communicates them to the PC in which the SLM-Studio software is present for data analysis. The measurements are driven through different measurement batches, which provide an alternating current, modulated in frequency, as inputs to the sensor and the collected data is inserted into a Zip folder containing an Excel file with the data.

IV. POST PROCESSING

The data analysis phase is performed through SLM-Studio which provides the possibility of training the various substances through three possible machine learning techniques developed directly by Sensichips in the case of the μDSP model and preloaded by the IT company UEKA in the case of the MLP and KNN models. At the end of the neural network training phase, SLM-Studio allows you to analyze the average accuracy of the

newly created model as well as the confusion matrix with the accuracies and standard deviations associated with the individual substances for which the model was trained. On the website, you can also consult the test data carried out with SLM-Studio in different experimental conditions that replicate the same training conditions and some small changes to the experimental setup.

V. CONCLUSIONS

The sensors described can be an integral part of an agricultural land where different chemical-physical properties of both the soil and the plants present can be monitored. This system can communicate the various data to SLM-Studio via microcontrollers that communicate with each other with a Wi-Fi or ZigBee system. The dataset available online on the Sensichips website can allow the creation of machine learning models with accuracies greater than 90% that allow the comparison of the raw data of the sensor with the training data available online. From this comparison it will therefore be possible to reveal any anomalies in the soil or plants, providing the farmer with important data on the possible timing of intervention to avoid damage to the plant and minimize losses to the harvest.

VI. ACKNOWLDGEMT

This work was co-funded by the European Union Horizon Europe program with the project URBAN M20 (GA 101180710).

REFERENCES

[1] F. Fina, F. Leccese, “The LED's spectrophotometry in precision agriculture: A brief survey”, Acta IMEKO, vol. 14 (2025) no. 1, pp. 1-15. DOI: 10.21014/actaimeko.v14i1.1970

[2] P. Musa, H. Sugeru, E. Prasetyo, “Wireless Sensor Networks for Precision Agriculture: A Review of NPK Sensor Implementations”, Sensors, 2023, doi: https://doi.org/10.3390/s24010051

[3] F. Leccese, “Remote-Control System of High Efficiency and Intelligent Street Lighting Using a ZigBee Network of Devices and Sensors,” in *IEEE Transactions on Power Delivery* , vol. 28, no. 1, pp. 21-28, Jan. 2013, doi: 10.1109/TPWRD.2012.2212215.

[4] F. Leccese, “A New Remote and Automated Control System for the Vineyard Hail Protection Based on ZigBee Sensors, Raspberry-Pi Electronic Card and WiMAX”, Journal of Agricultural Science and Technology, 2023, B. 3. 853-864.

[5] W. Bian, Y. Liu, Y. Wu, “A Bluetooth-Based Automated Agricultural Machinery Positioning System”, *Electronics* , *13* (24), 4902, (2024), DOI:https://doi.org/10.3390/electronics13244902

[6] K. Zajícová, T. Chuman, "Application of ground penetrating radar methods in soil studies: A review", Geoderma, Volume 343, 2019, Pages 116-129, ISSN 0016-7061, DOI: https://doi.org/10.1016/j.geoderma.2019.02.024 .

[7] D. Moret-Fernández, F. Lera, B. Latorre, J. Tormo, J. Revilla, “Testing of a commercial vector network analyzer as low-cost TDR device to measure soil moisture and electrical conductivity”, CATENA, Volume 218, 2022, 106540, ISSN 0341-8162, DOI: https://doi.org/10.1016/j.catena.2022.106540 .

[8] F. Adamo, F. Attivissimo, L. Fabbiano, N. Giaquinto, M. Spadavecchia, “Soil moisture assessment by means of compressional and shear wave velocities: Theoretical analysis and experimental setup, Measurement”, Volume 43, Issue 3, 2010, Pages 344-352, ISSN 0263-2241, DOI: https://doi.org/10.1016/j.measurement.2009.11.007 .

[9] M. Pischiutta, F. Linsalata, A. Mercuri, F. Salvini, F. Fina, G. Cultrera, L. Minarelli, G. Di Giulio, “Rock-site amplification on topography at Introdacqua, Central Italy: a rock-fracturing effect?”, *Acta Geophys,* (2025), DOI: https://doi.org/10.1007/s11600-024-01515-z

[10] M. Pischiutta, A. Mercuri, F. Fina, F. Salvini, L. Minarelli, G. Cultrera, G. Di Giulio, “Passive seismic survey across the fault-zone of Introdacqua, central Italy: the subsurface geological structure role on the site amplification pattern”, EGU General Assembly 2024, Vienna, Austria, 14–19 Apr 2024, EGU24-18014, https://doi.org/10.5194/egusphere-egu24-18014, 2024.

[11] H. Berger Roisenberg, F. Magrini, I. Molinari, L. Boschi, F. Cammarano, “Rayleigh wave attenuation and phase velocity maps of the greater Alpine region from ambient noise”, *Sci Rep* 14, 29164, (2024) , DOI: https://doi.org/10.1038/s41598-024-80729-z

[12] K. Grote, S. Hubbard, Y. Rubin, “Field-scale estimation of volu metric water content using ground-penetrating radar ground wave tech niques: Water Resources Research”, Volume 39, (2023), DOI: http://doi.org/10.1029/ 2003WR0245

[13] E. Pettinelli, A. Di Matteo, S. E. Beaubien, E. Mattei, S. E. Lauro, A. Galli, G. Vannaroni, “A controlled experiment to investigate the correlation between early-time signal attributes of ground-coupled radar and soil dielectric properties”, Journal of Applied Geophysics, Volume 101, 2014, Pages 68-76, ISSN 0926-9851, DOI: https://doi.org/10.1016/j.jappgeo.2013.11.012 .

[14] P. Tuccimei, M. Soligo. “Correcting for Interference in Soil Radon Flux Measurements.” Radiation Measurements 43, no. 1 (2008): 102–5. DOI: https://doi.org/10.1016/J.RADMEAS.2007.05.056

[15] A. Almaw, “Soil temperature Sensors in Agriculture and the role of Nanomaterials in Temperature Sensors Preparation”, (2018)

[16] Y. K. Kushwaha, R. K. Panigrahi, A. Pandey,

"Performance analysis of capacitive soil moisture, temperature sensors and their applications at farmer's field", *Environ Monit Assess* **196** , 793 (2024). DOI: https://doi.org/10.1007/s10661-024-12946-y

[17] M. Nadporozhskaya, N. Kovsh, R. Paolesse, L. Lvova, "Recent Advances in Chemical Sensors for Soil Analysis: A Review", *Chemosensors* **2022** , *10* , 35. DOI: https://doi.org/10.3390/chemosensors10010035

[18] U. B. Mahadevaswamy, N. Meghana, "Design and Development of Soil Moisture Sensor", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075 (Online), Volume-10 Issue-4, February 2021, DOI: https://doi.org/10.35940/ijitee.D8438.021042

[19] M. Mattia, G. Mondati, R. Mazza, C. Rosa, C. Di Salvo, P. Tuccimei, "Groundwater–River Water Interaction in an Urban Setting (Rome, Italy) Using a Multi-Method Approach (Hydrogeological and Radon Analyses)", *Water* , *17* (10), 1555, (2025), DOI: https://doi.org/10.3390/w17101555

[20] M. Vitelli, G. Cerro, L. Gerevini, G. Miele, A. Ria, M. Molinara, "SENSIPLUS-LM: A Low-Cost EIS-Enabled Microchip Enhanced with an Open-Source Tiny Machine Learning Toolchain", Computers. 12. 23. DOI: https://doi.org/10.3390/computers12020023

[21] A. Ria, M. Cicalini, G. Manfredini, A. Catania, M. Piotto, P. Bruschi, "The SENSIPLUS: A Single-Chip Fully Programmable Sensor Interface", In: Saponara, S., De Gloria, A. (eds) Applications in Electronics Pervading Industry, Environment and Society ApplePies 2021. Lecture Notes in Electrical Engineering, vol 866. Springer, (2022), DOI: https://doi.org/10.1007/978-3-030-95498-7_36

[22] M. Leccisi, M. Greco, G. Schirripa Spagnolo, E. De Francesco, F. Leccese, "Low Cost Multispectral Sensor for Monitoring Vine Leaf Density," *2024 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor)* , Padua, Italy, 2024, pp. 86-91, doi: https://doi.org/10.1109/MetroAgriFor63043.2024.10948775 .

[23] M. Greco, E. Giovenale, F. Leccese, A. Doria, "A Discrimination of Healthy and Rotten Hazelnuts Using a THz Imaging Scanner", 229-233, (2022), DOI: https://doi.org/10.1109/MetroAgriFor55389.2022.9964672.

[24] M. Greco, G. Schirripa Spagnolo, A. Lai, F. Bonfigli, E. Giovenale, F. Leccese, "Detection of Powdery Mildew Disease on Hazelnut Leaves by UV/VIS Fluorescence", (2024), 92-96, DOI: https://doi.org/10.1109/MetroAgriFor63043.2024.10948845.

[25] C.A. Santos, M. Lopo, R. N. M. J. Páscoa, J. A. Lopes. "A Review on the Applications of Portable Near-Infrared Spectrometers in the Agro-Food Industry." Applied Spectroscopy 67, no. 11 (2013): 1215–33. doi:https://doi.org/10.1366/13-07228.

[26] A. K. Mahlein, B. Arnal, G. Jayme, KS Chiang, E. M. Del Ponte, CH Bock, "From Detection to Protection: The Role of Optical Sensors, Robots, and Artificial Intelligence in Modern Plant Disease Management, Phytopathology", Vol 114, 1733-1741, 2024, DOI: https://doi.org/10.1094/PHYTO-01-24-0009-PER

[27] Y. Qm, W. Zhang, "Estimation of Leaf Water Content of Different Leaves from Different Species Using Hyperspectral Reflectance Data", Annals of Agricultural Sciences, Vol 7, 1111-2022, (2024), DOI:https://doi.org/10.26420/annagriccropsci.2021.1111.

[28] K. Zhang, W. Li, H. Li, Y. Luo, Z. Li, X. Wang, X. Chen, "A Leaf-Patchable Reflectance Meter for In Situ Continuous Monitoring of Chlorophyll Content", *Adv. Sci.* 2023, 10, 2305552. DOI: https://doi.org/10.1002/advs.202305552

[29] I. Kasajima, "Measuring plant colors", Plant *Biotechnology*, (2019), Vol 36, DOI: https://doi.org/10.5511/plantbiotechnology.19.0322a.

[30] N. Kirillova, T. Sileva, Colorimetric analysis of soils using digital cameras. Moscow University Soil Science Bulletin, Vol 72, 13-20, (2017) DOI: https://doi.org/10.3103/S0147687417010045

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Characterization of the lighting system of hospital rooms

Marian-Andrei Vieru[1], Cristian-Győző Haba[1],

[1] *"Gheorghe Asachi" Technical University of Iasi-Romania, Faculty of Electrical Engineering, Energetics and Applied Informatics, 21-23 Prof. D. Mangeron Blvd., 700050, Iasi, Romania,*
marian-andrei.vieru@student.tuiasi.ro, cristian-gyozo.haba@academic.tuiasi.ro

***Abstract* – Artificial lighting in the hospital areas is equally essential for medical staff, patients, and their visitors. Adequate lighting can have beneficial effects on human health by contributing to the patient's recovery process and also increasing the performance and efficiency of medical staff. In contrast, inappropriate lighting generates harmful effects on health and well-being. Thus, parameters such as luminous intensity, luminance, luminous flux, and illuminance provide information regarding the lighting system. Therefore, determining these parameters is a key factor in establishing and determining possible negative effects as a result of prolonged exposure to artificial light.**

In this context, a systematic study was carried out to determine the correlated color temperature, light intensity and UV index of artificial light from three hospital areas, including waiting areas, patient ward rooms, intensive care units, and operating rooms, and to evaluate the quality of the lighting system.

I. INTRODUCTION

Light is essential for human health and well-being because it enables vision and daily activities. Over time, artificial lighting has evolved, marking major changes in the source of illumination. This progression spans from the utilization of candlelight to incandescent light and, most recently, to the use of LEDs [1]. Today in the modern world, artificial light is near-ubiquitous, being used in the medical field for sterilization of medical equipment, phototherapy, endoscopy and laparoscopy, in phototherapy, ambient lighting, and many other fields.

Numerous studies highlight the significant role of light in human psychological and physiological health, having both harmful effects on vision (contrast perception, myopia, dry eye syndrome, macular degeneration, glaucoma) and non-vision (depression, anxiety, gastrointestinal and cardiovascular diseases, breast cancer) but also, beneficial effects in decreasing depression, reducing fatigue, increasing alertness and regulating circadian rhythms [2]. These effects are generated by different radiometric and photometric parameters.

Comparing natural outdoor light with artificial light, significant differences regarding radiometric and photometric information, including the dimmer, spectral distribution, and illuminance and luminance levels, can be observed [3]. Despite the progress made in recent years, it has been found that no lighting technology has been able to reproduce all the characteristics of natural light to date. For example, indoor lighting has illuminance values of less than 1000 lux [4], while the sunlight can reach values up to 100000 lux [5].

In hospitals, both patients and medical staff are exposed to artificial light for a long time. Light intensity is essential to ensure adequate visibility and reduce the risk of medical errors, but insufficient or excessive lighting can cause visual discomfort and eye fatigue [6]. Correlated color temperature (CCT) directly impacts alertness and psychological comfort. The study conducted by Cajochen et al.[7] shows that cool light (5000–6500 K), rich in blue components, increases alertness and cognitive performance, being suitable for active medical spaces, such as operating rooms or treatment rooms. On the other hand, Knez and Kers [8] show that warm light (2700–3500 K) is perceived as more relaxing and contributes to a state of emotional comfort, thus being suitable for wards and rest areas. The information is also supported by applied studies in human-centered lighting, conducted by the Lighting Research Center [9]. Regarding UV radiation, it must be kept at safe levels, as prolonged exposure can damage the skin and eyes. Although UV-C technology is effectively used for disinfection, it is strictly applied in the absence of people, according to the safety protocols recommended by the WHO [10]. Adapting these characteristics to each area of the hospital significantly contributes to patient safety, effective recovery, and the performance of medical staff.

Therefore, it is important to determine the correlated color temperature (CCT), variation in light intensity, and UV index through standard metrics and to evaluate the possible effects according to the characteristics of the light.

In this context, a systematic study was carried out to determine the correlated color temperature, light intensity, and UV index of artificial light from three hospital areas. The main focus was on operating rooms, intensive Care units, and patient ward rooms.

II. MATERIALS AND METHODS

The development of the final device involves a hardware part (the integration of the sensors and other components) and a software part (the application development).

A. System description - Hardware

The hardware part integrates an analog UV light sensor GUVA-S12SD, two RGB color sensors TCS34725, and one luminosity sensor TSL2561. Two TCS34725 sensors are placed at a 90° angle to each other. One of the sensors is oriented vertically to capture light from the ceiling, and the other is oriented laterally to capture light from the walls, thus ensuring complete coverage of the measurement field and obtaining accurate data on the color of the light from different angles. To perform 360° scan measurements automatically, the sensor platform is rotated by a stepper motor GM12-15BYC ratio 1:30, controlled by a motor drive module EasyDriver A3967. Using an HC-06 Bluetooth module and Arduino module, the entire system can be controlled remotely using a smartphone application specially designed for performing measurements, saving, and sharing the obtained data. The aim is to develop a portable system with low energy consumption that allows easy control, including remotely, to ensure the safety of operators in high-risk areas, such as radiology rooms, isolators, and other similar environments.

B. System description - Software

The part of the system responsible for data acquisition runs a custom code developed in Arduino IDE, which integrates all the hardware components, while the part responsible for setting parameters, receiving in real-time data, saving and sharing them also runs a custom code developed in MIT App Inventor. The mobile application has implemented sections that allow setting the scan rotation values (in degrees), setting the values for the delay between steps (in seconds), and entering the name of the location where the measurement was performed. The control button section contains: a BT-C button - to connect the application to the device, a BT-D button - to disconnect the application from the device, a START button - to start the scan after the scan values have been entered, a STOP button - to stop the scan instantly, a RESET button - to bring the device to the initial position 0, a SAVE button - to initiate saving documents to a .csv file, and a SHARE button - to be able to share the file via methods such as email, message, cloud, etc. The application also contains a status bar where notifications are displayed, such as connection status, measurement stages, data retrieval, data saving, data export, etc.

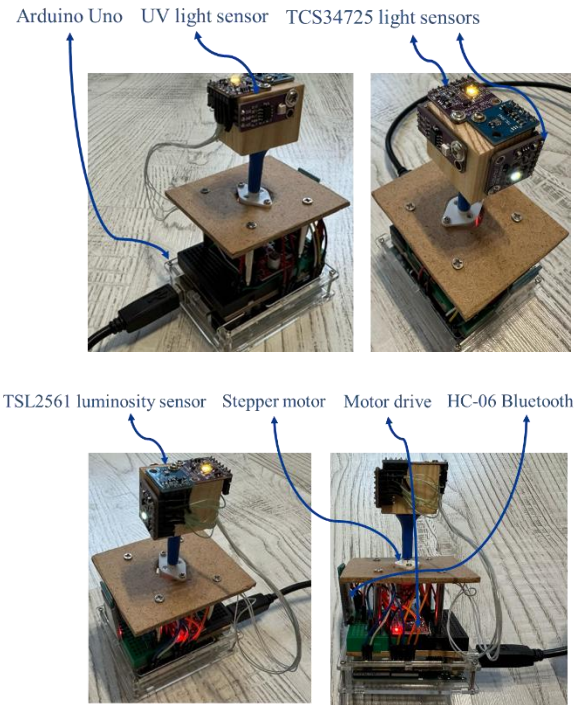


Fig. 1. The hardware components of the portable device.

C. Methods

The experimental measurements were carried out as a 360° scan of the space, with a one-second delay between steps. The device was set up at a distance of one meter from the light sources. It was done three measurements for each area, and one measurement takes around ten minutes.

III. RESULTS AND DISCUSSIONS

According to the European standard EN 12464-1:2021, the recommended illuminance levels vary significantly depending on the hospital area. For patient wards, the standard suggests an average illuminance between 100 and 300 lux, depending on the specific tasks performed and the space layout. Intensive care units typically require around 300 lux for general lighting to ensure adequate visibility for patient monitoring and staff activity. In contrast, operating theatres demand substantially higher lighting levels, with general lighting recommended at a minimum of 1,000 lux, and surgical field lighting often exceeding 10,000 lux. Regarding the correlated color temperature (CCT), lower values—typically in the range of 3,000 to 4,000 K—are preferred in patient recovery areas to support visual comfort and circadian alignment. In operating rooms, however, higher CCT values, often between 4,500 and 5,000 K, are used to improve visual acuity and accurate color rendering, which are critical during surgical procedures.

The experimental measurements were performed in three hospitals, including operating rooms, intensive care units, and patient ward rooms. Table 1 shows the main

characteristics of the rooms in which the determinations were performed.

Table 1. The general parameters of the hospital rooms.

Operating rooms (OR)				
Surface (m2)		Window	Type of light	Number of lighting
H1	56	3	LED	4
H2	60	4	LED	4
H3	59	3	LED	5
Intensive care units (ICU)				
Surface (m2)		Window	Type of light	Number of lighting
H1	20	2	LED Fluorescent	4
H2	25	2	LED Fluorescent	4
H3	22	2	LED Fluorescent	4
Patient ward rooms (PWR)				
Surface (m2)		Window	Type of light	Number of lighting
H1	10	4	Incandescent	2
H2	12	4	Incandescent	2
H3	12	4	Incandescent	2

A. Illuminance levels

According to international and national standards *[11]* the general lighting in operating rooms is necessary to be more than 1000 lx.

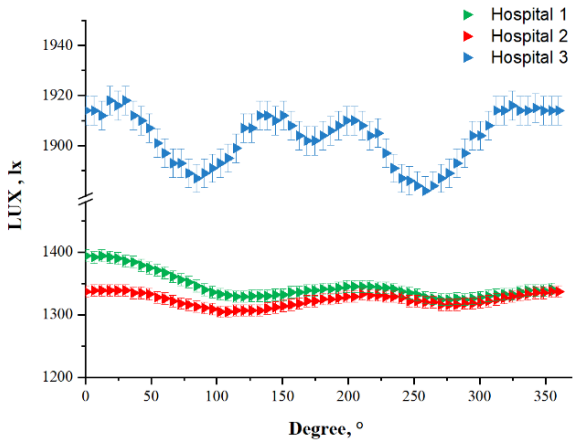


Fig. 2. The illuminance values for the operating rooms.

In the case of hospital 3, the illuminance values show the most significant oscillations, with the highest number of minima and maxima. These oscillations are not completely symmetrical, which may indicate significant variations in light sources.

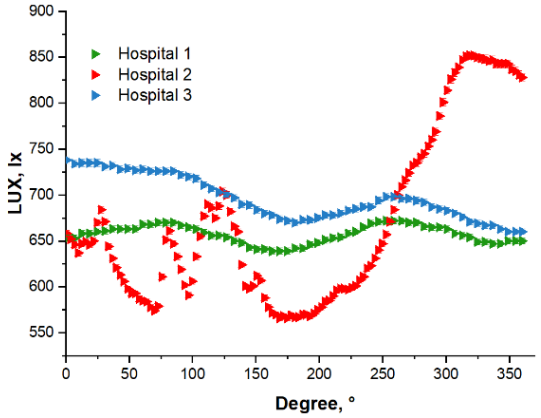


Fig. 3. The illuminance values for the intensive care units.

Fig. 3 presents the illuminance profiles from intensive care units in three hospitals. For hospital 1, the illuminance values show two clear cycles of light variation without abrupt fluctuations. The lighting exhibits a slightly oscillating behavior, which could affect the uniformity of light. The same behavior could also be identified in the case of hospital 3. In the case of hospital 2, the illuminance values show multiple peaks and sharp drops without a familiar periodic pattern. The lighting has a generally higher intensity but lacks consistency in terms of stability.

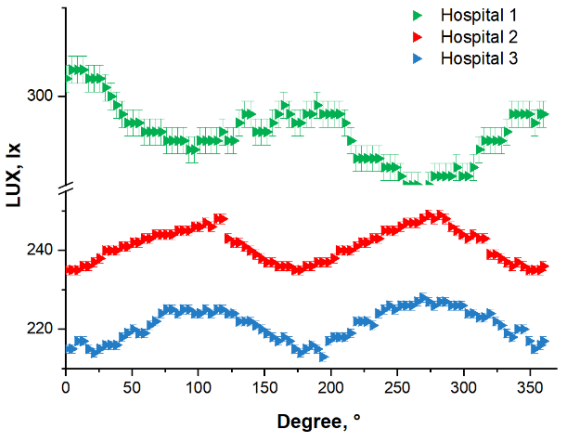


Fig. 4. The illuminance values for the patient ward rooms.

In the last examined area, hospital 1's illuminance values show a gradual decrease with slight fluctuations in the middle section. For hospital 2, the illuminance values indicate a double sinusoidal pattern, alternating exposure to a strong light source. The space has two intense light spots and transition zones between them, suggesting a directional lighting distribution. In the case of hospital 3, the illuminance values show a sinusoidal pattern, suggesting light coming from multiple directions, with

wide zones of consistent illumination.

B. Correlated color temperature

As can be seen from Fig. 5 a), the CCT values start high, followed by an abrupt drop, after which a stabilization appear, indicating a transition from cold to neutral light, while in the case of hospital 2 (Fig. 5 b)) the CCT values remain stable initially, followed by a sharp jump. The light presents a neutral color with a stable chromatic spectrum and intensity fluctuations. The CCT values of the operating room from hospital 3 (Fig. 5 c)) present a rapid drop at the beginning, followed by stabilization, indicating warmer lighting.

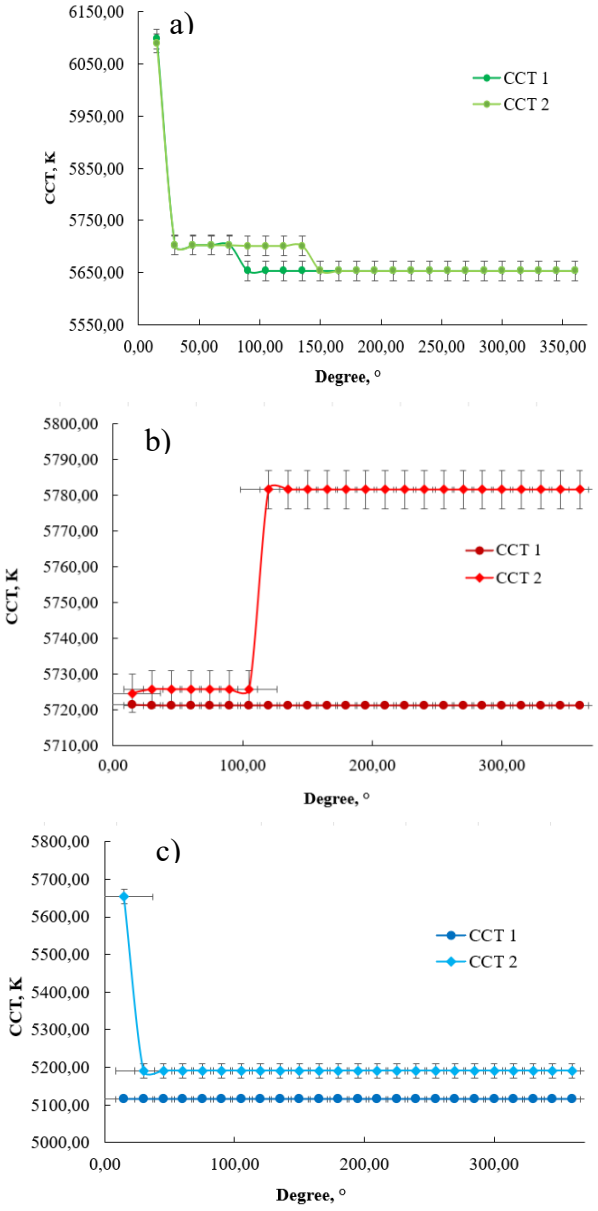


Fig. 5. The CCT values for operating rooms from a) hospital 1, b) hospital 2, and c) hospital 3.

Figure 6 presents the CCT values of the intensive care units determined in the second area. The CCT values for hospital 1 reflect a uniform lighting system, with small areas sensitive to color temperature. In the case of hospital 2, the CCT values reflect a very stable system with ideal behavior. In the last hospital (Fig. 6 c)) the CCT values suggest good and consistent color calibration of the lighting system.

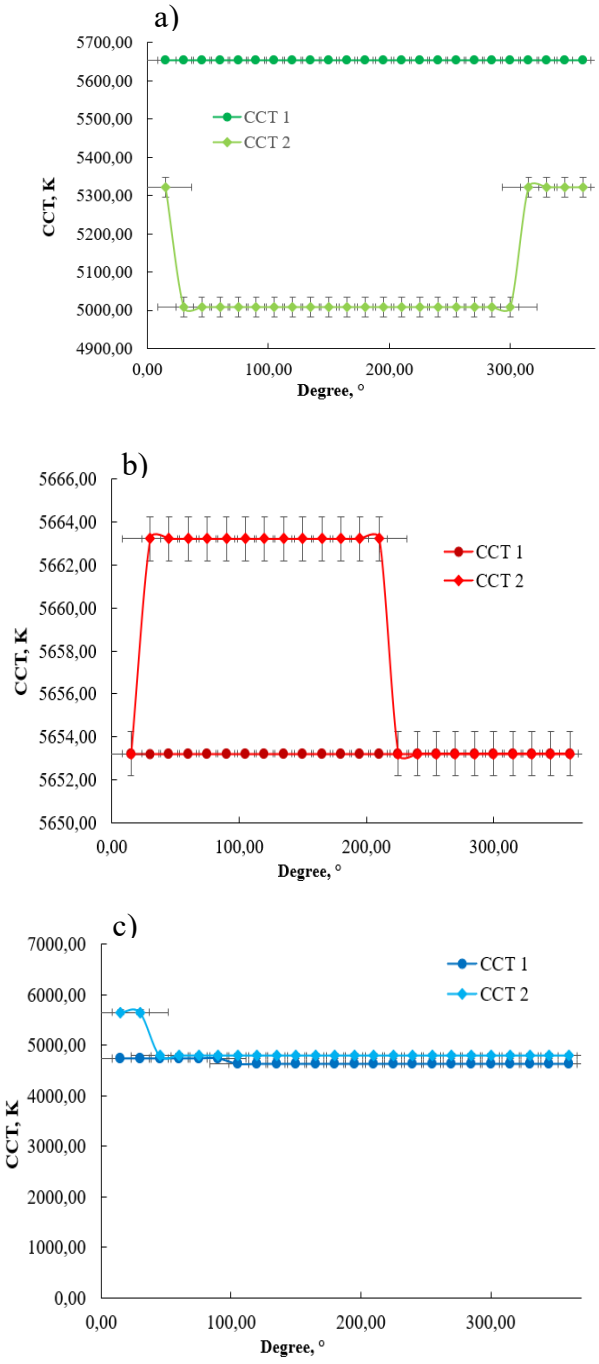


Fig. 6. The CCT values for intensive care units from a) hospital 1, b) hospital 2, and c) hospital 3.

In the patient ward room, for hospital 1, the CCT values initially show variation, followed by stabilization at constant levels, resulting in minimal changes. Fig.4 c) presents a high decrease followed by steady stabilization. This behaviour indicates a stable color system.

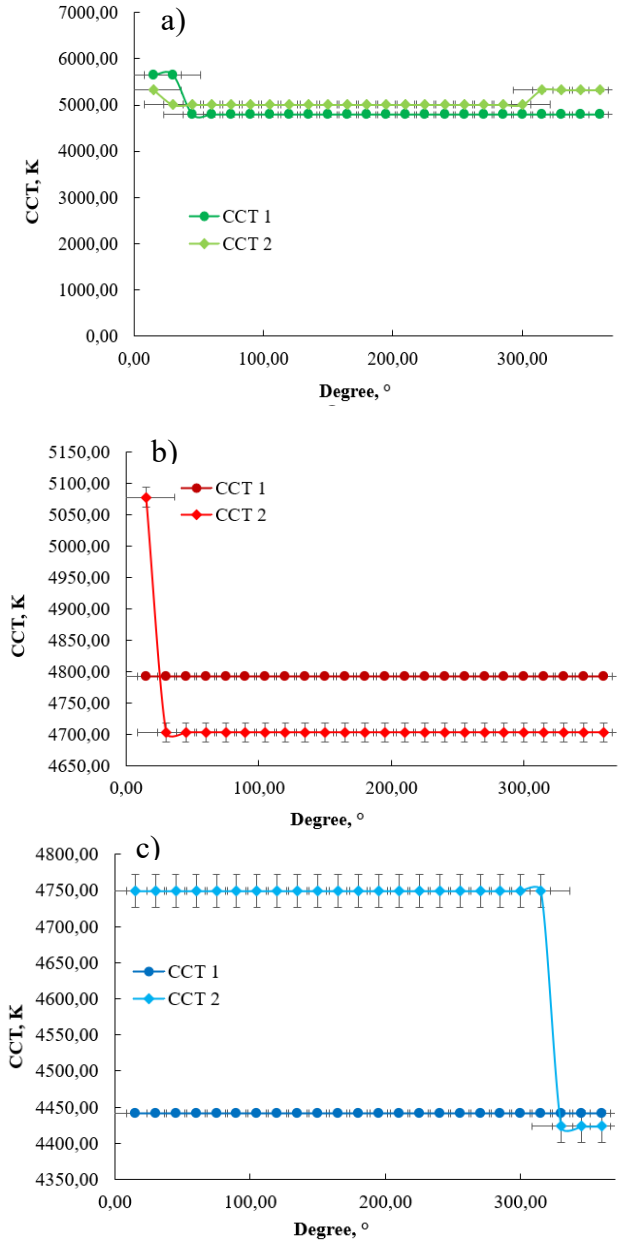


Fig. 7. The CCT values for the patient ward room from a) hospital 1, b) hospital 2, and c) hospital 3.

C. UV index

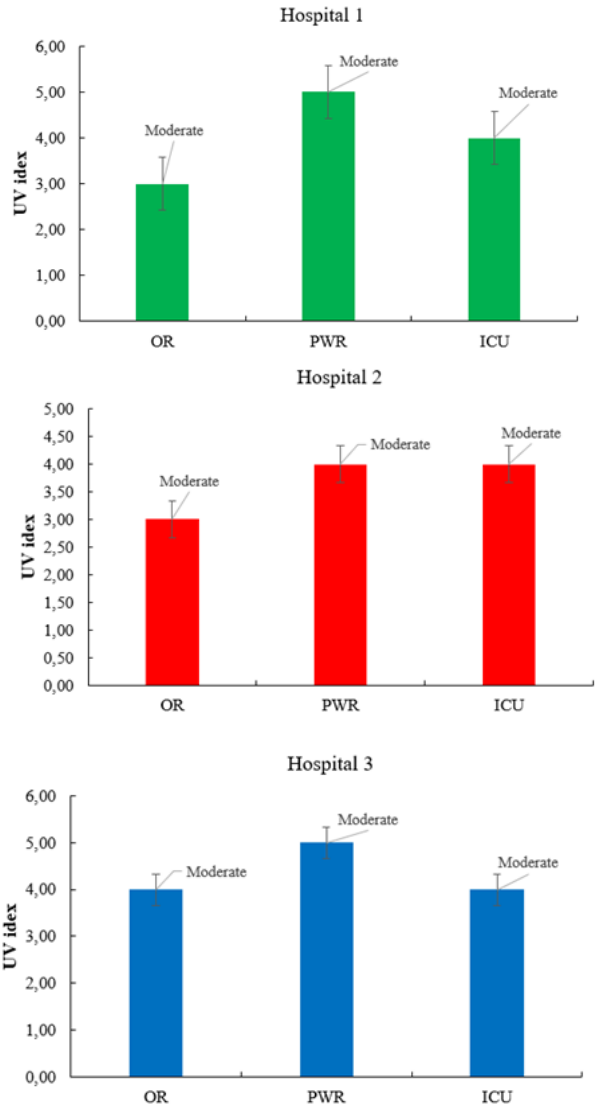


Fig. 8. The UV index in the examined areas.

As can be seen from the presented graphs, the UV index measured in each room from each hospital is in the moderate limits, according to the World Health Organization scale.

IV. CONCLUSIONS

A portable device was made using a microcontroller to be able to acquire data from the area of interest, using specific sensors added to the stepper motor. The acquired data is transmitted via Bluetooth in real time to the specially designed smartphone application. The application is responsible for setting the measurement parameters, receiving the data, saving them in the smartphone's internal memory, as well as sharing using applications as mail, cloud, message etc. The experimental

data show that Lighting levels, UV index, and Correlated color temperature are, with small variations, in accordance with International standards.

REFERENCES

[1] Moreno, O. *et al.* (2024). Redefining artificial lighting through spectral engineering of light sources for well-being. *Scientific Reports,* 14, 26298. https://doi.org/10.1038/s41598-024-78315-4.

[2] Mehrotra, S. *et al.* (2015). Effective Lighting Design Standards Impacting Patient Care: A Systems Approach. *Journal of Biosciences and Medicines*, 3(11), 54-61. https://doi.org/10.4236/jbm.2015.311006.

[3] Wong, N.A. and Bahmani, H. (2022). A review of the current state of research on artificial blue light safety as it applies to digital devices. *Heliyon*, 8(8), e10282. https://doi.org/10.1016/j.heliyon.2022.e10282.

[4] Daugaard, S. *et al.* (2019). Light Exposure during Days with Night, Outdoor, and Indoor Work. *Annals of Work Exposures and Health*, 63(6), 651-665. https://doi.org/10.1093/annweh/wxy110.

[5] Kim, C.-H. and Kim, K.-S. (2019). Development of Sky Luminance and Daylight Illuminance Prediction Methods for Lighting Energy Saving in Office Buildings. *Energies*, 12(4), 592. https://doi.org/10.3390/en12040592.

[6] (2011). European Committee for Standardization (CEN). EN 12464-1:2011 – Light and lighting – Lighting of work places – Part 1: Indoor work places. Brussels: European Standard.

[7] Cajochen, C. *et al.* (2005). High Sensitivity of Human Melatonin, Alertness, Thermoregulation, and Heart Rate to Short Wavelength Light. *The Journal of Clinical Endocrinology & Metabolism*, 90(3), 1311-1316. https://doi.org/10.1210/jc.2004-0957.

[8] Corraliza, J.A. and Berenguer, J. (2000). Environmental Values, Beliefs, and Actions: A Situational Approach. *Environment and Behavior*, 32(6), 832-848. https://doi.org/10.1177/00139160021972829.

[9] Lighting Research Center (LRC), Rensselaer Polytechnic Institute. Light and Health Program – Human-Centric Lighting Studies.

[10] World Health Organization (WHO). (2020). UV radiation and health – Use of UV disinfection technologies against COVID-19.

[11] Hemphälä, H. *et al.* (2020). Towards better lighting recommendations for open surgery. *Lighting Research & Technology*, 52(7), 856-882. https://doi.org/10.1177/1477153520903355.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - M4DConf 2025
Benevento, Italy, September 3-5, 2025

Machine-readable data for measurements of total luminous flux using goniophotometer: a step toward digitalized metrology

Khaled M. Ahmed [a,*], *Mohammad D. Almelfi* [a]

[a] National Measurement and Calibration Center, Saudi Standards, Metrology and Quality Org. (SASO-NMCC), Riyadh, Saudi Arabia,
*Corresponding author: k.abdelftah@saso.gov.sa / khaled55eg@gmail.com

Abstract **– The transition to digital metrology requires structured and machine-readable data for traceable measurement systems. This article presents a methodology for developing machine-readable data formats IES/LDT for total luminous flux measurements from Type C goniophotometers. A proposed data model aligns with digital calibration certificate (DCC) projects, enabling improved traceability, automation, and interoperability. IES (Illuminating Engineering Society) files are ASCII text files that contain photometric data for lighting fixtures. They are machine-readable and contain detailed information about how light is distributed from a fixture, including luminous intensity, distribution angles, and other relevant information. Lighting design tools and rendering software can use this information to accurately simulate how a fixture would behave in a virtual environment. An actual case study from an SASO-NMCC (National Metrology Institute of Saudi Arabia) illustrates how these methods enhance calibration and function with smart lighting systems.**

Keywords**:** Digital Metrology, Machine-readable, DCC, energy efficiency, modern lighting, total luminous flux, average luminous intensity, space measurements, IES format.

1. MOTIVATION

Since it shows how much light a lamp emits into a space, the measurement of luminous flux, or lumen, is essential for both economic growth and energy efficiency. Solid state lighting's (SSL) luminous efficiency is driving up the need for total luminous flux measurement. To replicate the amount of illumination and beam pattern on a surface, goniophotometry is used to create photometric data files that can be incorporated into lighting design software [1-3]. Light sources, luminaires, system realization, and lighting usage are some of the variables that affect a lighting system's energy efficiency. Goniophotometers, which can accommodate either 2 π or 4 π motion depending on their design, are commonly used to measure luminosity and luminosity intensity distributions. Goniophotometers come in three varieties: stationary light sources, fixed-head goniophotometers, and fixed-head goniophotometers that rotate the light source only on one axis. Each type has its limitations and can result in inaccuracies. To ensure accurate goniometric measurements, lamps must be mounted in the intended orientation and the burning direction remains constant with respect to gravity. By understanding these factors and using goniophotometers effectively, designers can optimize their lighting systems and improve overall energy efficiency [1-2].

A precise measurement of the total luminous flux is crucial for lighting performance. Traditional calibration certifications are not made to function well with digital systems and can be read by humans. The shift to machine-readable outputs is consistent with global digital metrology initiatives. By combining digital formats with goniophotometers, photometric traceability can be modernized.

IES files and digital metrology are crucial in modern lighting engineering, enabling accurate simulations and digitalization. IES files, developed by the Illuminating Engineering Society (IES), describe light distribution characteristics of luminaires, aiding in architectural lighting design [1-5]. Conversely, digital metrology is concerned with digitizing measurement procedures while guaranteeing data interoperability, accuracy, and traceability. It improves accessibility, decreases manual data transfer, and facilitates well-informed decision-making. Light meters and other measuring devices are guaranteed to be accurate by digital calibration certificates (DCC). Advanced lighting systems and automated quality control procedures are being produced as a result of the integration of digital metrology with smart manufacturing and the Internet of Things. When combined, they make the lighting engineering workflow more effective [6-15].

Lighting system digitization is a quickly developing trend that provides smart lighting solutions with new features, enhanced user experience, and energy savings. In order to enable features like human-centric lighting and Internet of Things integration, this shift entails switching from traditional, analog lighting technologies to digital ones like LEDs and intelligent controls. Energy efficiency and sustainability, smart lighting applications, and digitalization enablers like photonics, data analytics, artificial intelligence, and cybersecurity are important facets of digitalization. Smart street lighting networks in smart cities, urban digital twins, and smart highway lighting are a few instances of lighting that has undergone digitalization. Standardizing communication protocols, comprehending user preferences, cutting costs, and implementing strong security measures are among the difficulties. On the other hand, lighting

system digitization is a revolutionary trend that could completely change lighting solutions and result in more user-friendly, efficient, and sustainable options [5-13].

2. GONIOPHOTOMETRE MEASUREMENT PRINCIPLE AND DIGITALIZATION

A Type C goniophotometer is used to measure the luminosity intensity at different angles. The total luminous flux is determined by integrating the angular intensities over a sphere. The photometer head's spectral response, electrical stability, geometry, and ambient conditions all have an impact on accuracy. Calibration is done using standard lights that can be linked to national luminous flux standards.

IES and LDT files are essential in digital metrology of lighting, providing a traceable, structured, and machine-readable representation of lighting characteristics. They serve as digital counterparts to photometric measurements, enabling verification of luminous intensity distribution, integration into lighting simulation software, development of digital calibration certificates, and automated conformity assessment. Key contents of IES/LDT files include luminous intensity distribution, total luminous flux, beam angles and cutoff, luminaire geometry, and test lab and equipment metadata. These files offer data traceability, machine-readability, quality assurance, and integration with digital calibration certificates. Thus, interpreting and managing IES/LDT files within a digital metrology framework supports automation, transparency, and interoperability in lighting calibration and certification workflows.

3. SYSTEM DESIGN AND AUTOMATION SOFTWARE FOR MACHINE-READABLE DATA FORMATS TO FACILITATE DCC INTEGRATION

Our recently constructed goniometer system is made up of numerous parts and apparatus. This home-made type-C goniometer was created to meet the requirements of our laboratory applications, including the standard use at the National Metrology Institute (NMI) and the pertinent industry uses. Using C++ code and specifically designed mathematical algorithms offers full automation and control of the system, including the output files. Both photometer (f_1' <3.0%) and spectroradiometer (CCD based type) are included in the system, and their heads are positioned in a flexible manner that makes them suitable for a wide range of applications, such as average luminous intensity distribution (ALI) with high angle resolution (±0.1 deg). Geometrical design and occupied space dimensions of the system are shown in figure 1.

Among other parts, the system includes rotational motors, a DC extremely stable power supply, a standard shunt resistor, a digital multimeter, a sourcemeter, a photometer readout unit, a scanning controller, a connecting unit, and an automatic voltage regulator (AVR). The system's primary components and alignment processes using two integrated perpendicular lasers are depicted in Fig. 1, while Fig. 2 provides alignment process realization. Figure 3 presents screenshots of software graphical user interfaces for improved visibility of the software environment and employed equipment. With features like complete and stand-alone modes, data collection and visualization, and the ability to create 2D charts of the distribution of luminous intensity, the software is ready to enable full control and machine-readable output files.

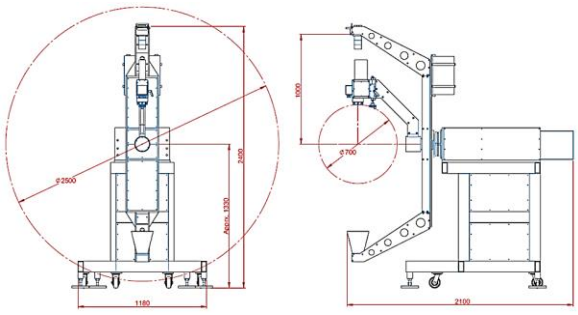


Figure 1. Geometrical design and dimensions of the custom-designed goniometer for light efficiency applications.

The change in luminous intensity with angle from a light source is measured by a goniophotometer. For lighting goods, goniophotometers are frequently used to create "photometric data files" in the IESNA (.ies) or EULUMDAT (.ldt) formats. The testing of general or architectural lighting goods, UVC germicidal lighting, vehicle lighting (headlight and marker), flashing visual alert systems, road traffic signals, and airfield and aviation lighting are among the uses for goniophotometers. There are motion goniometer stages of IES types A, B, and C available, and some models can be converted between motion types. Because of its intrinsic versatility, a single goniophotometer can be used to test directional vehicle lighting as well as general lighting goods.

The creation of photometric data files is among the most frequent tests we are asked to complete at Photometric Testing. The contents of an EULUIMDAT file are a mystery to many people. Photometric data files are used to measure the total luminous flux and luminous intensity of a light source. Total luminous flux is the total amount of light emitted from a light source, corrected for the human eye's spectral response. It is measured in lumens, while luminous intensity defines the amount of lumens in a given direction, per solid angle. A typical photometric file at Photometric Testing contains luminous intensity values for different angles and the input electrical power that the light source consumes (Watts).

There are three formats for photometric data files: IES, EULUMDAT, and TM-14. The IES file is the most common format in North America and widely used in Europe. EULUMDAT, also known as LDT, is the de-facto industry standard photometric data file in Europe. Despite its drawbacks, LDT files are still widely used and most customers request their data in both formats.

The IES and LDT formats can be formatted as absolute or relative files. Absolute photometric measurements record luminous intensity values, while relative measurements

measure candela values of the luminaire but the total luminous flux of the lamp is measured separately from the luminaire itself. This allows for the measurement of the efficiency of the luminaire by comparing the bare lamp lumens to the lumens of the luminaire. However, relative photometry can lead to misleading results due to the thermal performance of an LED. Luminous intensity data needed for photometric files is generated by a goniophotometer, which measures luminous intensity at each angle at a set distance. Most photometric files are generated by an automatic goniophotometer, which rotates the device under test through two axes. The data presented in photometric files is useful for lighting designers to observe both the total light output and the angular spread of the light output. Text files called IES and LDT (EULUMDAT) are used to characterize a light source's beam pattern. They can be used as inputs to determine the patterns of light intensity in complex lighting scenarios using a variety of free optical analysis applications.

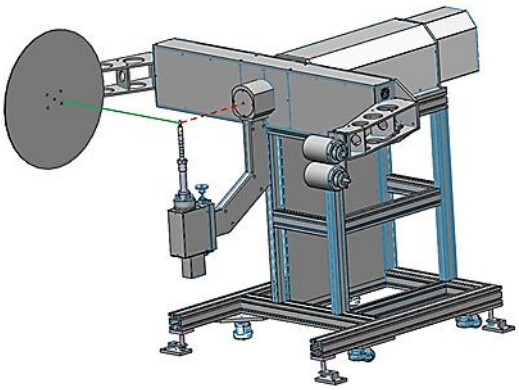


Figure 2. Alignment process visualization (automated and remotely controlled).

3.1. Use and investigation of IES/LDT files:

IES files are ASCII text files used to represent photometric data for lighting fixtures. They are machine-readable and contain detailed information about light distribution, including luminous intensity, distribution angles, and other characteristics. These files enable lighting design software and rendering programs to accurately simulate fixture behavior in a virtual environment. IES files are formatted in a specific format defined by the IES standard, with key components including file version, luminaire, manufacturer, photometric data, and angular data. Data elements include luminous intensity, angular data, lamp information, fixture geometry, and other relevant data. IES files are machine-readable, software compatible, accurate simulation, energy efficiency optimization, and reduced prototypes. They enable virtual prototyping, reducing the need for physical prototypes and mockups. Overall, IES files are a standardized, machine-readable format that enables lighting professionals to accurately simulate and analyze the photometric behavior of light fixtures in various applications. Examples of optical design software that employ IES and LDT files for scene design with various light sources are Dialux and Relux. Although LED component beam patterns can also be imported, the database repositories for these tools are mostly focused on luminaires. If an investigation does not require the near field features of the light source, IES files can be used as inputs to raytracing algorithms. Using these files (also known as apodization) has the benefit of allowing for arbitrary ray count adjustments.

4. SOFTWARE DESCRIPTION

The code of our software prepared using C++. Our software GUI, as shown in figure 3, is divided into four main areas:

1- Toolbar, 2- Current Status, 3- Command/Test and 4- Measurement.

Where, Toolbar comprises connect, configure, measure, open and save file and export IES files. Connect button is used to connect the devices and check for initialization. Config button is used for operating the photometer and spectrometer. Measure button is used to start the measurement based on the entered information. Open is used to load a saved file. Save is used to save the currently measured data. Export IES button is used to export the measured data in IES format.

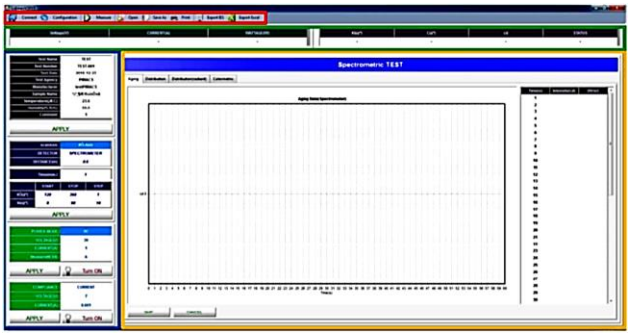


Figure 3. Graphical user interface (GUI) and its components. Red: Toolbar, Green: Status bar, Blue: Command/Test (Settings) modules and Yellow: Measurement results area.

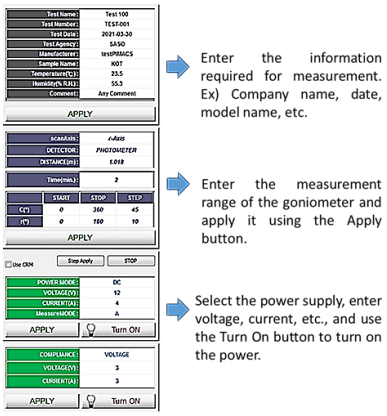


Figure 4. Command/Test modules for setting parameters with control buttons and boxes to set all parameters during scans and measurements.

Moreover, Current Status module displays various data, including the current goniometer position. While, Test Information/Parameter displays test information and sets various parameters to be used in the test. The Measurement

module displays measured light distribution data. The configuration varies depending on the selection of the photometer or spectrometer. As shown in figure 4, the Command/Test area comprises many control modules that are shown and described in the figure. Figures 5 through 8 show the GUIs of different modules and working functions of the software as well as snaps from the output files and their formats.

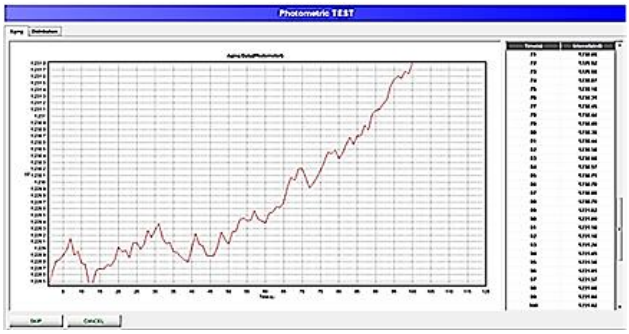


Figure 5. Aging window for optimal stability control.

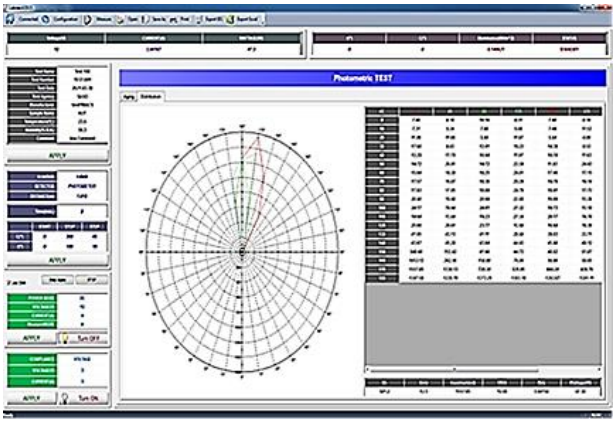


Figure 6. The measured intensity distribution data displayed in the grid and all data shown on the right columns, where each column represents azimuthal angle.

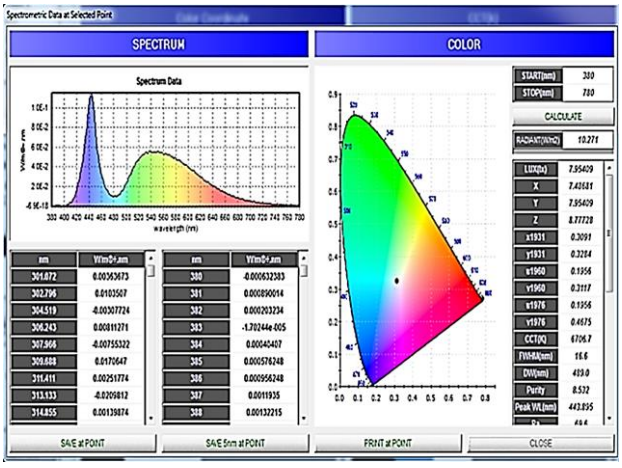


Figure 7. SpecData: the spectrum and color data concluded from the measured spectral data.

5. CONCLUSIONS

Modern lighting engineering relies heavily on digital metrology and IES files, especially when it comes to simulation and digitization. The Illuminating Engineering Society (IES) created IES files, which are standardized text files that carry light fixture photometric data. This allows software to simulate lighting accurately. Conversely, digital metrology is concerned with digitizing measurement procedures to guarantee data interoperability, accuracy, and traceability, particularly in the context of smart manufacturing and the Internet of Things.

For luminous flux, machine-readable calibration findings enhance digital integration, automation, and traceability. Advantages such as facilitating automated calibration processes - facilitating interaction with intelligent sensor networks Complies with future DCC structures can be counted. Future difficulties like; international schema harmonization required Data security version control and Limitations of legacy hardware need to be considered and challenged. For more openness, the results were connected to a traceability database built on the blockchain. The amount of time needed to generate reports dropped by 70%.

ACKNOWLEDGMENTS

Authors thank the “Research and Studies Center (RSC)” of SASO for their coordination and consideration of this work.

REFERENCES

[1] CIE 84-1989: **The Measurement of Luminous Flux**.

[2] CIE 121-1996: **The Photometry and Goniophotometry of Luminaires**.

[3] Khaled M. Ahmed, **Development of novel goniometric system for lumen realization and measurements of light sources with detailed uncertainty analysis**, Measurement: Sensors, Volume 38, Supplement, Feb. 2025, 101821.

[4] BIPM, **Digital Calibration Certificate Guidelines**, 2023.

[5] CIE 121-1996: Photometry — **The ClE System of Physical Photometry.**

[6] ISO/EC 17025:2017 **General requirements for the competence of testing and calibration laboratories**.

[7] IMEKO TC4, **"Metrology for Digital Transformation**," Conf. Proc., 2022.

[8] Khaled M. Ahmed, Hamad AlGhamdi, Homood AlOtaibi, “**An intelligent web-based platform for processing Digital Calibration Certificates utilizing the GEMIMEG tool and advanced AI capabilities**“, Measurement: Sensors, Volume 38, Supplement, May 2025, 101485.

[9] PTB, **SmartCom Project website**. Online [Accessed 15 April 2025] https://www.ptb.de/empir2018/smartcom/home/

[10] EURAMET, **European Metrology Programme for Innovation and Research (EMPIR).** Online [Accessed 15 April 2025] https://www.euramet.org/research-innovation/research-empir/

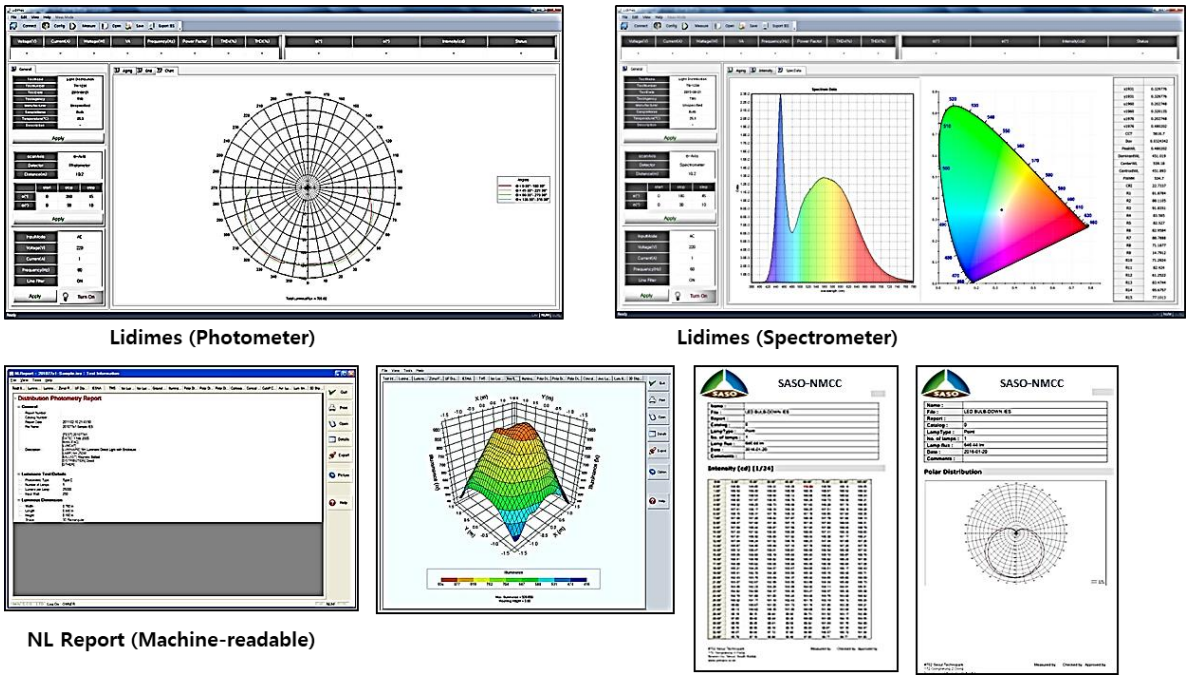


Figure 8. Software DAQ modules for photometer, spectroradiometer and 3D spectral charts with the output results in form of NL reports and IES (machine-readable) format.

[11] J. Cheney, A. Chapman, J. Davidson, A. Forbes, **Data provenance, curation and quality in metrology, in Advanced Mathematical and Computational Tools in Metrology and Testing** XII. World Scientific, 2022, pp. 167–187. DOI: 10.1142/9789811242380_0009

[12] S. G. Hackel, F. Härtig, J. Hornig, T. Wiedenhöfer, T. Schrader, A. Scheibner, J. Loewe, L. Doering, B. Gloger, J. Jagieniak, D. Hutzschenreuter, G. Söylev-Öktem, **The fundamental architecture of the DCC**, Measurement: Sensors, 18, 2021, 100354. DOI: 10.1016/j.measen.2021.100354.

[13] V. Galpin, I. Smith, J.-L. Hippolyte, **Supporting provenance of digital calibration certificates with temporal databases, IMEKO TC6 Int. Conf. on Metrology and Digital Transformation** (M4Dconf 2022), Berlin, Germany, 19-21 September 2022, 4 pp. DOI: 10.21014/tc6-2022.024

[14] Khaled M. Ahmed, Abdullah S. AlOsaimi and Khalid M. Amin, **Smart Infrastructure for Developing Digital Calibration Certificates at SASO-NMCC,** IMEKO TC6 International Conference on Metrology and Digital Transformation (M4DConf2022), September 19 ~ September 21, 2022, Berlin, Germany.

[15] BIPM, **The International System of Units (SI),** 9th ed.; Bureau International des Poids et Mesures: Sèvres, France, 2019. Online [Accessed 15 April 2025] https://www.bipm.org/documents/20126/41483022/SI-Brochure-9-EN.pdf

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - M4DConf 2025
Benevento, Italy, September 3-5, 2025

A Python-based graphical uncertainty calculator with optimal propagation of uncertainty and Monte-Carlo evaluation possibility

Khaled M. Ahmed [a,*] *and* Abdullah S. Alosaimi [a]

[a] National Measurement and Calibration Center, Saudi Standards, Metrology and Quality Org. (SASO-NMCC), Riyadh, Saudi Arabia,
*Corresponding author: k.abdelftah@saso.gov.sa / khaled55eg@gmail.com

Abstract **– A web-based uncertainty calculator for metrology professionals is presented in this paper. The system incorporates a Monte Carlo (MC) engine in accordance with GUM Supplement 1 and optimal first-order propagation that complies with the Guide to the Expression of Uncertainty in Measurement (GUM). A case study, algorithms, and architecture are described in detail. The findings demonstrate enhanced transparency, clear report generation (in PNG/PDF format) suitable for ISO/IEC 17025 documentation, and strong agreement between analytical propagation and MC evaluation. The benefits of integrating statistical simulation and rigorous analytical propagation in an intuitive web-based platform are emphasized. Users can view the installed script file and make changes to continuously improve its functionality using the editor-like text window that outputs the calculation results.**

Keywords**:** Digital Metrology, Uncertainty Calculator, Python-based algorithms, GUM, Monte Carlo Simulation.

1. MOTIVATION

When both the measured quantity and its uncertainty are provided, all measurements can be considered complete. Given the significance of this uncertainty, it makes sense that standards for expressing uncertainty have been developed and are widely applied. For general metrologists, evaluating uncertainty in accordance with these expression guidelines is not an easy task. When the number of independent variables corresponding to each uncertainty component increases or the model equation becomes complex, it becomes necessary to perform laborious partial differentiation in order to calculate the sensitivity coefficient for each component, even if the measurement uncertainty model equation can be identified and each uncertainty component can be estimated. If the probability distribution of the propagated uncertainty does not follow a normal distribution, the coverage factor corresponding to a certain confidence interval (e.g., 95%) is not simply twice the composite standard uncertainty (k = 2) [1-6]. Two program libraries for simply solving this mathematical problem have been released in the Python language. However, setting up and using the language environment is not an easy task for users who are unfamiliar with this language. This serves as a barrier to learning Python. The development of open source software is free because many helpful open libraries for Python are freely available. Python is a powerful, flexible, and accessible programming language that offers numerous advantages. It is easy to learn, open-source, and free, running on various platforms with minimal changes. It has numerous libraries for scientific computing, data analysis, and machine learning. Python's high-level nature allows quick prototyping and faster coding. It also has integration capabilities with other languages and tools. Its strong community and wide user base make problem-solving easier. Python is versatile, used in web development, automation, AI, and more. It supports object-oriented and functional programming and is ideal for automation and scripting.

Python is a widely used and open-source uncertainty calculator that offers several benefits. It is accessible to students, labs, and institutions without high software costs. Python automates complex calculations, reducing human error. Its scripts can be saved, versioned, and shared, ensuring traceability. Users can customize the code for specific models, such as GUM and Monte Carlo methods. Python also provides educational value by helping learners understand uncertainty analysis logic. It integrates easily with Excel, databases, and instruments, linking uncertainty analysis with real-time measurements or calibration workflows. Python-based tools contribute to digital transformation, aligning with metrology 4.0 initiatives [7-11].

Measurement science necessitates frequent statistical processing and numerical computations, just like other scientific domains. The uncertainty calculation module is only one example of using the Python engine built into this demo program. If scripts for other purposes are developed and shared in the future, this engine can function as a shared platform for calculations as its usability increases.

An essential component of metrology and scientific research is the assessment of measurement uncertainty. In order to make it easier to propagate measurement uncertainties through user-defined mathematical models, this paper presents uncertainty calculator, a lightweight, open-source Python application. This calculator provides an approachable platform for both instructional and real-world applications in uncertainty analysis by utilizing the SymPy library for symbolic mathematics and Tkinter/JSON for the graphical user interface. Through case studies, the tool's capabilities are illustrated, emphasizing both its potential to

advance digital transition in metrology and its usefulness in calibration operations [9-14].

To address uncertainty evaluation in calibration and testing, a calculator built with Python was created. As presented case study, the calculator ensures optical metrology realism and auditable outputs by combining a high-throughput Monte Carlo engine with GUM-compliant propagation. The calculator features domain-specific models, an online interface, and automatic PNG/PDF report generation. It has been improved to have a simple web user interface and MC capability. Real-world example includes optical metrology is presented. There is also comparison report that summarizes the methodology, benefits, and outcomes.

2. PREVIOUS WORK

A thorough list of the top software tools and uncertainty calculators for metrology, arranged by kind, features, common applications, and real-world restrictions. NPL software for GUM Supplement 1 examples, NMI-developed libraries and frameworks, and NMI-built web apps are the three primary categories into which the tools fall. NPL software for GUM Supplement 1 examples is a collection of Windows programs that replicate GUM S1 example problems, while NIST Uncertainty Machine (UM) is a free web-based tool that applies the Gauss formula and Monte Carlo (GUM S1) law of propagation. The sophisticated library METAS UncLib/metas-unclib offers linear, higher-order, numerical, and Monte Carlo propagation for multivariate, correlated, and complex-valued quantities. While Uncertainties uses first-order Taylor series to automate uncertainty propagation, MetroloPy is a Python toolkit for GUM-consistent computations. MUkit is a training ecosystem and software for chemical and environmental uncertainty that complies with ISO 11352 for water quality and Nordtest TR 537. GUM Workbench, GUMsim, and QMSys GU M are commercial desktop workbenches that are best suited for particular uses. The article highlights new developments in digital metrology, automated experiments for uncertainty, and the development of multivariate/MC capabilities while offering helpful selection advice for labs utilizing GUM and MC in their analysis tools [13, 15].

3. PROGRAM FEATURES AND CORE FUNCTIONALITIES

In order to ensure confidence in results and compliance with ISO/IEC 17025 and the Guide to the Expression of Uncertainty in Measurement (GUM), measurement uncertainty evaluation is essential in modern metrology. This Python-based graphical uncertainty calculator offers an adaptable, transparent, and user-friendly solution, in contrast to traditional calculators that frequently provide simple analytical propagation or restricted Monte Carlo capabilities. The calculator integrates analytical (GUM) uncertainty propagation using sensitivity coefficients, Monte Carlo simulation for non-linear models or complex uncertainty propagation, and an interactive, professional GUI for input, results, and report generation. The system is a multidisciplinary tool for labs and R&D facilities since it supports general measurement models in various fields. Its primary features include Monte Carlo simulation, the Analytical Propagation (GUM Method), and an online interface. Technical superiority, distribution flexibility, scalability, transparency, usability and accessibility, integration and customization, and automation readiness are among the benefits over current tools. Strategic benefits for management include improved decision-making, efficiency gains, cross-disciplinary use, and compliance and audit readiness. The calculator ensures measurement results have well-quantified uncertainty, facilitates risk-based thinking in line with ISO/IEC 17025 requirements, reduces time spent on manual uncertainty spreadsheets, and reduces calculation errors by automating repetitive processes. It is applicable in multiple divisions and reduces training overhead. Table 1 presents some comparative advantages of this new program over conventional ones.

4. PROGRAM STRUCTURE

4.1. Architecture of the System

Streamlit is web-based and compatible with all browsers provides model equation entry and dynamic input fields. Analytical and Monte Carlo results panels are distinct. Automatic creation of PDF reports. The Streamlit framework is used to implement the application in Python, allowing for browser-based interaction without the need for local installation. In addition to choosing from preloaded measurement models or defining custom functions, users can enter measured values, correlation coefficients, and standard uncertainties. Figure 1 presents the program structure.

4.2. The Engine of Uncertainty Propagation

There are two modes of calculation: (a) Monte Carlo evaluation with user-specified iteration counts, producing histograms and statistical summaries; and (b) first-order GUM propagation, which computes combined standard uncertainty using sensitivity coefficients. Analytical calculations are validated in real time by Monte Carlo results.

4.3. Complex Modelling

Nonlinear models, for instance Planck-law fitting for radiation thermometry, are incorporated into the tool. Users can upload instrument spectral response data or specify an effective wavelength (λ_{eff}). Next, nonlinear propagation is carried out directly within the physical model by the calculator.

4.4. Implementation details

The engine uses numerical central differences with adaptive step size to compute sensitivity coefficients. Monte Carlo sampling draws independent normal deviates by default; correlation can be added via a covariance matrix and a Cholesky factorization. For instance, in radiation thermometry, numerical quadrature (trapezoidal rule on a user-supplied wavelength grid) computes the expected

signal; inversion uses a robust scalar root finder over temperature.

4.5. Evaluation Cycle (simplified)

1) Read inputs (x, u, distributions, correlations).

2) Compute nominal outputs y = f(x).

3) Numerical sensitivities: $\boldsymbol{c_i = \frac{\partial y}{\partial x_i}}$

4) First-order $\boldsymbol{u_c^2 = \sum (c_i\, u_i)^2}$ (+ covariance terms).

5) Monte Carlo: draw N samples of x; evaluate y; compute mean, STDEV, percentiles.

6) Export GUI PNG, PDF report, and CSV (samples).

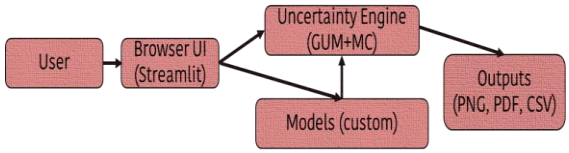


Figure 1. System architecture: browser UI interacts with the uncertainty engine and model modules. Outputs include PNG, PDF and CSV.

Table 1. Comparison with respect to conventional tools

ELEMENT	THIS SOFTWARE	CONVENTIONAL TOOLS
UNCERTAINTY	GUM + MC SIDE-BY-SIDE	ONE TECHNIQUE (USUALLY GUM ONLY)
MODELLING	LINEAR + NON-LINEAR POSSIBILITY, CUSTOM MODELS	PRIMARILY LINEAR
TRANSPARENCY	FORMULAS, SENSITIVITIES, VARIANCE BUDGET	BLACK-BOX OUTPUTS
EXTENSIBILITY	OPEN PYTHON CORE, CSVJSON I/O	RIGID FEATURES, CLOSED SOURCE
REPORTING	PNG GUI, PDF, CSV SAMPLES	MINIMAL TABLE, NO AUTOMATION
VALIDATION	MC VS. ANALYTICAL	RARELY IMPLEMENTED

4. CASE STUDY

Interferometric thickness measurements can achieve sub-micrometer accuracy, but the total uncertainty depends on laser wavelength calibration, fringe counting, air refractive index, mechanical alignment, and environmental stability. This Web-based tools can improve transparency and repeatability by coupling physics-based models with audit-ready reporting.

In metrology, interferometry is a technique that measures the thickness of thin films by examining how light waves interfere with reflections from a transparent film's top and bottom surfaces. A known refractive index is necessary for this procedure, which can be carried out with monochromatic or white-light sources. It enables accurate determination of the optical path difference and the thickness of the film. Interference, signal separation, and thickness calculation are all part of the process. Using the known refractive index and light wavelength, the thickness of the film is determined by calculating the optical path difference (OPD) between the interfering beams from the interference pattern.

A He–Ne source (λ = 632.8 nm) illuminates a Twyman–Green interferometer that measures a glass coupon's thickness inside a controlled chamber (T = 20.0 $^{\circ}$C, P=1013.2 hPa, RH=45%). The effective air refractive index n=1.0002713 with standard uncertainty 2.0^{-07}. The model equation and nominal fringe count are shown below. Alignment residuals are modeled with σ =0.05 μm. Uncertainty is propagated via linearization around nominal values and validated by Monte Carlo sampling.

Agreement between analytical and Monte Carlo estimates confirms near-linearity of the model at the stated uncertainties. The budget indicates fringe counting and alignment dominate; wavelength and air-index contributions are smaller under the stated priors. The GUI documents assumptions and enables direct export of figures and a structured PDF.

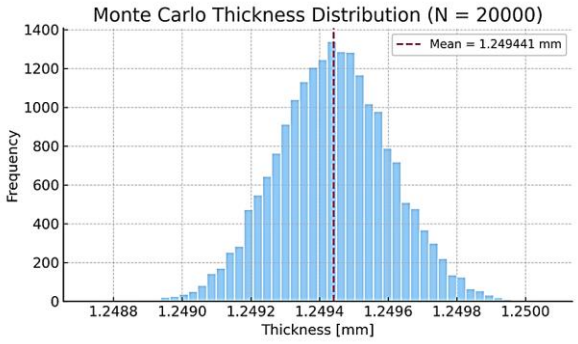


Figure 2. Monte Carlo Distribution yielded from thickness example execution.

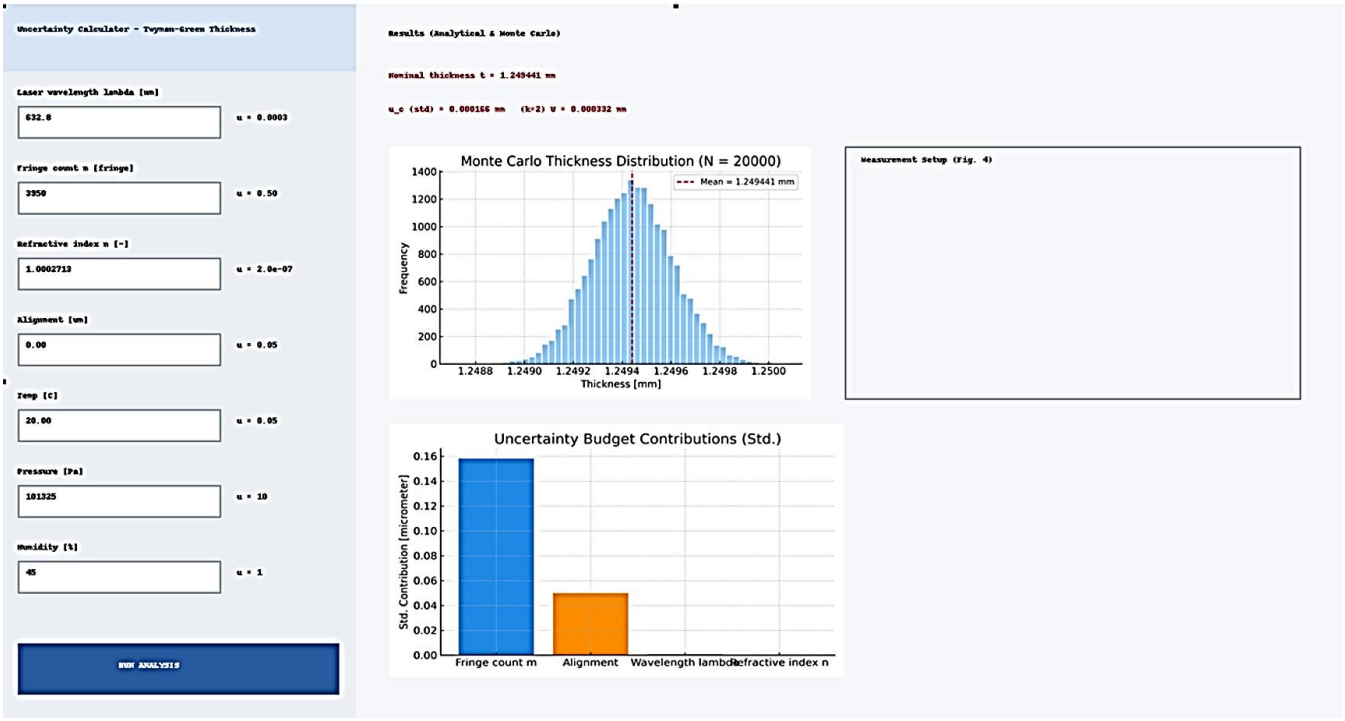


Figure 3. GUI of the program with thickness example implemented.

4.1. Measurement model, conditions and uncertainty calculator outcomes

$$t = \frac{\lambda}{2n}\left(m + \frac{\varphi}{2\pi}\right) \qquad (1)$$

where t is thickness in nm, λ is wavelength (nm), n refractive index, m fringe order (integer), and φ is measured phase difference in radians.

- Inputs (example):
  - λ = 632.8 nm ، σ_λ = 0.1 nm
  - n_{air}= 1.0002713
  - n = 1.500, σ_n = 0.005
  - m = 100 (exact integer)
  - φ = 1.234 rad, σ_φ = 0.02 rad
  - Monte Carlo samples : n = 20000
- Results:
  - Nominal thickness : t_0 = 21134.76004933929 nm
  - First order (numerical) σ_t ≈ 70.5315 nm (1σ)
  - Monte Carlo: mean ≈ 21135.344215809193 nm, std ≈ 70.3915 nm (1σ)
- Numerical partial derivatives (central differences):
  - ∂t/∂λ ≈ 33.3988 nm/nm
  - ∂t/∂n ≈ -14089.84 nm per unit n
  - ∂t/∂φ ≈ 33.57108 nm per rad

Because of the problem's modest nonlinearity (division by n and addition of φ/2π inside), Monte Carlo and first-order methods produce solutions that are similar but somewhat different; MC accounts for any slight sampling asymmetry.

The histogram aids in identifying any non-Gaussian characteristics or skewness in the t-distribution (in this case, it appears to be near-Gaussian). Figure 2 shows Monte Carlo chart of the presented example and figure 3 presents the GUI of the developed prototype (demo) uncertainty calculator, while this real-world example is implemented to demonstrate the applicability of the program.

5. CONCLUSIONS

The system provides quicker, error-free uncertainty evaluations by combining Monte Carlo simulation and the ISO GUM analytical method into a browser-based platform. It facilitates multidisciplinary testing, such as mechanical and optical metrology. Under ISO/IEC 17025, compliance-ready reports in PDF format can be tracked for audits. No licensing fees and flexibility to adjust to changing measurement requirements are made possible by the open-source foundation. Using this tool gives the metrology leader data-driven confidence in each measurement, strengthening their position.

The general Python-based graphical uncertainty calculator is a clever upgrade over conventional, fragmented tools. It offers a transparent, scientifically sound, and management-friendly solution that enables both analytical and simulation-based uncertainty evaluation on a single, user-friendly platform. Its open-source nature, which ensures long-term flexibility without licensing fees, makes it the ideal choice for modern metrology labs seeking efficiency, compliance, and traceability.

Combining Monte Carlo and GUM analysis, being web-based and installation-free, being multidisciplinary (mechanical, electrical, thermal, and optical), adhering to ISO/IEC 17025, eliminating spreadsheet errors, and being open-source with no license fees are some of its salient

features.

Among the advantages are quicker, error-free evaluations, handling intricate, non-linear models, and creating audit-ready PDF reports; Fair assistance for R&D and manufacturing; flexible to the organization's requirements; boosts measurement accuracy.

Finally, this work presents a Python-based graphical uncertainty calculator that provides a flexible and scientifically sound way to measure uncertainty quantification, making it a useful tool for labs and research and development facilities.

Normal distributions and uncorrelated inputs are now assumed by the defaults. Future releases will reveal correlation matrices and other distributions (rectangular, triangular, and U-shaped) with on-the-fly inversion. Batch mode and an API endpoint will support automated certificate generation.

ACKNOWLEDGMENTS

Authors thank the “Research and Studies Center (RSC)” of SASO for their coordination and consideration of this work.

REFERENCES

[1] Bich, W., Cox, M. G., & Harris, P. M. (2006). Evolution of the "Guide to the Expression of Uncertainty in Measurement." *Metrologia*, *43*(4), S161–S166. https://doi.org/10.1088/0026-1394/43/4/S02

[2] Cox, M. G., & Siebert, B. R. L. (2006). The use of a Monte Carlo method for evaluating uncertainty and expanded uncertainty. *Metrologia*, *43*(4), S178–S188. https://doi.org/10.1088/0026-1394/43/4/S03

[3] Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., ... Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

[4] Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

[5] Joint Committee for Guides in Metrology. (2008). *Evaluation of measurement data — Guide to the expression of uncertainty in measurement* (JCGM 100:2008). https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_E.pdf

[6] Joint Committee for Guides in Metrology. (2008). *Evaluation of measurement data — Supplement 1 to the "Guide to the expression of uncertainty in measurement" — Propagation of distributions using a Monte Carlo method* (JCGM 101:2008). https://www.bipm.org/documents/20126/2071204/JCGM_101_2008_E.pdf

[7] Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., & Willing, C. (2016). Jupyter Notebooks—a publishing format for reproducible computational workflows. *ELPUB*, *2016*, 87–90. https://doi.org/10.4000/proceedings.613

[8] Lebigot, E. O. (2010). uncertainties: a Python package for calculations with uncertainties. *Journal of Open Research Software*, *8*(1), 14. https://doi.org/10.5334/jors.238

[9] McKinney, W. (2010). Data structures for statistical computing in Python. In S. van der Walt & J. Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (pp. 51–56). https://conference.scipy.org/proceedings/scipy2010/pdfs/mckinney.pdf

[10] Millman, K. J., & Aivazis, M. (2011). Python for scientists and engineers. *Computing in Science & Engineering*, *13*(2), 9–12. https://doi.org/10.1109/MCSE.2011.36

[11] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830. https://jmlr.org/papers/v12/pedregosa11a.html

[12] Ramirez, A., & Barnes, J. (2020). GUI development for scientific applications: A comparative study of PyQt and Tkinter. *Journal of Computational Science Education*, *11*(2), 45–58. https://doi.org/10.22369/jcse.2020.1102.03

[13] Siebert, B. R. L., & Sommer, K. D. (2004). New developments of the GUM and Monte Carlo techniques. *Technisches Messen*, *71*(2), 67–80. https://doi.org/10.1524/teme.71.2.67.27066

[14] Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... van Mulbregt, P. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

[15] Willink, R. (2006). On using the Monte Carlo method to calculate uncertainty intervals. *Metrologia*, *43*(6), L39–L42. https://doi.org/10.1088/0026-1394/43/6/N01

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Metrological Data Application Solutions in the Field of Marine Power

Hao Wu[1], Xianglong Ma[1], Yuxin Sun[1], Jing Lin[1], Jiangang Ji[1]

[1] *Shanghai Marine Equipment Research Institute, No.10 Hengshan Road, Shanghai, China, zsjamxl@163.com*

Abstract **–This paper addresses critical metrological challenges in marine power systems—including measurement deficiencies for 23 key parameters, digital fragmentation hindering cross-system data integration, and inconsistent standardization across lifecycle phases. This paper propose a Cloud-Edge-Device Integrated Architecture with five functional strata: the Data Layer consolidating laboratory/vessel datasets, the Perception Layer enabling edge-computing acquisition via adaptive tools, the Platform Layer standardizing metadata through API services, and the Application Layer deploying predictive maintenance and lifecycle simulation. This framework establishes closed-loop data management for 53 critical parameters, resolving "unmeasurable/ inaccurate/ incomplete" data issues while implementing robust security via national encryption algorithms, RBAC/ABAC access controls, and blockchain-anchored trusted timestamps. Implementation demonstrates reliability improvement and accelerated design iteration, providing end-to-end metrological support compliant with international maritime standards.**

I. INTRODUCTION

Marine power systems serve as the "heart" of offshore engineering equipment, directly determining operational safety and economic efficiency. As complex coupled systems, the development, testing, and operation of marine power equipment require comprehensive support from metrological testing data throughout their lifecycle. To address the urgent demands of marine equipment advancement, it is imperative to strengthen the construction, aggregation, analysis, application, and governance of marine power metrological data, enhance data security management, and improve data governance capabilities and application effectiveness.

Under the driving force of new-generation technological revolution and industrial transformation [1-3], the global marine power equipment industry is experiencing rapid digital and intelligent transformation. Over the past decade, while significant progress has been made in metrological testing data for marine power systems, three critical challenges persist:

A. Critical Parameter Measurement Deficiencies

Current measurement capabilities remain limited to conventional instruments (pressure gauges, thermocouples, flow meters) under static laboratory conditions and normal measurement ranges. Significant gaps exist in dynamic parameter measurement, extreme environment testing, peak value detection, and online monitoring capabilities. These limitations result in unmeasurable, inaccurately measured or incomplete critical data, severely constraining scientific management of marine power systems and even leading to operational accidents.

B. Digital Fragmentation and Data Utilization Barriers

Despite the accumulation of substantial metrological testing data, low digitalization levels have created isolated data silos across different measurement devices. The strong correlation between operational parameters in marine power systems is hindered by this fragmentation, preventing comprehensive data integration and effective utilization for industrial upgrading.

C. Inconsistent Data Standardization Practices

The value transmission chain in marine power systems involves multiple components (calibration devices, sensors, dynamometers) with diverse measurement methodologies across different design phases and application scenarios. Lack of unified evaluation criteria frequently causes data inconsistency between design stages, posing significant challenges to equipment safety.

These metrological data challenges in acquisition, analysis, and application have created critical bottlenecks in marine power system development. Addressing these priorities through comprehensive metrological data system construction will fundamentally improve the reliability, safety and performance of next-generation marine power systems.

II. FRAMEWORK OF THE SOLUTIONS

The system adopts a Cloud-Edge-Device Integrated Data Circulation Architecture as illustrated, comprising five functional strata: Data Layer (Source Scenarios), Perception Layer, Network Layer, Platform Layer, and Application Layer.

A. Data Layer (Foundational Data Repository)

Serves as the multi-source data ecosystem encompassing:

- Laboratory metrological datasets
- Power equipment design verification data
- Operational performance metrics
- Marine vessel operational records

Data acquisition channels include experimental setups, sensing devices, sensor arrays, information systems, and physical/digital documentation.

B. Perception Layer (Intelligent Acquisition Gateway)

Implements edge computing-enabled data ingestion through:

- Multi-protocol edge data terminals
- Adaptive data collection tools

Key functionalities:

- Unified collection from heterogeneous sources
- Context-aware preprocessing
- QoS-guaranteed data transmission

C. Platform Layer (Data Middleware Infrastructure)

Operates as the distributed data orchestration hub featuring:

- Scalable storage architecture
- Data asset lifecycle management
- API-driven service provisioning

Core capabilities:

- Metadata standardization
- Multi-dimensional data modeling
- Secure service publication

D. Application Layer (Value-Creation Ecosystem)

Delivers intelligent analytics services through:

- Predictive maintenance models
- Operational diagnostics frameworks
- Lifecycle performance simulation

Implementation outcomes:

- Enhanced equipment reliability
- Optimized maintenance planning
- Data-driven design iteration acceleration

This framework establishes an industrial-grade metrological data value chain, delivering reliable and precise metrological support for the full lifecycle management of marine power systems - from conceptual design through operational deployment to predictive maintenance.

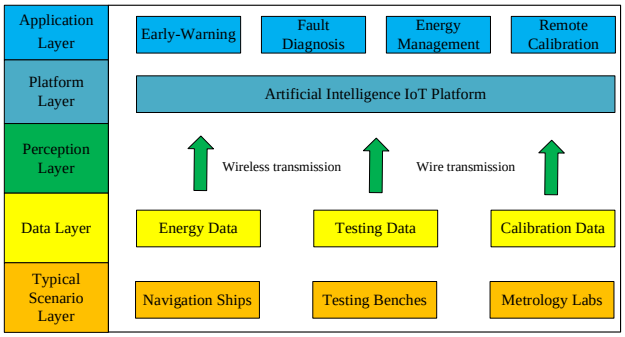


Fig. 1. Framework of the data application.

III. PARAMETER SYSTEM

In response to the problems of "inability to measure, inaccurate measurement, and incomplete measurement" of key parameters of marine power equipment, "data silos", and standardized management, we will investigate the mainstream testing and measurement technology requirements for marine strategic power equipment, strengthen top-level planning, establish a measurement and testing technology system that complements the commonality and individuality of marine power equipment, and support the systematic technological development of marine power equipment.

Preliminary research was conducted on the faults and measurement testing issues of the ship's power system in the early stage, and 53 parameters were sorted and summarized. 23 key parameters lacked testing capabilities, and 42 parameters lacked calibration capabilities.

IV. DATA MANAGEMENT ARCHITECTURE

The established measurement data platform promotes the classification, grading, and full lifecycle management of marine power equipment measurement data and resources. The data platform realizes the acquisition, aggregation, and development of measurement data throughout the entire chain, forming data assets and providing data services. It forms a data closed-loop management system from data collection and aggregation, data storage and integration, data processing and cleaning, data update and maintenance, data ownership and release, data service and use, data archiving and destruction. The data management architecture of the data center is shown in Figure 2.

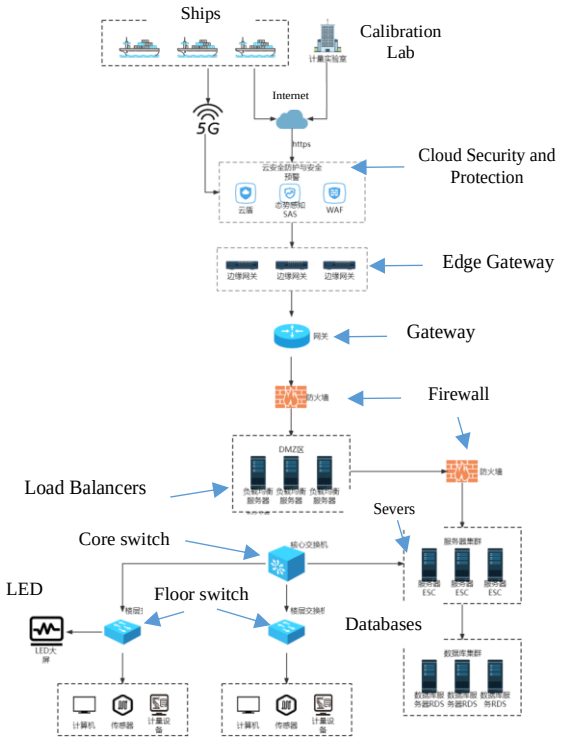


Fig. 2. Data management architecture.

V. MEASUREMENT DATA SECURITY AND COMPLIANCE

Encryption mechanism. Encrypt important and sensitive data using national encryption algorithms during storage and transmission to ensure data security.

A. Digital signature

Digitally sign measurement data, reports, certificates, etc. to ensure the integrity and authenticity of the data during transmission and storage.

B. Access control

Role based access control (RBAC) and attribute-based access control (ABAC) ensure that only authorized users and systems can access specific data.

C. Security audit

All data access operations require recording audit logs to ensure the traceability of data access, and regularly auditing the security and compliance of the system.

D. Data desensitization.

For sensitive data, anonymization and data perturbation techniques are used to ensure that personal privacy and sensitive information are not exposed during the data flow process.

E. Data compliance.

Ensure that data processing and storage comply with relevant international and national standards, especially in cross-border marine equipment metrology scenarios, and ensure the legality of data transmission and sharing

F. Electronic signature.

Verify the identity and protect the integrity of circulation documents such as testing reports, inspection reports, calibration certificates, etc. It includes digital signatures, encryption techniques, and authentication mechanisms to ensure the authenticity and immutability of files during the signing process.

G. Data trusted timestamp.

Generate a unique timestamp for each data record, identifying the time of data generation or modification. Verify the authenticity of timestamps through digital signatures and encryption techniques to ensure that data has not been tampered with.

VI. CONCLUSIONS

This article proposes a digital solution for measuring marine power equipment. By constructing a marine power measurement testing system, the data acquisition capability is systematically improved. Based on the data management architecture and data security scheme of the data platform, reliable data support is provided for marine power equipment.

REFERENCES

[1] Khaled M.Ahmed, Hamad AlGhamdi, Homood AlOtaibi, Khalid M. Amin, “An intelligent web-based platform for processing Digital Calibration Certificates utilizing the GEMIMEG tool and advanced AI capabilities”, Measurement: Sensors, 2025, vol.38, pp. 101485.

[2] Katharina Janzen, Daniel Heibelmann, “Development of digital calibration certificates for coordinate metrology applications”, 2025, vol.38, pp. 101478.

[3] Moritz Jordan, Shanna Schönhals, Sören Auer, “Enhancing interoperability in digital calibration data exchange: A case for ontology development”, Measurement: Sensors, 2025, vol 38, pp. 101459.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Barriers to Implementing Digital Twin Technologies in Industrial Settings

Józef Wiora[1], Alicja Wiora[1], Faisal Saleem[1,2]

[1]*Department of Measurements and Control Systems, Silesian University of Technology, ul. Akademicka 16, 44-100 Gliwice, Poland*
alicja.wiora@polsl.pl; jozef.wiora@polsl.pl
[2]*The Joint Doctoral School, Silesian University of Technology, 44-100 Gliwice, Poland*
faisal.saleem@polsl.pl

Abstract **– Recent review articles highlight an exponential rise in publications on Digital Twin (DT) technology. Despite its recognised potential, DTs have yet to achieve widespread practical use. Following a presentation of the state of the art and an original conceptual diagram illustrating the technology, this work presents the key factors contributing to the gap between conception and implementation. Using experimental data from a laboratory water system and a reliability-based perspective, this analysis examines the practical limitations of DT application. The findings indicate that broader use of DTs depends on the technological maturity and reliability of all system components, which still require further development.**
Keywords **– Industry, Reliability, Flow System**

I. INTRODUCTION

Michael Grieves introduced the concept of Digital Twin (DT) technology in 2002 during a presentation on industrial product lifecycle management [1, 2]. This concept extends earlier work on model-based adaptive control and state observers [1]. Today, DT technology evolves rapidly and ranks among the most strategically important trends [3]. However, these trends rarely translate into practical applications. In this study, we aim to identify the barriers, drawing on both a literature review and our own observations.

II. PROPERITES OF DIGITAL TWIN TECHNOLOGY

A. Components

Fig. 1 illustrates the DT concept. This approach creates a Virtual Model (VM), aka Digital Entity, of a Physical Asset (PA), aka Target Entity. This model enables real-time simulation and monitoring [2, 4, 5]. The VM synchronises with the PA and optimises its operation. The PA adjusts dynamically based on direct instructions from the VM [2]. Actions performed on the VM generate observations that support decision-making and influence the PA [6]. The model must maintain sufficient accuracy for the intended task.

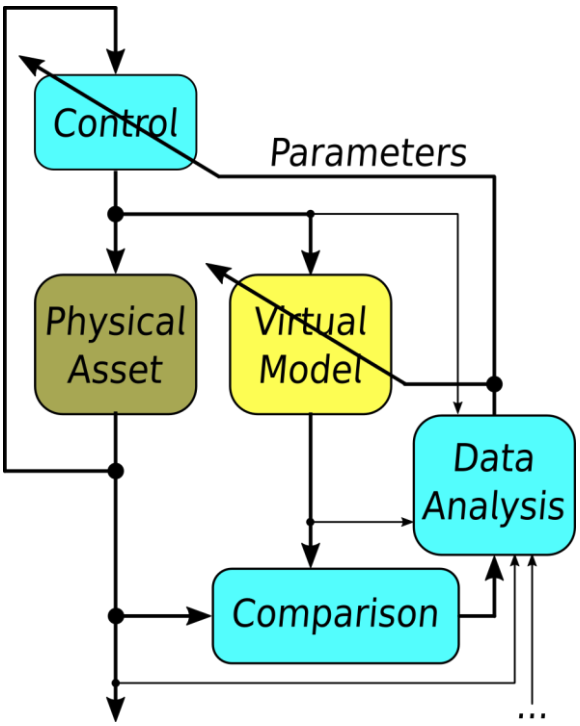


Fig. 1. Diagram of data flow in Digital Twins technology.

DT system also includes several key components: data analysis, computing, real-time communication, large-scale data processing, and servicing functions such as quality control and prediction [1, 3]. In essence, the system compares outputs from the VM and PA, integrates all available data, and calculates control parameters for the PA. In Fig. 1, bold arrows represent the main flow of information. In DT technology, data analysis leverages all available information (depicted with thin arrows) to fine-tune both the VM and the control parameters.

Although DT systems may appear autonomous, humans typically remain involved in decision-making or intervention [6]. If the VM does not affect the physical

world, the system is referred to as a Digital Shadow [1, 7]. In some cases, systems only collect raw data throughout their lifecycle without real-time modelling. This approach is known as the Digital Thread [7].

A similar concept appears in virtual, aka soft or software, measurements. Here, a model replicates a real measurement process. It replaces physical instruments by using available signals to simulate measurements. However, this model does not influence PA [6].

Measurements play an important role in the diagram presented in Fig. 1. These measurements capture most of the input and output data from PA [6]. Without them, constructing an appropriate model would not be possible. The system must acquire data with low communication latency and maintain synchronisation. This task becomes particularly challenging when PA is geographically distributed [1].

The data collection process must operate quickly enough to track the asset's behaviour in real time. At the same time, the algorithms must update the model promptly to detect and respond to significant changes [6]. For the correct operation of the DT system, live data from the physical system must be continuously available. Therefore, communication and coordination must be reliable and efficient [1].

B. Applications

The literature shows that DT technology was first implemented in the aerospace sector, particularly in aircraft and rocket systems. Its use has since expanded to a wide range of domains, including manufacturing, industrial systems, urban infrastructure, and military operations [1, 8, 9], as well as medicine, energy, healthcare, and agriculture [1, 3]. Nearly half of the existing publications focus on manufacturing, covering both production- and product-oriented applications [3].

Researchers have also explored the use of DTs to support operations in measurement laboratories [6]. In addition, DTs are emerging as a promising tool for intelligent power systems [1]. Owing to the strong correlation between the physical asset and its virtual counterpart, DTs offer effective solutions to complex industrial challenges [9].

Although the literature on DT is rapidly expanding, most applications remain at the conceptual stage [1]. There is still a lack of comprehensive and high-fidelity digital models, such as those representing the entire human body [3].

Information and communication technology includes cyber-physical systems (CPS), the Internet of Services, and cloud computing [2]. Within CPS, DT technology plays a key role due to its ability to capture, predict, and visualise both real and virtual states of human–system interactions. It also facilitates autonomous operation [1, 5]. However, CPS are inherently complex. As complexity increases, their vulnerability to faults and failures also rises [10].

DT technology plays a significant role in Industry 4.0. Its implementation enables faster identification of physical issues and contributes to the production of higher-quality products [11]. DT is also regarded as a key technology in the 5th industrial revolution. This new industrial paradigm integrates advanced sensors and the Internet of Things (IoT) to promote collaboration between humans and robots. It emphasises a human-centred perspective in the design and implementation of technological processes [7]. At the same time, reliability remains a key indicator of the operability and performance of production plants within manufacturing systems [10].

DT helps reduce time-to-market and optimises both plant equipment and product performance [11]. It also enables predictive maintenance, which significantly lowers the costs associated with production system failures. Moreover, it supports optimal utilisation of production resources and enhances operational safety [4].

Introduced in 2020, a policy document promoting DT development in China has significantly influenced both economic and social progress. The initiative supports the advancement of modelling, simulation, data fusion, and related technologies [9].

C. Design

The design of DT systems typically involves several key phases. It begins with the Modelling Phase, during which multiple models are developed and evaluated. This is followed by the Simulation Phase, where the DT simulates the behaviour of the real system. In the Deployment Phase, the system is installed in its target environment. The Monitoring Phase involves using the DT to track the system's performance in real time. Finally, the Optimisation Phase focuses on enhancing safety, efficiency, and overall effectiveness [7].

D. Model

The VM enables engineers to perform prognostics, optimisation, and control within a digital environment [3]. It also facilitates real-time monitoring and optimisation of PA. By creating a virtual replica, engineers can test and refine the system before physical implementation, thereby improving safety—particularly in applications such as robotic systems—by conducting trials in a virtual environment [7].

The VM serves as a comprehensive and autonomous digital mirror of the PA. It maintains an interactive connection with the physical system through simulation, verification, prediction, and control across the entire lifecycle of the process. It leverages both historical and real-time data to support decision-making. DT technology represents a dynamic process [9], driven by advanced communication and computational capabilities [1].

The DT model can be described across four key dimensions: geometry (including shape and assembly), physics (encompassing properties, characteristics, and

constraints), behaviour (capturing system dynamics), and rules (which apply historical data to enhance model intelligence). Developing such a model requires interdisciplinary expertise [3].

DT technology integrates all available data to support simulation through multidisciplinary, multiphysical, multiscale, and multiprobability approaches. This integration relies on physical models, real-time sensor data, and operational history to ensure accuracy and adaptability [11].

The design of a DT model can follow various methodological approaches, including white-box, grey-box, and black-box modelling techniques [1]. Advanced modelling often demands significant computational resources. These computations can be performed either in the cloud or at the edge. Edge computing offers advantages such as reduced latency and lower data transmission volumes. However, it also presents increased vulnerability to cyberattacks [1].

The PA typically evolves over time, which necessitates continuous updates to the model parameters to maintain accuracy [6]. Achieving sufficient model fidelity is essential, particularly when we need the real-time predictions and responses. Model reduction techniques enable faster response times and more efficient tuning, meeting these demands [1].

Standard practices such as validation, verification, and accreditation can further enhance model accuracy [1]. In particular, model verification plays a critical role in ensuring that the model meets the required performance and reliability standards [3].

E. DT in water pipe systems

Water systems consist of reservoirs, pumps, pipelines with valves, flow meters, and outlets. Both industries and households rely on them extensively [9]. These systems either supply drinking water or support industrial operations. Depending on the application, they may stabilize water pressure at a target level or control the flow rate.

In pump applications, DT technology enables continuous anomaly detection and monitoring of pump operations. They support maintenance by providing real-time insights into system behaviour. These models help analyse failures, assess performance, and identify optimization opportunities. By improving failure prediction and mitigation, DTs enhance system availability and reduce project costs [9].

Another application involves a model of a university laboratory flow system. In this case, 2D or 3D computational fluid dynamics proved suitable for simulating two-phase flows during pipeline filling and emptying procedures. The success was attributed to a relatively simple installation and the use of advanced modelling techniques [12].

DTs also show promise in improving water consumption forecasting. However, the sector currently lacks a unified framework capable of integrating advanced mathematical algorithms seamlessly and effectively [13].

III. BARIERS IN REAL APPLICATIONS

DT implementation presents several challenges, as mentioned in the literature. However, despite the growing number of publications on the topic, many authors tend to focus on the benefits while giving limited attention to the potential drawbacks of this emerging solution. These limitations include ensuring data security, integrating with existing processes, managing high initial investment costs, and addressing the skills gap, as operation and maintenance require well-trained personnel. Additional barriers include the lack of standardisation, ethical concerns such as potential job displacement, and resistance to change among employees [7].

Despite growing interest, most DT implementations remain application-specific and isolated, limiting their transferability across projects. Looking ahead, proprietary data regulations may further restrict data availability. Furthermore, the high data transmission rates required by DT systems contribute to increased energy consumption, raising concerns about long-term sustainability [1].

The above-presented properties of DT systems reveal a set of conflicting requirements. A DT performs effectively when it remains simple. However, we also expect it to deliver benefits in complex systems. Such demands frequently drive up system complexity, making it difficult to control costs and maintain efficient operation.

Another major limitation lies in modelling accuracy. High accuracy requires the inclusion of a large number of details, which is only feasible for relatively simple tasks. When modelling more complex processes, the number of input signals and the required computational power increase significantly. To address this, behaviour-based models—such as grey-box or black-box approaches—are often employed. However, these models may introduce substantial approximation errors.

An illustrative example is the modelling of a laboratory water flow system. In this setup, a pump drives the flow in a closed-loop system, controlled by an inverter. Several flowmeters measure the flow rate. Both the pump and flowmeters use digital communication protocols, enabling control of the inverter frequency and acquisition of flow rates. A PC manages the entire system, allowing for repeatable experiments.

Fig. 2 illustrates the flow rate dynamics in response to variations in inverter frequency. The system exhibits distinct time responses due to the combined effects of software, electronics, motors, fluid dynamics, and materials properties. Understanding the root causes of these variations would require considerable time and effort.

One possible approach is to explore correlations

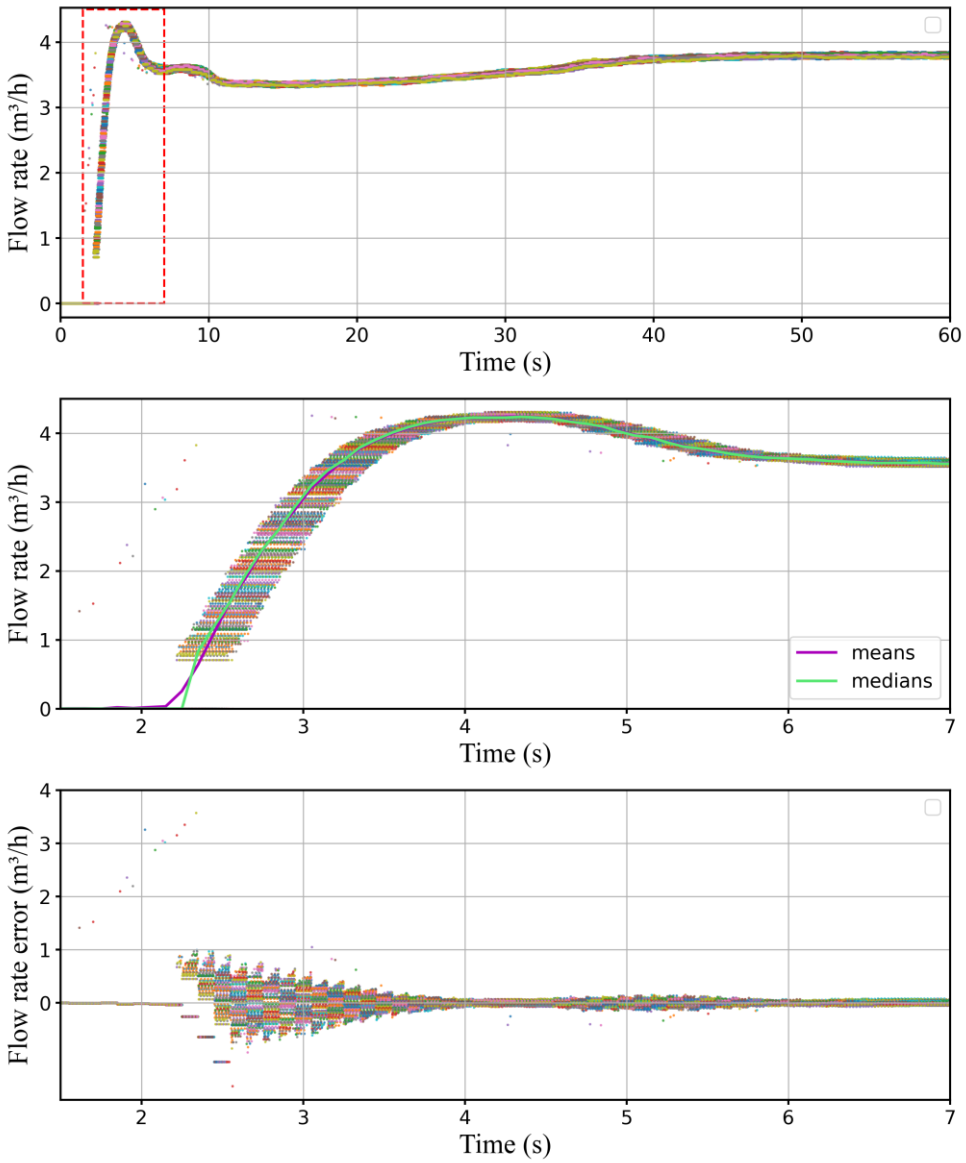


Fig. 2. Flow rates and flow rate errors vs. time in response to 89 repetivie inverter frequency changes (colour points) and the plot of averages (magenta) and medians (green) calculated for each time sample.

between input quantities and the resulting waveform. While this is an intriguing direction, it is also a complex one. Artificial intelligence techniques could assist in this analysis, but they require extensive datasets for effective training. Signal pre-processing methods such as filtering and thresholding may also be applied. However, while these techniques are effective in offline analysis, they can negatively impact system dynamics during real-time processing.

Accepting visible signal scatter introduces measurement errors. Whether this is tolerable depends on the specific application. In the collected plots, some exhibit diverse dynamic responses that result in significant errors. In the context of DTs, where the VM must adapt dynamically to the PA, such unrepeatability —especially in the absence of observable drift—can lead to unnecessary model tuning. This tuning may not yield any improvement in overall system performance.

No algorithm can correct random errors. Therefore, building a reliable virtual model based on data affected by such errors is not feasible. Moreover, it is often unclear whether the observed variability stems from measurement errors or is an inherent property of the system.

From a reliability perspective, the system shown in

Fig. 1 functions as a series connection of all its elements. These elements include both components shown on this figure and connections between them. This means it operates correctly only when every element remains operational. Its overall availability equals the product of the availabilities of each individual element. Industry demands extremely high availability, along with other reliability indicators. Consequently, for DTs to be usable in industrial settings, all components must demonstrate a high level of technological maturity. Until this maturity is achieved, widespread industrial utility will remain constrained. For simpler tasks, however, the current level of maturity is sufficient, and DTs are already being applied successfully.

IV. CONCLUSION

In this work, we began by defining the concept of a DT and illustrating it with a conceptual diagram. We then reviewed the most significant current and potential applications, noting that many remain largely hypothetical. Following this, we outlined the key requirements for constructing a reliable virtual model.

We provided background on the application of DTs in water pipeline systems and presented experimental data from our own laboratory setup. These results revealed non-repeatable system behaviour, prompting a discussion on the implications for developing virtual models based on such uncertain dynamics. We concluded that, in certain cases, further refinement of the model may not be economically justified.

Finally, we examined the limited real-world adaptation of DTs through the lens of system reliability. We argued that successful implementation depends on the technological maturity and reliability of all system components. Until these conditions are met, the widespread deployment of DTs will remain constrained. However, for simpler tasks where component maturity is sufficient, DTs are already proving to be effective.

ACKNOWLEDGMENTS

This research was funded by the Polish Ministry of Science and Higher Education 02/050/BK-25/0047 (JW, AW).

REFERENCES

[1] Z.Song, C.Hackl, A.Anand, A.Thommessen, J.Petzschmann, O.Kamel, R.Braunbehrens, A.Kaifel, C.Roos, S.Hauptmann, “Digital Twins for the Future Power System: An Overview and a Future Perspective”, Sustainability, vol.15, 2023, article 5259.

[2] X.Yao, N.Ma, J.Zhang, K.Wang, E.Yang, M.Faccio, “Enhancing wisdom manufacturing as industrial metaverse for industry and society 5.0”, Journal of Intelligent Manufacturing, vol. 35, 2024, pp.235-255.

[3] F.Tao, B.Xiao, Q.Qi, J.Cheng, P.Ji, “Digital twin modeling”, Journal of Manufacturing Systems, vol.64, 2022, pp.372–389.

[4] A.Ucar, M.Karakose, N.Kırımça, “Artificial Intelligence for Predictive Maintenance Applications: Key Components, Trustworthiness, and Future Trends” Applied Sciences, vol.14, 2024, article 898.

[5] ISO/IEC 30173:2023 Digital twin — Concepts and terminology.

[6] L.Wright, S.Davidson, “Digital twins for metrology; metrology for digital twins”, Measurement Science and Technology, vol.35, 2024, no.5, article 051001.

[7] M.Zafar, E.Langås, F.Sanfilippo, “Exploring the synergies between collaborative robotics, digital twins, augmentation, and industry 5.0 for smart manufacturing: A state-of-the-art review”, Robotics and Computer–Integrated Manufacturing, vol.89, 2024, article 102769.

[8] X.Fang, H.Wang, G.Liu, X.Tian, G.Ding, H.Zhang “Industry application of digital twin: from concept to implementation”, International Journal of Advanced Manufacturing Technology, vol.121, 2022, pp. 4289–4312.

[9] B.Yang, S.Yang, Z.Lv, F.Wang, T.Olofsson, “Application of Digital Twins and Metaverse in the Field of Fluid Machinery Pumps and Fans: A Review”, Sensors, vol. 22, 2022, article s22239294.

[10] S.Lazarova-Molnara, N.Mohamed, „Reliability Assessment in the Context of Industry 4.0: Data as a Game Changer”, Procedia Computer Science, vol.151, 2019, pp.691-698

[11] Z.Lv, “Digital Twins in Industry 5.0”, Research vol.6, 2023, article 0071.

[12] D.Paternina-Verona, O.Coronado-Hernández, V.Fuertes-Miquel, M.Saba, H.Ramos, “Digital Twin Based on CFD Modelling for Analysis of Two-Phase Flows During Pipeline Filling–Emptying Procedures”, Appl. Sci. vol.15, 2025, article 2643.

[13] T.Syed, M.Khan, S.Jan, S.Albouq, S.Alqahtany, M.Naqash, “Integrating Digital Twins and Artificial Intelligence Multi-Modal Transformers into Water Resource Management: Overview and Advanced Predictive Framework”, AI, vol.5, 2024, pp.1977-2017.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Metrological Validation of a Digital Model for a CMM including Digital Bias Correction

Marcel van Dijk[1], Walter Knulst[1], Devrim Nalbantoglu[1], Gertjan Kok[1]

[1] *Van Swinden Laboratory (VSL), Thijsseweg 11, 2629 JA Delft, the Netherlands, mvdijk@vsl.nl, wknulst@vsl.nl, dnalbantoglu@vsl.nl, gkok@vsl.nl*

***Abstract* – Virtual experiments and digital twins are virtual representations of simulation processes and can aid in crucial tasks, such as interpreting measurement data or evaluating uncertainties. Therefore, virtual experiments and digital twins are essential components in the digital transformation in metrology. However, it is important that the digital model is properly validated, to make sure its output is reliable. This research shows how to validate a virtual coordinate measurement machine (CMM) by means of reference measurements. The output of the digital model is compared to reference measurements of a ring gauge and its consistency is evaluated. The results show that the output of the digital model for the diameter of the ring gauge is consistent with the reference data. The roundness measurements seem to contain an inherent bias, introduced by the CMM. After correcting for this bias, the output of the digital model for the roundness of the ring gauge is also consistent with the reference data.**

I. INTRODUCTION

Virtual experiments and digital twins are detailed software-based simulation tools that can help in estimating the value of a measurand from measurement data or evaluate the uncertainty of such an estimate. See for example in [1], where a framework is established for virtual experiments and digital twins. These tools are seen as a crucial component in the digitalization of metrology, due to their potential uses. Therefore, a lot of research is aimed at developing and using virtual experiments and digital twins, such as in [2] and [3]. However, without properly validating the digital model, the results are not traceable and therefore difficult to use in practice. Typically, these types of digital models can be validated via reference measurements. By performing a reference measurement and processing the measurement data with the virtual experiment, one can test if the outcomes of the digital model are consistent with the known properties of the reference. See for example [4], where a finite element model is validated. In addition, it is also important to know how to deal with inconsistencies in the validation results. To this end a validation study is performed on the virtual coordinate measurement machine (CMM) developed at VSL. The aim of this work is to validate the virtual CMM, including the values and uncertainties of its parameters. Moreover, a general approach will be studied that is aimed at dealing with potential biases in the measurand estimates, which can improve the measurement results. First, the virtual CMM will be described, including the experimental determination of the virtual CMM's model parameters. Then, the method to evaluate the uncertainty of a new measurement, using the virtual CMM, will be explained, as well as the method to correct for potential biases. Hereafter, the details of the validation measurements are provided, after which the results are discussed. The paper concludes with a discussion of the results.

II. VIRTUAL CMM

The aim of this research is to validate a virtual CMM, which is a virtual representation of a CMM. In other words, a virtual CMM models how error sources influence the measurement data of a CMM, both random and systematic error sources. The measurement data of a CMM includes coordinates that describe the surface of an object. From these coordinates, geometrical features of the object can be derived, i.e. the radius of a circular object. Typically, the geometrical features are estimated via optimization procedures, such as least-squares or minimum zone optimization, see [5]. The remainder of this section describes the errors sources considered in the virtual CMM and how their values and uncertainties are determined experimentally. The order of the sections corresponds to the order in which the different error sources are applied in the virtual CMM.

A. Workpiece surface roughness

Roughness of the surface of the workpiece affects the measured coordinates by the CMM. Roughness can have a significant impact on certain measurands that depend on the coordinates of single data points, such as roundness or flatness, as the roughness effects cannot be averaged out in a single data point. The local roughness is typically unknown, but usually a sufficiently accurate estimate is available for the standard deviation of the roughness. This estimate can, under certain conditions, be determined via CMM measurements. Alternatively, the roughness can be estimated via roughness tables from literature that take the measurement process into account. In the virtual CMM,

the roughness is modelled as random deviations in the direction of the surface normal, where the standard deviation (and therefore the height of the simulated roughness) is determined by a roughness measurement.

B. Workpiece temperature

Changes in the temperature of the workpiece causes the workpiece to expand or shrink. It is therefore important to correct for the workpiece temperature, such that all values for the measurands are reported at the standard reference temperature of 20 °C, as described in [6]. This way, the different measurements can be compared. It is assumed that the workpiece is acclimatized to the room temperature during the waiting time before performing the measurements, such that temperature during the measurement determines the workpiece temperature. Moreover, it is assumed that the temperature of the workpiece is homogenous in both time and space, i.e. there is no temporal and spatial temperature gradient. The uncertainty of the temperature measurement, as well as the uncertainty of the coefficient of thermal expansion of the workpiece, determine the uncertainty contribution of this error source.

C. Kinematic errors

The kinematic errors of a CMM refer to rotations and translations the machine parts can exhibit along the different axes. Considering a 3D CMM this adds up to 18 errors (3 rotation and 3 translation errors per axis). In addition, between each of the axes a squareness error exists, which adds an additional 3 kinematic error sources. The 21 kinematic error sources were characterized (magnitude and uncertainty) by using a calibrated hole plate. This approach is similar to [7].

D. Probe error

The coordinates of the measured object are determined by contact with a probe. The CMM registers the coordinates at the centre of the probe and adds the radius of the probe (in the probing direction) to determine the coordinates of the object surface. However, errors in the radius estimate of the probe tip, as well as any form imperfections of the probe tip will result in errors in the measured coordinates. Furthermore, the electronics and sensors in the active probe head also add to the probe uncertainty. To account for this, the probe error is characterized (value and uncertainty of radius and directional error) by using a calibrated sphere, see also [8].

E. Random noise

In addition to the systematic error sources, there is always random noise present in the data, for example caused by random vibrations. To determine the standard uncertainty of the noise, the CMM measurements of the calibrated artefact were compared to the reference data. It is assumed that the random noise has a high frequency, as each data point contains random noise in a different direction. The low frequency components of the data therefore constitute the workpiece form deviations and systematic CMM errors. describes the random noise. The low frequency component is determined with the reference data (under the assumption that the noise is negligible in the reference data) and subsequently subtracted from the CMM data to obtain an estimate of the standard deviation of the noise. This analysis was performed for both scanning and touch trigger data, to determine the random noise uncertainty for both measurement techniques.

III. UNCERTAINTY EVALUATION

The use of the virtual CMM is twofold. On one hand, it can be used to correct for known systematic CMM errors affecting the measurement data, to improve the estimate of the measurand. On the other hand, the virtual CMM can be used for uncertainty evaluation. The uncertainties (or rather, probability distributions) of the different error sources can be propagated through the virtual CMM to obtain realistic samples of the measurement data. Subsequently, the uncertainty in the measurands can be determined by applying the fitting routines to the measurement data samples. In this section, the method to evaluate the uncertainty, using the virtual CMM, is described.

Propagation of distributions (PoD) is applied to evaluate the uncertainty in the measurands. To this end, the Monte Carlo method (MC) is employed. First, random samples of the error sources are drawn based on the respective uncertainties that are given by the characterizations. Then, for each random sample, the errors are applied to the measurement data, using the virtual CMM. Finally, the optimization procedure determines the measurands for each sample. This results in MC samples of the measurands, from which the standard uncertainty and other statistics can be derived.

Additionally, a method discussed in [9] was employed. Here, it is argued that potential biases can be recognized by the virtual experiment. The values of the measurands for the input data can be determined by applying the fitting routines. If these estimates differ from the means of the MC samples (which have the measurement data as input), then the errors introduce a bias in the estimates of the measurands. The estimate of this bias is equal to the difference between the estimate of the measurands based on the input data and the mean of the estimates provided by the MC samples. It stands to reason that this bias is also present in the measurement data, assuming that the virtual CMM is a realistic representation of the actual CMM. Therefore, the estimate of the measurands can be improved by subtracting the bias. In contrast to [9], in this paper the approach for bias correction is applied to real measurement data instead of synthetic data.

IV. MEASUREMENTS

In order to validate the virtual CMM and its uncertainty evaluation described in the previous sections, several measurements were performed on a ring gauge. In this validation, the diameter and roundness of the inner ring with a nominal diameter of 171 mm were studied. The roundness (RONt – peak-to-valley roundness deviation) is defined as the maximum deviation from the reference circle minus the minimum deviation from the reference circle, consistent with [5]. The reference circle is described by a maximum inscribed circle fit of the inner ring. Moreover, a low-pass Gaussian filter has been applied to the data before determining the roundness, to reduce the impact of random noise. This is also in line with [5].

First, reference measurements were performed. The diameter of the inner ring was measured with a laser interferometer, based on two points of the ring gauge (at 0° and 180°). In addition, a roundness meter (Mitutoyo Roundtest RA-H5100) measured the full profile of the inner ring. From this profile, the reference roundness of the inner ring was derived. Moreover, the average diameter over the entire ring gauge was derived from the profile. First the average profile deviation was set to zero at 0° and 180° and all other profile deviations were adjusted accordingly. Hereafter, the average deviation of the entire profile with respect to these two points was determined. See Figure 1 for an example of the measured ring gauge profile.

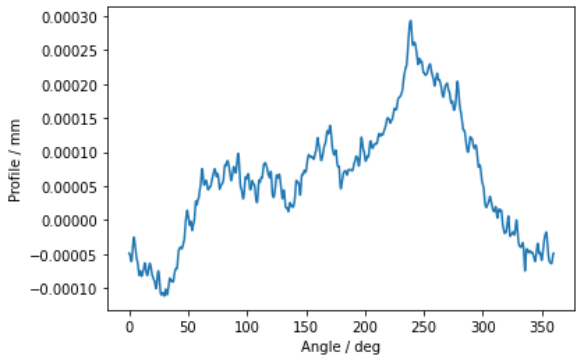


Figure 1: Profile of the ring gauge, as measured by the roundness meter.

Finally, the reference diameter was defined as the distance between the 0° and 180° points of the ring gauge, as measured by the laser interferometer, plus the average deviation from this distance of the entire profile, as measured by the roundness meter. This process is repeated three times: at the top, in the middle and at the bottom of the ring gauge.

The uncertainty of the reference roundness follows from the uncertainty budget of the roundness meter. The uncertainty of the reference diameter is a combination of the uncertainty of the laser interferometer measurement, as well as the uncertainty of a single point of the profile roundness measurement. Note that the uncertainty of the roundness meter is negligible compared to the uncertainty of the laser interferometer, so in practice the uncertainty of the reference diameter is equal to the uncertainty of the laser interferometer measurement.

In addition to the reference measurement, multiple measurements were performed on the CMM, measuring the calibrated ring gauge. By applying the virtual CMM to the measurement data, the diameter and roundness were determined and their uncertainties were evaluated. Consistency between the results of the reference measurements and the CMM measurements provides the validation of the virtual CMM. This consistency was evaluated with the well-known E_n value, which is given by

$$E_n = \frac{x_{\text{ref}} - x_{\text{meas}}}{k\sqrt{(u_{\text{ref}}^2 + u_{\text{meas}}^2)}}, \qquad (1)$$

where x_{ref} and x_{meas} are the estimates for the measurands of the reference and validation measurement, respectively, k is the coverage factor that was set equal to 2 and u_{ref} and u_{meas} are the standard uncertainties of the reference and validation measurement, respectively.

The idea behind the performed measurements is to test the performance limits of the CMM, in order to validate the virtual CMM to the largest extend. First of all, the measurements were repeated at different locations in the measurement volume of the CMM. Typically, measurements are performed in the middle of the CMM, but during this validation study additional measurements were performed at the corners of the CMM, as well as at elevated levels. By measuring at 7 different locations, the entire measurement volume of the CMM was validated. Moreover, the ring gauge was rotated during several measurements. This allows for making a distinction between the effects due to the shape of the ring gauge (which is rotated in these measurements) and the effects due to errors in the CMM (which remain constant between the measurements). The ring gauge was measured in 4 different orientations. Finally, all measurements are repeated twice, one scanning measurement and one touch trigger measurement, at 3 locations of the ring gauge (top, middle, bottom), leading to a total of 66 measurements.

V. RESULTS

In this section, the results of the virtual CMM validation study are discussed. First, the E_n values are discussed to provide a general overview of the validation results. Then, the results for the different measurement locations are compared, as well as the results where the ring gauge is rotated at different angles. Finally, the results after correcting for the bias corrections are discussed.

A. E_n values

The E_n values, calculated according to equation (1), of the diameter and roundness for the different measurements can be found in Figure 2 and Figure 3, respectively. For

the diameter measurements, all E_n values are between the thresholds of -1 and 1. This provides a strong indication that the diameter estimates are consistent with the reference measurements. The diameter seems to be a bit underestimated on average, as most E_n values are above 0. This is potentially due to the fact that the kinematic errors that are not corrected for, but only taken into account in the uncertainty. For the roundness, about 6 % of the measurements have an E_n value below -1. Moreover, all roundness estimates are higher than the reference roundness, as all E_n values are below 0. This is similar to the observations in [9], where it was found that there is a bias in the roundness estimate. Therefore, a correction for this bias, as discussed in section IV, seems appropriate.

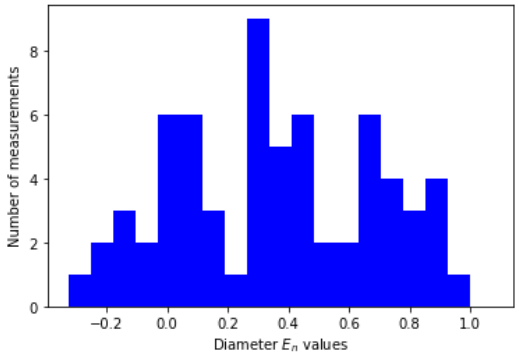


Figure 2: E_n values of the measured diameter.

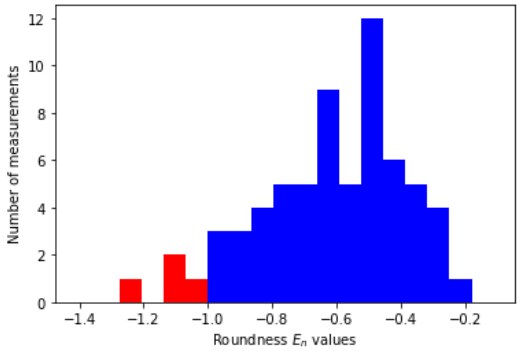


Figure 3: E_n values of the measured roundness.

B. Location dependence

The measurements were repeated in different locations of the measurement volume of the CMM. This way, the measurement location dependency could be studied. In Figure 4 a comparison is made between two different measurement locations. This figure shows the measurement residuals with respect to the reference circle of two different measurements at the centre of the CMM measurement volume, versus residuals in of one measurement in the corner of the measurement volume. It is clear that the measurements performed in the centre of the measurement volume are very comparable, while they are distinctly different from the measurement performed at the corner of the measurement volume. This means that the (kinematic) errors strongly depend on the measurement location and it is therefore good to take the location dependency into account when applying/correcting the kinematic errors.

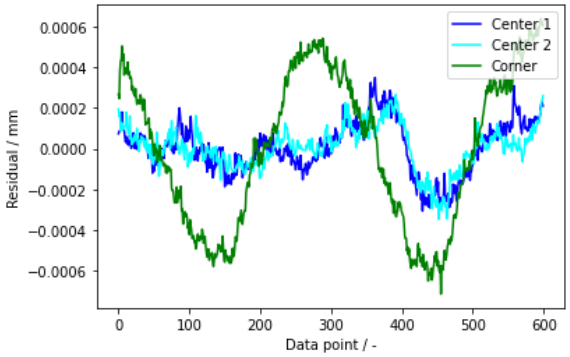


Figure 4: Comparison of residuals with respect to reference circle for measurements performed in the centre of the CMM versus the corner of the CMM.

C. Rotation

In Figure 5 the differences between the different rotation measurements are shown. The residuals with respect to the reference circle are very similar between the different rotation measurements. This shows that the measurement errors in the CMM are more dominant than the form deviations in the ring. Indeed, the measurement uncertainty for a single datapoint is approximately 0.8 micrometre for these measurements. The deviations shown in Figure 5 fall well within this uncertainty. Hence, for this reference object, with form deviations in the order of only 200 nm, the CMM is not able to reconstruct the profile. In case the form deviations of the measured object are in the same order of magnitude as the CMM errors, then it would probably be possible to detect a combined effect of the form deviations and CMM errors. When the form deviations are significantly larger compared to the CMM errors, then the rotated CMM measurements should be able to detect the profile of the object.

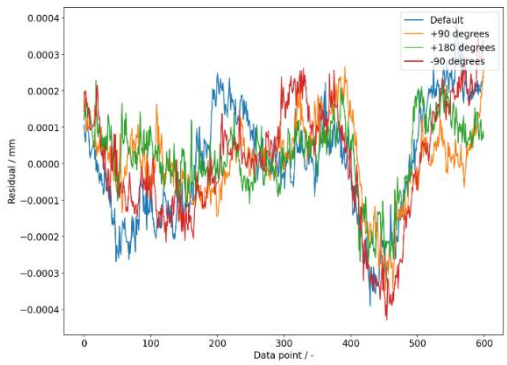


Figure 5: Residuals with respect to reference circle for the rotation measurements.

D. Bias correction

The mean diameter of the simulated MC samples is approximately equal to the diameter estimate based on the measurement data, so the diameter estimate is not biased. The mean roundness of the simulated MC samples is almost always higher than the roundness estimate of the measurement data, so the roundness estimate is biased. Therefore, the bias is subtracted from the roundness estimates for each measurement. The resulting E_n values can be found in Figure 6. The E_n values are all between -1 and 1, indicating consistency with the reference measurement. This provides a strong indication that the accuracy is improved after the bias correction. Applying the bias correction could result in a negative roundness estimate, which is physically impossible. It is most likely that the bias is overestimated due to an overestimated uncertainty of the error sources. In such a case, it is not recommended to correct the bias, but the high uncertainty will likely result in a sufficiently wide confidence interval, such that it captures the true roundness.

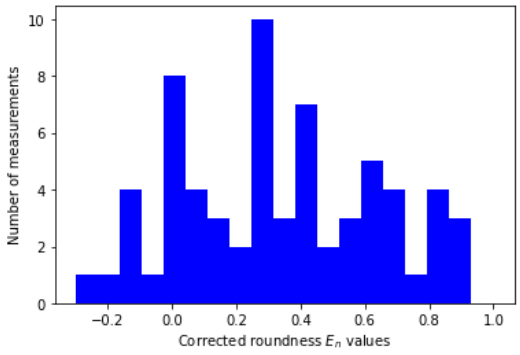


Figure 6: E_n values of the corrected roundness.

VI. CONCLUSION

In this research a digital model of a CMM was validated. These types of models are a crucial component in the digitalization of metrology and can only be properly used when validated. For different measurement locations and object orientations, the digital model produces outcomes are consistent with the reference. However, the roundness estimates seem to contain a bias. Assuming that the digital model is representative of the CMM, this bias can be estimated and corrected for to obtain more accurate results. Moreover, it was shown that the measurement location has a significant impact on the errors in the measurement data, while the orientation of the object had no significant impact in this study. However, the object orientation may become more relevant when the form deviations in the object are larger.

In addition to the work done in this research, it is also possible to correct for some of the CMM errors. This way, the estimated value of the measurand can be improved and the uncertainty of the estimate reduced. Finally, it is important to note that there is added value in validating a digital model in multiple ways. For example, the digital model can also be validated with different measurement objects. This may expose different aspects of the digital model, i.e. an object with sharp edges may introduce additional errors to the data that are not covered by studying a smooth ring. Hence, the validation of the digital model is only relevant if the objects that are measured have similar features as the object that was used for the validation. If this is not the case, it is advisable to repeat the validation procedure with a different object. In addition, it is also advised to validate the digital model by means of a comparison with other parties, in order to confirm consistency with the outside world.

FUNDING STATEMENT

The project 23IND12 ADAM has received funding from the European Partnership on Metrology, co-financed from the European Union's Horizon Europe Research and Innovation Programme and by the Participating States.

REFERENCES

[1] G. Maculotti, et al, "A Shared Metrological Framework for Trustworthy Virtual Experiments and Digital Twins", Metrology, vol.4, No.3, 2024, pp. 337-363.

[2] F. Hughes, et al, "Using virtual experiments to improve data analysis", Meas. Science and Technology, vol.36, No.4, 2025, 046005.

[3] L. Wright and S. Davidson, "Digital twins for metrology; metrology for digital twins", Meas. Science and Technology, vol.35, No.5, 2024, 051001.

[4] P. S. Anderson, et al, "Virtual experiments, physical validation: dental morphology at the intersection of experiment and theory", Journal of the Royal Society Interface, vol.9, No.73, 2012, pp. 1846-1855.

[5] ISO 12181-2, "Geometrical product specifications (GPS) – Roundness – Part 2: Specification operators", 2011.

[6] ISO 1, Geometrical Product Specifications (GPS) - Standard reference temperature for the specification of geometrical and dimensional properties, 2022.

[7] H. Kunzmann, E. Trapet, F. Wäldele, "A Uniform Concept for Calibration, Acceptance Test, and Periodic Inspection of Coordinate Measuring Machines Using Reference Objects", CIRP Annals, vol.39, No.1, 1990, pp. 561-564

[8] P.C. Miguel, T. King, A. Abackerli, "A review on methods for probe performance verification", Measurement, vol.23, No.1, 1998, pp.15-33.

[9] M. v. Dijk, G. Kok. "Comparison of uncertainty evaluation methods for virtual experiments with an application to a virtual CMM", Measurement: Sensors, 2025

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Towards Safety and Accuracy of Hydrogen Refuelling Stations through Digital Twins

Christian Klaus, Ricardo Soruco Aloisio, Claudia Koch, Matthias Prellwitz

Bundesanstalt für Materialforschung und -prüfung (BAM), Unter den Eichen 87, 12205 Berlin, Germany, {firstname}.{lastname1[-lastname2]}@bam.de

***Abstract* – This paper presents digital quality infrastructure methods for hydrogen refueling stations using the Asset Administration Shell as a standardized digital twin. Implemented at BAM's test platform, it integrates real-time sensor data, calibration certificates, and compliance documents to support traceable, interoperable asset management. In combination with AI and semantic tools, the system will enable predictive maintenance, remote audits, and improved safety. This approach reduces downtime, enhances transparency, and offers a scalable model demonstrating the potential of digital twins in advancing metrological traceability and operational efficiency in hydrogen technologies.**

I. INTRODUCTION

The quality infrastructure (QI) ensures the safety, reliability, and interoperability of products and services. Traditionally reliant on paper-based, oftentimes manual processes, QI is now being transformed by digital technologies. The German initiative QI-Digital brings together key stakeholders to develop integrated digital tools and ecosystems that modernize QI for today's production systems and global value chains [1]. Concrete use cases have been set up with dedicated R&D and test platforms established at the Federal Institute for Materials Research and Testing – Bundesanstalt für Materialforschung und -prüfung (BAM), one of the key players.

The use cases are centered around quality assurance in modern production and for technical facilities, whereas the latter is represented by a state-of-the-art hydrogen refueling station (HRS). As hydrogen applications and technologies scale up in response to the energy transition, the demand for solutions that ensure a safe and efficient operation is growing.

This paper presents approaches taken in our HRS reality lab to develop, test and demonstrate such digital QI solutions – applicable also for other types of technical facility and settings beyond hydrogen infrastructure. First, we introduce a vision for a digital QI in HRS and elaborate the role of digital twins. We then report on the implementation of digital QI in our HRS test platform building on the Asset Administration Shell (AAS). An outlook to potential extensions of the work concludes our paper.

II. QUALITY INFRASTRUCTURE FOR HRS: STATE OF THE ART AND VISION

Hydrogen refueling stations (HRS) are critical for the hydrogen mobility transition but pose significant safety and operational challenges due to risks like overpressure, leakage, and explosion. Moreover, an examination of the database of hydrogen-related incidents and accidents [2] emphasises the imperative for precise component specifications. Efficient access to information and tools is essential. Therefore, digital twins should integrate information on standards and relevant results from the quality infrastructure, including evaluation certificates.

HRS are subject to stringent safety, measurement accuracy and environmental compatibility requirements. International standards (e.g., SAE J2601, ISO 19880-1:2020) define strict operational limits, necessitating advanced monitoring and control systems. Independent testing organizations must verify that high-pressure and nozzle coupling systems, explosion-prone zones and calibrated, traceable dispensing measurements all meet technical and legal standards. Currently, these verifications are typically conducted through separate testing and certification processes, which often rely on extensive paper- or PDF-based documentation. Any modification, such as replacing a compressor, triggers a new time-consuming manual documentation process, which delays commissioning, increases costs, and reduces transparency for operators, authorities, and service partners. In addition, regular inspections of pressure vessels, hazardous area (Ex zone) assessments, and calibration of measurement and safety sensor systems are mandatory.

A digital QI that makes use of enhanced tools and processes as well as latest, standardized technologies can offer an efficient solution to these challenges. QI tools like machine-readable SMART standards, digital calibration and conformity certificates (DCC, D-CoC), test reports (DTR) and accordant digital trust system, process control system (PCS) data can be directly integrated into the AAS data system. All quality-, process-, and safety-relevant data – from temperature sensors to fill level monitoring – is collected in real time in a digital container. This data is structured as an AAS and is fully linked to the PCS. This eliminates discontinuities between different media, accelerates workflows and ensures interoperability, preventing costly issues from arising during operation.

Operators benefit from reduced downtime and real-time access to compliance status. Testing organizations can perform remote audits using machine-readable, real-time data in digital SI format, while service original equipment manufacturers (OEMs) can use verified original data for predictive maintenance algorithms. A digital QI thus can transform the complex safety and measurement chain of a HRS into a transparent, real-time data flow — an approach that can be applied to any technical system with high compliance requirements.

Digital twins (DT) offer a powerful means to model these systems virtually, enabling timely monitoring, predictive maintenance, and AI-driven optimization. The next section outlines the state of art of DT for HRS.

III. DIGITAL TWINS FOR HRS

A digital twin is defined as a digital representation of a physical product, system or service that is linked to it throughout its entire life cycle. The objective is to facilitate analysis, optimization, control and simulation of technical systems, with a particular emphasis on real-time applications. [3]

A. State of the Art

Digital twins facilitate data management and access, including in complex technical facilities such as hydrogen refueling stations.

Safety and Monitoring. Digital Twins are increasingly recognized as key enablers for improving HRS safety, efficiency, and reliability. They integrate real-time sensor data, IIoT (Industrial Internet of Things), and AI to simulate and monitor system behavior. Several studies propose DT frameworks that incorporate statistical process control, real gas modeling, and MATLAB simulations to monitor parameters such as hydrogen flow and temperature [4]. A mobile DT-based leak detection system combines IoT sensors with geospatial mapping (e.g., Google Maps) and time-series databases to enable location-specific, real-time safety monitoring [5].

AI-Enhanced Simulation and Risk Prediction. Advanced DT models leverage machine learning for predictive analytics. One approach integrates real-time sensor data into a 3D simulation using convolutional neural networks (CNNs) and Unity 3D to visualize and predict hazardous conditions, supporting rapid decision-making [6]. Another review highlights the role of AI techniques – such as deep learning, reinforcement learning, and federated learning –across the hydrogen value chain, emphasizing lifecycle coverage, model interpretability, and regulatory compliance [7].

Architectures and Technical Challenges. A comprehensive review outlines digital twin architectures comprising data, integration, and application layers, with features like real-time monitoring, predictive maintenance, and 3D visualization [8]. Key challenges include data scarcity, computational demands, and algorithm robustness. Standardization of data formats, cybersecurity, and scalability across diverse HRS configurations remain open research areas. Current efforts focus on interoperable platforms, federated data spaces, and AI-driven analytics to address these gaps.

B. Challenges

The effectiveness of DTs hinges on overcoming several key challenges, particularly regarding data quality and QI:

Data Integration and Interoperability. HRS components often originate from diverse manufacturers, requiring standardized data models and semantic to ensure consistent, high-quality data exchange across systems.

Sensor Reliability and Data Quality. DTs depend on accurate, real-time sensor data under extreme conditions (e.g., high pressure). Ensuring sensor calibration, redundancy, and fault detection is critical to maintaining model integrity and avoiding unsafe decisions.

Quantification of Uncertainties. Interoperability and traceability are important goals. Calibration of sensors is a very important step towards these goals; however, the quantification of uncertainties identified along the measurement chains is no less important to achieve these goals. Thus, in the future we should also quantify these uncertainties. This especially applies if an AI approach is implemented, as models should be trained and validated considering these uncertainties.

Model Validation and Responsiveness. High-fidelity simulations of thermodynamic and fluid processes must be continuously validated against real-world data. Balancing model accuracy with real-time performance remains a persistent challenge.

Quality Infrastructure and Metrological Traceability. A fundamental yet frequently disregarded challenge pertains to the incorporation of quality infrastructure – comprising metrology, standardization, certification, and conformity assessment – into digital twin ecosystems. A robust digital QI is essential for ensuring the traceability of key measurements (pressure, temperature, flow) to standards. DTs should support sensor calibration, model validation, and compliance with technical norms to ensure data reliability and regulatory acceptance – especially vital in safety-critical applications like hydrogen fueling.

Cybersecurity and Governance. As DTs operate within connected ecosystems, secure data handling, access control, and regulatory compliance are non-negotiable. Proprietary Solutions. Accessing data across stakeholders such as manufacturers, operators, or calibration labs requires familiarity with proprietary systems and their semantics, due to the lack of standardized, industry-neutral interfaces.

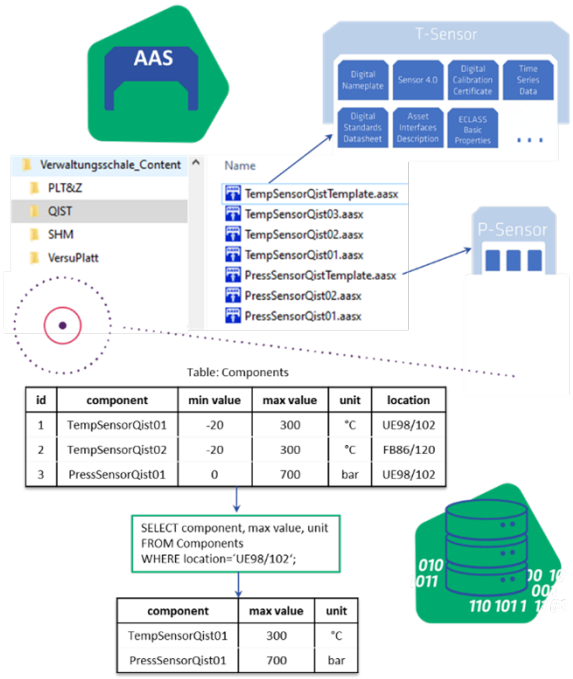


Fig. 1. Differences between Asset Administration Shell (top) and relational databases (bottom)

Many digital twins remain proprietary, tailored to specific tools (e.g., MATLAB, ANSYS) or platforms (e.g., Azure, AWS), often lacking semantic interoperability. The Asset Administration Shell employs defined interfaces, data models and semantics to facilitate interoperability.

C. Asset Administration Shell (AAS)

The AAS represents a standardized subset within the broader digital twin landscape, particularly suited for regulated, multi-stakeholder contexts requiring traceability and data sovereignty. As the standardized digital twin of a physical or logical asset, it enables semantic interoperability and consistent lifecycle management. It provides a structured, machine-readable interface for data exchange, control, and integration within Industry 4.0 environments such as HRS.

The AAS is structured modularly, comprising submodels that represent specific asset aspects—such as identification, documentation, or condition monitoring—using standardized semantics [9]. This allows flexible adaptation to various use cases while ensuring interoperability. For example, a temperature sensor's AAS may include submodels for identification, calibration, live data, diagnostics, and compliance (e.g., ISO/IEC 17025). To ensure compatibility, submodels should follow *Industrial Digital Twin Association (IDTA)* specifications, developed with industry and research partners, and aligned with international standards.

Relational databases are powerful for managing structured data, offering centralized querying and aggregation. However, they often lack standardized semantics and interoperable interfaces, making cross-organizational data exchange difficult and reliant on project-specific knowledge. The AAS takes a different approach by using a hierarchical, file-based XML structure to represent each asset (cf. Fig. 1). While it lacks centralized querying, its strength lies in providing a standardized, decentralized, and self-contained digital interface for asset data. Rather than replacing databases, the AAS complements them by adding a semantic layer that supports interoperable data exchange, system integration, and digital lifecycle management.

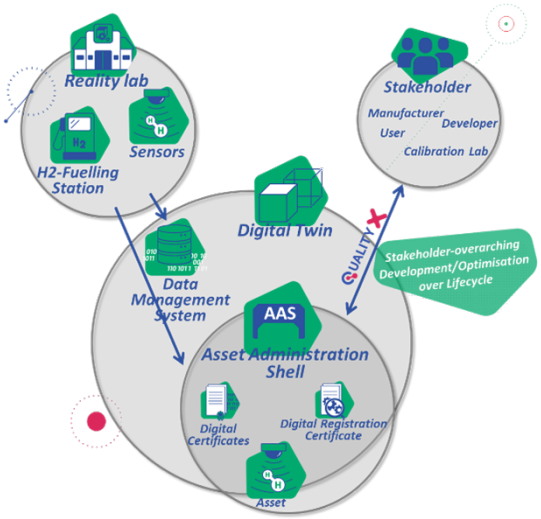


Fig. 2. Digital Twin of BAM's Reality Lab with AAS

With this, the AAS represents a valuable tool to enable the integration of digital QI processes and tools allowing for advanced quality assurance of HRS. The next section presents insights from the ongoing implementation in our HRS test platform.

IV. IMPLEMENTATION IN THE HRS TEST PLATFORM

The HRS at BAM functions as a testbed, thus facilitating collaboration with industry and research partners [10]. This enables the implementation of new or modified digital QI methods within a real-life technical facility. The initial notions of data retrieval, management and sharing are presented, in addition to the interplay between multiple processes.

A. Test Platform

Our QI-Digital reality lab comprises a complete state-of-the-art hydrogen refueling station (HRS) augmented with technologies such as sensors and monitoring systems (see Fig. 2). The test platform is supplied with green hydrogen and incorporates advanced process control technologies,

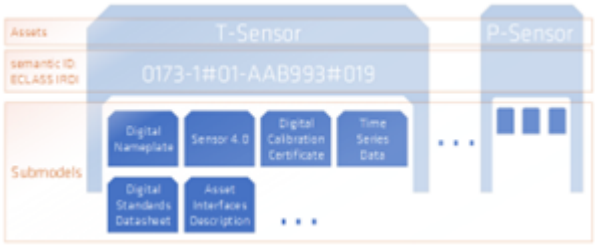


Fig. 3. Structure of the Asset Administration Shell

contributing to the development of robust digital standards and infrastructures that support the safe, scalable deployment of hydrogen technologies. Within this setup, various types of data are generated, including process data (time series from sensors), sensor metadata, and associated documentation and certificates. These data are stored in a proprietary data management system. To effectively manage, analyze, and utilize this heterogeneous, dynamic data, the implementation of a digital twin station – a virtual representation of the physical system – becomes essential and serves as a foundational tool for advancing quality assurance in digitally transformed technical systems.

B. Integrating the Asset Administration Shell

Our concept involves mapping selected assets and relevant documents in our proprietary data management system using AAS (cf. Fig. 2). For instance, this could be a temperature sensor used to monitor the refueling process, including the associated digital calibration certificate (DCC) [11]. We primarily use IDTA-published submodels for this purpose. This enables different stakeholders to access these standardized AAS without requiring knowledge of the proprietary data management systems of other organizations. The focusses here are on metadata, digital certificates, and process data.

C. Data Modeling through submodels

Submodel templates like Technical Data, Digital Nameplate, and Time Series Data were selected for the HRS test bed. The DCC submodel, central to *UseCaseH2*, includes calibration certificates and supports metrological standards, even though not yet officially published [12].

D. Semantics through ECLASS dictionary

ECLASS is a globally recognized semantic dictionary that enables machine-readable, interoperable asset descriptions [13]. Recommended by *Plattform Industrie 4.0* for AAS implementations, it ensures unambiguous, standardized semantics through unique identifiers (IRDIs), as shown in Fig. 3. The AAS natively integrates with the ECLASS catalogue, allowing direct import of assets and submodels using standardized classifications. This supports reusable, harmonized submodels across domains, enhancing semantic interoperability and digital quality infrastructure. It also facilitates traceable, verifiable, and auditable asset descriptions for compliance, certification, and lifecycle documentation. It is evident that the prototype under consideration employs the ECLASS vocabulary to facilitate interoperability among multiple stakeholders regarding the description of QI-related assets.



Fig. 4. AAS of a temperature sensor in AASX Package Explorer with Calibration Certificate

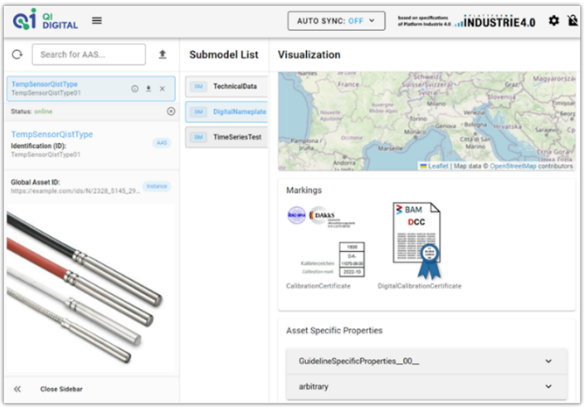


Fig. 5. BaSyx Web UI for AAS

E. Testbed at BAM

Various procedures exist for creating AAS, which should be used in the different implementation phases. Initially, the AAS are created manually via the Graphical User Interface (GUI) of the AASX Package Explorer software (see Fig. 2). To interact with and visualize the created AAS, an instance of the web GUI of the Eclipse BaSyx project has been implemented (cf. Fig. 5).

AAS templates with the relevant submodels are created for asset types (e.g. temperature sensors) and filled with metadata valid for sensors types (e.g. manufacturer). Then, an AAS instance is cloned from an AAS template and its submodels are manually filled with instance-specific metadata (e.g. the serial number) via the GUI. This creates ad-

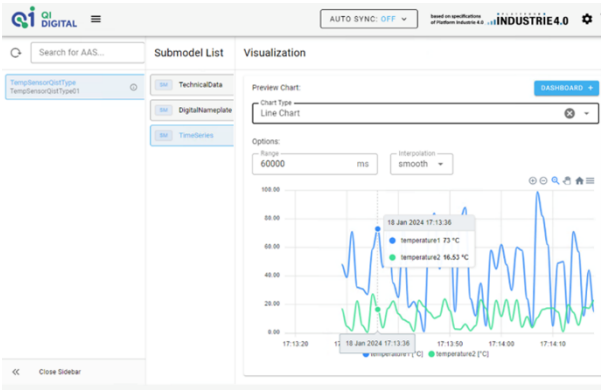


Fig. 6. Submodel Time Series Data

ministration shells for selected assets that contain all the relevant metadata available.

All the asset's existing metadata is stored in the Digital Nameplate submodel. This includes digital QI documents (e.g. a temperature sensor's calibration certificate). As soon as the DCC submodel becomes available, these documents will be stored there instead. Measurement data (e.g. H2 temperature during the fueling process) is stored in the Time Series Data submodel in the form of CSV files. These can be visualized interactively (see Fig. 6). Technically, an automated import of measurement data could also be implemented.

F. Use Case upon Quantitative Risk Assessment

The need of a standardized infrastructure and its support for process and safety optimization is demonstrated in a use case upon Quantitative Risk Assessment (QRA) at BAM [14]. Key of the assessment is building a database of different hydrogen refueling stations. The more data, the better the probabilistic statements can be supported. Therefore, the data must be accessible from the sensor field level and interpretable in terms of their semantics. Specially if stakeholders beyond other ecosystems are involved. In this regard a framework is deployed for transforming PCS data into the QRA database. The framework retrieves data from the Process Control System (PCS) via an API (application programming interface) embedded or described within the station's asset administration shell, which is accessible exclusively through a secure Virtual Private Network (VPN). The collected data include all executed control commands, process variables, automatically generated fault logs from system components, and triggered alarms. This data acquisition occurs at fixed intervals, with the system processing the data between each retrieval cycle. Further documentation on all sensors and actuators can be accessed via the AAS. On that account the framework is capable to work with other environments that support AAS and enhance the automation on processing data.

V. DIGITAL TWINS ENVISION QI APPLICATIONS

The complexity and safety-critical nature of hydrogen infrastructures demand innovative solutions for monitoring, control, and optimization. The integration of digital twins and artificial intelligence (AI) offers a transformative approach.

Digital twins – dynamic, virtual models synchronized with real-time plant data – combine current system states with historical, physical, and semantic data, creating a rich environment for training and deployment. Digital twins are essential for safety-critical systems, providing structured, interpretable data streams. Without them, application deployment would involve significant uncertainty and risk. Key scenarios include:

Condition Monitoring & Predictive Maintenance. DT replicate physical assets of a HRS – including storage tanks, compressors, valves, and sensors – in a virtual environment. AI techniques like anomaly detection with autoencoders or time-series analysis with long short-term memory (LSTM) networks enable fault detection at early stages and demand-driven maintenance, improving availability and reducing costs.

Energy & Resource Efficiency. Algorithms, supported by historical and real-time data, can optimize energy flows. Reinforcement learning helps reduce compressor energy use and adapt cooling strategies, lowering load peaks and emissions.

Real-Time Safety Management. AI can detect critical conditions through pattern recognition (e.g., pressure spikes) and triggers emergency protocols. DT also support forensic analysis by reconstructing past states.

Simulation & Training. Virtual environments allow staff to train in realistic scenarios. AI can personalize learning paths and simulate incidents to improve readiness.

Integration with Energy Systems. Coupling digital twins with smart grids and mobility platforms enables dynamic control of HRS within broader energy networks. Forecasting models support demand-driven energy supply and pricing – key for fleet operators and sector coupling.

VI. CONCLUSION AND OUTLOOK

As a key principle of the AAS is to support assets throughout their lifecycle, collaboration with suppliers or manufacturers is recommended. Given the high value of data and the AAS's interoperability, this approach offers significant potential for multi-stakeholder access. A team is developing predictive maintenance use case for high-pressure gas tanks, such as those at hydrogen refuelling stations, using AI-based acoustic and ultrasound technologies. Achieving this requires collaboration with software and hardware providers to integrate DCC of pressure sensors into industrial control systems and deliver traceable, metrology-validated DT-based subsystems to stakeholders.

VII. ACKNOWLEDGEMENTS

QI-Digital is an initiative of the key players in the German quality infrastructure (QI): BAM, DAkkS, DIN, DKE and PTB. The Federal Ministry for Economic Affairs and Energy (BMWE) recognizes the pivotal role of QI-Digital in safeguarding the success of innovative technologies, products and processes.

REFERENCES

[1] "QI-Digital initative". (accessed July 15, 2025). URL: https://www.qi-digital.de.

[2] European Commission. "The Hydrogen Incident and Accidents Database (HIAD 2.1)". (accessed July 15, 2025). URL: https://minerva.jrc.ec.europa.eu/en/shorturl/capri/hiadpt.

[3] Rainer Stark, Carina Fresemann, and Kai Lindow. "Development and operation of Digital Twins for technical systems and services". In: *CIRP Annals* 68.1 (2019), pp. 129–132. ISSN: 0007-8506. DOI: 10.1016/j.cirp.2019.04.024.

[4] Raduan Sarif. "Digital twin for hydrogen refueling stations: An analytical study for safe and efficient operations". In: *10. Fachsymposium des Vereins HybridSensorNet (HSN), KIT*. 2023. URL: https://opus4.kobv.de/opus4-bam/frontdoor/index/index/year/2023/docId/59034.

[5] Ha-Jin Hwang and Sung-Ho Hwang. "Developing an internet of things system for hydrogen leak detection at hydrogen refueling stations integrating LoRa and global positioning system". In: *Edelweiss Applied Science and Technology* 8.5 (Sept. 2024). DOI: 10.55214/25768484.v8i5.1874.

[6] Na Yeon An et al. "Digital Twin-Based Hydrogen Refueling Station (HRS) Safety Model: CNN-Based Decision-Making and 3D Simulation". In: *Sustainability* 16.21 (2024). ISSN: 2071-1050. DOI: 10.3390/su16219482.

[7] Zaid Allal et al. "A review on machine learning applications in hydrogen energy systems". In: *International Journal of Thermofluids* 26 (Mar. 2025), p. 101119. ISSN: 2666-2027. DOI: 10.1016/j.ijft.2025.101119.

[8] Zhiming Feng et al. "Review of Digital Twin Technology Applications in Hydrogen Energy". In: *CHAIN* 1.1 (2024), pp. 54–74. DOI: 10.23919/CHAIN.2024.000001.

[9] Federal Ministry for Economic Affairs and Energy (BMWi). "Structure of the Asset Administration Shell - Continuation of the Development of the ReferenceModel for the Industrie 4.0 Component". Apr. 2016. URL: https://industrialdigitaltwin.org/wp-content/uploads/2021/09/01_structure_of_the_administration_shell_en_2016.pdf.

[10] QI-Digital initiative. "Reliable Hydrogen Refuelling Station – A Digital Quality Infrastructure (QI) for Technical Systems". (accessed July 15, 2025). 2025. URL: https://www.qi-digital.de/en/reliable-hydrogen-filling-station.

[11] Siegfried Hackel et al. "The Digital Calibration Certificate (DCC) for an End-to-End Digital Quality Infrastructure for Industry 4.0". In: *Sci* 5.1 (2023). ISSN: 2413-4155. DOI: 10.3390/sci5010011.

[12] Kazeem Oladipupo Dr. Sebastian Käbisch. *Digital Calibration Certificate (InterOpera Abschlusspräsentation)*. Oct. 2023. URL: https://interopera.de/wp-content/uploads/2024/01/2023-10_AAS_Digital_Calibration_Certificate_Endpresentation.pdf.

[13] Alexander Belyaev et al. "Modelling the Semantics of Data of an Asset Administration Shell with Elements of ECLASS". June 2021. URL: https://www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/Whitepaper_Plattform-Eclass.html.

[14] Cristian Schaad, Carlo Tiebe, and Katrina M. Groth. "A Framework for Transforming Process Control System Data from a Hydrogen Fueling Station into HyCReD Data". In: *Proceedings of the 35th European Safety and Reliability Conference (ESREL2025) and the 33rd Society for Risk Analysis Europe Conference (SRA-E 2025)*. 2025. URL: https://rpsonline.com.sg/proceedings/esrel-sra-e2025/html/ESREL-SRA-E2025-P6264.html.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Metrology and Architectural Description: a case study of the Hybrid Comparison system

Diego Coppa (ORCID: 0009-0009-4187-5325)[1], Blair Hall (ORCID: 0000-0002-4249-6863)[2], Anjali Sharma (ORCID: 0000-0003-0389-046X)[3]

[1]*National Institute of Industrial Technology, Buenos Aires, Argentina, dcoppa@inti.gob.ar*
[2]*Measurement Standards Laboratory, Lower Hutt, New Zealand, blair.hall@measurement.govt.nz*
[3]*National Physical Laboratory, New Delhi, India, anjali@nplindia.org*

***Abstract* – An architectural approach has been explored as a means to describe a system for managing a metrological process known as a *hybrid comparison*. This top-down approach, based on the international standard for architectural description—ISO/IEC/IEEE 42010—focuses on stakeholder concerns and contrasts with the more common bottom-up development of tools, services, and file formats typically seen in the metrology community. By clearly separating stakeholder concerns, the approach also reveals structural relationships with other external metrological systems. This exploratory study suggests that the perspective provided by architectural descriptions supports the coordinated development of digital metrological systems.**

I INTRODUCTION

The international metrology community has initiated a digital transformation of national and international quality infrastructures (QIs), which deliver traceable measurement services where they are needed in society [1]. The international measurement system, including its associated QIs and other organisations, presents significant challenges for digital transformation because of its size and complexity. So, a Forum on Metrology and Digitalization (Forum-MD) has been established to advise the International Committee on Weights and Measures (CIPM) on the development of an SI Digital Framework and the wider implications of the global digital transformation for metrology [2, 3].

One of the Forum-MD's task groups is focused on metrological semantics. As part of its initial activities, the group looked at whether a top-down architectural approach could help to describe the complicated and, in some respects, unique nature of metrology as an enterprise. A relatively simple metrological process, known as the *hybrid comparison* (HC), was chosen as a case for study [4]. The architecture of a system for managing HCs is interesting, because the HC system must interoperate with other metrological systems, thereby revealing some of the unique architectural challenges in metrology.

The HC was developed by the Asia-Pacific Regional Metrology Organisation (APMP). Management of HCs currently relies on person-to-person exchanges and basic digital tools such as email and office productivity software. The semantics group's objective was to describe a high-level functional architecture for the HC system, which captures the essential conceptual features.

The ISO/IEC/IEEE 42010 standard for architectural descriptions of systems and software was used in this work [5]. The standard incorporates principles for architectural description that support a structured approach to producing meaningful and consistent documentation. The standard also allows specialised architectural description elements to be introduced for features that would otherwise be difficult to handle. Extensibility is attractive for metrology, as the field has some unique concepts, such as measurement uncertainty and metrological traceability.

This article summarises our experience of applying the principles in ISO 42010 to describe the hybrid comparison system. The remainder of the article is structured as follows. Section II introduces some of the main architectural description elements in the context of the HC. Section III looks in more detail at the challenges posed in describing the result of a comparison. Section IV discusses our conclusion that architectural description is a useful tool for metrology during and after the digital transformation process. A short glossary is provided at the end.

II THE HYBRID COMPARISON

A HC evaluates the metrological compatibility of results obtained independently during the comparison by two national metrology institutes (NMIs), each measuring a specific property of a comparison artefact. One NMI will already have an appropriate internationally recognised technical competency. So, the outcome of a HC may provide evidence of a specific technical competency for the other participating NMI. The HC results are published in a final report, which may be used to support a future Calibration and Measurement Capability (CMC) claim under the CIPM Mutual Recognition Arrangement (CIPM MRA). A detailed description of the HC process is provided in an APMP guide [4].

The following subsections introduce some of the principal architectural description ideas (Fig. 1) within the context of the HC.

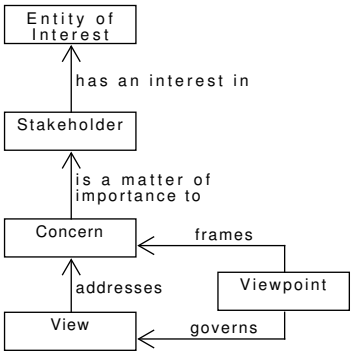


Fig. 1. Relationships between some of the principal elements used in architectural descriptions [5]. The entity of interest in this work is a system for managing hybrid comparisons.

A. Stakeholders

Architectural requirements depend on the various needs and interests of *stakeholders* in the entity of interest. Among the stakeholders in a HC are those typically referred to as *participants*. A HC is bilateral, with two NMIs participating in a comparison. One of them—henceforth, I-NMI—has an existing CMC entry in the CIPM Key Comparison Database (KCDB) [6], recognising that its measurement capability has been assessed and accepted by other NMIs. The other—henceforth, A-NMI—does not have a CMC entry. Accordingly, A-NMI participates in the hybrid comparison to obtain evidence of their competence in a specific measurement capability (the future application process to obtain a CMC belongs to a different metrological system and is not considered here).

Other stakeholders include the chair of the technical committee (TC-Chair) and the regional metrology organisation (RMO) in charge of the HC. The RMO is responsible for effective management of hybrid comparisons and the publication of final reports. The TC-Chair has several responsibilities: they review the initial HC application to determine its acceptability and later evaluate the comparison outcome, assessing the metrological compatibility of the results. Variants of the HC process allow the TC-Chair to appoint a third-party to assist in the technical evaluation of results, and a member of the associated RMO technical committee may be called upon to independently review the final report. However, the TC-Chair may carry out these roles themself; for simplicity, these alternatives are not discussed.

B. Architectural concerns

The particular interests of each stakeholder with regard to the HC must be documented. These architectural *concerns* may be shared among multiple stakeholders. An associative table, which matches concerns with stakeholders, is an acceptable way to document this information, as it allows for easy pairing with other architectural elements. However, during our initial analysis of the HC, flexible tools, such as mind-mapping software, were useful for collecting and organising concerns.

A HC progresses according to a definite sequence. We found it useful to divide this progression into three distinct phases. In the first phase, A-NMI gathers the information required to apply for a hybrid comparison. The I-NMI must agree to this proposal before it is submitted to the TC-Chair for approval. So, these stakeholders share many concerns related to the information that will ultimately determine the configuration of the subsequent HC.

The second phase involves measurements and the reporting of results by each NMI. In this phase, concerns about impartiality arise. A strict protocol is followed to ensure that all results are submitted to the TC-Chair before A-NMI and I-NMI are permitted to access each other's results. An example of a shared concern here is the need to know precisely what stage the comparison is at. Additionally, there are many concerns related to the representation of scientific information that must be addressed. Reporting formats must ensure the safe transmission of all important details. Concerns here include: formats for unambiguous representation of measured quantities, units of measurement, and measurement uncertainties.

The third and final phase involves analysis and reporting. The A-NMI will analyse the comparison using information from its own results and those reported by the I-NMI. This analysis includes an evaluation of the metrological compatibility of the measured values from the NMIs. A final report will be prepared by the A-NMI, incorporating this analysis, and submitted to the TC-Chair for approval. The TC-Chair will make a final assessment of the comparison based on the report. The RMO will then publish the final HC report. Some concerns at this stage carry over from earlier stages. For instance, unambiguous interpretation of technical details reported by each NMI is a prerequisite for the comparison analysis and the evaluation of metrological compatibility. There are also concerns related to publication of the final report. For instance, it should be prepared in accordance with the FAIR principles, with particular attention to the associated metadata [7].

C. Architectural views and models

Among the more valuable lessons learned in this project is an appreciation for the role played by architectural views and the concepts they entail. Views address one or more

stakeholder concerns and are intended to facilitate reasoning about them. They are representations of relevant features of the entity of interest. Each view has a *viewpoint*, which defines how the view is constructed, including notations, modelling languages, and types of models used. While views are often graphical, other forms are possible.

Although this description of views may seem dry, the underlying concept is familiar—much like maps. A map presents certain features of a landscape depending on its intended purpose, such as navigation or land use planning. The conventions used to create a map—such as symbols, scale, and projection—define its viewpoint. A map's legend summarises the viewpoint by explaining what those symbols and elements represent. Most readers will know from experience that a good map—whatever the context—eases navigation, whereas a poor one can be deeply frustrating. The same applies to views, which should be designed to facilitate reasoning about particular concerns.

A view that summarises HC progression is shown in Fig. 2. The viewpoint for this diagram is based on the standardised definition of *sequence diagrams* in the Unified Modelling Language (UML) [8]. Summary information is shown in a legend at the top of Fig. 2, which identifies the stakeholders and distinguishes between synchronous and asynchronous events. The view depicts a comparison's evolution in an intuitive way. It can frame participant concerns, such as what has happened so far, what will happen next, and which upcoming event will involve a particular participant. A view may also serve to draw attention to elements that have been overlooked. For example, in the third phase, should the I-NMI approve the draft HC report before it is submitted to the TC-Chair?

A view is composed of architecture view components, each governed by a model kind or a legend defined in the view's architectural viewpoint. Model kinds determine the conventions used to depict information. For example, the UML defines rules for constructing sequence diagrams, some of which are mentioned in the caption to Fig. 2. Understanding these conventions enables the view to be analysed effectively. Using well-established, standardised model kinds facilitates collaboration.

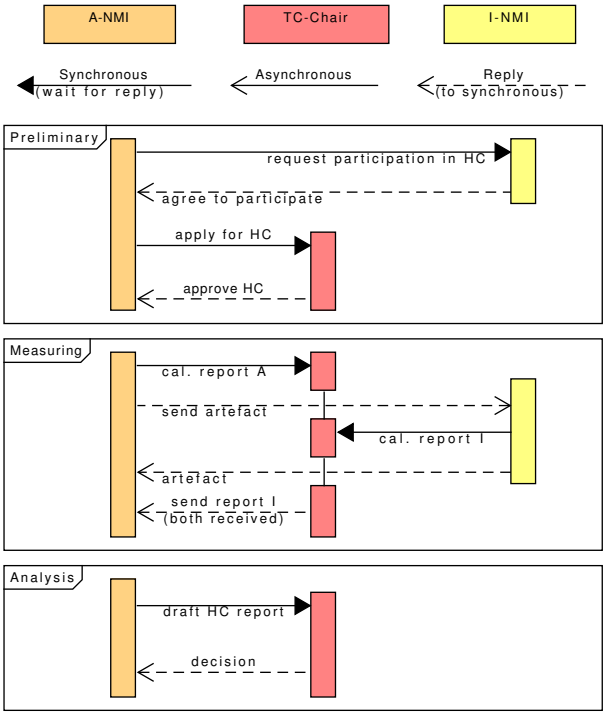


Fig. 2. An architectural view of a HC in the form of a UML sequence diagram. Time progresses from top to bottom. The three main stakeholders are represented by rectangles in vertical lanes. Events are shown as horizontal arrows between stakeholders. HC phases are identified in the top left corner of the corresponding frames.

III THE HC RESULT

The final analysis of a HC must assess the metrological compatibility of results reported by the NMIs. Usually this should involve a calculation of *normalised error*,

$$E_n = \frac{|y_\mathrm{A} - y_\mathrm{I}|}{k_{95}\, u(y_\mathrm{A} - y_\mathrm{I})} \,, \tag{1}$$

where y_A and y_I are the results reported by A-NMI and I-NMI, respectively, $u(y_\mathrm{A} - y_\mathrm{I})$ is the standard uncertainty in their difference, and k_{95} is a *coverage factor* for a 95 % level of confidence that depends on the *degrees of freedom* each NMI has reported for their result [9]. A value of $E_n > 1$ suggests that the uncertainty associated with one or both results has likely been underestimated. However, a precondition of the HC is that the competency of I-NMI is accepted without further scrutiny. So, $E_n > 1$ will be interpreted as a problem with A-NMI's measurement capability.

An architectural description should include a representation of the information required to evaluate the normalised error. This requirement exposes features of common interest in metrology, and indeed in scientific communication more broadly. The features are reflected in the architectural concerns of the NMIs and the TC-Chair throughout the HC.

For example, the property measured in a comparison—the measurand—must be unambiguously identified. This is a concern initially, when the application to carry out a HC is prepared and assessed, as the CMC service category attributed to I-NMI's competency must be validated for a measurand. During the measuring phase, the reported results must be clearly associated with the measurand. In the final phase, results must be extracted from the NMI reports for evaluation of normalised error. At present, however,

there is neither a standard model kind nor a widely accepted convention for expressing specific measurands. Nor are unique digital identifiers for measurands yet available.

Equation (1) requires evaluation of the standard uncertainty in the difference $y_{\mathrm{A}} - y_{\mathrm{I}}$. However, this entails fusion of data from A-NMI and I-NMI, which depends on the interoperability of their respective reporting formats. This is a concern for A-NMI, who drafts the report, and the TC-Chair who assesses the validity of the results. Again, there is neither a standard model kind nor a widely accepted convention for expressing the information required. Indeed, we believe the uncertainty in the difference is not calculated in practice. Instead, the denominator in Equation (1) is replaced with

$$\sqrt{U_{\mathrm{A}}^2 + U_{\mathrm{I}}^2} \tag{2}$$

where U_{A} and U_{I} are *expanded uncertainties* reported by A-NMI and I-NMI, respectively [9]. This calculation is only correct when the same coverage factor, k_{95}, was used to derive the expanded uncertainties, and when the measurements are free from common influence factors (uncorrelated). These conditions must be verified, yet there is currently no standardised form or convention for recording the information necessary to enable such verification.

The reporting of measured physical quantities should conform to the International System of Units (SI) [10, 11]. This is necessary to reconcile unit usage during data fusion. An unambiguous reporting format for measured quantities is the concern of NMIs and the TC-Chair, and may also be a concern of anyone who consults the published HC report. However, there is a problem with compliance to SI norms [12], and in any case, official SI formats cannot be reliably parsed by digital systems [13]. Yet again, we encounter challenges to a clear architectural description of these concepts.

IV DISCUSSION

Even the relatively small and simple HC system requires substantial documentation to adequately capture its functional architecture. It was not our objective to do this. Rather, the aim of this project was to explore the benefits of an architectural analysis of a metrological system in the context of widespread digital transformation. Our conclusion is that such an analysis certainly can be beneficial. This section will outline our reasons for this view.

An architectural description uses a set of concepts to express the architecture of a system. This is an abstract representation that incorporates suitable architectural elements to build useful descriptions of a system, which may be perceived in different ways by different stakeholders. The description is anchored to the needs of those stakeholders, whose concerns must be taken into consideration in the requirements for system design.

To ensure our analysis remained grounded, we found a simple pattern—suggested by Ross and Schoman—to be very helpful [14]. It involves systematically, and often recursively, posing three questions about the system element under consideration: **why** is the element needed and for what purpose; **what** will be done to meet that need; and **how** will the outcome be implemented or designed.

The analysis in this work started at a high level—the purpose of a HC—and became increasingly fine-grained. This ensures that any particular element of the architectural description can be related to a specific need or purpose, typically associated with a stakeholder concern. This is important to enable coordination of widespread digital transformation.

Many of the concerns identified here will likely be addressed by other systems in the wider quality infrastructure environment. A formal mechanism to capture such relationships in an architectural description is known as *element correspondence*. Element correspondences can be used to establish architectural relationships between elements in different parts of a description, or across the descriptions of distinct systems. In this way, the HC system's relationship to another distinct system could be recorded, and those responsible for other systems can be made aware of the HC requirements.

An architectural description can also highlight points where architectural decisions need to be made. For example, some readers may have felt, while reading Section III, that solutions to some of the concerns raised are already available. That may be true, though we might disagree. However, where alternatives solutions exist, it is very useful to capture the reasoning behind the choice of one solution over others. Doing so contributes to a much more resilient system design. Therefore, it is important to document concerns and subsequently capture the specific design decisions made to address them.

To conclude, the top-down architectural description is an insightful and valuable tool for analysis of metrological systems. The ISO 42010 standard is a useful tool for architectural descriptions in the metrological context. Further work is needed to define suitable representations—model kinds—of core metrological concepts, such as measurable quantities, units, measurement modelling, and measurement uncertainty. Nevertheless, the development of architectural descriptions of metrological systems need not be delayed until these core concepts are fully defined, because, as highlighted here, concerns and correspondences can be employed to manage design elements that remain unspecified.

ACKNOWLEDGEMENTS

The authors are grateful to Ryan White, Chair of the Forum-MD Semantics Task Group, for suggesting this topic, and to other members of the Task Group for comments and suggestions. Blair Hall acknowledges the sup-

port of the New Zealand government and thanks Peter Saunders for careful review of the manuscript.

GLOSSARY

This section summarises specific terms used in this paper, with references to their authoritative definitions.

architecture view [5, §3.7]: a way of expressing a portion of the architecture of an entity of interest from a particular viewpoint.

architecture viewpoint [5, §3.8]: rules, methods, and criteria used to create architecture views of an entity of interest relevant to one or more concerns. Includes model kinds, modelling languages, notations, and analytic techniques that can frame a specific set of concerns.

concern [5, §3.7]: a matter of importance or relevance to a stakeholder.

entity of interest [5, §3.12]: The subject of the architectural description—often a system of interest. An Entity of Interest operates within a defined environment. In this paper, the entity of interest manages hybrid comparisons and operates in an environment with various services provided by the international quality infrastructure and regional metrology organisations.

metrological compatibility [15, §2.47]: property of a set of measurement results for a specified measurand, such that the absolute value of the difference of any pair of measured quantity values from two different measurement results is smaller than some chosen multiple of the standard measurement uncertainty of that difference.

model kind [5, §3.15]: a category of model distinguished by a set of conventions defined for a type of view component.

stakeholder [5, §3.17]: an individual, or organisation, with interests in the entity of interest.

view component [5, §3.19]: a view is comprised of view components (models), each in accordance with a model kind.

REFERENCES

[1] General Conference on Weights and Measures, "Resolution 2 – on the global digital transformation and the international system of units," November 2022, 27th CGPM.

[2] CIPM Forum on Metrology and Digitalization, https://www.bipm.org/en/committees/fo/forum-md, (accessed: 4 March 2025).

[3] S. J. Chalk, D. N. Coppa, F. Flamenco, A. Forbes, B. D. Hall, R. J. Hanisch, K. Hosaka, D. Hutzschenreuter, J. Park, and R. M. White, "International development of the SI in FAIR digital data," *Measurement: Sensors*, vol. 18, p. 100293, 2021.

[4] APMP Guideline for using Hybrid Comparison as CMC Evidence, November 2023, https://www.apmpweb.org/portal/list/index/id/15.html, (accessed: 20 March 2025).

[5] ISO, IEC, and IEEE, *Systems and software engineering – Architecture description*, 2nd ed. International Organisation for Standardisation, 2022.

[6] "The BIPM Key Comparison Database," https://www.bipm.org/kcdb/, (accessed: 11 March 2025).

[7] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg *et al*, "The FAIR Guiding Principles for scientific data management and stewardship," *Sci. Data*, vol. 3 160018, 2016. DOI: 10.1038/sdata.2016.18

[8] Object Management Group, "Unified modeling language specification (version 2.5)," 2015, https://www.omg.org/spec/UML/2.5 (accessed 28 August 2023).

[9] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML, *Evaluation of measurement data – Guide to the expression of uncertainty in measurement JCGM 100:2008 (GUM 1995 with minor corrections)*, Paris, Sèvres: BIPM Joint Committee for Guides in Metrology, 2008.

[10] BIPM, *The International System of Units*, 9th ed. BIPM, Sèvres, France, 2019.

[11] A. Thompson and B. N. Taylor, "The NIST Guide for the use of the International System of Units," NIST, Gaithersburg, MD, NIST Special Publication 811, 2008.

[12] R. Hanisch, S. Chalk, R. Coulon, S. Cox, S. Emmerson, F. J. Flamenco Sandoval, A. Forbes, J. Frey, B. Hall, R. Hartshorn *et al.*, "Stop squandering data: make units of measurement machine-readable," *Nature*, vol. 605, no. 7909, pp. 222–224, 2022.

[13] M. P. Foster, "Disambiguating the SI notation would guarantee its correct parsing," *Proc. R. Soc. A: Mathematical, Physical and Engineering Sciences*, vol. 465, no. 2104, pp. 1227–1229, 2009.

[14] D. T. Ross and K. E. Schoman, "Structured analysis for requirements definition," *IEEE Trans. on Softw. Eng.*, no. 1, pp. 6–15, 1977.

[15] BIPM, IEC, IFCC, ILAC, IUPAC, IUPAP, ISO, and OIML, "The international vocabulary of metrology—basic and general concepts and associated terms (VIM)," BIPM, Sèvres, France, 2012.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Bridging Knowledge Gaps: A Requirement Elicitation Use case for Digitalizing Calibration Certificates

Dr. Anjali Sharma[1], Niharika Bhatia[1,2]

[1]*CSIR-National Physical Laboratory, New Delhi, India, anjali@nplindia.res.in*
[2]*Vellore Institute of Technology, Vellore, India, niharikabhatia09@gmail.com*

Digitalization in metrology is not merely substituting manual processes with software but also incorporating the tactical knowledge of experts into digital systems. This use case explores the application of structured knowledge elicitation methods—such as unified modeling language (UML) diagrams, decision trees, and iterative feedback loops—to formally elicit the tacit knowledge of metrologists and integrate it into a digital system. For example, a calibration certificate typically records the measured value of an instrument, its associated uncertainty, and the date of calibration. These models, combined with cooperative sessions and iterative feedback loops, help translate tacit knowledge into well-defined workflows, ensuring that the resulting system maintains technical accuracy and expert insights. It also helps in ensuring accuracy, regulatory compliance, and stakeholder trust in the digital evolution of metrology.

I. INTRODUCTION

The growing digitalization of scientific and technological processes increasingly requires the collaboration of practitioners from different disciplines [1]. Within metrology this means an expanding requirement for software developers and metrologists to work hand-in-hand with each other to transform conventionally manual and expert-oriented procedures into organized, digital processes [2]. But these two groups have fundamentally distinct worldviews, terminologies, and working methods [3]. The absence of mutual understanding and a common conceptual framework frequently leads to communication barriers that make system design and implementation more difficult [4].

This cross-disciplinary challenge is especially apparent when computerizing intricate manual procedures, like the issuance of calibration certificates. Historically, calibration work is informed by formal procedures and tacit knowledge acquired over many years of practical experience [5]. The result of this process—a calibration certificate—needs to record not only measurement outcomes, but also contextual choices, adjustments, and expert judgments. While systems are transitioning to create Digital Calibration Certificates (DCCs), the problem does not only involve digitizing measurement data, but also formalizing the decision-making mechanisms preceding certificate creation [6].

A key problem in such attempts is the elicitation and formalization of tacit knowledge. Much of metrologists' work relies on intuition, situational perception, and adaptable judgment but not on strict rules [7]. These types of competence cannot be easily conveyed in the formal logic necessary for software development. For instance, operations like drift correction rely on the subjective judgment of what an "unusual" reading is—a judgment not formally stated but in practice. Software developers, on the other hand, need well-defined conditions, flows, and data models to encode such decisions into executable code [3].

Solving these problems calls for a conscious bringing together of the views. Methods like Unified Modeling Language (UML) diagrams, process flowcharts, decision trees, and recursive feedback sessions can close the knowledge gap between metrology skills and technical design [8]. These methods facilitate co-modeling of processes so that metrologists can represent their tacit knowledge, and developers can optimize system requirements. With systematic cooperation and inclusive design, one can develop systems that embody both technical expertise and everyday know-how [9].

One valuable initiative addressing this cross-disciplinary knowledge divide is the strategic CABUREK Global Programme 2024-25 under the project Metrology for Digital Transformation in International Cooperation (M4DT-IC), coordinated by PTB [10].

At the heart of CABUREK lies its innovative methodology—**Capacity Building in Technical and Scientific Organizations Using Regional Experience and Knowledge**—which promotes the alignment of quality infrastructure (QI) services with user needs through regional collaboration, stakeholder involvement, and user-oriented innovation. Within this general framework, **Working Group 3 (WG3)** is tasked with the

specific issue of Digital Process Management in which 10 NMIs/DI/QI institutes are tasked with developing digital process flows that promote transparency, accessibility, and responsiveness in customer engagement. The WG3 platform has enabled participants—mostly from metrology institutes but not necessarily from IT backgrounds—to learn from each other's diverse experiences and project environments stakeholders.

This has proven to be essential not only in understanding the functional requirements of digital calibration systems but also in identifying critical non-functional aspects like security, system performance, deployment choices and compliance with national and international policies. Interactive CABUREK Talks were also organized to promote these discussions, encouraging reflection on deployment trade-offs and stakeholder-specific constraints. These peer interactions supplement traditional requirement elicitation techniques by bringing to the surface domain-specific knowledge and cross-cutting issues early in the project life cycle [10].

This paper illustrates a DCC development use case as representative example of cross-domain digital change within metrology. It explains the communication and knowledge elicitation problems faced, methodological tactics employed to cope with them, and the technical lessons learned by the process. The conclusions state the need for ongoing interaction among domain specialists and developers and provide a set of lessons for future efforts [9].

II. USE CASE: DIGITAL CALIBRATION CERTIFICATES AT NPL INDIA

The DCC initiative at CSIR-NPL exemplifies the interdisciplinary challenges involved in mapping metrological competence into ordered digital frameworks. Although DCCs seek to substitute human-readable, paper-based certificates with ordered, machine-readable documents, producing such systems involves both extracting structured data and the experts' implicit reasoning.

A. *Key Challenges in Knowledge Elicitation*

A.1. Tacit Expertise and Implicit Decisions

Metrologists tend to use tacit knowledge—intuition, experience, and situational judgment—whereas system developers need formal, rule-based logic.

For example, as indicated in Figure 1, properties like measure_uncertainty and uncertainty are related to calibration choices that do not necessarily have defined thresholds. The vagueness of terms such as "unusual reading" makes them difficult to represent in structured systems.

A.2. Schema Complexity and Interoperability

The DCC ecosystem has complex stakeholder relationships. Figure 2, the stakeholder diagram, exhibits several interconnected parties that range from calibration fees and services through to BIPM/NPL services and CMCs. There is a need for shared schema that reflects both mandatory requirements as well as laboratory- specific differences to achieve interoperability between NMIs.

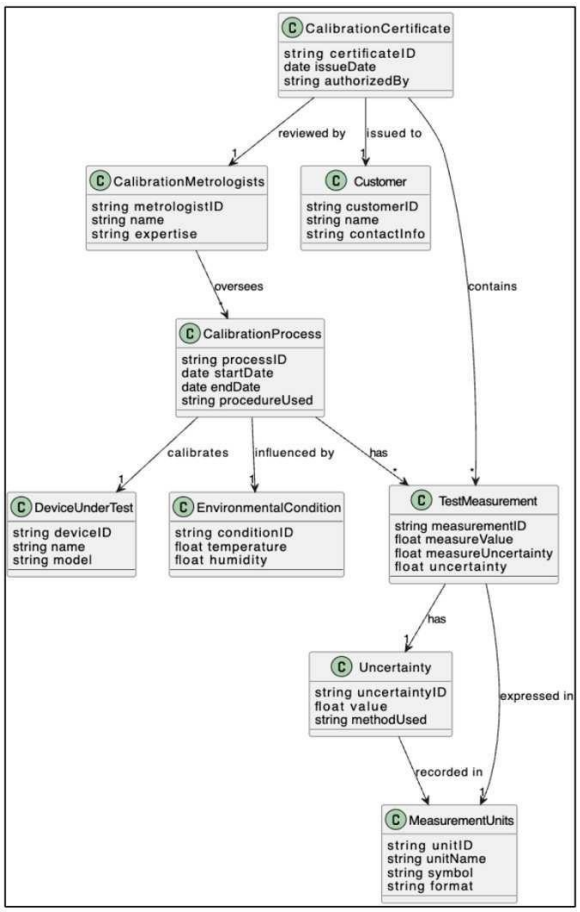


Figure 1 Class Diagram depicting the entities involved in the calibration certificate

A.3. Resistance to Automation

Metrologists are concerned with losing control over decision-making in automatic systems. Since calibration entails high-level judgments (e.g., choosing correction techniques, assessing drift), reliance on system transparency and expert override provisions becomes critical.

A.4. Workflow Complexity

Figure 2 describes a multi-stakeholder calibration process that includes equipment, responsible individuals, customers, services, and environmental information. Representation of these interactions in a modular software

architecture demands precise modeling of relationships, roles, and procedures.

Figure 1 also assists by demonstrating how test measurements, uncertainty, and calibration procedures are logically organized.

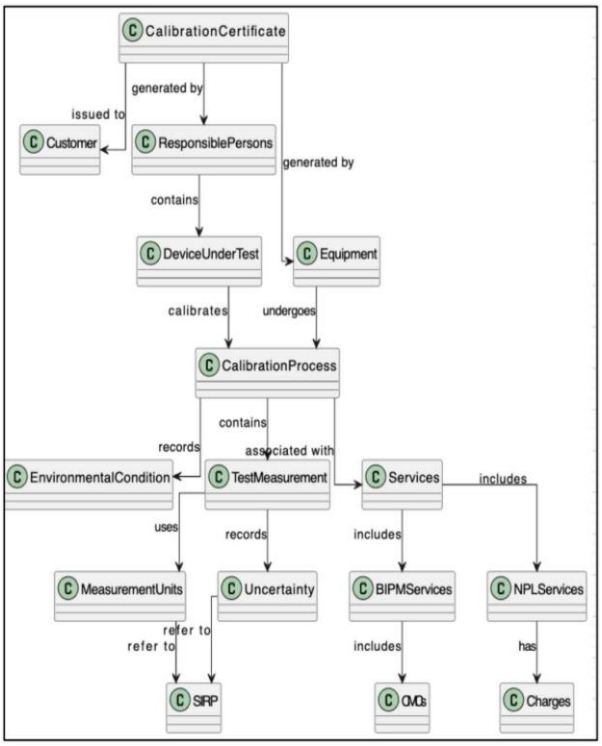


Figure 2 Stakeholder Diagram showing how entities interact in the ecosystem of a digital calibration certificate

B. *Addressing the Challenges: Modeling and Implementation*

B.1. Structured Modeling of Expert Processes

In order to make decision-making explicit, Unified Modeling Language (UML) diagrams and decision trees are utilized. Figure 1, the class diagram, gives us a template to lay out relationships like CalibrationProcess → TestMeasurement → Uncertainty. Whereas Figure 2 represents the larger stakeholder interactions. Recursive co-modeling sessions enable metrologists to check and tailor representations to align with real-world practices.

B.2. Schema Standardization Using ISO-Compliant Formats

Based on the object model in Figure 1, A unified XML/JSON schema is created which aligns with ISO 17025 requirements. CertificateID, processID, and uncertaintyID attributes were mapped for traceability and platform interoperability.

B.3. Hybrid Reasoning and User-Centric Adoption

To maintain the autonomy of experts, a hybrid reasoning system is used. Rule-based elements performed deterministic logic (e.g., certificate formatting), while AI elements supported anomaly detection (e.g., drift outside normal limits), always needing user consent before execution. Training sessions and phased adoption plans also aid in minimizing resistance by engaging metrologists in hands-on feedback loops.

B.4. Modular Architecture Based on Model-View-Controller Approach (MVC)

With reference to Figure 2, the DCC system is implemented using the **Model-View-Controller (MVC)** architecture—an architectural pattern that separates concerns into three interconnected components: the *Model* (e.g., DeviceUnderTest, EnvironmentalCondition) for handling data and logic, the *View* for user interfaces, and the *Controller* for decision logic and certificate generation. This modular implementation enhances flexibility, simplifies maintenance, and facilitates scalability as the system evolves [11].

B.5. Cross-Disciplinary Knowledge Elicitation through CABUREK

The CABUREK M4DT-IC WG3 offers a collaborative environment where participants, predominantly from metrology institutes, share knowledge and learn from various project experiences. The project enables participants to comprehend not just the functional elements of digital calibration systems but also important non-functional elements, including security, performance, deployment options, and standards compliance.

Through CABUREK M4DT-IC, discussion sessions were arranged to address these issues, facilitating consideration of deployment trade-offs and stakeholder-specific constraints. Peer discussion reinforces standard requirement elicitation methods to guarantee that domain-specific knowledge and cross-cutting concerns are resolved early in the project life cycle.

B.6. Steps Ahead

Numerous initiatives are planned to ensure the continuous development and resilience of the DCC system. Schema compatibility can be confirmed using interoperability testing, and it will facilitate smooth platform-to-platform certificate exchange. Data contextualization can be improved through the integration of semantic web technologies and standardized ontologies. This will allow for better semantic interoperability and intelligent querying. To guarantee the coherence of rule-based and AI-assisted reasoning components, formal verification techniques like model checking can be explored.To reduce the audit burden, compliance automation tools can be created to expedite adherence to ISO/IEC 17025 standards. In conclusion, CABUREK-led partnerships can be extended through workshops, shared repositories, and co- development environments [13]. This can promote inter- institutional knowledge exchange and assist the long-term viability of digital metrology systems.

III. IMPACTS AND LESSONS LEARNED

The development the DCC system at CSIR-NPL is a step towards a technological revolution and a challenge to knowledge management. As revealed in Figures 1 and 2, the shift required reimagining calibration not only as a technical process but also as a structured, interoperable data ecosystem. The main impacts and synthesized lessons learnt from this use case are highlighted below.

A. *Practical Impacts*

Structured knowledge representation and traceability by employing object-oriented models facilitated tracing relationships among calibration actors and objects like devices, environmental conditions, uncertainties, and measurement units [4]. Organized representation made traceability of calibration choices more effective, which is crucial for internal audits as well as for international comparisons [6].

Interoperability between metrology systems to develop a schema compatible with ISO/IEC 17025 and compliant with international initiatives like PTB's DCC framework provided semantic interoperability [13]. This provides avenues for cross-border acceptability and automated validation of certificates [5].

Transparency and reproducibility of calibration choices by making formal expert decisions decision trees and logical models, the system enhanced more transparency regarding uncertainty and measurement corrections derivation. This is particularly vital in fields where reproducibility underpins scientific and legal metrology [8].

Modular and maintainable design to keep the concerns separate through MVC-based design made the logic, interface, and data layers cleanly separated. This modularity makes future modifications—like updating the calibration process or adding new types of services—much more viable.

Empowerment through digital skill-building piloting the system with active metrologist involvement builds trust and empowered domain experts to participate in system design. It turns the digital transformation into a co- created experience rather than a top-down imposition [10].

B. *Generalized Lessons for Digital Transformation in Metrology*

Iterative knowledge elicitation is critical, and tacit knowledge cannot be drawn out in one workshop [5]. Rather, iterative cycles with UML models and decision trees—checked by metrologists—guarantee that domain understanding will not be lost in translation. This is in line with knowledge engineering best practice [7].

As observed in Figure 2, calibration data engages with varied services (e.g., BIPM, SIRP, NPL charges). The DCC schema, therefore, needs to be modular and extensible [2]. Utilizing linked data principles and metadata tagging guarantees durability and reusability of digital records [9].

Digitalization must never replace human judgment, particularly in edge cases or uncertain situations. Permitting expert validation at decision points—like uncertainty estimation or environmental anomaly detection—is important for trust [3].

Conformity of the DCC architecture to globally accepted standards such as ISO 17025, brought inbuilt adaptability to the system across NMIs and regional quality infrastructures [2].

Even a technically effective system may break if users do not adopt it. Pilot deployment, incremental training, and incorporation of user feedback at every phase broke resistance and promoted acceleration of adoption—a lesson with wide applicability to digital public infrastructure initiatives [12].

The CABUREK project was instrumental in bridging the knowledge gap between software developers and metrologists. Through interactions among the professionals from different metrology institutes, CABUREK allowed the participants to gain a better understanding of functional and non-functional requirements at the beginning of the project. In this collaborative setting, issues such as security, performance, and compliance with national and international standards were found to be valuable.

Also, CABUREK Talks created a formal forum where these challenges could be addressed, allowing participants to consider in-depth real-world deployment trade-offs. These discussions not only enriched the requirements elicitation but also verified that the resulting system would fulfill the specific needs of each metrology institute while satisfying overarching digital transformation objectives [10].

IV. CONCLUSION

The digital revolution of metrology, with its potential to enhance precision, traceability, and interoperability, brings challenges and cross-disciplinary issues. This research addressed those challenges in the context of the DCC project at CSIR-NPL, in order to make software developers work together with experts in metrology to develop a structured, standard, and forward-looking system.

At the heart of the transformation was the need to bridge the communication divide across two disciplines with inherently different knowledge paradigms—one based on experiential, tacit knowledge and the other on explicit, rule- based representation. Through systematic knowledge elicitation methods like UML diagrams, decision trees, and continuous verification, we can re-express subtle metrological processes as formal models executable in a digital system. This approach enabled common ground and co-creation, reducing requirements ambiguity and enhancing the fidelity of digital models.

The use case illustrates how object-oriented modeling and modular architecture (e.g., MVC) facilitate the system to capture calibration workflow complexity, ranging from uncertainty estimation and environmental conditions to user interaction and secure certificate generation. Most importantly, by using joint modeling and validation

process, metrologists had more confidence in the system, whereas developers obtained domain knowledge that guided improved design decisions.

Application of internationally harmonized standards (e.g., ISO/IEC 17025, JSON/XML schema) guarantee that the resulting DCC system would be compatible with worldwide metrology projects and regulatory schemes [13].

Overall, this research illustrates that successful digital transformation of metrology is more than a technological undertaking but a socio-technical process encompassing repeated knowledge elicitation, co-modeling, and establishment of trust. The integration of digital technologies into metrology requires a systematic approach that balances technical innovation with the preservation of expert knowledge [14].

Through the capture of expert knowledge in an organized and interoperable way, the DCC project paves the way for more intelligent, more transparent, and harmonized metrological services. The methodology and results outlined here are part of the expanding body of global effort toward digital calibration, and provide a guide to follow for subsequent projects that wish to traverse the tangled intersection of measurement science and digital engineering.

V. ACKNOWLEDGEMENT

This work was made possible with the support of the CABUREK M4DT-IC project coordinated by PTB. The insights and collaborative efforts provided by coaches and participants from various metrology institutes have significantly contributed to bridging the knowledge gap between metrology and software development. Special thanks to the WG3 coaches, participants and speakers of various CABUREK Talks and workshops, which created valuable forums for discussion and knowledge-sharing, fostering cross-disciplinary collaboration.

VI. REFERENCES

[1] S. Eichstädt, A. Keidel, and J. Tesch, “Metrology for the digital age,” Measurement: Sensors, vol. 18, p. 100232, 2021. Available: https://www.sciencedirect.com/science/article/pii/S2665917421001951 [Accessed: 6 Apr. 2025].

[2] P. Brézillon and J. C. Pomerol, “Dynamics Between Contextual Knowledge and Proceduralized Context,” Springer, 1999. Available: https://link.springer.com/chapter/10.1007/3-540-48315-2_22 [Accessed: 18 Mar. 2025].

[3] I. Kim, R. S. Peak, and S. K. Sitaraman, “ROM: A reliability knowledge representation for collaborative system design,” Engineering with Computers, vol. 26, pp. 11–33, 2010. https://doi.org/10.1007/s00366-009-0135-4

[4] B. D. Hall, “An opportunity to enhance the value of metrological traceability in digital systems.” Available: https://ieeexplore.ieee.org/document/8792841 [Accessed: 23 Apr. 2025].

[5] JCGM 200:2012, International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM), 3rd ed., BIPM, 2012. Available: https://www.bipm.org/documents/20126/2071204/JCGM_200_2012.pdf [Accessed: 2 Apr. 2025].

[6] D. Norman, The Design of Everyday Things, Rev. and Exp. ed., MIT Press, 2013. Available: https://ia902800.us.archive.org/3/items/thedesignofeverydaythingsbydonnorman/The%20Design%20of%20Everyday%20Things%20by%20Don%20Norman.pdf [Accessed: 21 Mar. 2025].

[7] I. Nonaka and H. Takeuchi, The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation, Oxford University Press, 1995. Available: https://lumsa.it/sites/default/files/UTENTI/u95/LM51_ITA_The%20Knowledge-Creating%20Company.pdf [Accessed: 27 Apr. 2025].

[8] W. S. Melo, A. A. A. Rocha, and M. S. Silva, “Smart Manufacturing and Digitalization of Metrology: A Systematic Literature Review and a Research Agenda,” Sensors, vol. 22, no. 16, p. 6114, 2022. Available: https://doi.org/10.3390/s22166114

[9] M. Uschold and M. Gruninger, “Ontologies: Principles, methods, and applications,” The Knowledge Engineering Review, vol. 11, no. 2, pp. 93–136, 1996. Available: https://www.researchgate.net/publication/302937543_Ontologies_Principles_methods_and_applications [Accessed: 12 Apr. 2025].

[10] PTB, “CABUREK M4DT-IC – Metrology for Digital Transformation in International Cooperation.” Available: https://www.m4dt.ptb.de [Accessed: 9 Apr. 2025].

[11] Mozilla Developer Network (MDN), “MVC (Model-View-Controller),” Available: https://developer.mozilla.org/en-US/docs/Glossary/MVC [Accessed: 10 Mar. 2025].

[12] S. Lim, A. Henriksson, and J. Zdravkovic, “Data-driven requirements elicitation: A systematic literature review,” SN Computer Science, vol. 2, no. 1, p. 16, Jan. 2021. Available: https://doi.org/10.1007/s42979-020-00416-4

[13] International Organization for Standardization (ISO), “ISO/IEC 17025:2017 General requirements for the competence of testing and calibration laboratories Available: https://www.iso.org/standard/66912.html [Accessed: 10 Jun. 2025].

[14] N. Varshney, K. Garg, T. S. Nagla, S. K. Jaiswal, S. Yadav, and D. K. Aswal, “Challenges in Sensors Technology for Industry 4.0 for Futuristic Metrological Applications,” MAPAN, vol. 36, pp. 215–226, May 2021. Available: https://doi.org/10.1007/s12647-021-00453-1

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Communities of practice in metrology

Blair Hall (ORCID: 0000-0002-4249-6863)[1]

[1]*Measurement Standards Laboratory, Lower Hutt, New Zealand, blair.hall@measurement.govt.nz*

***Abstract* – Digital transformation in metrology will entail digitalisation of work practices in national metrology institutes (NMIs). This article argues that *communities of practice* (CoPs)—informal social structures—play a key role in how NMIs operate, making their dynamics a critical consideration in managing digitalisation. NMI work is empirical in nature, and the knowledge required to perform measurements at the highest level is embedded in practice-based forms within CoPs. However, digitalisation necessitates the development of abstract representations of processes and information. Moreover, scientific knowledge is usually communicated in a formal language using abstract concepts. So, digitalisation of NMI work products may be challenged by the distinction between empirical and abstract modes of knowledge capture. This article explores the conceptual tension between abstract and empirical modes of thought in metrology and the challenge of aligning them in the context of digital transformation. It draws on published anecdotal accounts and the author's professional experience.**

I INTRODUCTION

National metrology institutes (NMIs) are centres of excellence in measurement that operate to the highest standards in their respective economies. Their activities span a range of disciplines, naturally organising into broad categories that subdivide into specialised areas—often mirroring the structure of the International Committee for Weights and Measures (CIPM) Consultative Committees. For instance, an NMI's Electricity and Magnetism section might include groups of people responsible for different DC and AC quantities.

Within NMIs, the practice of metrology has distinctive characteristics compared to other scientific organisations. Metrologists typically work in a specialised area and develop a high level of expertise. They often remain in the same field for extended periods—sometimes for their entire careers. Depending on the size of the institute, they may be somewhat isolated from colleagues in other branches. Moreover, opportunities for metrologists to engage with technical peers in counterpart organisations are often limited, as such interactions typically require international travel.

People typically enter an NMI with no prior experience in metrology. Most have undertaken advanced academic studies—often to doctoral level—in a related scientific field, usually with a focus on experimental research. However, the foundational *concepts* of metrology are not generally taught in universities. As a result, new staff acquire metrological knowledge primarily through interpersonal interaction, with experienced metrologists guiding them in understanding the specific aspects of their field. Importantly, this learning process is embedded in the shared experience of metrology in practice, creating a highly pragmatic context for knowledge acquisition. This contrasts with modes of learning supported more directly by a well-defined theoretical curriculum.

This article considers the particularities of metrology groups and the people in them through the lens of certain sociological factors. Addressing social aspects in a field grounded in physical science is relevant now because of the digital transformation recently initiated by the international metrology community. Digitalisation, the driver of digital transformation, is inherently a process of abstraction. It involves the systematic development of logical representations for concepts and processes. Since digital technology operates on strictly logical systems, digitalisation in metrology requires the elicitation and development of abstract models. However, when confronted with informal, practice-based knowledge held within metrological communities, digitalisation faces a significant challenge.

The distinction between abstract and empirical perspectives among scientists is well recognised. However, the tendency toward abstraction is reinforced by the formal language of science, which underpins both the teaching and communication of scientific concepts. Abstract models are powerful representations of physical phenomena; however, their validity ultimately rests on empirical evidence. Of particular interest here is the empirical nature of the work carried out by NMIs, and the practice-based knowledge held by metrolgists, which is vital in maintaining the international measurement system. The work of metrologists tends to be grounded in specialised instrumentation, measuring processes, and raw data processing. Unsurprisingly, local, experience-based interpretations of terminology and concepts can develop within small insular groups. These modes of knowledge representation may be difficult to express outside the context in which they developed.

The close-knit teams working together in NMIs exhibit characteristics of Wenger's concept of a *Community of Practice* (CoP) [1], which builds on earlier work with Lave exploring how knowledge is generated and sustained

through social interaction in shared practice [2]. In a CoP, members engage in a joint enterprise, developing a shared repertoire of tools, language, and routines, as well as a strong collective sense of identity. Active participation in the community is fundamental. Different roles and trajectories of participation can be observed: newcomers typically begin on the periphery of a group and gradually move toward full membership. A distinctive feature of CoPs is that knowledge is often held tacitly, which complicates the communication of ideas in interdisciplinary collaboration.

The remainder of this article explores the nature of NMI groups in relation to key concepts of a CoP and the effect of abstract and empirical modes of knowledge representation. The central question is how to facilitate communication between metrology groups and digital technology experts to support effective digitalisation. The article is an informal account intended to share ideas with managers and leaders guiding the international metrology community toward a successful and fruitful digital transformation. The discussion draws on anecdotal evidence and on the author's experiences working with groups involved in metrology's digital transformation.

II METROLOGICAL COMMUNITIES OF PRACTICE

In a Community of Practice, a group of individuals engage in a common domain of activity and develop practices, terminology, and a sense of identity through sustained interaction [1]. CoPs are likely to be widespread within metrology. These informal groups emerge spontaneously and are sustained through collaboration related to a specific activity, fostering the creation of knowledge within a community. CoPs are not created by formal structures; however, they may be supported within an organisation's framework.

One of the central ideas of a CoP is that knowledge develops as individuals participate in shared activities. Competency is central to this; meaningful knowledge cannot be asserted without participation, making competency and knowledge inseparable. This emphasis aligns closely with NMI work practices, where tasks must usually conform to international standards, such as ISO/IEC 17025, which requires technical competence to be demonstrated [3]. The competency requirement is fundamental to ensuring the reliability of measurements.

Individuals participate in CoPs at different levels and in different roles that evolve as they gain experience and expertise. Newcomers start on the periphery of a CoP, gradually moving inward as they absorb practices and terminology. Individuals contribute to the co-creation of knowledge, language, and understanding within the community as they engage in shared practices. The idea of *negotiated meaning* is central in CoPs, as knowledge is expressed as part of a dynamic and adaptive process. The negotiation of meaning within a CoP will tend to blend new ideas with existing practice and terminology. So, a newcomer may bring new and useful ideas that contribute to shared activities, but as the meaning of this new knowledge is negotiated, it will be shaped by existing practice and terminology.

The theory of CoPs identifies a duality of *reification* and *participation*, viewing them as complementary features. Reification makes knowledge tangible by codifying it in documents, databases, protocols, and other artefacts to crystallise shared practices and understanding. Participation, on the other hand, provides continuity of meaning over time and across contexts, compensating for the limitations inherent in reification. For example, teachers and students can negotiate meaning from a reified subject curriculum.

The representation of information and knowledge in reified artefacts is an important consideration, because such artefacts are likely to be identified as objects for digitalisation. Reification occurs both inside and outside metrological CoPs. Internally reified artefacts—such as laboratory notebooks, calibration procedures, and data acquisition software—are primarily intended for use within the CoP, although they may also be scrutinised by external parties as part of quality assurance processes. In contrast, work products, such as calibration reports, are primarily for external use. Some reified artefacts originate outside NMIs but are intended to inform internal practice. Examples include documents intended to harmonise terminology (VIM) [4] and provide guidance for best practice in the evaluation and reporting of measurement uncertainty (GUM) [5].

A strong sense of collective identity develops within a CoP, and is likely a feature of metrological communities due to their stability, long tenures, and specialisation. Stability preserves expertise and ensures continuity in measurement science, but the conceptual foundations of practices and terminology held by enduring CoPs may be conservative. Senior members, having invested heavily in their own understanding of terminology and practices, may be unreceptive to alternative perspectives or explanations. When this occurs, communication across the CoP boundary can become challenging.

CoPs may exploit the role of 'brokers' to address communication challenges. Brokers, who have peripheral access to several CoPs, can facilitate connections between them, enabling coordination and even opening up new possibilities for meaning. Brokership arises from an individual's multiple memberships in CoPs and depends on the individual's status and levels of influence within those communities. A common difficulty in establishing shared terminology is often cited by people with a background in information technology (IT) who have had difficulty engaging with metrology teams. In such contexts, the broker role may be valuable. A broker can facilitate the establishment of a terminological common ground, whereas an

outsider, lacking insight into or understanding of a CoP, may struggle to engage effectively.

To sum up, CoPs are likely to be common in NMIs, making an understanding of their general characteristics essential for facilitating interdisciplinary activities related to digitalisation. Metrological CoPs will hold a wealth of detailed scientific and technical knowledge. However, this knowledge may be held in ways that make access to it challenging. A significant amount of information may remain unreified or be reified in *ad hoc* forms—such as process-related knowledge that underpins specialised technical procedures.

III ABSTRACTION AND EMPIRICISM

Scientific knowledge can be thought of in both abstract terms—representing general concepts—and in empirical terms—grounded in observable phenomena and the methods by which they can be observed. Maxwell's unit systems, which are the foundation of the modern SI and central to the work of NMIs, illustrate this distinction well [6]. Maxwell provided an empirically based description of how unit systems facilitate unit conversion. In his approach, quantities are expressed as a product of factors: a numeric value and a symbol referring to a unit. Maxwell explained that the unit symbol can be thought of as a numeric variable, which is assigned a value according to the desired unit. For example, in the expression $19\,\mathrm{m}$, 19 represents the value of a length measured in metres. To convert this to feet, one assigns $\mathrm{m} = 3.28084$, the measure of one metre in feet, and evaluates the product

$$19 \times \mathrm{m} = 19 \times 3.28084 = 62.33596\ .$$

This explanation is grounded in empirical ideas, relying on the existence of reference physical quantities (units) and the measurement of relative amounts of quantities.

Nowadays, an abstract description of unit and quantity systems is more familiar. This treats the idea of *quantity* as fundamental: *quantity* is an abstract notion that is simply assumed [7]. Units are individual quantities that can be compared by ratio with other quantities of the same kind to obtain measures (numbers). Algebraic rules of *quantity calculus* can be applied to abstract quantities, enabling *quantity equations* to express relationships between quantities independently of any specific choice of units. These equations can be manipulated and analysed algebraically, providing a foundation for modelling physical systems. This view is grounded in abstract ideas, relying on a logical framework of mathematical concepts.

The distinction between the abstract and empirical concepts of quantity is important. It would be a mistake to assume that abstract ideas have simply replaced empiricism. Indeed, attempts to form a complete axiomatic description of unit systems based on abstract quantities have so far been unsuccessful. So the empirical explanation given by Maxwell is still arguably more complete. De Courtenay explains how the meticulous quantification of measurement error (uncertainty) in metrology reconciles the inevitable discord between empirical data and abstract theoretical models [8]. Metrology spans the divide between these two domains. It is through measurement that we obtain information about the actual physical world, so metrology is inherently tied to an empirical world view. However, modern scientific language is formal, and so metrology must adapt to modes of expression that are universally understood. This necessity to align practice with common forms of expression is the one of the motivations for guidance documents like the GUM and the VIM [5, 4].

A sense of the difficulty in adapting empirical experience to abstract concepts is given in a remarkable account of Maxwell's researcher assistants' use of moving-coil instrumentation. In a private communication to Maxwell, reproduced by Schaffer, we read [9, page 36]:

> "... we are now going ahead with the comparison of coils, and several men are able to rig up the various combinations out of their own heads, now that they know by inspection the difference between a
>
> $$\begin{aligned}&\text{shove} \propto \text{current}\\ &\text{kick} \propto \int \text{current}\, dt\\ &\text{jerk} \propto \int\int \text{current}\, dt\, dt\end{aligned}$$
>
> I have just got a clear case of double jerk from zero to + and back to zero and stop there".

The empirical representation here is striking, yet Schaffer notes that these men were learning to see effects in the laboratory as "instances of more general principles of changing current and field strength"—in other words, in abstract terms. He also says, in a remark particularly relevant to this study, that "The problem the Cavendish faced was defining and distributing this fragile notion of practice."

The role of measurement in linking abstract theories and meaningful empirical observations received further attention in the mid-20th century. Stevens coined the 'schemapiric principle' to refer to the duality where empirical procedures generate data that is interpreted through the lens of abstract theoretical concepts [10]. In a handbook for experimental scientists, he emphasised the importance of conveying meaning through careful use of syntax and semantics when reporting observations [11]. These issues remain pertinent today, as digitalisation progresses.

Stevens also discussed something he called *pragmatics*, referring to how scientists respond to the use of formal symbolic expressions. He may have sought to highlight a human element in the abstract-empirical tussle, writing "... the paralytic terror instilled by those clever souls who can calculate in those who cannot is a problem in applied pragmatics" [11, page 4]. Indeed, some metrologists express

reluctance toward mathematical formalism, while theorists may be prone to dismissing experimental complexities as secondary. This tendency suggests a bias toward theoretical representations over experimental practice. However, as already emphasised, the empirical knowledge embedded within NMIs is of fundamental importance. The fact that it does not always lend itself to abstract expression must not be mistaken for a lack of value.

In summary, metrology occupies a unique position between empirical and abstract representations of knowledge. Metrologists themselves may not be aware of this distinction. Similarly, experts in digital technology—whose discipline is inherently focused on abstraction—are unlikely to recognise these differences. This lack of awareness could present challenges for communication.

IV TOWARDS SUCCESSFUL DIGITALISATION

Sections II and III outline challenges to digitalisation that stem from the social structure of metrological groups. Communities of practice, staffed with dedicated NMI scientists, will likely find themselves at odds with abstract thinkers intent on codifying their practices. Digitalisation, approached in these terms, risks becoming what might euphemistically be called 'disruptive'—and unnecessarily arduous for all involved. Strong leadership from within metrology communities is essential. The approach to digitalisation must acknowledge the unique challenges faced by metrology and avoid viewing digitalisation as merely a case of 'change management'.

A Terminology

To achieve the goals of digital transformation in science, standardisation of terminology is essential. Broad guidelines are available to produce 'FAIR' data (Findable, Accessible, Interoperable, and Reusable) [12, 13]. However, metrological CoPs may be operating with local empirically-based interpretations. So, strategies are needed to deal with this.

Issues related to codifying processes and knowledge are not unique to metrology; they arise in other sectors as well. So, the challenges to metrology are to some extent already familiar to IT professionals. In business settings, the role of *business analyst* is typically tasked with gathering, documenting and formalising detailed requirements for digitalisation projects. The analyst acts to bridge the gap between business needs, which are likely practice-based, and digital designs, which are abstractions. However, the challenge in metrology is of an entirely different scale. Harmonisation on a global level must ensure that scientific information can be shared and reused as widely as possible.

To address this, the metrology community should consider a role analogous to the business analyst, one that would facilitate the harmonisation of terminology and concepts embedded within CoPs. The activities of people in this *broker* role (as described in section II) could negotiate changes to align with conventional abstract forms (our laboratory is exploring this by offering tuition in mathematical modelling to early-career metrologists—see below).

The metrology community should consider strategies for managing terminology. Neumann observed a large and specialised terminology in use—over 20,000 terms in just one large NMI [14]. She proposed ways that working vocabularies could be reduced in size by structuring them inside well-defined contexts. To support this, she introduced the concept of a dynamic controlled vocabulary, which actively manages terminology while preserving nuances of meaning across different contexts.

The challenge of managing terminology is not unique to metrology; it is a common issue in large IT development projects and the IT industry has also developed strategies to address it. For instance, the *Domain-Driven Design* process formalises domain-specific languages in projects [15]. This establishes working vocabularies that both domain experts and technical team members agree on. This domain-specific language is adopted throughout the project—in code as well as documentation—improving communication and creating more meaningful and accurate work products. The idea sounds quite similar to Neumann's dynamic vocabularies. Are these ideas that could be exploited more widely in metrology?

B Modelling

Digitalisation in metrology requires the development of suitable models. Modelling is a process of abstracting and representing knowledge in a way that can be codified as tangible artefacts. Whether in the form of written procedures, data structures, mathematical formulations, or software, reified artefacts are based on an underlying model representing relevant aspects of the domain.

However, CoPs may reify models locally that diverge from more widely understood formal representations. A striking example of this is described in the report of an international measurement comparison of microwave quantities [16]. The report states, on page 13, "We have the impression that many participants are not willing to completely disclose their method of uncertainty determination, which might be due to proprietary calibration techniques that are applied." This situation led the pilot laboratory to abandon use of uncertainty data in the comparison analysis, as explained on page 15, "... it is questionable if the uncertainties would indeed represent the measurement capabilities of the laboratories and in this case it is better to abandon the uncertainties in the calculation of the [comparison reference value]."

The microwave metrology community has taken steps to address this issue. First, guidelines have been developed that describe the most common mathematical measurement models applicable to their tasks and explain how

to evaluate the corresponding uncertainties [17]. Second, software facilitating the use of these models in the laboratory is available [18], which enables metrologists to adopt standard models with little change to their laboratory practice. Could this type of response to the diversification of CoP artefacts be useful elsewhere?

The Measurement Standards Laboratory recently offered staff tutorials on the formulation of mathematical measurement models and the use of software for model evaluation. The initiative was well received, prompting the development of an e-Learning module to make these ideas more widely accessible. Tuition was deliberately offered to metrologists recruited within the last ten years, as their roles are still taking shape within CoPs. These metrologists are open to new ideas and can negotiate a place for those ideas in their groups.

V FINAL COMMENTS

The scientific knowledge held in NMIs is likely retained in a multitude of metrological CoPs, much of it embedded in empirical, practice-based forms. In contrast, external scientific communication and digitalisation efforts depend on abstract representations, which may not align with reifications developed by CoPs. As robust social structures with strong group identities, CoPs require careful consideration in any digital transformation. Abrupt change is neither realistic nor desirable, as it risks disrupting valuable knowledge and established practice. Effective transformation must actively involve CoP members and evolve through the continued negotiation of meaning within shared practice.

This article suggests several ways to influence CoPs, all of which involve members acting in peripheral roles—either as relative newcomers or as brokers who participate in multiple communities. A common theme in these approaches is that change is most effectively driven from within, by CoP members themselves.

ACKNOWLEDGEMENT

This work was funded by the New Zealand government. The author thanks Ellie Molloy for careful review of the manuscript.

REFERENCES

[1] E. Wenger, *Communities of practice: Learning, meaning, and identity*, Cambridge Univ. Press, 1998.

[2] J. Lave and E. Wenger, *Situated learning: Legitimate peripheral participation*, Cambridge Univ. Press, 1991.

[3] ISO and IEC, *ISO/IEC 17025:2017 General requirements for the competence of testing and calibration laboratories*, 3rd ed., International Organization for Standardization, 2017.

[4] BIPM, IEC, IFCC, ILAC, IUPAC, IUPAP, ISO, and OIML, "The international vocabulary of metrology: basic and general concepts and associated terms (VIM)," BIPM, Sèvres, France, 2012.

[5] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML, *Evaluation of measurement data – Guide to the expression of uncertainty in measurement JCGM 100:2008*, Paris, Sèvres: BIPM JCGM, 2008.

[6] J. C. Maxwell, *A Treatise on Electricity and Magnetism*, Clarendon Press, Oxford, 1873.

[7] J. de Boer, "On the history of Quantity Calculus and the International System," *Metrologia*, vol. 31, no. 6, pp. 405–429, 1995.

[8] N. de Courtenay, "The double interpretation of the equations of physics and the quest for common meanings," in *Standardization in measurement: philosophical, historical and sociological issues*, Routledge, pp. 53 – 68, 2015.

[9] S. Schaffer, "Late Victorian metrology and its instrumentation: A manufactory of Ohms," in *Science as Practice and Culture*, Univ. Chicago Press, pp. 23–56, 1992.

[10] S. S. Stevens, "Measurement, statistics, and the schemapiric view," *Science*, vol. 161, no. 3844, pp. 849–856, 1968.

[11] ——, "Mathematics, measurement, and psychophysics." in *Handbook of Experimental Psychology*, Wiley, pp. 1–49, 1951.

[12] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, *et al*, "The FAIR Guiding Principles for scientific data management and stewardship," *Sci. Data*, vol. 3 160018, 2016. DOI: 10.1038/sdata.2016.18

[13] S. Hodson, "FAIR Data Action Plan: Interim recommendations and actions from the European Commission Expert Group on FAIR data," 2018. (archived: https://zenodo.org/record/1285290)

[14] J. Neumann, "Cross-domain dynamic vocabulary in metrological use cases: Linkage, automatization and implementation," *Meas. Sens.*, 101493, 2024.

[15] E. Evans, *Domain-driven design: tackling complexity in the heart of software*, Addison-Wesley Professional, 2004.

[16] J. Ruefenacht and M. Zeier, "EUROMET.EM.RF-S16 Final Report: Comparison of scattering parameter measurements in the coaxial 2.4 mm line system," *Metrologia*, vol. 42, no. 1A, 01001, 2005.

[17] M. Zeier, D. Allal, and R. Judaschke, *Guidelines on the evaluation of vector network analysers (VNA)*, 3rd ed., EURAMET Calibration Guide cg-12, 2018.

[18] M. Wollensack, J. Hoffmann, J. Ruefenacht, and M. Zeier, "VNA Tools II: S-parameter uncertainty calculation," in *79th ARFTG Microw. Meas. Conf.*. IEEE, pp. 1–5, 2012.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Provenance information in metrological traceability: application and modeling

Ryan White[1], Julia Neumann[2], Jean-Laurent Hippolyte[3], Blair Hall[4], Thiago Menegotto[5]

[1] *National Research Council Canada, Metrology Research Centre 1200 Montreal Road Bldg. M-36 Canada, ryan.white@nrc-cnrc.gc.ca, ORCID: 0000-0003-3589-5900*
[2] *Physikalisch-Technische Bundesanstalt, Abbestraße 2-12 Berlin Germany*
[3] *National Physical Laboratory, Hampton Rd, Teddington TW11 0LW UK*
[4] *Measurement Standards Laboratory of New Zealand, Lower Hutt, New Zealand*
[5] *Instituto Nacional de Metrologia, Qualidade e Tecnologia, Av. Berlim, 627, Porto Alegre, Brasil*

***Abstract* – This paper proposes to use provenance information to describe processes in metrology. The PROV data model is used as an example to showcase a conceptual analysis about how to improve quality, reliability, and overall interoperability within cross-domain applications that are based on metrological timeline data. The conceptual analysis will be used as a foundation for further contributions to the topic of improving metrological traceability with provenance data models.**

I. INTRODUCTION

In information technology, provenance describes the history of a piece of data, including its origin and modifications over time [1]. It is fundamental to establishing the quality and reliability of data, especially as data interoperability becomes essential for cross-domain applications. In metrology, provenance metadata documenting measurement assurance and calibration activities could underpin metrological traceability claims, enabling machine-actionable assessments of measurement data quality. This paper proposes a provenance-based approach to describing measurement assurance processes in metrology, aligning with the FAIR principles (Findable, Accessible, Interoperable, and Reusable) that are driving global digital transformation in measurement science [2].
We demonstrate the application of the W3C PROV data model to capture key events and entities in metrological processes, such as calibration and validation of measurement standards across different stages within a traceability chain [3]. Integrating provenance into the ecosystem of metrology digital documents, including Digital Calibration Certificates (DCCs), ensures compliance with specific traceability requirements and establishes a robust, machine- actionable framework for a digital quality infrastructure [4]. Provenance-enabled systems not only enhance cross-disciplinary data usability but also provide a scalable foundation for metrological traceability and digitalization in industry, where reliable measurement assurance data are critical to maintaining global standards and supporting innovation.

II. BACKGROUND AND RELEVANCE OF PROVENANCE IN METROLOGICAL TRACEABILITY

Metrological traceability ensures the comparability of measurement results across different times, locations, and measuring instruments, thus providing a foundation for reliable decision making when dealing with information about physical quantities. Establishing traceability requires the coordination of metrological processes through a sequence of intermediate stages—called a traceability chain—ensuring that information is available where needed. The staged nature of a traceability chain enables stage-specific information to accumulate at each step, while intermediate results are passed along the chain. As a result, downstream process participants typically lack access to details about activities carried out upstream. To ensure the reliability and integrity of staged activities in practice, metrological traceability stipulates that each stage must conform to international standards for the competence of testing and calibration laboratories, such as ISO/IEC 17025:2017 [5]. Conformance to such standards involves regular independent assessments of a laboratory's capabilities, which, when deemed satisfactory, lead to accreditation. Such quality assurance requires meticulous documentation management. Maintaining a coherent system of traceable measurements requires sustainable, verifiable processes to assess quality and reliability.
Ehrlich and Rasberry noted that, at the turn of the 20th century, NIST defined traceability as follows [6, 7]:

"The property of the result of a measurement or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons all having stated uncertainties. It is noted that traceability only exists when scientifically rigorous evidence is collected on a continuing basis showing that the measurement is producing documented results for which the total

measurement uncertainty is quantified."[1]

Recognizing the importance of the second sentence, which distinguishes metrological traceability as a scientific principle, Ehrlich and Rasberry introduced the concept of metrological timelines—explicit sequences of key measurement events—to strengthen the collection of evidence of traceability [7].

Timelines are used to gather data about the systems contributing to a traceable measurement. Results from independent measuring systems are compared with those obtained from systems within the traceability chain. This comparison helps assess whether the systems in the chain are under metrological control, a crucial requirement for ensuring that the accuracy of results is objectively justifiable.

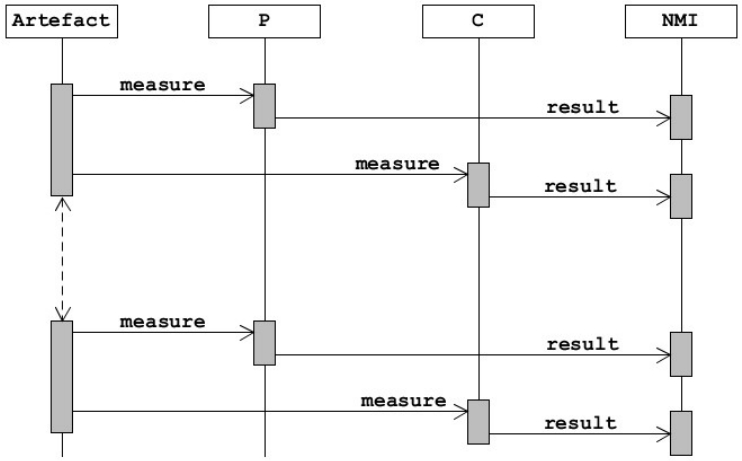


Fig. 1 A UML sequence diagram for the metrological timeline discussed here. Time progresses from top to bottom with horizontal lines used to represent events. The artifact is measured using systems P and C on two occasions (a substantial time apart). The NMI collects the results for quality assurances purposes.

The scenario considered in this paper is a simple timeline described by Ehrlich and Rasberry [6]. It involves a primary measurement standard, *P*, maintained by a national metrology institute (NMI) and an ensemble of control standards that perform at a level commensurate with *P* but are not considered primary standards. For simplicity, we will refer to a single control standard, *C*. The NMI also has an artifact that can be characterized using the systems *P* and *C*.

The timeline we are interested in describes a sequence of measurements of this artifact, using both *C* and *P*. Fig. 1 shows a UML sequence diagram of the scenario considered. The artifact is measured on two occasions using both systems P and C. The measurement results are retained by the NMI and used to check that P is operating at a satisfactory level, providing evidence of traceability claims on an ongoing basis.

III. APPLYING PROVENANCE IN METROLOGICAL TRACEABILITY

Generally speaking, provenance is a set of structured metadata that provides information about an object of interest, the origin of the object, the processes applied to the target object over time, including sources from which data have been derived and characterization or justification of the creation and transformation processes [8]. Fig. 2 shows the core information of the W3C PROV family of specifications:

- Agents, that perform processes on or own the objects of interest in the provenance description, whether they are software, hardware or human.
- Entities, the virtual or physical objects of interest in the provenance description.
- Activities, digital or physical manipulations applied to the objects of interest.
- Relationships that link agents, entities and activities, such as association of an agent with an activity or derivation of an entity from another entity.

[1] The International Vocabulary of Metrology defines metrological traceability as a property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty. [14, §2.41]

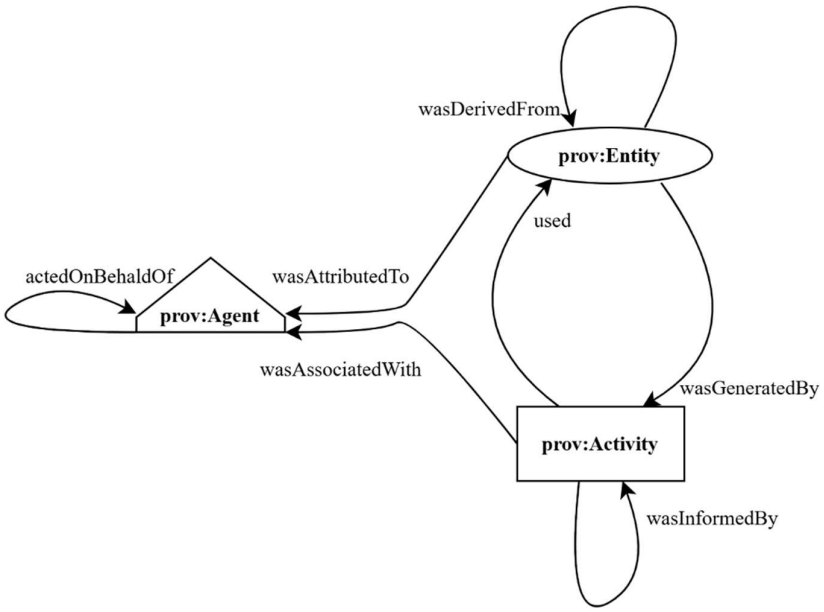


Fig. 2. Figure based on the starting point classes of W3C PROV-O ontology [13].

Provenance metadata take the form of a "graph describing the relationships among all the elements (sources, processing steps, contextual information and dependencies) that contributed to the existence of a piece of data [9]." Since provenance metadata offers a formal way to track data transformations and ownership, it is essential to support semantic procedures to assess the quality, verify authenticity, and evaluate the trustworthiness of data [10]. As a specification of provenance, the W3C PROV semantic model must accommodate different uses of provenance given that different perspectives on provenance may be applicable, resulting in different information that can be captured in metadata records [11].

- *Agent-centered provenance* describes the people and organizations involved in generating or manipulating the object of interest.
- *Object-centered provenance* traces the origins of the different elements of an object of interest to other entities.
- *Process-centered provenance* captures the actions and steps taken to generate the object of interest.

In measurement assurance, provenance could track the generation of, and updates to, a chart monitoring the stability of a measurement standard across its metrological timeline. Measurement assurance data provenance would include critical details of calibration activities, control standards, and associated measurement uncertainties. This metadata effectively describes a metrological timeline at a stage in a traceability chain, ensuring all relevant measurement and quality assurance activities are accurately represented, universally interpreted and machine-actionable.

IV. METROLOGICAL TIMELINE ─ PROV-O USE CASE

The concept of the metrological timeline is a compelling use case for provenance metadata. The use case details the provenance of internal measurement assurance data, supporting traceability to a primary standard. The provenance metadata links documented measurement results to a relevant quality assurance event from an object-centered perspective, documenting the origins of measurement assurance data to a reference. The provenance metadata is generated using the PROV ontology (PROV-O) specification of the PROV data model, and the model is illustrated graphically following the PROV graph layout convention [12, 13]. The intention of the provenance layout is to show the essence a set of provenance descriptions for this particular timeline scenario.

The object-centered provenance of measurement assurance data for an event in a metrological timeline, illustrated in Fig. 3, originates from the first stage of a traceability chain, measuring a quantity X. The primary measurement standard, P, is used (`prov:used`) in a primary calibration activity (`prov:Activity`) to measure the value X of a physical artifact (`prov:Entity`), part of an NMI's internal measurement assurance. This activity generates (`prov:generated`) a measurement result, X_P, linking the measured property of the artifact with the primary standard. A control standard, C (`prov:Entity`), is used in a control calibration

activity (`prov:Activity`) to measure the value *X* of the physical artifact to serve as an independent check on the stability of *P* over time. This activity generates a second measurement result X_C. Together, X_P and X_C contribute measurement assurance data that can be displayed on a chart, the entity of interest (`prov:Entity`) in this provenance record, which is derived from both the measurement results (`prov:wasDerivedFrom`).

One example for Fig. 3 could be the traceability chain for silicon trap detectors used in radiometry. The primary calibration of trap detectors (measurement artifact) is done by using a cryogenic radiometer (primary standard). One possible approach for measurement assurance is to use a set of control standards that are of the same type as the measurement artifact, i.e., silicon trap detectors. The control standards are used to check the stability of the cryogenic radiometer scale, under specific conditions, when this is used to calibrate another silicon trap detector. The provenance layout provides guidance for implementing provenance descriptions for silicon trap detectors in radiometry.

The use case demonstrates a laboratory's internal application of provenance metadata to document metrological timelines. However, the value of documenting metrological timelines is to communicate quality assurance information at any point in the traceability chain as supporting evidence of traceability claims. The timeline scenarios that introduce the comparison of standards along a traceability chain need to be modeled and validated with specific examples. The resulting provenance layouts would introduce an entity representing calibration certificates. The layouts that incorporate calibration certificates would illustrate sets of provenance descriptions that relate downstream calibrations to upstream measurement assurance activities.

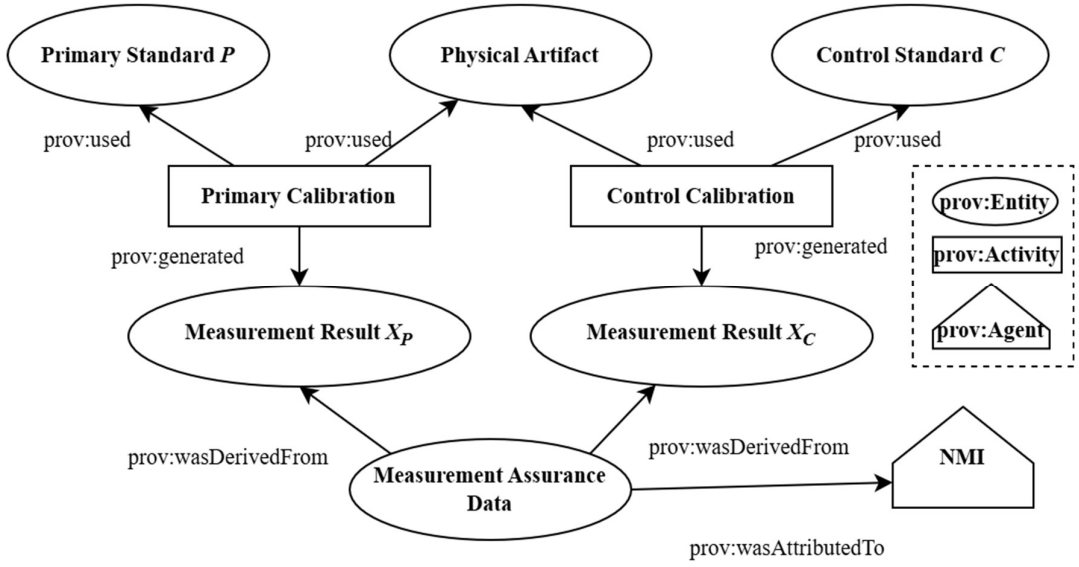


Fig. 3. PROV-O data model example applied to a metrological timeline describing the provenance of measurement assurance data. The subscripts P and C denote primary and control standards, which are used to measure the property X of a physical artifact.

V. DISCUSSION: PROVENANCE AND TRACEABILITY

The value of provenance metadata depends on the ability to distinguish the concept of provenance from metrological traceability, and model provenance metadata to adequately capture details of sequences of key measurement events to support metrological traceability.

Provenance captures the historical context and record of processes that influence a piece of data or thing, including what happened, why it happened (e.g., triggering entities), and the conditions under which changes occurred. Processes can include information about the creation, update, transcription, abstraction, validation and transferring ownership of data [1].

Traceability is the ability to verify and validate each step in a chain of custody through documented evidence, establishing accountability and reliability. Metrological traceability is a property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each

contributing to measurement uncertainty [14]. The stated measurement uncertainties should be backed by rigorous scientific evidence collected over time.

Measurement assurance data provenance documents its origin from a reference and its evolution over time to support traceability claims. While accreditation to ISO 17025 provides assurance to clients of calibration providers, provenance metadata can accompany calibration certificates, providing detailed records of measurement assurance data over time at any stage in the traceability chain. The ability to exchange and interoperate provenance metadata, to accompany digital calibration certificates, can facilitate visualization of measurement relationships, measurement assurance data and sequences of measurement events Along a traceability chain, provenance metadata linked to references provides access to quality assurance details both downstream and upstream in the chain. The capabilities that provenance metadata provides can increase trust and reliability of digital measurement data, and support metrological traceability as a scientific principle

PROV-O classes provide structured data to support evidence of traceability claims collected by quality assurance systems [12]. Mapping components and events to PROV-O classes enables the capture of metrological timelines. Additionally, the core classes of PROV-O serve as flexible containers for detailed, metrology-specific metadata, which further enhances the provenance record. For example, measurement results can be described with digital representation of quantity and units and attributed to a measurand [15, 16]. Measurands can be fully specified within a controlled taxonomy and quantity and units uniquely identified from authoritative metadata repositories to enhance interoperability.

Process-centered provenance metadata may provide information of the specific steps or actions taken at each calibration activity. The details of the calibration method and specific software used and their provenance (e.g. the documented changes to those methods and software) can be documented with extended components of PROV-O, such as `prov:Plan` (a subclass of `prov:Entity`). The provenance of the software used may give details of changes over time with the accounting of measurement errors and quantification of measurement uncertainties.

Agent-centered provenance documents the roles and responsibilities within the NMI's organizational structure, assigning specific activities to roles such as metrologists, technicians, and quality managers. Provenance metadata attributes changes in calibration methods, procedures and software to measurement experts, technicians or quality managers.

This paper advocates for the implementation of object-centered provenance to support the scientific principle of metrological traceability. Provenance metadata documents the relationship between the metrological entities ─ measurement results, measurement standards and calibration certificates. Provenance metadata also documents the historical context of measurement assurance of a standard which can be provided, along with the calibration certificate derived from the measurement result, at any point in a traceability chain.

VI. CONCLUSION

The W3C standard for provenance, the PROV data model provides a domain-agnostic reference model for describing certain aspects of metrological processes. This paper demonstrates that metrological timelines can be adequately represented to capture the details of activities and events that continuously monitor the stability of measurement standards and contribute to the generation of measurement assurance data. As a specification, provenance provides a digital-ready and interoperable foundation to support various metrological and quality processes that need to represent, exchange and integrate information generated from different systems and different perspectives. Interoperability remains a challenge without standardization of measurement concepts, their representation and relationships with provenance concepts. PROV-O can be used as reference model to develop metrology-specific provenance ontology.

ACKNOWLEDGEMENTS

The authors acknowledge Michael Chrubasik and Frédéric Tessier for taking the time and effort to review this work. We sincerely appreciate the valuable comments and suggestions, which have improved this manuscript.

REFERENCES

[1] ISO/IEC 8000-2:2022, Data Quality Part 2 - Vocabulary, International Organization for Standardization, 2022.

[2] M. Wilkinson, M. Dumontier, I. Aalbersberg and et al., “The FAIR Guiding Principles for scientific data management and stewardship,” *Scientific Data,* vol. 3, no. 1, p. 160018, March 2016.

[3] L. Moreau and P. Missier, “PROV-DM: The PROV Data Model,” W3C Recommendation, 30 April 2013. [Online]: https://www.w3.org/TR/2013/REC-prov-dm-20130430/.

[4] F Haertig, J. Hornig, T. Wiedenhofer, and S. G. Hackell, “The digital calibration certificate”, *PTB-Mitteilungen,* no. 127, p. 7, 2017.

[5] ISO/IEC, 17025:2017 General requirements for the competence of testing and calibration laboratories, 3rd ed., International Standards Organisation, 2017.

[6] R. E. Hebner, “Calibration Traceability: A Summary of NIST's View,” 1996.

[7] C. D. Ehrlich and S. D. Rasberry, “Metrological timelines in traceability,” *Metrologia,* vol. 34, pp. 503 - 514, 1997.

[8] P. Missier, K. Belhajjame and J. Cheney, “The W3C PROV family of specifications for modelling provenance metadata,” Genoa, Italy, 2013.

[9] L. Carata, S. Akoush, N. Balakrishnan, T. Bytheway, R. Sohan, M. Seltzer and A. Hopper, “A primer on provenance,” *Communications of the ACM,* vol. 57, no. 5, p. 5, 2014.

[10] J. Cheney, A. Chapman, J. Davidson and A. B. Forbes, “Data provenance, curation and quality in metrology,” in *Advanced Mathematical and Computational Tools in Metrology and Testing XII*, vol. 90, Singapore, World Scientific Publishing Co Pte Ltd, 2022, pp. 167-187.

[11] Yolanda Gil; Simon Miles; eds. “PROV Model Primer”, 30 April 2013, W3C Note. URL: http://www.w3.org/TR/2013/NOTE-prov-primer-20130430/.

[12] T. Lebo, S. Sahoo and D. McGuinness, “PROV-O: The PROV Ontology,” W3C Recommendation, 30 April 2013. [Online]. Available: https://www.w3.org/TR/2013/REC-prov-o-20130430/.

[13] W3C PROV Working Group, “PROV Graph Layout Conventions”, 2012. [Online]. Available: http://www.w3.org/2011/prov/wiki/Diagrams.

[14] Joint Committee for Guides in Metrology (JCGM), *The international vocabulary of metrology--Basic and general concepts and associated terms (VIM,* Third ed., JCGM, 2012.

[15] B. D. Hall and M. Kuster, "Representing quantities and units in digital systems," *Measurement: Sensors,* vol. 23, p. 100387, 2022.

[16] M. Kuster, “A guide to standardizing digital calibration and measurement capabilities,” *Metrologist,* vol. 15, no. 1, pp. 36-47, January 2022.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

On-Line Frequency Forecasting using Convolutional Neural Networks

Théo Chacou Bertoldi[1], Viktoriya Mostova[2], Silvia Iuliano[2], Alfredo Vaccaro[2]

[1]*Electrical Engineering Department, Polytech Clermont, Campus des Cézeaux, Aubière, 63100, France, theochacoubertoldi@gmail.com*
[2]*Department of Engineering, University of Sannio, Piazza Roma 21, 82100, Benvento, v.mostova@studenti.unisannio.it, s.iuliano@studenti.unisannio.it, vaccaro@unisannio.it*

***Abstract* – In modern power systems the replacement of large synchronous generators with distributed inverter-based resources is lowering the power system inertia, making electrical grids more vulnerable to dynamic perturbations. For this purpose the European Network of Transmission System Operators for Electricity promoted the enhancement of the grid operation tools with specific functions for on-line estimation of the power system inertia, which are extremely useful in detecting critical thresholds and triggering proper mitigation strategies. For this purpose, the development of reliable frequency predictive models represents an essential requirement for enabling system inertia estimation. To try and address this issue, this paper explores the role of Convolutional Neural Networks (CNN) for on-line frequency estimation from grid measurements. The main idea is to model the grid frequency deviations by a CNN-based identification technique, which allows inferring the main parameters ruling the power system dynamics. Detailed simulation results obtained on several case studies demonstrate that the CNN model is able to detect data patterns, and discover hidden relationships maintaining low estimation errors even for multi-step ahead frequency predictions, thus offering a valuable tool for inertia estimation in modern power systems, especially with the increasing share of renewable energy.**

I. INTRODUCTION

The large-scale integration of renewables and high-voltage direct current transmission lines, the development of increasingly competitive electricity markets, and the expansion of power systems have posed significant challenges in modern electrical grids operation [1] [2]. In this context, the need for enabling secure and reliable power systems operation has fostered the enhancement of conventional operation tools with monitoring and forecasting functions aimed at estimating the power demands in critical areas, the active power generated by renewable generators, and the active power flows on strategic interconnections [3]. However, as power grids evolve, the overall level of uncertainty increases significantly, requiring the analysis of an ever-increasing number of potentially relevant variables [4]. Among the different variables that can be predicted, it is beneficial to focus on the frequency forecast, which is a key measure to ensure the power system resilience to dynamic perturbations, and to assess the power system adequacy [5]. In particular, frequency forecasting is of primary importance for two main reasons: it allows to identify the location of possible faults, and it allows detecting the errors in measurement or data transmission, thus acting as a backup system in case of significant discrepancies between the measured and the expected value [6]. Two distinct computing paradigms can be adopted to address the problem of frequency prediction, which rely on statistical and machine learning (ML)-based approaches, respectively. The statistical approach is based on the analysis of the system past behavior to identify patterns and trends that allow the prediction of the future values of the variables of interest, typically through the use of time series [7]. Among these methods it is worth mentioning the application of the state space model and basis functions [6], extended Kalman filter [8], Newton-type algorithms [9], discrete Fourier transform [10] and least squares techniques [11]. Although the statistical approach is well-established and widely used in several fields, it has some limitations, including the inability to capture complex non-linear relationships and reduced robustness to variations in data distribution. In this context, the deployment of ML techniques represents an interesting alternative, which is recently attracting increasing interest due to their ability to identify complex correlations that would otherwise be difficult to detect with statistical methods [7]. Indeed ML-based frequency forecasting methods are based on adaptive mathematical algorithms enabling the knowledge discovery from large amounts of data. Through this process, the analyst can identify recurring patterns, correlations, and complex relationships between the input variables. Therefore, this information can be used to build predictive models capable of generating accurate frequency forecasts and dynamically adapting to new data, continuously improving the forecasting performance. References to studies using artificial neural networks [12], genetic algorithms [13] and support vector regression [14] for frequency forecast-

ing are available in the literature. These papers demonstrated that the deployment of ML-based methods could be the most promising enabling methodology for power system frequency prediction. Anyway, in the Authors' opinion, several open problems still need to be addressed for deploying ML-based frequency prediction models in realistic operation scenario. In particular, new methods for enhancing the adaptivity and generalization capabilities of ML-based models are required to effectively address the complex time-varying phenomena characterizing frequency variations in modern power system operation. To try and address this issue, this paper explores the potential role of CNN to progressively learn the complex features of the frequency variations, starting from low-level elements to more abstract concepts, through a series of convolutional layers. It is expected that the distinctive features of CNN, and in particular their highly adaptivity levels, allow inferring the main parameters ruling the frequency variation dynamics. Detailed simulation results obtained on several case studies demonstrate that the CNN model is able to detect data patterns, and discover hidden relationships maintaining low estimation errors even for multi-step ahead frequency predictions, thus offering a valuable tool for inertia estimation in modern power systems, especially with the increasing share of renewable energy.

II. A CNN-BASED MODEL FOR FREQUENCY PREDICTION

A CNN-based method is proposed to predict the frequency evolution over several time frames. The main idea is to deploy a CNN to extract features from frequency time series using convolution and pooling operations. Convolution is a the first step in feature extraction, inspired by signal analysis, which consists of applying a filter (also called kernel) on the input matrix in order to identify local patterns. The result of this operation is a feature map, which highlights the most relevant aspects of the frequency variations. For this purpose, several variants of convolution can be deployed to improve both the efficiency and accuracy of the knowledge extraction process, including standard convolution, deformable convolution, dilated convolution, and padding convolution. In particular, the convolution formula, shown in Eq.(1), has been adopted in this study:

$$S(i,j) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I(i+m, j+n) \cdot K(m,n) \qquad (1)$$

where $S(i,j)$ is the frequency value of the convolution, $I(i+m, j+n)$ is the value of the input variable, $K(m,n)$ is the value of the kernel (or filter) at (m,n), which are the dimensions of the filter. After applying convolution, the resulting feature maps contain a large number of features, which can lead to a risk of overfitting. To reduce this risk, pooling (or downsampling) is introduced, which acts as a dimensionality reduction technique, keeping only the most important information. There are two main types of pooling: Max Pooling and Average Pooling. Max Pooling selects the maximum value in each region of the feature map as shown in Eq.(2):

$$P(i,j) = \max_{m,n \in R(i,j)} I(m,n) \qquad (2)$$

where $P(i,j)$ is the max pooling output value at position (i,j) and $R(i,j)$ is the input $f \times f$ size region considered for pooling. Average Pooling calculates the average of the values present within each region of the feature map, as shown in Eq.(3):

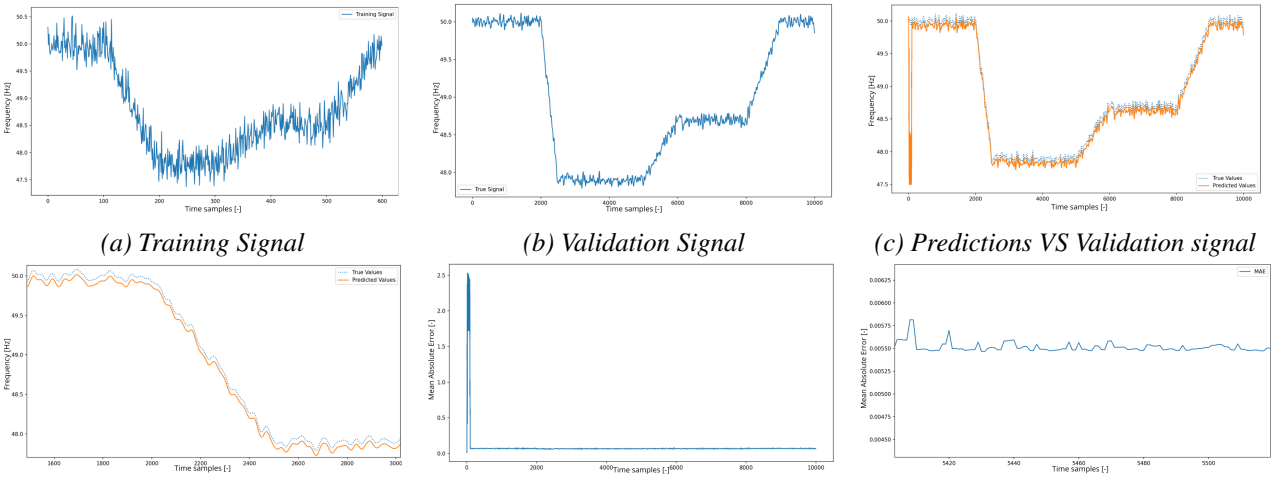
$$P(i,j) = \frac{1}{f^2} \sum_{m,n \in R(i,j)} I(m,n) \qquad (3)$$

where $P(i,j)$ is the average pooling output value at i,j, $R(i,j)$ is the region of dimension $f \times f$ and $1/f^2$ it is the normalization to obtain the mean.

Hence the overall computational process can be summarized as follows:

1. The raw frequency data, reflecting an experimental profile, are received by the input layer;
2. The convolutional layers apply filters to detect features, whose parameters are updated during the learning process. A key step in this process is the selection of the activation function, which introduces nonlinearity into the model, allowing the network to represent more complex relationships;
3. The Pooling Layer reduce the dimensionality of the feature maps, improving computational efficiency;
4. The feature maps are flattened into a vector, which is then processed by a feedforward neural network in the so-called fully connected layers;
5. The output layer produces the final result.

The expected advantages of using this computational process mainly derive from some key features characterizing CNN-based data processing. First, there are local connections; instead of connecting each neuron to all those in the previous layer, each neuron is connected to only a tiny portion of the neurons, which significantly reduces the number of parameters and accelerates the model convergence. Furthermore, CNNs exploit weight sharing, a principle according to which a group of connections can share the exact weights, reducing the number of parameters to be learned. Another important aspect is dimensionality reduction through downsampling; the pooling layers apply the principle of local correlation, which allows reducing the amount of data to be processed and preserving the relevant information.



(a) Training Signal *(b) Validation Signal* *(c) Predictions VS Validation signal*

(d) Zoom on predictions *(e) MAE profile* *(f) Zoom on warm-up zone*

Fig. 1. Ultra Short-term Frequency Prediction: First Approach

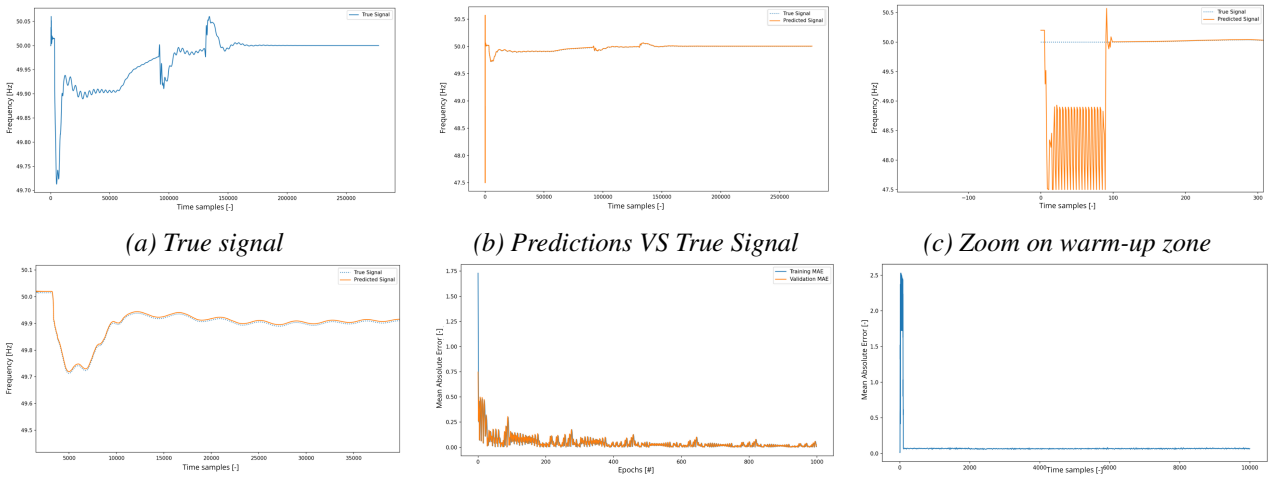
III. CASE STUDY

The CNN-based frequency prediction model has been tested in the task of analyzing sequential frequency data, reflecting an experimental profile, receiving, as input to the neural network, the frequency value at time t. The first convolutional layer uses 16 filters to identify patterns in the data, followed by a subsampling layer, MaxPooling, that reduces complexity by keeping only the most relevant information. This process is repeated with convolutional layers of 32 and 64 filters, alternating with subsampling, progressively refining the analysis. Next, GlobalMaxPooling summarizes the data in one value per filter, and the fully connected layers use this information to recognize patterns and generate predictions or classifications. Therefore, the CNN takes a frequency as input and predicts the next using the direct method.

A. *Ultra Short-term Frequency Prediction*

During the first experimental activity, the training and validation process are separated. Training involves creating, compiling, and training the CNN on a complex signal that simulates variations. The trained model is then saved for validation and tested with accurate data. This method does not allow for hyperparameter optimization, but it does reduce the development time. In particular, the training process starts by generating signals of different natures combined to generate a complex time series. A CNN, composed of multiple convolution and pooling layers, is trained on this data to predict frequency variations, minimizing the mean absolute error (MAE). Once training is complete, the model is saved, and its performance is evaluated through graphical representations that illustrate the trend of the predicted frequencies and the MAE evolution, providing an in-depth analysis of the model's ability to anticipate frequency variations in the system. The signal used to train the CNN varies between 47.5 Hz and 51.5 Hz, representing the physical limits of the actual system. Furthermore, the signal is intentionally characterized by a certain degree of noise to ensure that the CNN can make predictions even in more challenging operating conditions than the standard ones. The graphical representation of the noise signal is shown in Fig. 1a. After completing the training process, the CNN has been validated by generating a variable frequency signal characterized by variations and adding noise to simulate realistic conditions. A deviation of the frequency profile has been introduced into the model that reflects realistic operating conditions, ensuring representative coverage of probable scenarios. In particular, a uniform noise profile has been chosen to simulate random noise, to represent uncertainty without introducing any polarization into the profile. After that, the signal is then filtered using a low-pass filter to attenuate the noise, and the result is represented in Fig. 1b. During the validation phase, the model predicts the subsequent frequency from the filtered signal segments, ensuring the predictions fall within a physically plausible range. The results, including forecasts, actual values, and mean absolute errors, are recorded and analyzed to evaluate the model's performance. In addition, explanatory graphs illustrate the evolution of the signal, the predictions made, and the MAE over time, allowing a visual assessment of the model's accuracy.

At the end of the compilation process, the following results emerge: the graph in Fig. 1c shows the prediction curve (in orange) superimposed on the actual signal (in dotted blue), initially in its entirety and then with an enlargement of the frequency descent phase from 50 Hz to 47.9 Hz, Fig. 1d. The analysis highlights how the predictions correctly follow the validation signal. A relevant aspect is the trend of the MAE as a function of the pre-



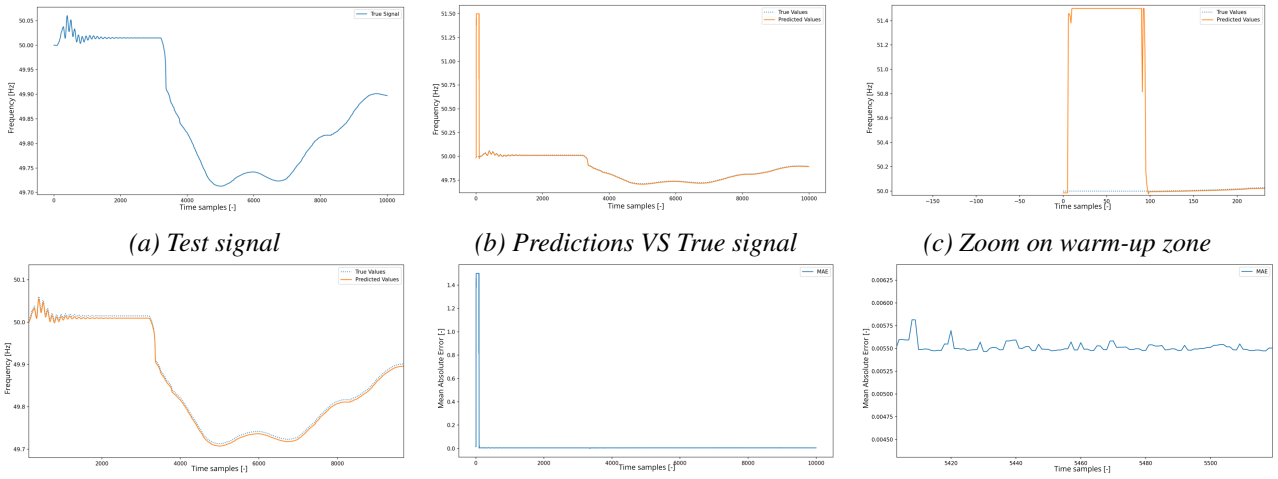
(a) True signal *(b) Predictions VS True Signal* *(c) Zoom on warm-up zone*

(d) Zoom after the warm-up zone *(e) MAE profile* *(f) Training and validation MAE*

Fig. 2. Ultra Short-term Frequency Prediction for a realistic operation scenario

dictions. In particular, by enlarging the area characterized by the highest error rate, Fig. 1f, it is observed that this corresponds to the first 100 points of the signal. This interval represents the warming up phase of the CNN, during which the model calibrates itself and learns the characteristics of the input signal. A synthetic analysis shows that the CNN presents significant errors in the first 100 points out of 10,000, corresponding to 1% of the entire signal. Once the calibration phase is completed, the CNN demonstrates a high predictive capacity, generating forecasts with a constant error close to zero, as highlighted in Fig. 1e. Indeed, after the first 100 forecasts, the maximum error found equals 0.02790. Once the CNN model has been trained and validated, it is essential to verify its prediction ability under more realistic conditions. To this end, the frequency evolution of a real system was analyzed. The compilation process took a significant amount of time due to the length of the signal, which includes 270,000 points, resulting in a processing time of approximately six hours. At the end of the compilation, the Fig. 2a and Fig. 2b were obtained. The warm-up phase occurs again in the first 100 points, as the enlarged Fig. 2c highlights. Beyond this interval, the CNN demonstrates a high predictive capacity, providing accurate estimates of the different frequency variations. Further experimental analysis have been developed by solving the same prediction problem applying the conventional approach used in neural network applications, which involves splitting the data into training, validation, and testing. At the end of each training epoch, the model is evaluated on the validation set to optimize hyperparameters and monitor performance, using metrics such as accuracy and F1 score. This approach prevents overfitting and improves the model ability to generalize the unseen data, thus ensuring reliable and accurate predictions. The input signal to the CNN consists of 80% of the training set, while a validation set distinct from the training set represents the remaining 20%. At the end of the compilation phase, the following results were obtained.

The Fig. 2f illustrates the evolution of the mean absolute error for the training and validation sets during the CNN training epochs. In the initial phases, the errors are high, but they quickly decrease, indicating that the model is effectively learning the characteristics of the signal. The convergence of training and validation errors around 0.1 suggests a good model generalization capability, with no apparent signs of overfitting. However, regular oscillations are observed, especially in the final stages of training, which could arise from fluctuations in the data batches or the nature of the validation set itself. Overall, the model appears stable and efficient after about 200 epochs. After the CNN training and validation phase, it is essential to verify its predictive capabilities under more realistic conditions. For this purpose, the frequency evolution of a real system was recovered. Since the system under consideration includes a high number of samples (270,000), to optimize the design time, it was decided to perform the predictions on the first 10,000 samples (Fig. 3a). The Fig. 3b compares the predicted signal with the actual signal, clearly analyzing the model performance. During the initial warm-up phase, which extends over the first 100 points, the model presents a higher error rate due to the calibration and adaptation process to the input data. However, after this phase, the predictions align closely with the actual signal, demonstrating the model effectiveness in capturing the underlying patterns and variations present in the data. Furthermore, a constant offset in the predictions is observed, suggesting a possible systematic bias. This deviation can be manually corrected to further improve the model accuracy.



(a) Test signal *(b) Predictions VS True signal* *(c) Zoom on warm-up zone*

(d) Zoom after warm-up zone *(e) MAE profile* *(f) Zoom after warm-up zone*

Fig. 3. Ultra Short-term Frequency Prediction for a realistic case study: Second Approach

B. Result Discussion

The analysis of the obtained results confirmed that after an initial adaptation phase, the CNN model exhibits a high accuracy in predicting the actual signal. The strong alignment between the predicted and the actual signals in the post-warm-up phase highlights the model ability to effectively learn from the input data and generalize the acquired information. The average error observed after the warm-up phase is tiny, equal to 0.0054044, suggesting a potential margin for further optimization of the model accuracy. The model demonstrates robustness and reliable performance, confirming its effectiveness in analyzing and predicting frequency variations in a realistic application context. It should be pointed out that the choice of using CNN is related to their high capacity to capture patterns from the signal, unlike recurrent neural networks, which are also used in frequency estimation and allow better understanding of the time dynamics of the signal. There are also hybrid approaches that tend to improve the extraction capability of both spatial and temporal dynamics.

C. Short-term frequency predictions

Further experimental analysis have been performed to assess the performance of the CNN-based model for frequency prediction over longer time frames. For this purpose, we chose to limit the predictions to frequency $k+100$ since this interval already allows us to identify a significant trend. Indeed, the prediction error tends to increase as the temporal distance of the predicted frequencies increases. For a realistic For this analysis, the mean squared error (MSE) on the test signal is calculated using two distinct approaches:

1. **The Persistence Method:** This method relies on a simple temporal shift to estimate future frequency values. In particular, the current value of the signal is used as a prediction for a future time step. The implementation involves shifting the test signal by T steps and then calculating the MSE between the original and the shifted signal. This calculation is performed for different values of T, ranging from 1 to 100.
2. **CNN predictions:** To evaluate the CNN ability to predict frequencies up to $k+100$, a pre-trained model is used. For each time step k, it generates frequency predictions for the following k steps, with k varying from 1 to 100.

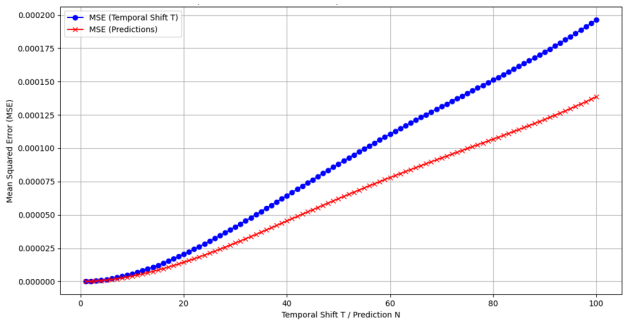


Fig. 4. Comparison of MSE based on temporal shift T and MSE for predictions

The results obtained using the two approaches are reported in Fig. 4. Specifically, the blue curve represents the MSE trend as a function of the temporal shift T for the persistence method, while the red curve illustrates the MSE associated with the CNN predictions up to $N+100$. The graph analysis shows that the MSE of the CNN predictions is lower than that obtained with the persistence method, suggesting that the neural network can capture signal trends

more accurately than a simple model based on the temporal shift. This comparison confirms that, although the error increases with the temporal distance of the predictions, the CNN model provides more accurate results than the persistence method.

IV. CONCLUDING REMARKS

The planned CNN has proved highly effective in predicting frequency changes in power systems. Indeed, the results show that the model can accurately estimate frequency in a variable time scenario. Validation tests performed on several signals have further confirmed the accuracy of the CNN, providing results in line with theoretical expectations. In particular, the approach based on splitting the data into training, validation, and test sets has proven particularly effective, enabling the model optimization by hyperparameter tuning. This resulted in improved generalization ability and predictive accuracy. Overall, from the analysis carried out, the implementation of identification models based on CNN could represent a reliable tool for the estimation of inertia in modern energy systems, providing a strategic tool in the context of increasing integration of renewable energies. The model's ability to predict frequency variations with high precision could help improve situational awareness, which is essential for improving the resilience of the grid to dynamic disturbances. Future developments will focus on implementing and comparing recursive and hybrid neural networks, which combine convolutional and recursive. The goal is to exploit the strengths of both architectures: the ability of CNNs to identify patterns in the data, and the ability of recursive networks to extract information, particularly in the case of dynamic series, to improve predictive accuracy.

V. REFERENCES

[1] I. Jayawardene and G. K. Venayagamoorthy, "Cellular computational extreme learning machine network based frequency predictions in a power system," in *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2017, pp. 3377–3384.

[2] Y. Zhang, X. Wang, and L. Ding, "Lstm-based dynamic frequency prediction," in *2020 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 2020, pp. 1–5.

[3] N. Tomin, "Intelligent monitoring and forecasting of the expected operating conditions of electric power system," in *Proceedings of the 2011 3rd International Youth Conference on Energetics (IYCE)*. IEEE, 2011, pp. 1–8.

[4] M. P. Garcia and D. S. Kirschen, "Forecasting system imbalance volumes in competitive electricity markets," *IEEE Transactions on Power Systems*, vol. 21, no. 1, pp. 240–248, 2006.

[5] O. Yurdakul, F. Eser, F. Sivrikaya, and S. Albayrak, "Very short-term power system frequency forecasting," *IEEE Access*, vol. 8, pp. 141 234–141 245, 2020.

[6] J. Dong, X. Ma, S. M. Djouadi, H. Li, and Y. Liu, "Frequency prediction of power systems in fnet based on state-space approach and uncertain basis functions," *IEEE Transactions on Power Systems*, vol. 29, no. 6, pp. 2602–2612, 2014.

[7] V. Negri, A. Mingotti, R. Tinarelli, and L. Peretto, "Comparison between the machine learning and the statistical approach to the forecasting of voltage, current, and frequency," in *2023 IEEE 13th International Workshop on Applied Measurements for Power Systems (AMPS)*. IEEE, 2023, pp. 01–06.

[8] A. Routray, A. K. Pradhan, and K. P. Rao, "A novel kalman filter for frequency estimation of distorted signals in power systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 51, no. 3, pp. 469–479, 2002.

[9] V. V. Terzija, M. B. Djuric, and B. D. Kovacevic, "Voltage phasor and local system frequency estimation using newton type algorithm," *IEEE Transactions on Power Delivery*, vol. 9, no. 3, pp. 1368–1374, 1994.

[10] T. Tyagi and P. Sumathi, "Comprehensive performance evaluation of computationally efficient discrete fourier transforms for frequency estimation," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 5, pp. 2155–2163, 2019.

[11] A. Pradhan, A. Routray, and A. Basak, "Power system frequency estimation using least mean square technique," *IEEE transactions on power delivery*, vol. 20, no. 3, pp. 1812–1816, 2005.

[12] L. Lai, W. Chan, C. Tse, and A. So, "Real-time frequency and harmonic evaluation using artificial neural networks," *IEEE Transactions on power delivery*, vol. 14, no. 1, pp. 52–59, 1999.

[13] D. V. Coury, A. C. Delbem, J. R. De Carvalho, M. Oleskovicz, E. V. Simoes, D. Barbosa, and T. V. da Silva, "Frequency estimation using a genetic algorithm with regularization implemented in fpgas," *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1353–1361, 2012.

[14] Q. Bo, X. Wang, and K. Liu, "Minimum frequency prediction of power system after disturbance based on the v-support vector regression," in *2014 International Conference on Power System Technology*. IEEE, 2014, pp. 614–619.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Metrological Characterization of a Vehicle's Charging Profile for Smart Charging Applications

Chiara Franzoni[1], Antony Vasile[1], Davide Astolfi[1], Dmitrii Vasenin[1], Alessandro Musatti[1], Marco Pasetti[1], Stefano Rinaldi[1]

[1]*University of Brescia, Dept. of Information Engineering, Via Branze 38, chiara.franzoni, antony.vasile, davide.astolfi, d.vasenin, alessandro.musatti, marco.pasetti, stefano.rinaldi@unibs.it*

***Sommario* – In the latest years, the concept of smart charging has emerged as a necessity for the optimal management of power grids with high penetration of electric vehicles and renewable power generation. Based on this, the aim of the present work is filling a gap in the literature about the metrological characterization of electric vehicles connected at the charging station. In particular, a Charging Point Management System based on the Open Charge Point Protocol is developed and then integrated into an ICT system to support the charging of electric vehicles via wallboxes installed at the University of Brescia (Italy). Subsequently, smart charging functionalities are implemented to control current-based charging and these are investigated statically, as for example by characterizing the power factor as a function of the requested current, and dynamically, by inquiring the response time of the communication and the current resolution limits.**

I. INTRODUCTION

Integrating renewable energy sources into the charging pipeline of Electric Vehicles (EVs) [1] is essential to meeting Greenhouse Gas (GHG) emission reduction [2] targets and effectively combating climate change. In fact, the energy sector is the primary contributor to GHG emissions, with the transportation sector being a significant energy consumer. For instance, in 2022, transport activities accounted for 31.0% of the European Union's final energy consumption. The growing share of renewable energy production and the transition to Electric Vehicles (EVs) introduce new challenges for the operation and management of power grids [3, 4], particularly due to uncertainties in both energy generation and consumption. The uncoordinated charging of vast amounts of EVs might negatively affect peak load and transformer loading and might advocate for grid upgradation costs. The smart charging [5, 6] has hence evolved as a game changer, in that the adaptivity of the charging process not only might limit the above drawbacks, but might also contribute to the smooth functioning of the power grid, as for example by stabilizing the power grid [7], by optimizing the exploitation of renewable power sources [8],by minimizing costs and optimizing tariffs [9].

A. Related Works

In such a context, measurements are fundamental in order to characterize the responsiveness of the communication between the EV and the Charging Station (CS) and to evaluate the real energy efficiency during the recharging process. Despite this, to the best of the authors' knowledge, there is a very limited number of works related to metrological aspects of the charging process. The work in [10] deals with the realization and testing of field equipment for metrological verification of the metering subsystem inside electric vehicle supply equipment and an accurate estimation of the energy absorbed by a battery electric vehicle is provided. In the following work [11], the underlying principle of smart charging (forcing the EV to draw different current levels) is contemplated, but the core of the study is on the communication between the EV and CS and a Man In the Middle approach is proposed and tested. These works are part of an European project, outlined in [12], in the context of which a devoted laboratory has been developed (described in [13]).

B. Objectives

Based on the identified research gap, the objective of this work is developing a CPMS (Charging Point Management System), a software for managing charging processes based on the OCPP (Open Charge Point Protocol) which is then integrated into an ICT system to support the charging of electric vehicles via wallboxes installed at the University of Brescia (Italy). Subsequently, smart charging functionalities are implemented to control current-based charging, and algorithms are developed to achieve a metrological characterization of a vehicle connected to the charging network. This characterization aims to understand the vehicle's features and determine how it can be utilized in a smart grid context to help balance the grid when needed.

II. MATERIALS AND METHODS

A. The ICT Infrastructure

The diagram in Figure 1 illustrates the ICT infrastructure used for charging, which consists of:

- **The CPMS**, the software developed for managing the charging process;

- **The CS** connected to the grid to enable the charging of electric vehicles;
- **Smart Meter, UPM209 Network Analyzer**: The charging station provides information on the energy supplied via OCPP, but these are not sufficient for a metrological characterization. Therefore, an additional meter was added at this aim. In our case, we monitored values such as current, phase shift, active, reactive, and apparent power, as well as absorbed and supplied energy, to perform a characterization of the vehicle. The UPM209 network analyzer supports a measurable voltage range from 10 V to 285 V RMS (phase-to-neutral) and from 17 V to 495 V RMS (phase-to-phase), depending on the wiring configuration, and enables FFT-based harmonic analysis only above 20 V RMS for voltage and 200 mA for current. To ensure data reliability, measurements are automatically disabled outside these ranges, with overflow conditions clearly indicated in compliance with EN 61010-2-030. From a metrological perspective, the device guarantees typical accuracies within specified operating ranges: ±0.2% of the reading for voltage measurements between 10% and 100% of full scale, ±0.4% for current measurements between 5% and 100% of full scale, ±0.5% for active power (at power factor = 1), and ±0.1% for frequency. It complies with Class 1 accuracy for active energy according to IEC/EN 62053-21 and Class 2 for reactive energy as per IEC/EN 62053-23. The accuracy of the device guarantees that the experiments indicated in Section D. are distinguishable the one with respect to the other.

In Figure 2, the structure of the CPMS is reported. Specifically, the developed CPMS consists of:

- **A main block**, which is the main Charging Management System (CMS).
- **A relational database**, with which the CMS interacts using the Python library `psycopg2`.
- **A time-series database**, used to store data from smart meters via the `InfluxDB` library.
- **The pymodbus library**, which allows the implementation of the `Modbus protocol` in Python to enable communication with smart meters.
- **The OCPP 1.6 library by Mobility House**, based on JSON, to enable communication between the CPMS and the EVSE.
- **Functions for generating charging profiles**, capable of calculating the current value to be supplied to the vehicle.

B. The Open Charge Point Protocol

The OCPP protocol, based on WebSocket, represents the standard for communication between the CPMS and the CS, defining all possible messages that the two parties can exchange and the payload of each. In particular, since the version of OCPP implemented by Mobility House was used, the exchanged messages are simply JSON. Each communication is based on two messages:

- A **req** message, through which a request is sent.
- A **conf** message, through which the response is sent.

This protocol also includes smart charging features, allowing the CPMS to set charging profiles that impose current limits.

C. Measurement Protocol And Smart Charging Characterization

The procedure is the following:

- The vehicle is connected to one of the available charging stations via a Mennekes cable. In this way, the signal sent from the station to the car through the Control Pilot (a wire in the connector) changes from a voltage of 12V to 9V (according to the IEC 61851), indicating that a vehicle is connected.
- Next, the RFID tag is scanned: the charging station can then read the identification and send an Authorize request via OCPP to the CPMS, which will verify the presence of the identification in the database and send a response to the authorization request.
- If the authorization is granted, the signal voltage will drop to 6V, indicating that charging can begin.
- The desired current limit is calculated and sent to the charging station via the appropriate OCPP functionality.
- Once the station receives the limit, it will impose it on the vehicle: specifically, the current value for charging is described by the duty cycle value of the signal exchanged through the Control Pilot, following the IEC 61851 protocol.
- Once the desired limit has been set, the charging parameters are read using UPM209 devices. The measurements typically included currents, active power, reactive power, apparent power, and the power factor.
- Once these values are acquired, they are stored both in the time-series database to keep track of the parameters associated with the transaction and in a .csv file to facilitate future analyses.

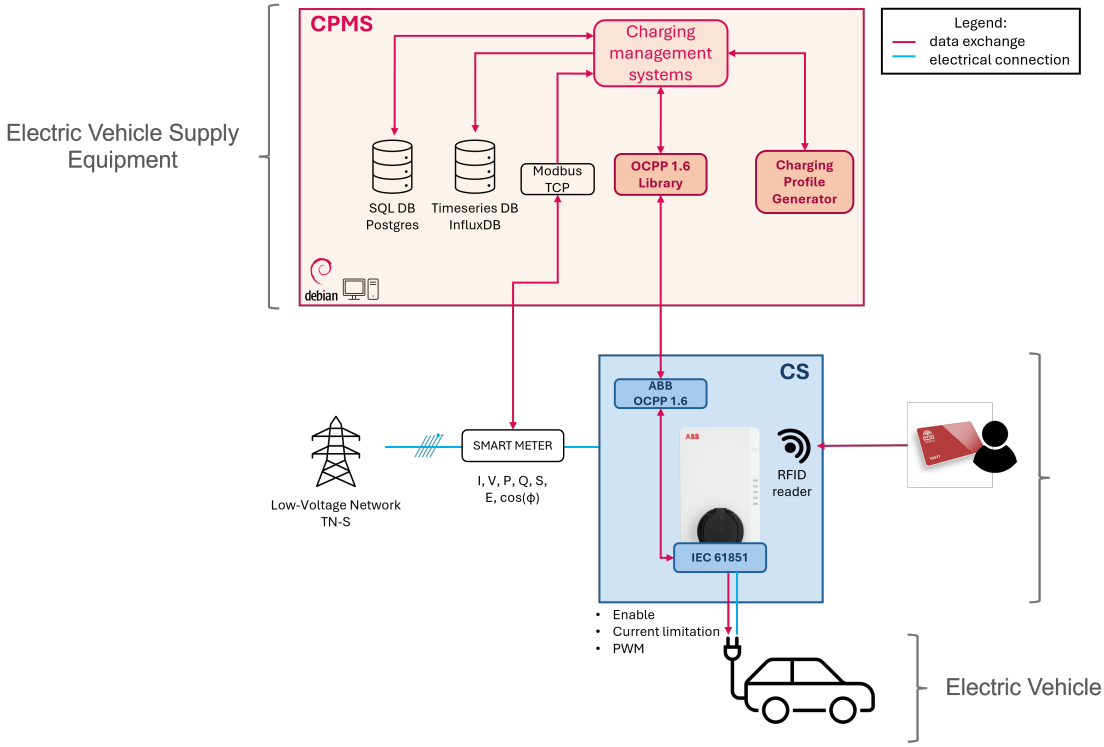


Fig. 1. Conceptual layout of the ICT infrastructure.

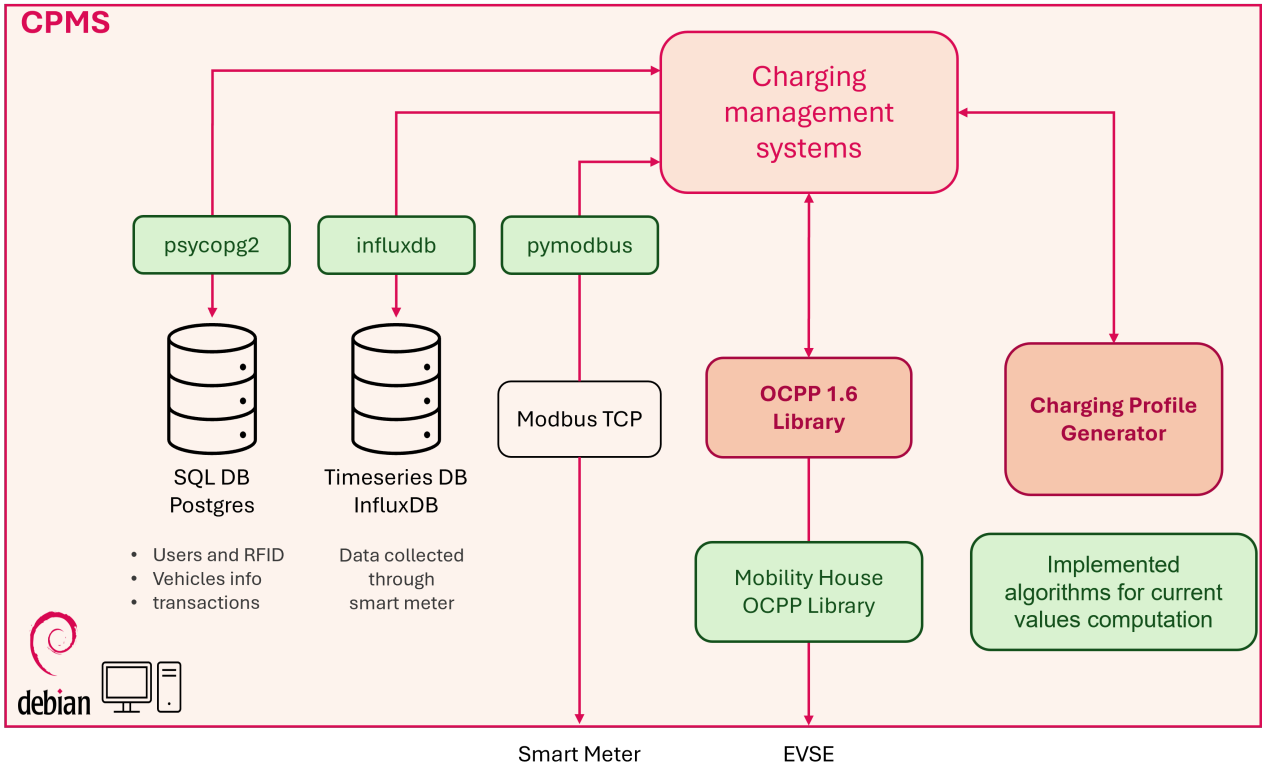


Fig. 2. The structure of the CPMS.

D. *The Experiments*

Several tests are conducted aimed at obtaining a metrological characterization of the charging profile of a specific vehicle. An electric Renault Zoe has been employed, and two ABB wallboxes, one single-phase and one three-phase, located in the university parking lots. In particular, two types of experiments have been carried out:

- **Stationary charging profile**: Sweep tests with linear steps of 1 A of imposed current are performed, in order to characterize the upper and lower limits of current absorption and the power factor associated with each current value. The objective is the creation of power-based charging profiles. The current set point is updated each 30 seconds.
- **Dynamical charging profile**: An essential characterization of the charging station has conducted to determine the resolution and accuracy of the current response, the response time to the commands sent, and the behavior in response to sudden changes in the imposed current limits, i.e., the step response.

III. RESULTS

A. *Stationary Charging Profile*

The results for the sweep test in the single-phase case are reported in Figures 3 and 4. From Figure 3, it arises that there is a lower current threshold, which is 6 A. The upper limit, however, contrary to what is stated in the wallbox specifications, turns out to be 16A instead of 32A, indicating that some settings of the station may have been modified. A power factor higher than 0.9 can be achieved only by requesting more than order of 8 A.

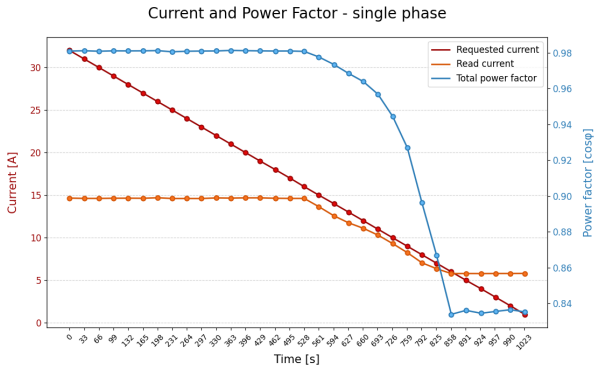


Fig. 3. Requested and actual current and power factor for the descending sweep test in the single-phase case.

The results for the three-phase sweep tests are reported in Figures 5 and 6. For requested currents lower than 7 A, the vehicle behaves as a generator rather than as a load, as it has been confirmed by further discharge tests which will be reported in the full paper version. For requested currents higher than approximately 9 A, the power factor has

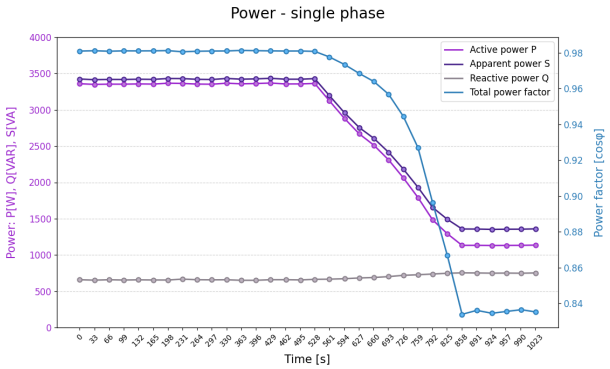


Fig. 4. Active, reactive, apparent power and power factor for the descending sweep test in the single-phase case.

quite high values, with a minimum of 0.86 for 12 A. From Figure 5, a mismatch between requested and measured current is highlighted especially as long as the current increases. This result motivates the further analyses conducted in Section B..

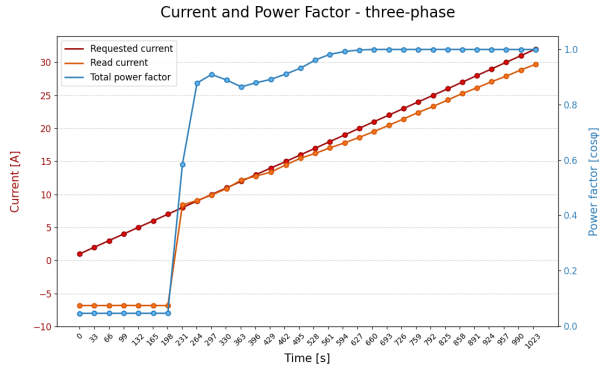


Fig. 5. Requested and actual current and power factor for the ascending sweep test in the three-phase case.

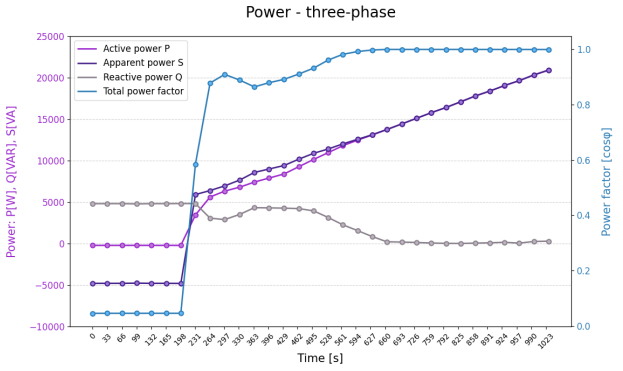


Fig. 6. Active, reactive, apparent power and power factor for the ascending sweep test in the three-phase case.

B. *Dynamical Charging Profile*

The results reported in Figure 7 refer to a test during which the requested three-phase current has been varied from 0 to 32 A every 10 seconds, with steps of 0.1 A. By varying the current with a finer grain compared to what was done in the sweeps, a step-like behavior in the current was observed, as different set current values corresponded to the same measured reading. In this test, for set limit values between 8 and 13 A, the measured current sometimes appears to exceed the limit of requested current. However, when plotting the current value multiplied by the power factor, it becomes evident that the current actually contributing to the active power never actually exceeds the set limit. This implies that if algorithms are to be developed to control charging based on power, the active power P used for charging would never exceed the limit value calculated using the formula $V \cdot I$.

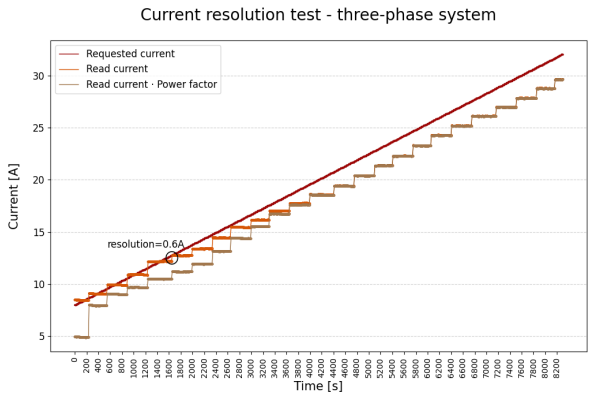


Fig. 7. Current resolution test.

IV. DISCUSSION

Future developments of the present work include the integration of a Blockchain into the implemented ICT system. This would facilitate local transaction management while enabling the adoption of smart contracts, which could dynamically adjust charging power based on demand, ultimately helping to reduce costs. Another possible extension of this work could be the development of Vehicle-to-Grid (V2G) regulation algorithms [14, 15, 16, 17], allowing the vehicle to function as a generator, supplying power to the grid to help balance it when needed. For these applications and in general for managing a fleet of vehicles, it is fundamental to have a reliable metrological characterization of the power flows, as done in this work.

ACKNOWLEDGMENTS

This study was carried out within the MOST-Sustainable Mobility National Research Center and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR)-MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4-D.D. 1033 17/06/2022, CN00000023), Spoke 5 “Light Vehicle and Active Mobility”. This manuscript reflects only the authors’ views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

RIFERIMENTI BIBLIOGRAFICI

[1] C. Filote, A.R. Felseghi, M.S. Raboaca, I. Aschilean, “Environmental impact assessment of green energy systems for power supply of electric vehicle charging station”. International Journal of Energy Research, 44(13), 2020, pp.10471-10494.

[2] T. Wei, J. Wu, S. Chen, “Keeping track of greenhouse gas emission reduction progress and targets in 167 cities worldwide”. Frontiers in Sustainable Cities, 3, 2021, 696381.

[3] S.A.A. Rizvi, A. Xin, A. Masood, S. Iqbal, M.U. Jan, H. Rehman, “Electric vehicles and their impacts on integration into power grid: A review”, 2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2), pp. 1-6, IEEE

[4] M. Nour, J.P. Chaves-Avila, G. Magdy, A. Sánchez-Miralles, “Review of positive and negative impacts of electric vehicles charging on electric power systems”. Energies, 13(18), 2020, 4675.

[5] S. Deb, M. Pihlatie, M. Al-Saadi, “Smart charging: A comprehensive review. IEEE Access”, 10 2022, pp. 134690-134703.

[6] P. Barman, L. Dutta, S. Bordoloi, A. Kalita, P. Buragohain, S. Bharali, B. Azzopardi, “Renewable energy integration with electric vehicle technology: A review of the existing smart charging approaches”, Renewable and Sustainable Energy Reviews, 183, 2023, 113518.

[7] A.Y. Lam, K.C. Leung, V.O Li, “Capacity estimation for vehicle-to-grid frequency regulation services with smart charging mechanism”. IEEE transactions on smart grid, 7(1), 2015, pp. 156-166.

[8] A.U. Rehman, Z. Wadud, R.M. Elavarasan, G. Hafeez, I. Khan, Z. Shafiq, H.H. Alhelou, “An optimal power usage scheduling in smart grid integrated with renewable energy sources for energy management”. IEEE access, 9, 2021, pp. 84619-84638.

[9] F. Daneshzand, P.J. Coker, B. Potter, S.T. Smith, “EV smart charging: How tariff selection influences grid stress and carbon reduction”, Applied energy, 348, 2023, 121482.

[10] A. Delle Femine, D. Gallo, C. Iodice, C. Landi, M. Luiso, “On-field Metrological Verification of EVSE”, 2023 IEEE International Workshop on Metrology for Automotive (MetroAutomotive), pp. 132-136, IEEE.

[11] A. Delle Femine, D. Gallo, C. Iodice, C. Landi, M. Luiso, “EVSE Metrological Verification Throu-

gh IEC 61851 Protocol Hacking", 2024 IEEE International Workshop on Metrology for Automotive (MetroAutomotive), pp. 130-135, IEEE.

[12] A. Delle Femine, C. Iodice, J. Kucera, E. Dierikx, A. Mariscotti, I.H. Torres, J.D. de Aguilar, T. Van Wijk, "Metrology for Electric Vehicle Charging Systems: An Overview of the European Research Project", 2024 IEEE International Workshop on Metrology for Automotive (MetroAutomotive), pp. 130-135, IEEE.

[13] J.D. de Aguilar, M.L. Romero, L. Matias, D. Peral, Y. Alvarez, M. Cervantes, R.M. Plaza, A.P. Balsells, "Metrology for electric vehicle charging stations", 2024 Conference on Precision Electromagnetic Measurements (CPEM), pp. 1-2, IEEE.

[14] S.S. Ravi, M. Aziz, "Utilization of electric vehicles for vehicle-to-grid services: Progress and perspectives". Energies, 15(2), 2022, 589.

[15] Y. Yang, W. Wang, J. Qin, M. Wang, Q: Ma, Y. Zhong, "Review of vehicle to grid integration to support power grid security". Energy Reports, 12, 2024, 2786-2800.

[16] D. Astolfi, A. Vasile, S. Iuliano, M. Pasetti, "Recent Developments on the Incentives for Users' to Participate in Vehicle-to-Grid Services". Energies, 17(21), 2024, 5484.

[17] A. Vasile, D. Astolfi, M. Pasetti, G. Sciumé, G. Zizzo, E. Riva Sanseverino, A. Flammini, "Technical Feasibility of Modular Energy-Saving Control Strategies of Vehicle-to-Grid Charging Stations for Frequency Regulation. Smart Grids and Sustainable Energy", 10(1), 2025, 4.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

A Comprehensive Methodology Based on SCADA Data Analysis for Diagnosing Static Errors Affecting Wind Turbine Performance

Davide Astolfi[1], Silvia Iuliano[2], Antony Vasile[1], Alessandro Canali[3], Marco Pasetti[1], Francesco Castellani[3], Alfredo Vaccaro[1]

[1]*University of Brescia, Dept. of Information Engineering, Via Branze 38, Brescia, Italy, davide.astolfi, antony.vasile, marco.pasetti@unibs.it*
[2]*University of Sannio, Dept. of Engineering, Piazza Roma 21, Benevento, Italy, s.iuliano@studenti.unisannio.it, vaccaro@unisannio.it*
[3]*University of Perugia, Dept. of Engineering, Via G.Duranti 93, Perugia, Italy, alessandro.canali@studenti.unipg.it, francesco.castellani@unipg.it*

***Abstract* – One of the most important root causes of non-optimal efficiency in the conversion of wind kinetic energy into electricity is represented by static errors affecting wind turbines. These can simply involve sensors (like the anemometers) or might deal with the set up of the rotor (as when the three blades do not have the same pitch reference). Based on such premise, in this work, a comprehensive framework is proposed for diagnosing static errors affecting wind turbine performance, based on SCADA data analysis. The three types of considered static errors are: static yaw error; absolute or relative blade pitch misalignment; anemometer bias. It is desirable to exploit as much as possible the SCADA-collected data for diagnosing such types of static errors. Yet, given the mediation of the control system, the SCADA data might even be misleading and, for example, might indicate a correct alignment of the rotor to the wind flow, while it is not correct. Hence, the objective of the work is identifying phenomena observable through SCADA data which can be related to such static errors. Acknowledged that some manifestations can be similar for the various errors, a work flow is set up for distinguishing between them. The selected phenomena are: under-performance, which is associated to all the errors; increased vibrations; shift in the nacelle anemometer measurements. Real-world test cases, namely wind farms operated by the ENGIE Italia company, are discussed. A combination of single wind turbine and fleet-wide methods is proposed. The latter involve dividing the fleet in potentially anomalous and normal wind turbines and modeling quantities of interest (like the power) as a function of the corresponding quantities of the healthy wind turbines. The collected case studies indicate that the proposed method is effective and allows distinguishing between the various types of errors related to the anemometer, the yaw and the pitch systems.**

I. INTRODUCTION

In the near-future scenarios [1, 2, 3], wind turbines will likely represent the most important technology for the production of electricity. Hence, it is fundamental that wind farms operate at their maximum possible efficiency but, unfortunately, this is typically not the case in real-world environments.

In particular, an overlooked cause of wind turbine efficiency loss is the presence of static errors. At least three main examples of such errors can be identified:

- The static yaw error [4, 5], meaning that, when the control system records that the rotor plane and the wind flow are perpendicular, indeed there is a constant angular shift γ;
- The blade pitch misalignment β [6, 7, 8], which can be absolute or relative;
- The anemometer bias [9], meaning that, at a certain point in the lifetime of the turbine, a constant shift δ occurs in the measurements collected by the nacelle anemometer.

The first two types of static error cause a loss of producible energy up to some percents of the Annual Energy Production [10, 11], while the anemometer bias might or might not be associated to a performance worsening, depending on the logic of the control system. Typically, for wind turbines with less than 2 MW of rated power, the wind speed measured by the nacelle anemometer is used as input for setting the rotational speed and blade pitch set points and, thus, a measurement bias leads to a non-optimal operation point [9].

It is desirable to exploit as much as possible the SCADA-collected data [12, 13] for diagnosing such types of static errors. Yet, given the mediation of the control system [14, 15], the SCADA-collected data might even be

misleading. Furthermore, in the literature it has not been systematically elaborated what manifestations (observable through SCADA data) are associated to the various types of static errors.

Hence, the contribution of this study is a comprehensive framework for employing SCADA data to diagnose static errors and distinguish between them. Namely, the main innovation of the present work is in the following aspects:

- Application of multivariate SCADA data analysis for wind turbine performance monitoring;
- Identification of the observable phenomena associated to the most common static errors;
- Formulation of guidelines for fixing the static errors and improving the power production of a fleet.

Real-world test cases, namely wind farms operated by the ENGIE Italia company, are discussed.

II. METHODOLOGY

In this work, a comprehensive analysis is carried out regarding the following phenomena, which can possibly be associated to the various static errors:

- Under-performance;
- Shift in the nacelle anemometer measurements;
- Increased level of vibrations.

Single wind turbine (Section A.) and fleet-wide (Section B.) can be employed. A brief explanation is provided here on and the flowchart of the method is reported in Figure 1, which is discussed as well in the related paper [16].

A. Single Wind Turbine Methods

The single wind turbine methods can be employed on their own for characterizing the behavior of wind turbines affected by static errors, and hence for diagnosing, but can as well be employed as preliminary steps for dividing a fleet in potentially anomalous and likely healthy wind turbines.

A..1 Nacelle Anemometer Measurement Analysis

A devoted analysis has to be dedicated to the nacelle anemometer measurements, as they are evidently associated to the anemometer bias and, less evidently, to the static yaw error [17]. If there are two sensors sited at the opposite lateral ends of the nacelle, it is possible to analyze the relation between the so collected wind speed measurements v_1 and v_2. By posing a linear relation, as in Equation 1:

$$v_2 = kv_1, \tag{1}$$

it has been observed in [17] that the Least Squares estimate of k changes as the static yaw error changes and, in particular, if the two anemometers are of the same type, the expected baseline if $k \simeq 1$ if $\gamma \simeq 0$.

A..2 Operation Curve Analysis

Based on the above observations, it is more reliable to analyze the power and vibration measurements without putting them in relation with the nacelle anemometer measurements, because those are altered by the presence of the errors. Hence, the method of bins can be applied for characterizing various operation curves. Consider for example to group the measurements in rotor speed intervals of 0.5 rpm from cut-in rotor speed to rated. For each interval, average the corresponding set of power or absolute value of vibration measurements. The following might occur:

- Decreased power extracted for a certain rotational speed means under-performance (which can be associated to all the considered static errors);
- Increased level of vibrations for a certain rotational speed should be associated to blade pitch misalignment and static yaw error.

B. Fleet-Wide Approaches

The fleet-wide approach can be implemented as follows:

- Based on the single wind turbine methods, divide the fleet in anomalous and normal wind turbines;
- Set up a data-driven regression for, respectively, the power, the wind speed and the vibrations of the target wind turbines as a function of the corresponding quantities of the normal wind turbines;
- Analyze how the behavior of the residuals between measurements and model estimates changes in time.

III. RESULTS

A. Case Studies

Due to space constraints, it is prohibitive to comprehensively discuss all the cases indicated in Figure 1. Hence, two meaningful test cases have been selected, noticeably with the same wind turbine model (Senvion MM92). The former is a case of relative blade pitch misalignment, ascertained through laser measurements, and the latter is a static yaw error case, ascertained by upwind spinner anemometer measurements. For the former test case, the label "No PM" refers to laser measurements indicating a pitch misalignment below a threshold (0.3°). For the latter test case, data are available for the anomalous wind turbine and for three normal wind turbines belonging to the same fleet.

The measurements at disposal for both test cases are:

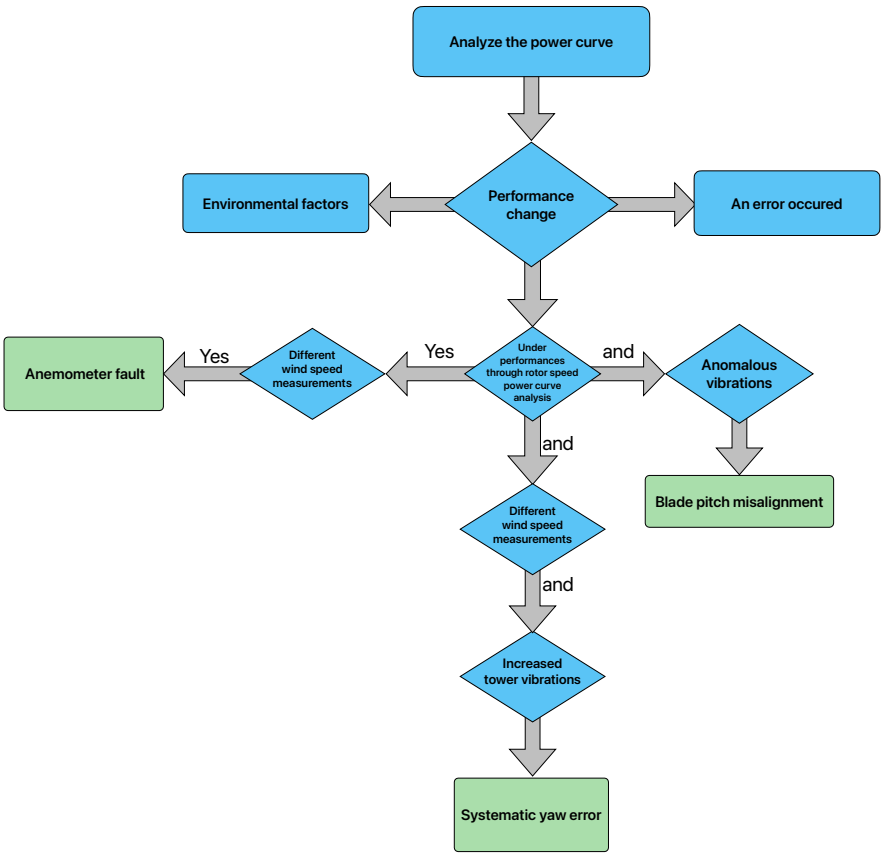


Fig. 1. Flow chart of the approach

- Wind speed sensor 1 (m/s);
- Wind speed sensor 2 (m/s);
- Grid Power (kW);
- Blade Pitch (°).

The measurement channel indicated as wind speed in the SCADA-collected data is the average of the measurements from the two physical sensors (cup anemometers mounted at the opposite lateral ends of the nacelle). For the former case study, the tower vibration measurements (mg) are available, while for the latter, there are the drivetrain vibrations (mg) which presumably are collected at a meaningful component of the drivetrain (unknown to the authors).

B. *Pitch Misalignment Analysis Through Single Wind Turbine Methods*

In Figure 2, the power curve is reported for the wind turbine of interest for various portions of the data set, associated to the presence of various values of relative pitch misalignment. Referring to the flowchart in Figure 1, Figure 2 provides a qualitative evidence of performance change, from which the further analyses follow for identifying what type of static error possibly affects the wind turbine.

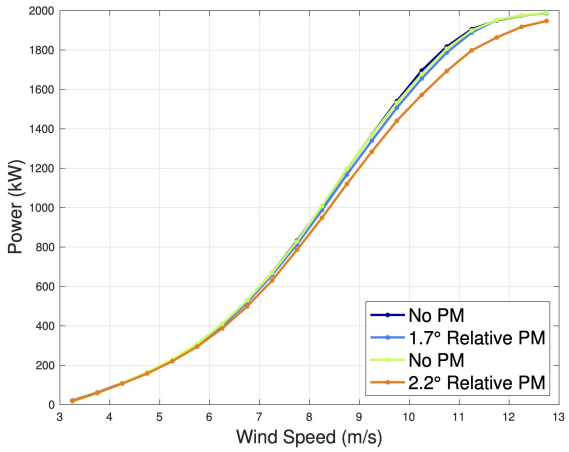


Fig. 2. The average power curve, for various values of relative pitch misalignment, ascertained through laser measurements.

Hence, from Figures 3 and 4, it arises that, for a certain rotational speed, the average absolute tower vibrations remarkably increase with the blade pitch misalignment and the extracted power decreases.

In Table 1, the Least Squares estimates of k (Equation 1) are reported. The variation as the blade pitch misalignment varies is much more limited with respect to the static yaw error case [17]. Hence, it can be argued that such type

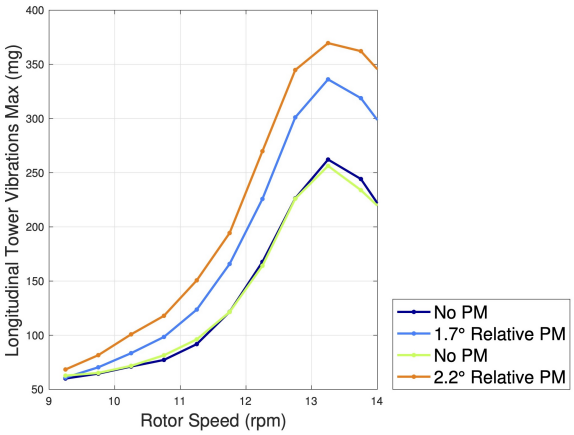


Fig. 3. The average curve of the tower vibrations as a function of the rotational speed, for various values of relative pitch misalignment, ascertained through laser measurements.

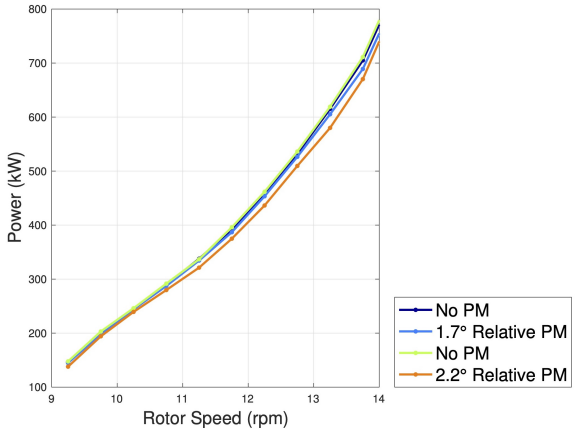


Fig. 4. The average curve of the extracted power as a function of the rotational speed, for various values of relative pitch misalignment, ascertained through laser measurements.

of error does not impact the nacelle anemometer measurements, thus confirming the hypothesis summarized in the flowchart in Figure 1.

Pitch Misalignment	**Estimated** k
No pitch misalignment (1)	0.971
1.7° pitch misalignment	0.977
No pitch misalignment (2)	0.975
2.2° pitch misalignment	0.976

Table 1. Estimates of k (Equation 1), for various values of relative pitch misalignment, ascertained through laser measurements.

C. Static Yaw Error Through Fleet-Wide Approaches

The data set at disposal regards one wind turbine (analyzed in [17]), for which the static yaw error passes from order of $-14°$ to less than $1°$ in correspondence of the vertical line in Figures 6, 7 and 8. Correspondingly, the power curve changes, as observable from Figure 5, which represents the starting point of the flow diagram in Figure 1.

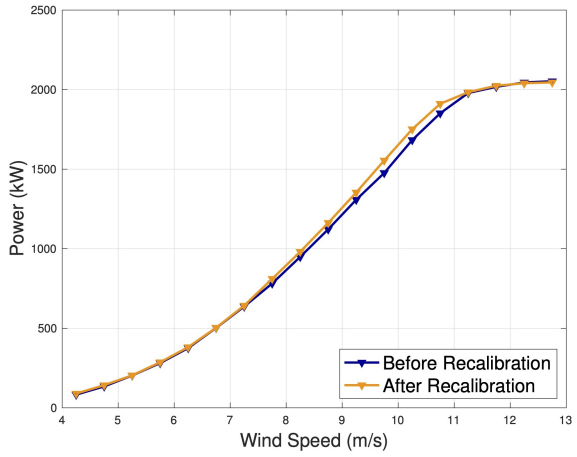


Fig. 5. The average power curve of wind turbine of interest, before and after the correction of $-14°$ of static yaw error.

According to the line of reasoning of Section B., the wind speed, the power and the drivetrain vibrations at the target wind turbine are modeled through a data-driven multivariate regression as a function of the corresponding quantities of the three normal wind turbines. It arises that, as the static yaw error decreases, the wind speed measurements change in relation to those of the nearby wind turbines, the power increases (error fixing, hence performance improvement) and the vibrations decrease as well (although drivetrain vibrations are less responsive than tower vibrations). Hence, the hypothesis formulated in Figure 1 about the manifestations of the various static errors is confirmed.

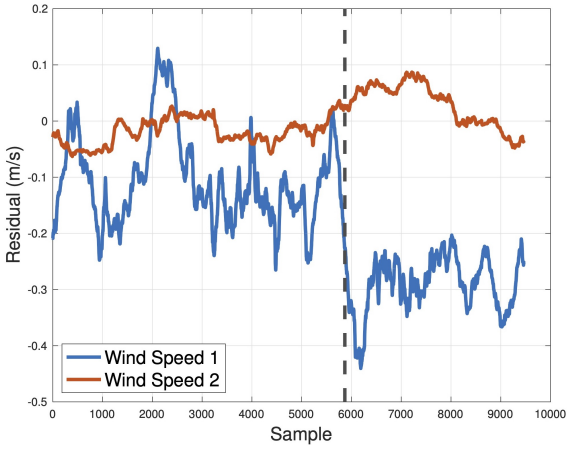


Fig. 6. The moving average of the residuals between measurements and model estimates for the two wind speed sensors. The vertical line indicates the date of correction of the yaw offsets.

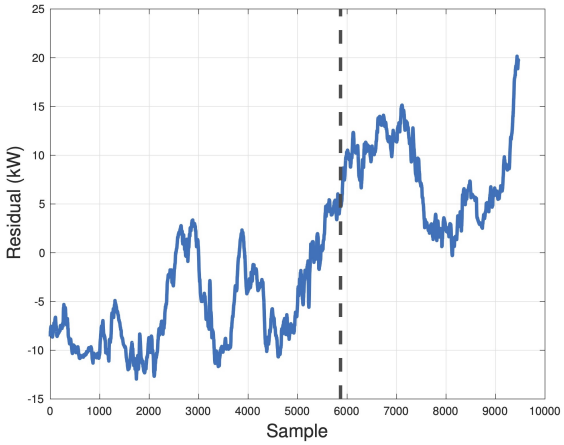


Fig. 7. The moving average of the residuals between measurements and model estimates for the power. The vertical line indicates the date of correction of the yaw offsets.

IV. CONCLUSIONS

A novel comprehensive framework for diagnosing static wind turbine errors through SCADA data analysis has been proposed. The collected case studies indicate that the proposed method is effective and allows distinguishing between the various types of errors related to the anemometer, the yaw and the pitch systems, by targeting appropriate phenomena like nacelle anemometer measurement shifts,

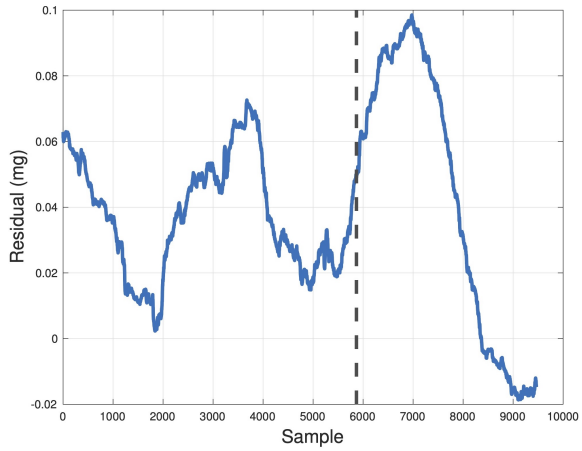


Fig. 8. The moving average of the residuals between measurements and model estimates for the drivetrain vibrations.

reduced power production and increased vibrations. Future developments of the present work regard incorporating further types of anomaly (like, for example, blade cracks) in the framework, especially for taking into account what happens in harsh environments [18].

ACKNOWLEDGMENTS

The authors acknowledge the ENGIE Italia company for sharing the data and supporting this research.

REFERENCES

[1] O. Summerfield-Ryan, S. Park, "The power of wind: The global wind energy industry's successes and failures". Ecological Economics, 210, 2023, 107841.

[2] C. Jung, "Recent Development and Future Perspective of Wind Power Generation". Energies, 17(21), 2024, 5391.

[3] K. Franke, J.F. Garcia, C. Kleinschmitt, F. Sensfuss, "Assessing worldwide future potentials of renewable electricity generation: installable capacity, full load hours and costs". Renewable Energy, 226, 2024, 120376.

[4] D. Choi, W. Shin, K. Ko, W. Rhee, "Static and dynamic yaw misalignments of wind turbines and machine learning based correction methods using lidar data". IEEE Transactions on Sustainable Energy, 10(2), 2018, 971-982.

[5] L. Gao, J. Hong, "Data-driven yaw misalignment correction for utility-scale wind turbines". Journal of Renewable and Sustainable Energy, 13(6), 2021.

[6] M. Saathoff, M. Rosemeier, T. Kleinselbeck, B. Rathmann, "Effect of individual blade pitch angle misalignment on the remaining useful life of wind turbines". Wind Energy Science, 6(5), 2021, 1079-1087.

[7] S. Milani, J. Leoni, S. Cacciola, A. Croce, M. Tanelli, "Automatic Detection and Intensity Classification of Pitch Misalignment of Wind Turbine Blades: a Learning-based Approach". In Journal of Physics: Conference Series, 2767(3), 2024, 032010

[8] S. Milani, J. Leoni, S. Cacciola, A. Croce, M. Tanelli, "A machine-learning-based approach for active monitoring of blade pitch misalignment in wind turbines". Wind Energy Science, 10(3), 2025, 497-510.

[9] D. Astolfi, F. De Caro, M. Pasetti, F. Castellani, A. Vaccaro, A. Flammini, "Assessing the effects of anemometer systematic errors on wind generators performance by data-driven techniques". Sustainable Energy, Grids and Networks, 39, 2024, 101417.

[10] B. Jing, Z. Qian, Y. Pei, L. Zhang, T. Yang, "Improving wind turbine efficiency through detection and calibration of yaw misalignment". Renewable Energy, 160, 2020, 1217-1227.

[11] D. Astolfi, "A study of the impact of pitch misalignment on wind turbine performance". Machines, 7(1), 2019, 8.

[12] D. Astolfi, F. Castellani, L. Terzi, "Mathematical methods for SCADA data mining of onshore wind farms: Performance evaluation and wake analysis". Wind Engineering, 40(1), 2016, 69-85.

[13] R. Pandit, J. Wang, "A comprehensive review on enhancing wind turbine applications with advanced SCADA data analytics and practical insights". IET Renewable Power Generation, 18(4), 2024, 722-742.

[14] F. Castellani, D. Astolfi, A. Garinei, S. Proietti, P. Sdringola, L. Terzi, U. Desideri, "How wind turbines alignment to wind direction affects efficiency? A case study through SCADA data mining". Energy Procedia, 75, 2015, 697-703.

[15] D. Astolfi, F. Castellani, F. Natili, "Wind turbine yaw control optimization and its impact on performance". Machines, 7(2), 2019, 41.

[16] D. Astolfi, S. Iuliano, A. Vasile, M. Pasetti, S. Dello Iacono, A. Vaccaro "Wind Turbine Static Errors Related to Yaw, Pitch or Anemometer Apparatus: Guidelines for the Diagnosis and Related Performance Assessment". Energies, 17(24), 2024, 6381.

[17] D. Astolfi, R. Pandit, A. Lombardi, L. Terzi, "Diagnosis of wind turbine systematic yaw error through nacelle anemometer measurement analysis". Sustainable Energy, Grids and Networks, 34, 2023, 101071.

[18] F. Castellani, D. Astolfi, E. Piccioni, L. Terzi, "Numerical and experimental methods for wake flow analysis in complex terrain". Journal of Physics: Conference Series, 625(1), 2015, 012042

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Partitioning Algorithm for Integrating Electric Vehicles into Residential Renewable Energy Communities

Manuela Minetti[1], Andrea Bonfiglio[1], Maria Martino[1], Renato Procopio[1]

[1] *University of Genoa, Genova*

***Abstract* – The rapid growth of Electric Vehicles (EVs) and Renewable Energy Communities (RECs) introduces new challenges in energy management and requires innovative solutions for enhancing local self-consumption. This paper introduces a Virtual Partitioning Algorithm (VPA) for the energy storage systems of EVs, aiming to improve EVs integration within RECs. The VPA improves the use of the available EV storage capacity to increase the economic benefits, and the shared energy of a REC equipped with residential photovoltaic systems or other renewable energy sources. The proposed approach is tested on a realistic case study involving a multi-family housing with a shared photovoltaic installation. Results demonstrate the effectiveness of the VPA in increasing overall gains and increasing both self-consumption and environmental sustainability.**

Keywords-Electric Vehicles, Renewable Energy Communities, Vehicle-to-Home.

I. INTRODUCTION

The spread of electric vehicles (EVs) emerged as a promising solution to reduce pollutant emissions and to ensure a better coupling between energy generation and consumption, mitigating Renewable Energy Sources (RESs) fluctuations [1-3] as well as providing economic advantages. EV can play an important role in ensuring grid stability, in reducing power outage episodes, in increasing global energy efficiency and providing additional grid services [4, 5]. EVs' batteries can be seen as a rolling accumulation system that allows peak reduction, absorbing or providing energy to the grid, as required in each moment. This service is called Vehicle to Grid (V2G) and its variants are Vehicle to Home (V2H) and Vehicle to Building (V2B), all of them being a two-way system that allows energy to be injected from the car into the grid and vice versa [6, 7]. In this context, EVs are identified as active contributors to the energy transition, not only associating EV with mere transportation, but by making them active players in the flexibilization of the electricity system to facilitate RES integration in the grid while still involving the final user.

Focusing on V2H application, the stored energy is at disposal of residential user to be used for final user power plant and building [8]. This is particularly convenient for prosumers with PV installations that can use the EV battery to store excess PV production during peak hours that will later feed domestic electricity demand [9-11]. As reported in [12], this strategy is beneficial both for the prosumer's self-consumption and for the resilience of electric distribution grid with high penetration of RES. In this context, a great portion of the research body on V2H has focused on developing optimization strategies to integrate EV into the building's energy system to reduce the electricity supply cost [13, 14].

Notably, recent studies have investigated the inclusion of EVs in Renewable Energy Communities (REC)[15, 16], as in [17] that proposes an optimization model with benefits descending from integrating external EV owners into an energy community. An optimal energy management strategy for a local energy community composed of DC microgrids with electric vehicle charging stations is presented [18]. The impact of EV charging stations on the electrical load of smart buildings, with a particular focus on energy communities is analysed in [19].

To the best of the authors' knowledge, the research body on V2H technology has primarily focused on the combination between EV ST and traditional renewable energy production systems, investigating energy-related aspects and developing management algorithms and optimized charging strategies for multi-energy systems in buildings, addressing primarily cost optimization for different daily scenarios. These studies present the EV battery as a single unit, without involving the end user in its management. Therefore, the following literature gaps have been highlighted:

- Lack of applicability of the algorithms for both single-family homes and residential REC.
- The insufficient attention to the users' willingness to participate in the services
- No possibility to dedicate an a priori portion of EV energy for V2H service.

Building on this, the aim of the article is to define an algorithm based on the creation of virtual partitions of an electric vehicle (EV) battery, allowing a user-defined

allocation of available energy for V2H services and support to the Renewable Energy Community (REC), while the remaining energy is fully reserved for user mobility. Specifically, the article applies the Virtual Partitioning Algorithm (VPA) strategy for EV battery management within a realistic residential energy community. In this context, it is highlighted that the effective implementation of virtual partitioning of EV batteries for V2H services within a REC depends on accurate measurements and reliable digital infrastructure, as detailed in [20, 21].

The paper is structured as follows: in Section II, the proposed VPA is generalized for a REC. Section III describes the study case and Section IV provides the simulation results of VPA application in a in a residential, multi-family housing with shared PV. Finally, some conclusive remarks are included in Section V.

II. VIRTUAL PARTITIONING ALGORITHM

The idea behind the VPA is to create a virtual partition of the physical storage capacity so that each portion can be dedicated to different purposes, specifically for mobility and to provide V2H service and support to the REC.

The EV State of Charge SOC_{EVT} at time t can be expressed as the ratio of the actual energy E_{EVT} to the total EV battery $E_{EVT,max}$, in percentage. SOC_{EVT} evolves according to the following equation:

$$SOC_{EVT}(t) = SOC_{EVT\_0} + \int_0^t \tilde{P}_{EV}(\tau)\, d\tau \tag{1}$$

Where P_{EV} is the power of the EV charging station in kW and SOC_{EVT_0} the initial SOC of the EV. From (1) on, power is expressed in the passive sign convention (positive power for charging and vice versa).

Within the VPA the percentage of the EV battery energy to dedicate to V2H services can be defined by the user and is indicated with the coefficient $\varepsilon_{V2H} \in [0,1]$. Therefore, the partial virtual capacities of the EV battery for V2H service $\tilde{E}_{V2H,\max}$ and actual mobility $\tilde{E}_{EV1,\max}$ can be expressed as:

$$\begin{cases} \tilde{E}_{V2H,\max} = \varepsilon_{V2H} E_{EVT,\max} \\ \tilde{E}_{EV1,\max} = (1-\varepsilon_{V2H}) E_{EVT,\max} \end{cases} \tag{2}$$

Hereafter, quantities marked with a tilde will represent virtual values, to differentiate them from physical counterparts. The two virtual SOC can, thus, be defined considering $\tilde{E}_{V2H}$ and $\tilde{E}_{EV1}$, referring to the virtual energy allocated to the two respective partitions expressed in kWh, and the maximum values $\tilde{E}_{V2H,max}$ and $\tilde{E}_{EV,max}$. The virtual power signals associated with the EV mobility $\tilde{P}_{EV1}$ and to the V2H service $\tilde{P}_{V2H}$ that charge or discharge each virtual partition must comply with the actual power of the physical charging station P_{EV} all expressed in kW:

$$P_{EV}(t) = \tilde{P}_{EV1}(t) + \tilde{P}_{V2H}(t) \tag{3}$$

The maximum power assigned to charge the mobility virtual partition ($\tilde{P}_{EV1,max}$) can be configured by the owner based on their charging needs, following the standard procedure used in state-of-the-art EV, where this is typically managed through the vehicle-to-human interface. The value must remain below the maximum power capacity of the charging infrastructure $P_{EV,max}$ to ensure sufficient margin for V2H services. The user's choice directly affects the power reserved to the V2H service. The virtual powers of the VPA are linked to their corresponding virtual SOC dynamics as follows:

$$\begin{cases} \tilde{SOC}_{EV1}(t) = \tilde{SOC}_{EV1\_0} + \int_0^t \tilde{P}_{EV1}(\tau)\, d\tau \\ \tilde{SOC}_{V2H}(t) = \tilde{SOC}_{\mathrm{V2H\_0}} + \int_0^t \tilde{P}_{V2H}(\tau)\, d\tau \end{cases} \tag{4}$$

with $\tilde{SOC}_{EV1\_0}$ and $\tilde{SOC}_{V2H\_0}$ indicating the initial SOC of the virtual partitions which are proportional to EV initial SOC SOC_{EVT_0}.

Typically, $\tilde{P}_{EV1}$ is assigned the maximum value defined by the user, as long as it aligns with the current SOC of the respective virtual partition, regarding this virtual partition, it only can charge (i.e. the power can only be positive or null), therefore:

$$\tilde{P}_{EV1}(t) = \begin{cases} 0 & if \;\; \tilde{SOC}_{EV1}(t) = 100\% \\ \tilde{P}_{EV1,\max} & elsewhere \end{cases} \tag{5}$$

the value of $\tilde{P}_{V2H}$ is subject to V2H needs. Let $\tilde{P}^*_{V2H}$ be the V2H partition power request, the actual power is linked to the relative SOC from the relations:

$$\tilde{P}_{V2H}(t) = \begin{cases} 0 & if \;\; (\tilde{P}^*_{V2H}(t) > 0 \; and \; \tilde{SOC}_{V2H}(t) = 100\%) \\ 0 & if \;\; (\tilde{P}^*_{V2H}(t) < 0 \; and \; \tilde{SOC}_{V2H}(t) = 0\%) \\ \tilde{P}^*_{V2H}(t) & elsewhere \end{cases} \tag{6}$$

Finally, the physical power of the EV, to which the two partitions are related must satisfy the CI bound:

$$\left|P_{EV}(t)\right| \le P_{EV,\max} \tag{7}$$

When violating (7) the V2H power request is curtailed so not to interfere with the EV charging need, which has priority between the two virtual partitions. Therefore, if the battery power limit is exceeded, the $\tilde{P}_{V2H}$ contribution is curtailed according to:

$$\begin{cases} \tilde{P}_{V2H}(t) = P_{EV,\max} - \tilde{P}_{EV1}(t) & if \;\; \tilde{P}^*_{V2H}(t) > 0 \\ \tilde{P}_{V2H}(t) = -P_{EV,\max} - \tilde{P}_{EV1}(t) & if \;\; \tilde{P}^*_{V2H}(t) < 0 \end{cases} \tag{8}$$

A. Virtual partitioning for a residential REC

This paper proposes the implementation of the VPA in the context of a REC consisting of a multi-user residential building with N apartments and a common PV system, as illustrated in the schematic of Fig. 1.

For this scenario it is considered that every residential unit has a EV, to which VPA is applied individually. The methodology remains applicable even when only a portion of the residents have EVs, with no impact on its general validity. In this case the VPA is used to maximize the REC revenue, generated by Shared Energy (SE), defined as the minimum between total energy production and total energy consumption, on an hourly basis [22].

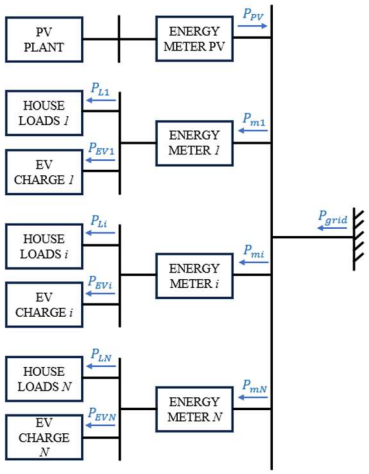


Fig. 1. Layout and power convention of the considered multi-family residential building.

The SE in the generic time interval Δt can therefore be calculated for the test case as:

$$SE(t) = \min\left(P_{PV}(t), \sum_{i=1}^{N} P_{mi}(t) \right) \cdot \Delta t \tag{9}$$

When deploying VPA for a REC, a centralized management system that collects N units' measurements performs the definition of the V2H set points. Therefore, relations (1)-(8) take subscript *i* (from 1 to N) for the *i-th* user.

For the *i-th* residential unit, the overall consumption profile is expressed by:

$$P_{Ltoti}(t) = P_{Li}(t) + \tilde{P}_{EV1i}(t) \tag{10}$$

With P_{Li} and $\tilde{P}_{EV1i}$ representing the load profile and the EV charging profile of the *i-th* residential unit, respectively. This represents the user power demand without the V2H partition, which is accounted for in the total measured power P_{mi}:

$$P_{mi}(t) = P_{Li}(t) + \tilde{P}_{EV1i}(t) + \tilde{P}_{V2Hi}(t) \tag{11}$$

The deviation from perfect SE is represented by ΔP_{SE}:

$$\Delta P_{SE}(t) = P_{PV}(t) - \sum_{i=1}^{N} P_{Ltoti}(t) \geq 0 \tag{12}$$

The following operating conditions (mutually exclusive) dictate the functionality of the V2H:

(i) When there is null PV production, SE is null, the set point $\tilde{P}^{*}_{V2Hi}$ can be calculated independently from SE maximization as:

$$P^{*}_{V2Hi}(t) = -P_{Li}(t) - \tilde{P}_{EV1i}(t) \tag{13}$$

(ii) With non null P_{PV} and positive (12) there is an amount of PV production that exceeds the load demand, which leads to SE not being maximized. Accordingly, to increase SE, the V2H should ensure that each EV absorbs power proportionally to its maximum available power allocated for the V2H service.

$$\tilde{P}^{*}_{V2Hi}(t) = \frac{P_{EV,\max i} - \tilde{P}_{EV1i}(t)}{\sum_{i=1}^{N}\left(P_{EV,\max i} - \tilde{P}_{EV1i}(t)\right)} \Delta P_{share}(t) \tag{14}$$

The power requested in (14) for V2H is positive, coherently with the sign convention adopted for the EV charging station, there is a need for power absorption.

(iii) With non null P_{PV} and negative (12), SE is maximized, but there is an excess of load demand. In this case, V2H service implies that all users should limit their power absorption, with the sum of all contributions not exceeding ΔP_{SE} to stick to maximum SE. Accordingly, the V2H contribution of the *i-th* unit is calculated as follows:

$$\tilde{P}^{*}_{V2Hi}(t) = \max\left(\frac{-P_{EV,\max} - \tilde{P}_{EV1}(t)}{\sum_{i=1}^{N}\left(-P_{EV,\max} - \tilde{P}_{EV1}(t)\right)} \Delta P_{share}, -P_{Li}(t) - \tilde{P}_{EV1i}(t) \right) \tag{15}$$

The power requested in (15) for V2H is negative, coherently with the sign convention adopted for the EV CI, there is a need of power production.

(iv) With non null P_{PV} and null (12) SE is maximized and V2H is not requested, setting:

$$\tilde{P}^{*}_{V2Hi}(t) = 0 \tag{16}$$

In all the four cases described, the V2H power request must be limited to (8) whenever (7) is not satisfied.

III. CASE STUDY

The VPA performance is tested in the context of V2H service provision within a multi-family residence organized as a REC. Specifically, the setup involves a three-unit apartment operating as a REC, powered by a 13.8 kW peak photovoltaic installation on the rooftop. Each apartment includes a 6 kW bi-directional charging station and an EV featuring a 45 kWh battery.

Fig. 1 reports loads and PV production for the analysed week. Both the load and PV production profiles are based on real-world data collected from three separate households located in northern Italy.

The connection of the three EV to the charging system varies depending on the daily routines of their respective users. Table 1 outlines the usage patterns, specifying for each vehicle and day the time intervals during which it is

unplugged and reconnected, along with the energy consumed for each trip. Note that energy consumed while the EVs are not connected to charging infrastructures is obtained from the virtual storage portion reserved for mobility purposes.

The analysis proposes a comparison with a reference scenario in which V2H functionality is not implemented, and the EVs are charged at the same power levels as those assumed in the V2H configuration.

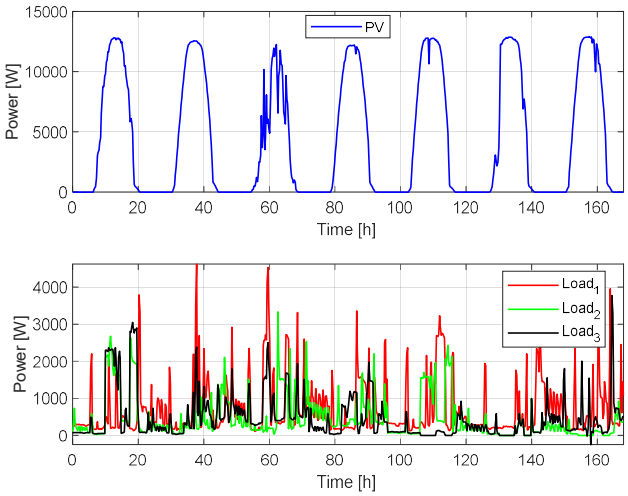


Fig. 1. Load and PV profiles.

TABLE I
EV TRANSIT SCHEDULING AND ENERGY CONSUMPTION

	EV_1	EV_2	EV_3
Day 1	07:45-13:30 13 kWh 18:00-19:15 3 kWh	12:15-13:15 3 kWh 17:15-18:15 3 kWh	07:45-14:45 13 kWh
Day 2	07:45-13:30 13 kWh 18:00-19:15 3 kWh	07:15-17:15 13 kWh	07:45-14:45 13 kWh
Day 3	07:45-13:30 13 kWh 18:00-19:15 3 kWh	07:15-17:15 13 kWh	12:15-13:15 2 kWh 17:15-18:15 2 kWh 20:15-22:15 10 kWh
Day 4	07:45-13:30 13 kWh 18:00-19:15 3 kWh	12:15-13:15 3 kWh 17:15-18:15 3 kWh	08:15-15:15 13 kWh
Day 5	07:45-13:30 13 kWh 18:00-19:15 3 kWh	07:15-17:15 13 kWh	08:15-15:15 13 kWh
Day 6	11:45-13:15 2 kWh 19:00-22:15 10 kWh	08:15-12:15 10 kWh 14:15-17:15 10 kWh	08:15-23:15 13 kWh
Day 7	-	13:15-20:15 13 kWh	14:15-23:15 13 kWh

IV. SIMULATION RESULTS

Fig. 2 displays the physical and virtual SOCs of the three EVs, considering the EV disconnection from the charging infrastructure. As one can observe, the V2H virtual partition significantly supports the exploitation of PV generation in the REC and is used to cover loads demand. Notably, the mobility virtual SOC (black lines) is not affected by VPA, and the user can increase the V2H service share if needed. Fig. 3 also shows the PV system power profile and the overall power profile P_m, which is the sum of the measured active powers P_{mi}.

The VPA enhances the exploitation of PV generation, especially when production exceeds the total REC load, as shown by the frequent overlap between the red and blue curves in the upper panel of Fig. 3. Additionally, it allows to provide service to user in the evening hours. This dual action boosts self-consumption within the REC and supports the REC objective of maximizing the self-consumption. Fig. 4 illustrates the power demand at the three EV charging stations. It can be observed that the VPA enables greater utilization of the charging infrastructure, which, at certain times during the week, operates at full capacity. In this configuration, the VPA serves both shared REC objectives and individual user needs.

In order to estimate the financial advantage introduced by the VPA in the REC context, the analysis assumes a compensation rate of 0.25 € per kWh for the energy shared within the community and a market price of 0.10 € per kWh for the sold energy. The REC self-consumption reaches 520.1 kWh, compared to 228.0 kWh without VPA. This results in a remuneration of 130.03€ with VPA, while it is only 57.00 € without it. Conversely, the energy sold by the REC is equal to 204.2 kWh with VPA and to 496.3 kWh without VPA, leading to earnings from energy sales of €20.42 and €49.63, respectively. Overall, the total gain amounts to €150.45 with VPA and €106.63 without it. Therefore, the implementation of the VPA, which increases self-consumed energy within the REC by reducing the amount fed into the distribution grid, results in higher economic returns for REC members.

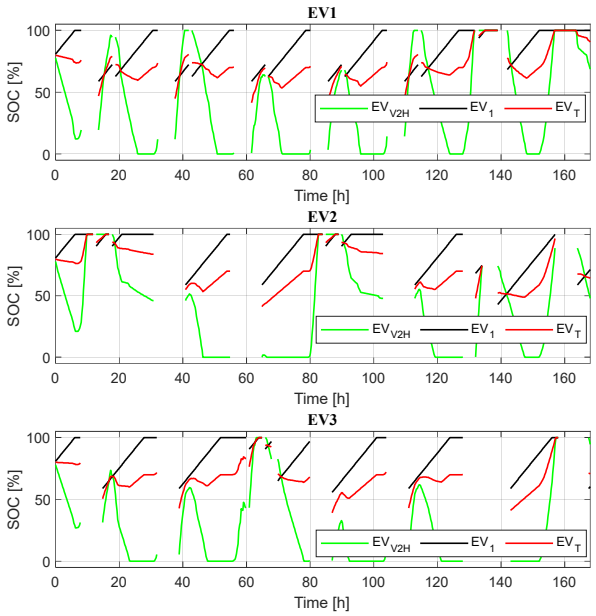


Fig. 2. SOC profiles of the mobility and virtual partitions.

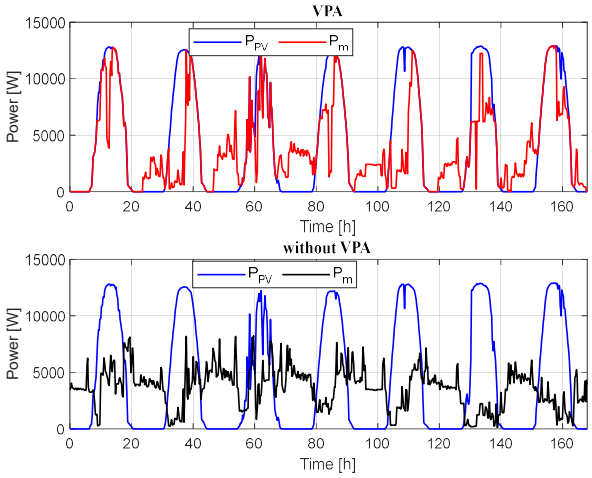


Fig. 3. P_m and P_{pv} power profiles.

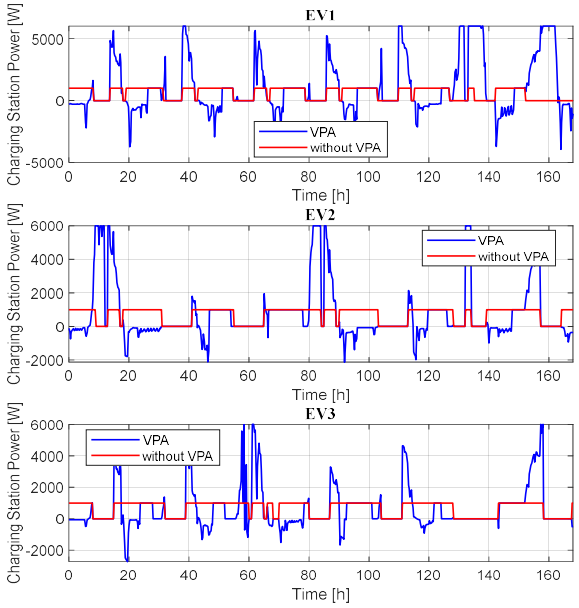


Fig. 4 Charging stations power profiles.

V. CONCLUSION

This paper introduces an innovative approach for the virtual partitioning of EV batteries, aimed at efficiently providing V2H services. The concept of virtual partitioning is first presented and then applied to a specific scenario involving multi-family housing with a shared PV system, forming a REC.The feasibility of the method was assessed through simulations using realistic REC, with actual data for residential load and PV production in a multi-family housing setup. The simulation results demonstrate the VPA effectiveness in providing both economic advantages and technical benefits to REC participants, such as enhanced exploitation of RESs and reduced power fluctuations at the grid connection point.

Future work will focus on developing optimization approach to determine the most efficient energy allocation for V2H services, considering the EV usage patterns.

REFERENCES

[1] J. A. Sanguesa, V. Torres-Sanz, P. Garrido, F. J. Martinez, and J. M. Marquez-Barja, "A review on electric vehicles: Technologies and challenges," *Smart Cities,* vol. 4, no. 1, pp. 372-404, 2021.

[2] S. Bruno *et al.*, "Fast frequency support through led street lighting in small non-synchronous power systems," *IEEE Transactions on Industry Applications,* vol. 59, no. 2, pp. 2277-2287, 2022.

[3] M. A. Amin, A. La Fata, R. Procopio, M. Invernizzi, M. Petronijevic, and I. R. Mitic, "Photovoltaic Power Nowcasting Using Decision-Trees Based Algorithms and Neural Networks," in *2024 11th International Conference on Electrical, Electronic and Computing Engineering (IcETRAN)*, 2024, pp. 1-6: IEEE.

[4] N. S. Pearre and H. Ribberink, "Review of research on V2X technologies, strategies, and operations," *Renewable Sustainable Energy Reviews,* vol. 105, pp. 61-70, 2019.

[5] A. Vaccaro, S. Iuliano, V. Galdi, V. Calderaro, and G. Graber, "Achieving consensus in self-organizing electric vehicles for implementing v2g-based ancillary services," *IEEE Access,* 2024.

[6] M. Fresia, M. de–Simon–Martin, and S. Bracco, "Energy Management System for Optimal Operation of a Prosumer with Renewable Generation and Electric Vehicle Fleet with Vehicle-to-Building," in *2024 IEEE International Conference on Environment and Electrical Engineering and 2024 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe)*, 2024, pp. 01-06: IEEE.

[7] G. Ellemund, M. Fresia, and S. Bracco, "Energy Management System for a Prosumer with Vehicle-to-Home: a Case Study in Italy," in *2024 International Conference on Smart Energy Systems and Technologies (SEST)*, 2024, pp. 1-6: IEEE.

[8] T. Higashitani, T. Ikegami, A. Uemichi, and A. Akisawa, "Evaluation of residential power supply by photovoltaics and electric vehicles," *Renewable Energy,* vol. 178, pp. 745-756, 2021.

[9] B. Zafar and S. A. B. Slama, "PV-EV integrated home energy management using vehicle-to-home (V2H) technology and household occupant behaviors," *Energy Strategy Reviews,* vol. 44, p. 101001, 2022.

[10] A. Bonfiglio, M. Minetti, and R. Procopio, "Vehicle-to-Home Service Via Electric Vehicle Energy Storage Virtual Partitioning," in *2024 IEEE/IAS 60th Industrial and Commercial Power Systems Technical Conference (I&CPS)*, 2024, pp. 1-6: IEEE.

[11] A. La Fata, M. A. Amin, M. Invernizzi, and R. Procopio, "Structurally Tuned LSTM Networks to Nowcast Photovoltaic Power Production," in *2024 IEEE International Conference on Environment and Electrical Engineering and 2024 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe)*, 2024, pp. 1-6: IEEE.

[12] C. Villante, S. Ranieri, F. Duronio, A. De Vita, and M. Anatone, "An energy-based assessment of expected benefits for V2H charging systems through a dedicated dynamic simulation and optimization tool," *World Electric Vehicle Journal,* vol. 13, no. 6, p. 99, 2022.

[13] L. Almeida, A. Soares, and P. Moura, "A systematic review of optimization approaches for the integration of electric vehicles in public buildings," *Energies,* vol. 16, no. 13, p. 5030, 2023.

[14] D. Borge-Diez, D. Icaza, E. Açıkkalp, and H. Amaris, "Combined vehicle to building (V2B) and vehicle to home (V2H) strategy to increase electric vehicle market share," *Energy,* vol. 237, p. 121608, 2021.

[15] S. Ahmed, A. Ali, and A. D'angola, "A review of renewable energy communities: concepts, scope, progress, challenges, and recommendations," *Sustainability,* vol. 16, no. 5, p. 1749, 2024.

[16] A. Bonfiglio, S. Bruno, M. Martino, M. Minetti, R. Procopio, and A. Velini, "Renewable energy communities virtual islanding: A novel service for smart distribution networks," in *2024 IEEE/IAS*

60th Industrial and Commercial Power Systems Technical Conference (I&CPS), 2024, pp. 1-8: IEEE.

[17] B. Velkovski, V. Z. Gjorgievski, B. Markovski, S. Cundeva, and N. Markovska, "A framework for shared EV charging in residential renewable energy communities," *Renewable Energy,* vol. 231, p. 120897, 2024.

[18] F. Marasciuolo, C. Orozco, M. Dicorato, A. Borghetti, and G. Forte, "Chance-constrained calculation of the reserve service provided by EV charging station clusters in energy communities," *IEEE Transactions on Industry Applications,* vol. 59, no. 4, pp. 4700-4709, 2023.

[19] E. Boasso, M. Manganelli, and L. Martirano, "Impact of ev charging on the electric load of smart buildings and energy communities," in *2021 IEEE Industry Applications Society Annual Meeting (IAS)*, 2021, pp. 1-7: IEEE.

[20] A. Mokarim, G. B. Gaggero, and M. Marchese, "Evaluation of the impact of cyber-attacks against electric vehicle charging stations in a low voltage distribution grid," in *2023 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, 2023, pp. 1-7: IEEE.

[21] G. B. Gaggero, M. Marchese, A. Moheddine, and F. Patrone, "A possible smart metering system evolution for rural and remote areas employing unmanned aerial vehicles and internet of things in smart grids," *Sensors,* vol. 21, no. 5, p. 1627, 2021.

[22] T. Bauwens *et al.*, "Conceptualizing community in energy systems: A systematic review of 183 definitions," *Renewable Sustainable Energy Reviews,* 2022.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Pathways to digitalisation at the Measurement Standards Laboratory of New Zealand

Blair Hall (ORCID: 0000-0002-4249-6863)[1] and Annette Koo (ORCID: 0000-0001-9390-345X)[1]

[1]*Measurement Standards Laboratory, Lower Hutt, New Zealand*

***Abstract* – Several strategic initiatives undertaken by the Measurement Standards Laboratory of New Zealand in response to the digital transformation of metrology are discussed. While the laboratory does not currently perceive an external demand for the digital delivery of its metrological services, it recognises that change will be required as global support for digital services becomes the norm. To support the longer-term vision of semantic representation of metrological data, a shift will be required from procedural, process-based knowledge capture towards abstract scientific forms. MSL has several initiatives in place to prepare for and facilitate this transition.**

I INTRODUCTION

The international metrology community recently announced its intention to digitally transform national and international measurement quality infrastructures (QIs) [1]. This will have far-reaching effects, as these QIs underpin metrologically traceable measurement services worldwide. All modern economies stand to benefit from increased efficiencies and new ways of using measurement data enabled by digital technologies. Achieving a successful outcome from digital transformation will require exemplary leadership and coordination at the highest levels, together with wholehearted commitment and engagement in digitalisation activities throughout the metrology community. To support this, the International Committee for Weights and Measures (CIPM) established a Forum on Metrology and Digitalization (Forum-MD) in 2019, to advise on the broader implications of global digital transformation in metrology [2, 3].

There are both international and national perspectives to this digital transformation in metrology. Internationally, a shared digital infrastructure is needed to support the exchange and unambiguous interpretation of measurement data. To enable increasingly autonomous machine use, this infrastructure must address and resolve ambiguities and inconsistencies that exist in current normative standards. Nationally, there is a need to adapt to a diversity in the rules and regulations imposed by local authoritative bodies. While each economy maintains sovereignty over its affairs, digitalisation will have to negotiate a greater degree of harmonisation across jurisdictions.

This conference paper describes the challenges and opportunities presented by digital transformation in metrology from the perspective of a small and geographically isolated economy—New Zealand. New Zealand is a member the international Metre Convention, which establishes a world-wide system of measurement units.[1] The country's national metrology institute (NMI)—the Measurement Standards Laboratory of New Zealand (MSL)—plays a key role in ensuring that the nation's measurement system is internationally recognised.

By act of parliament, the Chief Metrologist at MSL is the verifying authority of the country's measurement standards [4, 5]. The laboratory is signatory to the CIPM Mutual Recognition Arrangement (MRA) [6], which provides a framework for NMIs to demonstrate the performance of their measurement capabilities, thereby supporting the global dissemination of measurement services. MSL's services are also assessed against the international quality standard ISO/IEC 17025 for calibration and testing laboratories [7]. The assessment is organised by New Zealand's accreditation body, International Accreditation New Zealand (IANZ), which operates in accordance with internationally recognised standards and is a signatory to the International Laboratory Accreditation Cooperation (ILAC) Mutual Recognition Arrangement. The ILAC MRA ensures that IANZ accreditations are recognised by other ILAC MRA signatories and complements the CIPM MRA in supporting the international acceptance of measurement and calibration results. ILAC is one of nine signatories to a joint statement of intent with the CIPM, which seeks to involve stakeholders in international scientific and quality infrastructures in the on-going digital transformation in metrology.

In 2019, MSL allocated a small resource—about one half of a full-time equivalent (FTE)—to explore emerging metrology projects in digitalisation. Among the international developments that drew attention were two European-funded projects launched in 2018: SmartCom [9], and Metrology for the Factory of the Future [10]. Also, a few years earlier, an eloquent vision and strategy for dig-

[1]The Metre Convention (Convention du Mètre), also known as the Treaty of the Metre, is an international treaty signed in Paris on 20 May 1875. The treaty set up an institute for the purpose of coordinating international metrology and for coordinating the development of the metric system. In 1960, the system of units it had established was overhauled and relaunched as the "International System of Units" (SI). New Zealand became a full member of the Metre Convention in 1991.

ital transformation in metrology had been articulated in the United States by Kuster [11, 12]. MSL's strategic plan states that the laboratory will . . . *engage with international colleagues to share our expertise, as pre-normative standardisation takes shape, and [. . .] assess new technologies, as they emerge, for relevance to New Zealand*. So, it was appropriate to investigate these new digitalisation initiatives. Since 2019, the resource allocated to this area has grown to approximately 1 FTE.

MSL is not a large NMI by international standards, employing around 30 scientific and technical staff across a range of disciplines. Its expertise is world class, as reflected by membership in four of the CIPM Consultative Committees: those for mass (CCM), temperature (CCT), electricity and magnetism (CCEM), and photometry and radiometry (CCPR). MSL has approximately 180 Calibration and Measurement Capabilities (CMCs) published in the Key Comparison Database maintained by the International Bureau of Weights and Measures (BIPM). These CMCs span electricity and magnetism, photometry and radiometry, mass and related quantities, temperature and humidity, length, and time and frequency. Much of MSL's activity involves maintaining the metrological references and infrastructure that underpins these CMCs, as well as delivering measurement services to customers (approximately 260 measurement reports were issued to customers in 2024).

This paper discusses MSL's approach to digitalisation and digital transformation from a local perspective. While MSL has also been active internationally, and is a member of the Forum-MD, those activities are outside the scope of this paper. The local environment for digitalisation has presented challenges quite unlike those being addressed through international collaborations and research focused on the scientific principles of metrology.

The structure of the remainder of the paper is as follows. The next section outlines the environment in which MSL operates and how this has influenced the laboratory's approach to digitalisation. Section III describes two initiatives intended to better align the laboratory with the future demands of digital transformation. Section IV briefly discusses our ongoing strategy.

II THE ENVIRONMENT AT MSL

The environment in which MSL operates is shaped by the regional metrology organisation, Asia Pacific Metrology Programme (APMP), the national accreditation body (IANZ), and the needs of its customer base. In 2021, the APMP responded to the CIPM's growing interest in digitalisation by setting up a focus group to coordinate regional activities—the DXFG [13]. However, no digital policy decisions have yet been made by the APMP, nor has ILAC—IANZ's coordinating body—indicated any response to digitalisation. MSL surveys its customers regularly and has monitored for indications of emerging demand for digital services. To date, no such signals have been observed, and recent discussions with IANZ have confirmed our impression that there is no significant local 'pull' toward digitalisation. Accordingly, there is no immediate pressure on MSL to change its practices and offer digital services externally.

All MSL staff use digital systems for aspects of their duties. The laboratory has access to a corporate digital infrastructure that provides enterprise services such as email, networking, file storage, and office software. Specialised software can also be used as required; however, there is a reliance on commercial spreadsheet software for data analysis and word processing software for reporting. While these tools are familiar and well-established, they are not well suited to the rigorous quality control required by ISO/IEC 17025. The burden of developing and maintaining spreadsheets, and verifying their content, is recognised as a challenge across most sections of MSL, yet many metrologists are reluctant to move away from familiar tools. Although more appropriate digital systems may exist for certain activities, the perceived benefits are outweighed by individual preferences and practical considerations. Accordingly, there is little impetus within MSL to change practices.

Historical events may explain the reluctance at MSL to adopt new digital technology. MSL is embedded within a much larger organisation, where corporate priorities take precedence over the specific needs of the laboratory. In particular, the need for an NMI to maintain tight control of its technical records for decades does not sit comfortably with the much faster pace of change in commercial corporate IT systems. In the last ten years, MSL has endured three major corporate digital transformations, unrelated to metrology, which have been disruptive to the NMI. Given the impact of these changes, it is easy to imagine that even paper-based technical records may seem more effective and enduring than digital technologies.

Digitalisation, and ultimately digital transformation, are about much more than better record-keeping. The goal of digitalisation is to capture and codify information and metrological processes in forms compatible with digital systems—that is, in logical representations with well-defined semantics. Once such representations are available, metrology business models can be reconsidered, leading to a full digital transformation.

In a companion paper to this conference [14], we draw attention to challenges to digitalisation that arise within the metrology workforce. Many metrologists appear to favour empirical, experience-based forms of knowledge representation over abstract forms. Indeed, we suspect that the predilection for spreadsheet use at MSL is a reflection of this tendency—a preference for imperative rather than declarative software tools. If so, this may hinder efforts to

digitalise.

Imperative software development defines tasks step by step—an approach that aligns with procedure-based experimental work. In contrast, declarative programming focuses on outcomes. Digitalisation will produce representations in structured, formal terms that are amenable to automation. This is a declarative approach, which may feel unnatural to metrologists accustomed to hands-on and procedural thinking. Yet it is precisely this shift—from tacit procedural know-how to explicit, formalised representations—that digitalisation requires.

III MSL'S APPROACH TO DIGITALISATION

The digitalisation strategy at MSL aims to help metrologists transition their work practices toward the more abstract declarative programming paradigm. The focus is not on adopting new technology *per se*, but on introducing a shift in the way metrologists approach their science. Incremental change, rather than large and disruptive shifts, is proposed. While we regard widespread digitalisation as an inevitable consequence of global trends, we do not perceive a need for substantial change at MSL. Instead, we see an opportunity to better position the laboratory for informed decision-making when the requirements for digitalisation and alignment with new international digital infrastructures become clearer.

A. Calibration reports

One example of our softly-softly approach concerns the process of generating customer calibration reports. The preferred tool for this task at MSL is Microsoft Word™. The software is used in an imperative workflow, where the author has full control over the appearance and contents of the report. There are strict guidelines for report formatting, so careful checking is essential before issuing each report. To save time, metrologists often use a previous report as a template for a new one. While expedient, this can perpetuate errors unless they are detected by careful checking.

An alternative to what-you-see-is-what-you-get word processing, LaTeX-based tools separate the specification of a document's appearance from its structure and content.[2] This is a declarative approach to document production where authors describe the content and structure of a document but not its appearance.

Recent enhancements to the LaTeX core have added support for documents produced in the PDF/A-3 format [15], which is an international standard for long-term archiving and preservation of documents. PDF/A files are self-contained and can be accurately reproduced regardless of the software or hardware used in the future [16].

A LaTeX class specifying the appearance of MSL calibration reports has been developed.[3] The class is declared at the beginning of calibration report text files to configure the LaTeX system, allowing authors to use the prescribed environments and macro commands (see Fig. 1).

```
DocumentMetadata{
    pdfversion=1.7,
    pdfstandard=A-3b,
}
\documentclass[CIPM]{MSLCalCert}

 \reportDate{2018-02-19}

 \reportNumber{Electrical/2018/Sxxx}
 \fileRef{J01xxx}

 \title{A report on the calibration of an ???}

 \workerChecker{Blair Hall}[RF Standards]
 \workerChecker{Neil Swift}
 \chiefMetrologist{Annette Koo}
```

Fig. 1. This shows the first few lines in a calibration report file. `\documentclass` declares the MSL style, `MSLCalCert`, and selects a title-page option to include the CIPM MRA logo. Then various report details are declared, such as the report title and the staff involved in its preparation and validation.

LaTeX report production is straightforward, and on efficiency grounds alone, there is an incentive to transition to LaTeX-based report generation. Because source files are plain text, content can be entered manually or automatically. Automation offers the opportunity to streamline workflows and reduce the risk of manual data-handling errors.

However, LaTeX also opens up digital enhancements at almost no extra cost to the metrologist. For example, PDF files support a feature called XMP (Extensible Metadata Platform), which allows XML metadata to be added to the file. This metadata can be accessed by document management systems for records management. In the MSL class, information processed by LaTeX—such as the title, date, report number, and other details—is written into the document's XMP metadata. The XMP format is extensible, allowing additional *ad hoc* information to be added. We plan to explore this capability further as a way of handling metadata required by ISO 17025.

Another advantage of the PDF/A-3 format is its ability to embed other digital files—such as calibration data—directly within the final report. These embedded files may use various formats, including spreadsheets or *ad hoc* formats specific to particular application software. We an-

[2] Maintained by the open-science community, these tools are freely available on various platforms and are renowned for their stability and long-term compatibility with legacy documents—an important feature for metrology. They are particularly well-suited for scientific reporting, with support for mathematical notation, units of measurement, and numerical data.

[3] A tutorial on the development of similar LaTeX classes is available online `https://apmp-dxfg.github.io/dxfg-pdfa-tut`.

ticipate that PDF reports containing embedded data will become a preferred pathway for MSL customers to transition from human-readable reports to machine-readable ones. Once again, the declarative approach greatly simplifies adoption of this enhancement by our metrologists: embedding an external file requires a single line in the LaTeX source.

B. *Measurement modelling*

A second MSL initiative encourages staff to develop appropriate scientific models of measuring procedures. Here again, the intention is to favour a declarative outlook over an imperative mindset. The connection between this initiative and digitalisation is indirect, as we shall explain.

When measurement results are reported outside MSL, the procedures followed must be documented and carefully controlled to comply with the requirements of ISO 17025. Such procedures can be, and often are, described as a sequence of actions—much like a recipe for the measurement in question. However, this procedural form of description is not easily reconciled with a scientific account of the same measurement.

ISO 17025 requires the measurement uncertainty associated with a result to be evaluated. This necessitates a formal scientific measurement model that accounts for all significant influence quantities—factors that cause a result to deviate from the measurand. A measurement model is inherently declarative, expressed in mathematical notation where the terms represent abstract physical quantities. Some staff prefer to elide the step of preparing such a model, opting instead to assert a relationship for evaluating the uncertainty. In some cases, this calculation is embedded in a spreadsheet, and it is claimed that "the spreadsheet *is* the model"!

To encourage better practice, MSL now offers tutorials on measurement modelling to staff who joined the laboratory in the last ten years. The syllabus introduces a straightforward and systematic modelling approach, previously published [17], that draws on familiar and fundamental ideas about quantities. This approach leads to a detailed description of how quantities in a measurement—some of which are undesirable influences—contribute to a result. Vague notions of *uncertainty* are avoided until a clear interpretation can be made in relation to residual measurement errors.

Models developed using the MSL approach are immediately compatible with special software for uncertainty evaluation [18]. In this approach, terms appearing in a model correspond directly to instances of an abstract data type called an uncertain number [19]. Mathematical expressions in the model can be replicated by identical source code calculations involving uncertain numbers. This design effectively separates the evaluation of measurement uncertainty into two tasks: first, the formulation of a model, and second, its evaluation. The software available at MSL takes care of evaluation. As a result, metrologists adopting this approach can spend more time on understanding their measurements and identifying sources of error, and less time on the mechanics of uncertainty propagation or the validation of associated software. The approach enables rigorous handling of complicated measurement set-ups and provides more accurate uncertainty evaluations than would otherwise be possible [20, 21, 22].

The self-discipline of formally modelling measurements leads to robust and streamlined data processing. When explicit models are included in measurement procedure documentation, they serve as independent tools for reasoning about the measurement process and the effects of influencing factors. The direct coupling between a model and uncertainty evaluation can be used to enhance the quality of measurement procedures and to support software validation. Furthermore, once a model is established, measurement uncertainty can be evaluated using different computational approaches, if desired, providing additional flexibility.

IV DISCUSSION

MSL's strategy is to encourage different ways of working with digital technology, some examples of which have been presented above. Our focus is on deep-seated issues rather than short-term technological trends. The strategy has several objectives: reducing exposure to disruptive external change, improving internal efficiency, and promoting approaches to metrological work that are suited to abstract, declarative expression. The transitions involved must proceed at a pace determined by metrologists themselves—they cannot be rushed.

Technologies, such as LaTeX tools, offer long-term stability. By facilitating the use of LaTeX, we are adopting a well-maintained and enduring digital system—one that can be relied upon for decades, if necessary. Producing LaTeX documents is a declarative way of working, helping metrologists to become accustomed to this way of thinking. At the same time, the PDF/A-3 format offers opportunities to move beyond conventional human-readable certificates. Its XMP metadata feature can be extended to hold metrological data, and its ability to encapsulate diverse file types is potentially very useful. In this way, we aim to be well-positioned to respond to external developments in the international quality infrastructure, as well as to meet customer expectations as demand grows for more direct delivery of digital data.

Our strategy recognises the fundamental role of semantics in digitalisation. Clear and consistent semantic definitions for metrological concepts will be essential—without these, semantic interoperability will not be achieved. Metrologists will increasingly need to engage with semantic digital technologies as the digital transformation of

quality infrastructure takes shape. So, we are encouraging a shift from procedural, process-based knowledge capture toward abstract scientific representations—a shift reflected in our support for modelling.

More effort is needed in the area of semantics, which we plan to explore further. The terminology used within each metrology discipline must be systematically organised. Identifying and rigorously managing a domain vocabulary is a proven strategy in software development [23]. A well-defined domain language enhances collaboration by reducing jargon and minimising misunderstandings of discipline-specific terms. Adopting controlled vocabularies will improve the semantic alignment between technical documentation and quality-related processes such as software development and validation. We expect that careful attention to harmonised terminology will enhance the quality of work in the laboratory.

ACKNOWLEDGEMENTS

This work was funded by the New Zealand government. The authors are grateful to Peter Saunders for careful review of the manuscript.

REFERENCES

[1] General Conference on Weights and Measures (CGPM), "Resolution 2 – on the global digital transformation and the international system of units," 27th CGPM, November, 2022.

[2] CIPM Forum on Metrology and Digitalization, https://www.bipm.org/en/committees/fo/forum-md, (accessed: 4 March 2025).

[3] S. J. Chalk, D. N. Coppa, F. Flamenco, A. Forbes, B. D. Hall, R. J. Hanisch, K. Hosaka, D. Hutzschenreuter, J. Park, and R. M. White, "International development of the SI in FAIR digital data," *Measurement: Sensors*, vol. 18, p. 100293, 2021.

[4] Measurement Standards Act 1992 http://www.legislation.govt.nz/act/public/1992/0052/latest/whole.html (accessed: 15 April 2025).

[5] Measurement Standards Regulations 2019 http://www.legislation.govt.nz/regulation/public/2019/0091/latest/whole.html (accessed: 15 April 2025).

[6] CIPM, "Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes," http://www.bipm.org/en/cipm-mra/, October 1999. (accessed 27 April 2025)

[7] ISO and IEC, *ISO/IEC 17025:2017 General requirements for the competence of testing and calibration laboratories*, International Organisation mfor Standardisation, 2017.

[8] CIPM, "Joint Statement of Intent On the digital transformation in the international scientific and quality infrastructure," https://www.bipm.org/documents/20126/42177518/Joint-Statement-Digital-Transformation.pdf (accessed: 19 April 2025).

[9] "Communication and validation of smart data in IoT-networks (SmartCom)," https://www.ptb.de/empir2018/smartcom/project/, 2021, (accessed: 15 April 2025).

[10] "Metrology for the Factory of the Future (Met4FoF)," https://researchoutreach.org/articles/metrology-factory-future/, 2021, (accessed: 15 April 2025).

[11] M. Kuster, "Toward a measurement information infrastructure," in *Proc. MSA 11th Bienn. Australas. Meas. Conf.*, 2015, pp. 13–16.

[12] M. Kuster, "Toward a measurement information infrastructure: Smart accreditation scopes," *Metrologist*, 2015.

[13] APMP-DXFG, https://apmp-dxfg.github.io/ (accessed 19 April 2025).

[14] B. D. Hall, "Communities of practice in metrology," IMEKO International Conference on Metrology and Digital Transformation, *submitted* 2025.

[15] F. Mittelbach, U. Fischer, D. Carlisle, and J. Wright, "Automatically producing accessible and reusable PDFs with LATEX," in *Proc. ACM Symp. Doc. Eng.*, 2024, pp. 1–4. 2024.

[16] T. Zellmann and H.-J. Hübner, "PDF/A-3, the newest part of the ISO standard 19005," in *Archiving Conference*, vol. 11. Society for Imaging Science and Technology, 2014, pp. 123–125.

[17] B. D. Hall and D. R. White, *An introduction to measurement uncertainty*. Measurement Standards Laboratory of New Zealand, 2017.

[18] B. D. Hall, "The GUM Tree Calculator: A python package for measurement modelling and data processing with automatic evaluation of uncertainty," *Metrology*, vol. 2, no. 1, pp. 128–149, 2022.

[19] B. D. Hall, "Computing with uncertain numbers," *Metrologia*, vol. 43, pp. L56–L61, 2006.

[20] E. Molloy, P. Saunders, and A. Koo, "Effects of rotation errors on goniometric measurements," *Metrologia*, vol. 59, no. 2, p. 025002, 2022. [Online]. Available: http://iopscience.iop.org/article/10.1088/1681-7575/ac438e

[21] E. Molloy, "Metrology of scattering distributions," Ph.D. dissertation, Te Herenga Waka-Victoria University of Wellington, 2023.

[22] E. Molloy and P. Saunders, "Uncertainty Analysis of MSL's CCT-K7.2021 Key Comparison Measurements Taking Account of Correlations," *Metrologia*, submitted.

[23] E. Evans, *Domain-driven design: tackling complexity in the heart of software*. Addison-Wesley Professional, 2004.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

"SASO Uncertainty Machine" - Advanced Pythonic ML Algorithm for Estimating Uncertainty in General Calibration Services at Saudi Standards, Metrology, and Quality Organization-SASO-KSA

Saad A. Bin Qoud [1,], Fahad A. AlMuhlaki[1], Rayan A. AlYousefi [1]I. AlFaleh[1], N. Qahtani, Lama M. AlBugami, Rawan A. AlMutairi, Khaled G. AlGhizzi, Raed H. AlShabatat, A. El-Matarawey[2]

1. NMCC-SASO- Saudi Arabia
2. National Institute of Standards (NIS-Egypt)

***Abstract* – The missions of the National Metrology Calibration Centre (NMCC) at Saudi Standards, Metrology, and Quality Organization (SASO) as a national metrology institute are to maintain, disseminate, develop and realize the International SI units. One of the introduced services by the laboratories is the routine calibration that covers nearly all fields and sectors such as industrial, medical, environmental, etc.... This work aimed to digitized a web-based software application to evaluate the measurement uncertainty associated with a measurement's quantity for all kind of model functions (linear, non-linear, polynomial, logarithmic, etc..) that related to the calibration, testing, accreditation, verification and validation services. Pythonic machine learning algorithm was built based on the requirements of JGCM-100 to give the normal user a full information report about the experiment and measurement uncertainty during three main steps. The validation of the automated algorithm is carried out on five examples that are reported by JGCM standard with completely numeric matching. The article will provide open-source step by step algorithm as a part of digital transformation trend around the world.**

***Keywords*:** SOSO, NMCC, Digitalization, Uncertainty, JGCM-100, Calibration, Testing, AI.

I. INTRODUCTION

The term "uncertainty" signifies doubt, "uncertainty of measurement" denotes doubt regarding the validity of a measurement result [1]. Due to the absence of distinct terms for both the general concept of uncertainty and the specific quantities, such as standard deviation, that quantitatively represent it . Measurement uncertainty typically involves multiple factors. Some of these factors can be assessed through the statistical analysis of a series of measurement results and are defined by experimental standard deviations. Other factors, which can also be defined by standard deviations, are evaluated using assumed probability distributions based on experience or other information. Uncertainty in general consists of two main parts [1], type A which is related to the statistical analysis of series of observations, type B is related to sources that affect the measurement process. Combined standard uncertainty is the standard uncertainty of a measurement result that is derived from the values of several other quantities. It is calculated as the positive square root of the sum of terms, with each term representing the variances or covariances of these other quantities. These terms are weighted according to how the measurement result changes with variations in these quantities. Finally, expanded uncertainty is a measure that defines a range around the result of a measurement, which is expected to cover a significant portion of the distribution. In order to estimate the uncertainty, it should be following the process (figure 1) and determine the sources of uncertainty Table 1. The Type A and Type B classification [1, 2] serves to differentiate between two distinct methods of assessing uncertainty components. This distinction does not imply any inherent differences in the nature of the components derived from each evaluation method. Both methods rely on probability distributions, and the resulting uncertainty components are quantified through variances or standard deviations [1, 2].

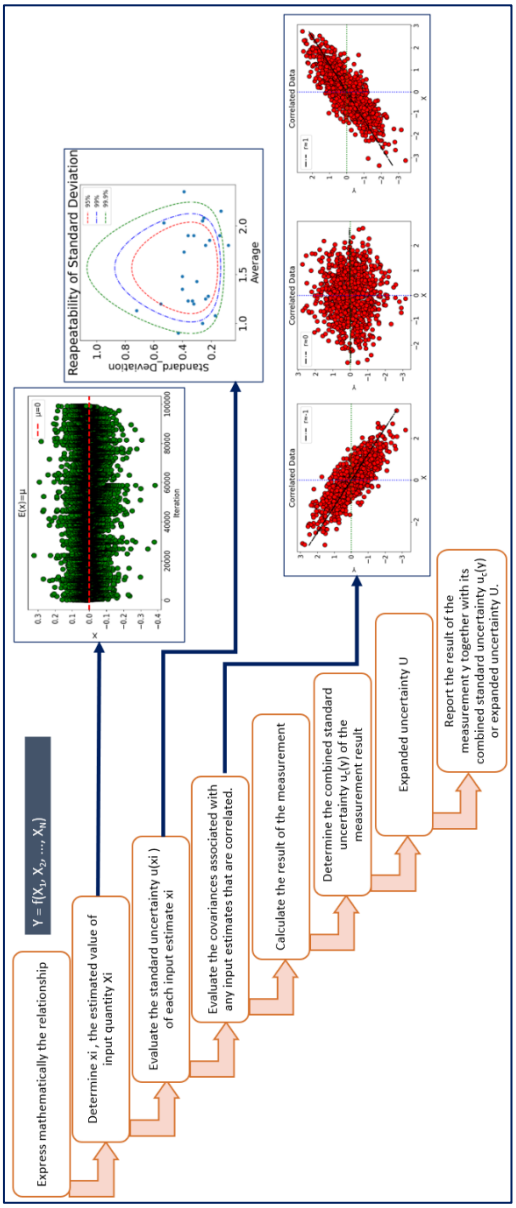


Fig 1. Summary of procedure for evaluating and expressing uncertainty

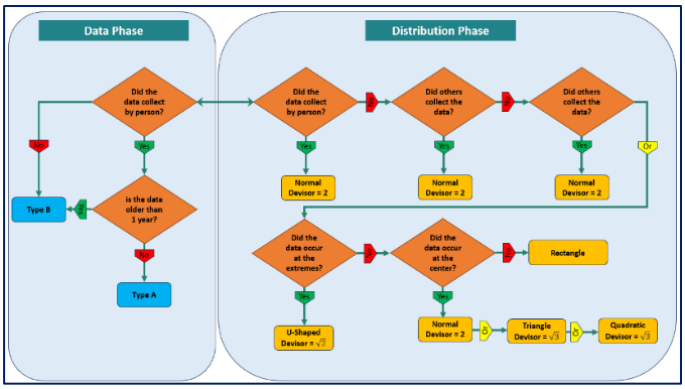


Fig 2. Classification of data and distributions

Table 1. The uncertainty sources that may including (but not limited to) experiment

Method	Equipment/ Reference Material	Personnel	Environment
• Model function • Preparation • Testing • Calibration • Analysis • Calculations • Conversions • Rounding	• Repeatability • Stability • Bias • Drift • Resolution • Certificates • Stability • Linearity • Hysteresis • Homogeneity • Technical Documents • Manufacture Manuals • Specifications • Accuracy • Precision	• Reproducibility • Reading analogue instruments	• Temperature • Humidity • Pressure • Gravity • Vibration • Noise • EMI • Illumination

In order to classify the sources into Type A or B which is driven from **C**lassification **A**lgorithm in **M**achine **L**earning (CAML), we can summarize the concepts as shown in figure 2. Some time is useful to investigate the behind the measurements error, it can be classified into tow main sources, the systematic and random error (figure 3.a). Drift and bias also considered errors (figure 3.b), both are related to the used equipment, with variable and constant behavior respectively [1, 2, 3, 4].

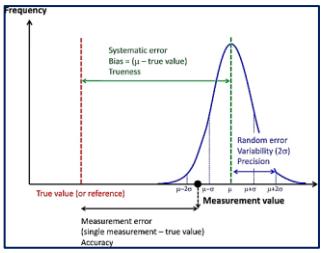


Fig 3.a. Systematic and random error concept

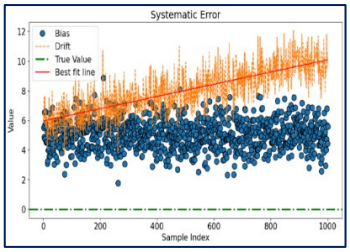


Fig 3.b. Drift and bias concept

II. SASO UNCERTAINTY MACHINE (SASO-UM)

(SASO-UM) is a web-based software tool designed to assess, evaluate and estimate the measurement uncertainty related to an output quantity as specified by a measurement model function. The code architecture is based on python programming code that used several libraries to complete ethe technical report to the users. The process starts with input the results and end with reporting the results, full techniques will be discussed in the following sections. The source code used under python and Jupiter note book environment [5]. Jupyter lab [5, 6] is used as an interactive environment with python, different sets of libraries used for building the algorithm. NumPy [7], which is short for Numerical Python, is a robust Python library that is widely utilized in scientific computing. It offers assistance for vast, multi-dimensional arrays and matrices, as well as an

extensive selection of mathematical operations such as algebra and matrices that can be performed on these structures. Pandas [8] is a data manipulation tool that operates at a high level and is constructed on top of NumPy. It offers user-friendly data structures and analysis tools for managing tabular data in Python. Pandas has two primary classes, namely Series and Data Frame, for managing data. Additionally, it has robust capabilities for indexing, filtering, selecting, and transforming subsets of data. Matplotlib [9] is a data visualization library built on top of NumPy. It provides a variety of plotting functions to create complex visualizations and graphs in Python. Plotly [10] is a data visualization library that provides interactive and dynamic visualizations. It can be used with NumPy to create informative graphs in Python. SciPy (Science Python) [11] is a library that builds on NumPy to provide a wide range of scientific computing tools. It includes modules for optimization, integration, linear algebra, and more. SciPy can perform complex mathematical operations with ease and efficiency. Seaborn [12] is a data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics. Statsmodels [13] is a Python module that provides classes and functions for the estimation of many different statistical models, as well as for conducting statistical tests, and statistical data exploration. An extensive list of result statistics is available for each estimator. Scikit-learn (Sklearn) [14] is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including *classification*, *regression*, clustering and dimensionality reduction via a consistence interface in Python.

III. MATHEMATICAL MODEL

1. MODEL FUNCTION

Consider an estimate of the measurand Y, denoted by y and an estimate of the input X_N, denoted by x_N

$$Y = f(X_1, X_2, \dots\dots\dots X_N) \quad 1$$

$$y = f(x_1, x_2, \dots\dots\dots x_N) \quad 2$$

The arithmetic mean or average

$$y = \bar{Y} = \frac{1}{n}\sum_{k=1}^{n} Y_k = \frac{1}{n}\sum_{k=1}^{n} f(X_{1,k}, X_{2,k}, \dots, X_{N,k}) \quad 3$$

The estimate of the expectation or expected value μ_x of a quantity x for n independent observations x_k

$$\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k \quad 4$$

The experimental variance of the observations, which estimates the variance σ^2 of the probability distribution of x.

$$s^2(x_k) = \frac{1}{n-1}\sum_{j=1}^{n} (x_{j-}\bar{x})^2 \quad 5$$

The experimental variance of the mean is the best estimate of $\sigma^2(\bar{q}) = \sigma^2/n$

$$s^2(\bar{x}) = \frac{s^2(x_k)}{n} = \frac{1}{n(n-1)}\sum_{j=1}^{n} (x_{j-}\bar{x})^2 \quad 6$$

The standard uncertainty

$$u(\bar{x}) = s(\bar{x}) = \sqrt{s^2(\bar{x})} = \frac{s(x_k)}{\sqrt{n}} \quad 7$$

2. COMBINED STANDARD UNCERTAINTY

Combined standard uncertainty of the estimate y is denoted by $u_c(y)$, the combined standard uncertainty $u_c(y)$ is the positive square root of the combined variance $u_c^2(y)$

3. UNCORRELATED INPUT QUANTITIES

Combined uncertainty In term uncorrelation

$$u_c^2(y) = \sum_{i=1}^{N} \left(\frac{\partial f}{\partial x_i}\right)^2 u^2(x_i) \quad 8$$

The partial derivatives $\frac{\partial f}{\partial x_i}$ are called sensitivity coefficients

$$c_i = \frac{\partial f}{\partial x_i} = \frac{\partial f}{\partial x_i}\Big|_{x_1, x_2, \dots, x_N}, \; u_c^2(y) = \sum_{i=1}^{N} [c_i u(x_i)]^2 \quad 9$$

4. CORRELATED INPUT QUANTITIES

$$u_c^2(y) = \sum_{i=1}^{N} c_i^2 u^2(x_i) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} c_i c_j COV(x_i, x_j) \quad 10$$

The relation between correlation and covariance is

$$\rho(x_i, x_j) = \frac{COV(x_i, x_j)}{u(x_i)u(x_j)} = \frac{COV(x_i, x_j)}{\sigma_{x_i}\sigma_{x_j}} \quad 11$$

The covariance between components

$$COV(x_i, x_j) = \frac{1}{n(n-1)}\sum_{k=1}^{n}(x_{ik} - \mu_{x_i})(x_{jk} - \mu_{x_j}) \quad 12$$

Combined uncertainty in term correlation

$$u_c^2(y) = \sum_{i=1}^{N} c_i^2 u^2(x_i) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} c_i c_j u(x_i) u(x_j) \boldsymbol{\rho}(x_i, x_j) \quad 13$$

If the model function in form $Y = cX_1^{P_1}X_2^{P_2} \dots\dots X_N^{P_N}$, the equation 13 can be expressed as $\left[\frac{u_c(y)}{y}\right]^2 = \sum_{i=1}^{N}\left[P_i \frac{u(x_i)}{x_i}\right]^2$

5. EFFECTIVE DEGREE OF FREEDOM

A t-distribution (eq. 14) with an effective degrees of freedom v_{eff} obtained from the Welch-Satterthwaite formula, $v_{eff}\ eq(15)$.

$$t = \frac{(\bar{x} - \mu)}{s(x)} = \frac{(\bar{x} - \mu)}{u(x)} \quad 14$$

$$v_{eff} = \frac{u_c^4(y)}{\sum_{i=1}^{N}\frac{u_i^4(i)}{v_i}} \quad 15$$

6. COVERAGE FACTOR AND LEVELS OF CONFIDENCE

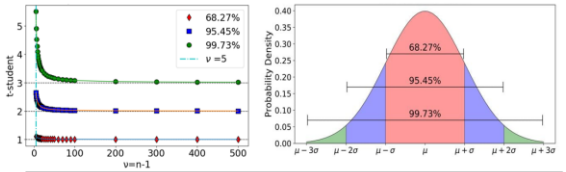
The relation between t-student and coverage factor k is

$$As\ v \to \infty \cdots t_p \approx \left(\sqrt{1 + \frac{2}{v}}\right) k_p \quad 16$$

In order to determind the specific value of coverage factor, SASO-UM able to detect the distribution of data if it was manualy entered or satisfied by user.All possiple distribution are introduced to SASO-UM algorithm but we will discuss in details the most common distributions in the uncertainty in general.

7. NORMAL DISTRIBUTION

For a quantity x is cosider a normal distribution with expectation μ_x and standard deviation σ, the value of k_p that produces an interval $\mu_x \pm k_p\sigma$ [1] is shown in figure 4.

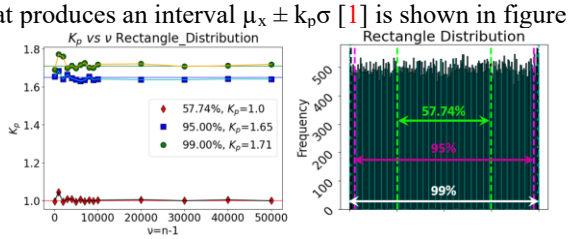


Fraction p in percent							
ν	50	68.27	90	95	95.45	99	99.73
∞	0.67	1.000	1.645	1.960	2.000	2.576	3.000

Figure 4. Confidence intervals and levels in normal distribution.

8. RECTANGLE DISTRIBUTION

If a quantity x is defined by a rectangle distribution, the expectation μ_x and standard deviation σ, the value of k_p that produces an interval $\mu_x \pm k_p\sigma$ [1] is shown in figure 5.



Fraction p in percent						
ν	50	57.74	90	95	99	100
∞	0.864156	1.000	1.555955	1.643018	1.71129	$1.73 \approx \sqrt{3}$

Figure 5. Confidence intervals and levels in rectangle distribution.

Recall the estimator and expectation of moment about the mean, the mean can be estimated as

$$\mu = E[x] = \int_a^b xf(x)dx. \qquad 17$$

Where a and b is the limits of rectanglar distribution and $f(x)$ is its function, $f(x) = \frac{1}{b-a}$.

$$\mu = \int_a^b x\frac{1}{b-a}dx = \frac{1}{b-a}\frac{x^2}{2}\Big|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2}. \qquad 18$$

The variance is estimated as $\sigma^2 = E[x^2] - \mu^2$

$$E[x^2] = \int_a^b x^2\frac{1}{b-a}dx = \frac{1}{b-a}\frac{x^3}{3}\Big|_a^b = \frac{b^3-a^3}{3(b-a)} == \frac{b^2+ab+a^2}{3}$$

$$\sigma^2 = E(X^2) - \mu^2 = \frac{(b-a)^2}{12} \qquad 19$$

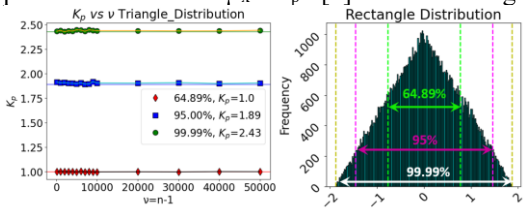
The standard deviation is estimated as

$$\sigma = \sqrt{\sigma^2} = \frac{b-a}{2\sqrt{3}} \qquad 20$$

For symetrical distrtibution, $a = b, b-a = 2a = 2b$ and the standerd deviation becomes $\sigma = \frac{a}{\sqrt{3}}$

9. TRIANGLE DISTRIBUTION

If a quantity x is defined by a triangle distribution, the expectation μ_x and standard deviation σ, the value of k_p that produces an interval $\mu_x \pm k_p\sigma$ [1] is shown in figure 6.



Fraction p in percent						
ν	50	64.89	90	95	96.69	99.99
∞	0.719573	1.000	1.665737	1.894298	2.000	2.429255

Figure 6. Confidence intervals and levels in Triangle distribution.

The mean can be estimated as equation18. The variance and the standatd deviation are $\frac{(b-a)^2}{24}$ and $\frac{b-a}{2\sqrt{6}}$ respectively. For symetrical distribution, $a = b, b-a = 2a = 2b$ and the standerd deviation becomes $\frac{a}{\sqrt{6}}$

10. EXPANDED UNCERTAINTY

SASO-UM is exporting report contains three confidence levels and intervals, the expanded unceratinty is expressed as $Yy \pm ku_c(y)$ with confidence interval $y - ku_c(y) \leq y \leq y - ku_c(y)$ and confidence levels at 68%, 95% and 99%

IV. INPUT STAGE

The front of the application is a simple data entry form, as shown in Figure 7. It begins with the model function for the measurement process, which should be written in Pythonic mathematical syntax [1]. The variables included in the model function are counted to generate (Number of Model Function Variables) × 7 cells. Each row describes one variable, with 7 columns representing the following details: variable symbol, unit, mean, standard deviation (or standard uncertainty), type of distribution, source type, and degree of freedom. The distribution dropdown list includes the options: Normal, U-Shape, Rectangle, Triangle, Quadratic, Log-Normal, and Rayleigh. If the variables are correlated with each other (as per Equation 13), the "Generate Matrix" button is used. The user is able to generate a matrix of size n × (Number of Model Function Variables), where n is the number of measurements for those variables. An example is provided in the validation section.

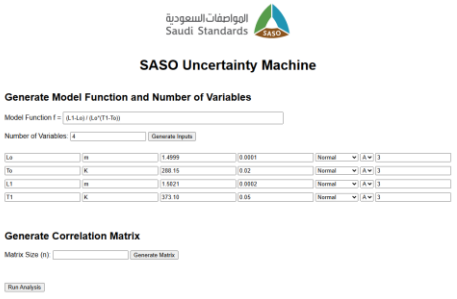


Figure 7. Front page of SASO-UM Application.

V. PROCESSING STAGE

The data are collected from the Input-Stage as a JSON file and posted to FastAPI (Application Programming Interface). The data are then directed to the Python script, as illustrated in the data pipeline in Figure 8.a. After completing the mathematical calculations and statistical analysis, the script exports four types of data, as shown in Figure 8.b.

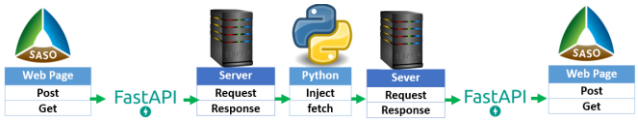


Figure 8-a. Data pipeline

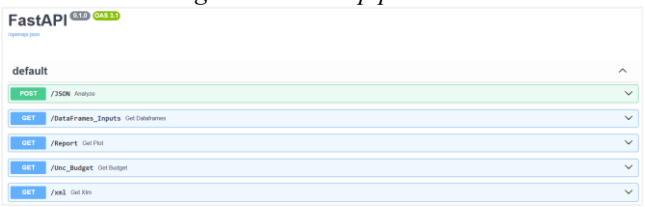


Figure 8-b. Post and Get data in FastAPI

VI. OUTPUT STAGE

In this section, we will discuss one of the four outputs mentioned earlier: the statistical report. The statistical report consists of seven tables and a plotted graph, providing a detailed explanation for the user about the internal workings of the ML algorithm. It starts with an introduction to the model function and variables and concludes with a table detailing the uncertainty budget.

The first table presents the sensitivity coefficients, including the symbol of the partial derivative, its derivative expression, and its computed value. The second table addresses uncertainty Type A, which follows the usual format of an uncertainty budget. The third table describes the model function, while the fourth table outlines uncertainty Type B. The fourth table will display zeros if the user inputs all data entry variables as Type A.

Table 6 combines the uncertainties, and the coverage value (K) is calculated based on the effective degrees of freedom at different confidence levels. The statistical report concludes by plotting the confidence interval as a normal distribution, centered on the mean, figure 9.

VII. SASO-UM VALIDATION

To judge the performance of the machine, it would be better to apply the same examples and numeric that was mentioned in JCGM 100 then compare the output results from API with JCGM-100 text and NIST Uncertainty Machine [15]

1. RELIABILITY OF STANDARD ERROR (UNCERTAINTY OF UNCERTAINTY)

The reliability of the standard error, often referred to as the "uncertainty of uncertainty," is a critical aspect of evaluating measurement processes. It reflects the confidence in the calculated uncertainties and ensures that the reported values are not only precise but also accurate representations of the underlying variability. The close alignment of these data points, as shown in Figures 10 and 11, highlights strong agreement across different methodologies. Additionally, the percentage values decrease with larger sample sizes, effectively illustrating that larger samples reduce uncertainty, leading to more reliable statistical measurements, as demonstrated in Equations 21 and 22.

$$\frac{\sigma[s(\bar{q})]}{[\sigma(\bar{q})]} = \frac{1}{\sqrt{2(n-1)}} \ (approximation) \qquad 21$$

$$\frac{\sigma[s(\bar{q})]}{[\sigma(\bar{q})]} = \frac{\sqrt{V\left(\frac{s}{\sqrt{n}}\right)}}{E\left(\frac{s}{\sqrt{n}}\right)} \qquad 22$$

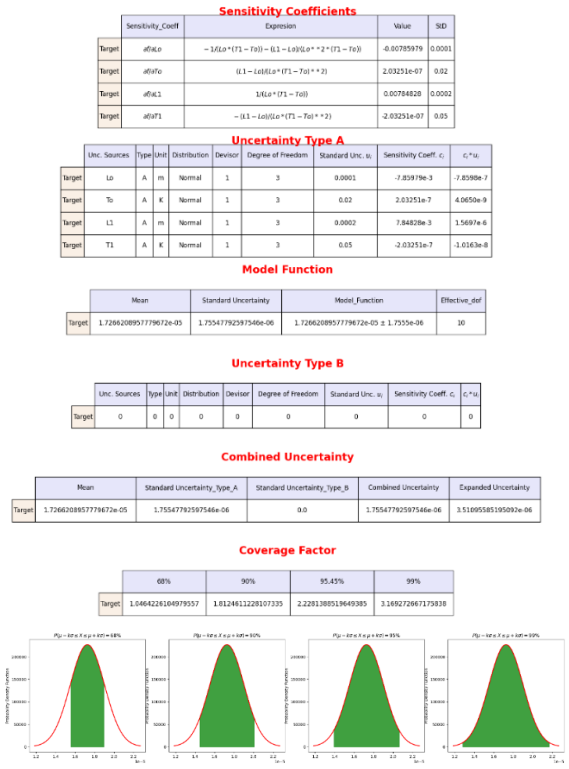


Figure 9. Statistical Report

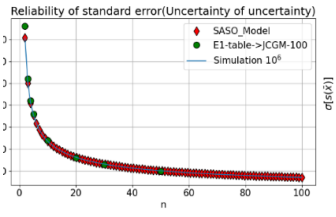


Figure 10. Reliability vs n

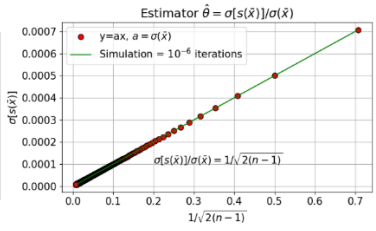


Figure 11. Reliability vs $\frac{1}{\sqrt{2(n-1)}}$

2. SIMULTANEOUS RESISTANCE AND REACTANCE MEASUREMENT

In Example H.2 of the GUM, the measurement model for the resistance of an element of an electrical circuit is $R = \frac{V}{I}\cos \emptyset$.The inputs data and output statistical report are described in detailed in figures 12. The example here is includes correlated parameter, so we generate 5 x 3 matrix (three variables with five observations as given by GUM)

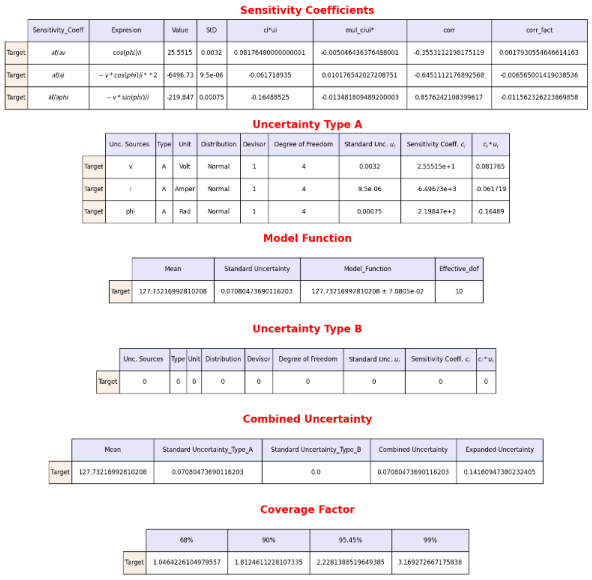


Figure 12. Statistical Report of H.2

The comparison between the GUM, SASO and NIST machines yielded values of $50000838\,nm \pm 32\,nm$, $50000838 \pm 31.71\,nm$ and $50000838 \pm 31.7\,nm$ respectively.

3. DENSITY MEASUREMENT

In the following example, the experiment was carried out at NMCC SASO using regular xls file. The model function is

$$\rho_L = \frac{\{\rho_T[1+\beta(T_T - T_0)] - \rho_{a2}\}(\alpha O_2 + \Gamma_L/g)}{\alpha O_2 - \alpha O_3 + \frac{\Gamma_L}{g} + m_s - V_s\rho_T[1+\beta_S(T_s - T_0)]} + \rho_{a2}$$

The density measurements by xls file were 0.702182 gm/cm^3 with expanded uncertainty 7.49 E-05, SASO uncertainty machine gives as the following figure 13.

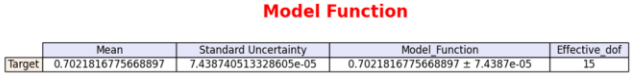


Figure 13. Density measurements by SASO-UM

CONCLUSION

A web-based application has been developed to interact with JCGM-100:2008. The SASO team has automated the standards to ensure simplicity of use, helping all users determine, evaluate, and build their own uncertainty budgets with ease. The architecture of the SASO machine is based on a decision tree algorithm, trained continuously for 112 days to achieve a robust model with high accuracy. The application was built using the FastAPI framework, like the one used by BIPM KCDB [16]. The machine has been validated using examples from JCGM-100 to compare the obtained results against the standard and the NIST machine. As the next step soon, the XML extension will be linked with the DCC, marking an important milestone in digital transformation.

REFERENCE

1. Guide to the expression of uncertainty in measurement - JCGM retrieved from
2. ISO/TR 13587 Three statistical approaches for the assessment and interpretation of measurement uncertainty.
3. ISO 7870-2 Control charts —Part 2: Shewhart control charts
4. ISO 16269-7 Statistical interpretation of data — Part 7: Median — Estimation and confidence intervals
5. ISO IEC 23053 Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)
6. Jupyter Documentation (2025) retrieved from https://docs.jupyter.org/en/latest/
7. NumPy Documentation (2025) retrieved from https://numpy.org/news/
8. Pandas Documentation (2025) retrieved from https://pandas.pydata.org/docs/
9. Matplotlib Documentation (2025) retrieved from https://matplotlib.org/
10. Plotly Documentation (2025) retrieved from https://plotly.com/
11. SciPy Documentation (2025) retrieved from https://scipy.org/
12. Seaborn Documentation (2025) retrieved from https://seaborn.pydata.org/
13. statsmodels Documentation (2025) retrieved from https://www.statsmodels.org/stable/index.html
14. scikit-learn Documentation (2025) retrieved from https://scikit-learn.org/stable/
15. NIST Uncertainty Machine — User's Manual
16. ABI KCDB Swagger UI

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

The DX schema as a modular concept for metrological certificates and reports

Justin Jagieniak[1*], Shan Lin[1], Moritz Jordan[1], Muhammed-Ali Demir[1], Lutz Doering[1], Thomas Engel[2], Wiebke Heeren[1], Jan Loewe[1], Shanna Schönhals[1], Siegfried Hackel[1]

[1]*Physikalisch-Technische Bundesanstalt (PTB), Bundesallee 100, 38116 Braunschweig, Germany, justin.jagieniak@ptb.de, shan.lin@ptb.de, moritz.jordan@ptb.de, muhammed.demir@ptb.de, lutz.doering@ptb.de, wiebke.heeren@ptb.de, jan.loewe@ptb.de, shanna.schoenhals@ptb.de, siegfried.hackel@ptb.de*

[2]*Siemens AG, Friedrich-Ludwig-Bauer-Straße 3, 85748 Garching, Germany, engelthomas@siemens.com*

[*]*Corresponding author: justin.jagieniak@ptb.de*

***Abstract* – The Digital Calibration Certificate (DCC) has been the first metrological quality related document that has undergone a thorough analysis and requirements assessment process in order to develop a fully machine-actionable digital process that fulfills the requirements of the ISO 17025. During this, still ongoing, process, the demand for a more modular approach to digital quality documents has become obvious. The paper introduces the DX schema, which is the conceptual basis for the future DCC version 4 and other digital certificates, reports, and documents related to metrology, conformity assessment and a process oriented digital quality infrastructure. In addition the paper introduces some examples of digital certificates and reports which will potentially benefit from the DX.**

I. INTRODUCTION

The concept of DX schema has its origins in several projects for digitalization in metrology, including the GEMIMEG, GEMIMEG-II [1], and ISAN [2] projects. In the GEMIMEG and GEMIMEG-II projects, participants identified the need for digital documents within the industrial supply chain, such as Digital Calibration Certificate (DCC), Digital Calibration Request (DCR), and Digital Calibration Answer (DCA). In the ISAN project, a digital accident report (D-AR) was developed by a team at Physikalisch-Technische Bundesanstalt (PTB). The D-AR was designed to transfer data from a sensor system to a rescue center.

Both developments share the commonality of being based on metrological measurements, which are provided by DCC or Digital Test Report (DTR), respectively. Unfortunately, the current structure of the DCC is not modular enough to generate other reports or certificates. This limitation led to the development of a digital document schema, the DX schema. The DX schema incorporates more generic types that are common to the certificates, reports or requirement exchange documents and thus should be defined in a unified way. The DX schema refers to the variable "X", symbolizing its ability and aspiration to serve as structured basis to digitalize all other schemas related to metrology, which means that it is the foundational or basic but modular schema to derive other exchange documents.

Through extensive discussions and reconsideration, it was concluded that DX should offer greater flexibility and compatibility with various ISO standards. While the DCC was designed for ISO 17025 [3], the DX schema aims to support additional ISO standards.

In this paper, the concept of the DX schema and its relevant ISO standards are presented in Section II. Section III presents the DX ontology alongside the corresponding DX XML implementation, and proposes a workflow for co-development of ontology. Based on DX, three examples of digital quality documents are given in Section IV, including DCC, digital calibration request, and digital test certificate. Finally, Section V outlines the DX roadmap, such as the DX Wiki and DX-related software tools.

II. DX AND INTERNATIONAL STANDARDS

The fundamental concept of the DX is to provide a versatile generic and modular structure characterized by recurring types. The DX can be viewed as a systematic and structured collection of various types of documents, such as administrative data, metrological data, statements and findings from the examiner, and concluding conformity statements. In order to ensures flexibility, compatibility, and compliance with independent digital certificates and reports, individual components (i.e. types) must consider all relevant norms, standards, nomenclatures, and technical regulations.

The DX schema inherits the representation of measurement data from the Digital System of Units (D-SI) [4] for the representation of measurement data, which ensures seamless integration and consistency within existing sys-

tems. The D-SI also traces back to the BIPM reference point of metrological units, which incorporates the fully traceable history of all seven SI metrological base units and the 21 derived units.

The ISO 17000 series of standards are related to conformity assessment, which provides guidance and requirements for a variety of activities, such as operation, calibration, testing, certification, validation, and accreditation. The DCC is defined by ISO/IEC 17025, while other digital certificates and reports are defined by various other ISO standards. The DX will be designed in accordance with the ISO 17000 series and developed using the following specific documents.

- ISO/IEC 17000:2020 Conformity assessment – Vocabulary and general principles [5].
- ISO/IEC 17025:2017 General requirements for the competence of testing and calibration laboratories [3].
- ISO 17034:2016 General requirements for the competence of reference material producers [6].
- ISO/IEC 17043:2023 Conformity assessment – General requirements for the competence of proficiency testing providers [7].
- ISO/IEC 17065:2012 Conformity assessment – Requirements for bodies certifying products, processes and services [8].

Moreover, the DX will also be designed based on the fundamental metrological norms and standards. The most important of these are:

- SI Brochure of the BIPM [9],
- International Vocabulary for Metrology (VIM) [10],
- Guide to the Expression of Uncertainty in Measurement (GUM) [11].

The aim is to use the modular DX schema to represent various digital certificates and reports related to metrology, including the Digital Calibration Certificate (DCC), Digital Calibration Request (DCR), Digital Test Certificate (DTC), Digital Test Report (DTR), Digital Reference Material Certificate (DRMC), Digital Certificate of Conformity (DCoC), and others. Each certificate or report must be developed according to one or more specific ISO standards. Based on the DX, the relevant ISO standards for each document are shown in Fig. 1. The details of selected certificates and reports, i.e. DCC, DCR and DTC, are provided in Section IV.

In the future, the development of the DX should consider not only the ISO 17000 series of standards but also additional series of ISO standards and guidelines. For example, ISO 9000 series of standards are related to quality

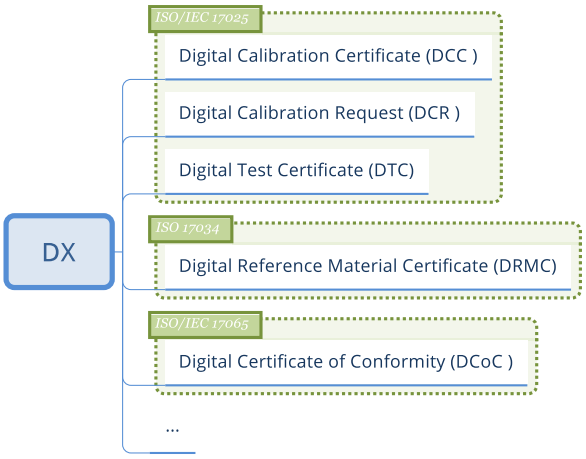


Fig. 1. DX based schema and the relevant ISO standards.

management. As already mentioned in Section I, the DX can also be used for many other ISO or IEC based standards to create appropriate reports or certificates, like the D-AR. The modular structure helps to cover a broad range of content with dedicated modular types related to the specific content of a respective standard. To incorporate a new standard, the rules of DX have to be followed in order to enable full cross-functionality an to leverage all synergies.

III. SEMANTICS AND ONTOLOGY OF DX

The semantics and ontologies are the basis of the digital language used in digital documents (see [12] and [13]). Simply speaking, semantics defines the word and the ontology the relation between different term of the semantics. This enables a shared and unambiguous understanding of key concepts across communities, systems, and domains. In the context of the ISO 17000 series - which includes standards for calibration laboratories (ISO 17025), reference material producers (ISO 17034), and other conformity assessment bodies - the development of machine-interpretable data structures is critical. The DX schema initiative aims to support this need by providing a generalized XML implementation for digital documents across these standards. However, to unlock the full potential of semantic interoperability and machine reasoning, we propose the parallel development of a DX semantics and ontology.

A. *The role of ontologies in metrology*

Ontologies provide a formal, machine–understandable specification of domain knowledge. In metrology, they help to ensure that fundamental terms such as quantity, unit, uncertainty, or measurement result are consistently understood and used - not just by humans, but also by software systems. Concise and precise semantic definitions for every term used are essential for building trustworthy,

interoperable digital systems that can reason over measurement data, link documents across standards, and enable advanced search and analytics.

Recent efforts in the metrology domain underscore the importance of ontologies:

- The SI Reference Point [14], developed under the auspices of the BIPM, provides persistent identifiers (PIDs) and semantic definitions for SI units, prefixes, and defining constants, along with a semantic model for compound units. It forms a semantic backbone for referencing standard quantities in a machine-readable way.
- Another significant ontology development in metrology is the D-SI Ontology [15], developed by PTB. It provides a semantic model for quantity values and uncertainties and was derived from the D-SI XML schema definition, ensuring alignment between schema and ontology for improved interoperability. The D-SI Ontology reuses the semantic model of units provided by the SI Reference Point to ensure consistency and compatibility in the representation of units across different systems.

The latter initiative demonstrates the value of coupling XML schema definitions with corresponding ontologies, especially when aiming for long-term sustainability and interoperability in digital infrastructure.

B. Why Develop the DX Ontology Alongside the DX XML Schema Definition

We strongly recommend that the DX schema be developed in tandem with a corresponding DX Ontology. This co-development approach offers multiple benefits, particularly in ensuring semantic alignment and long-term interoperability. Designing the ontology alongside the XML schema definition allows for early alignment of concepts, structures, and naming conventions. This integrated approach helps to prevent inconsistencies and mismatches that often arise when ontologies are developed retroactively or without consideration for the underlying schema.

Moreover, the DX Ontology can be intentionally standardized and structured to reuse and extend existing ontological models, such as the SI Reference Point, the D-SI Ontology, or QUDT (Quantities, Units, Dimensions and Types). The latter is a set of "ontologies modeling physical quantities, units of measure, and their dimensions in various measurement systems" [16]. This promotes modularity and supports a wide range of use cases across the ISO 17000 series of standards, enabling the schema to serve as a foundation for interoperable and future-proof data models.

A common DX schema with standardized and defined semantics and ontology could also serve as a basis to harmonize and align technical terms for identical content in standards on their iteration when transforming to smart digital standards to leverage synergies even across standards.

A further advantage of ontology-based modeling is the support it provides for smart documents and reasoning. The use of the Web Ontology Language (OWL) [17] enables powerful querying and inferencing capabilities, allowing systems to answer complex competency questions. Examples include identifying which documents include calibration results for a specific measurement standard, determining which laboratories are accredited for a particular calibration method, or detecting when reported uncertainties exceed acceptable thresholds.

The DX Ontology also facilitates improved data integration and linking. By representing documents in a structured, machine-readable format and using standardized terms from an ontology to describe their structure and content, it becomes possible to connect information across different systems and to enrich DX-based data with external resources - such as accreditation registries, instrument databases, or calibration hierarchies.

In terms of validation and data quality assurance, ontologies provide a richer set of tools than XML schema alone. While the schema ensures that documents follow the correct structure, the DX Ontology - especially when combined with SHACL (a rule-based language for validating data) [18] - can also define and enforce rules about the meaning and consistency of the content. This allows for validating more nuanced conditions, such as whether a quantity value includes both a numeric value and a corresponding unit, or whether a measurement result correctly references a defined measurand.

Finally, the combination of DX XML and ontology design directly supports the FAIR principles (Findability, Accessibility, Interoperability, and Reusability) [19]. Embedding semantics into the structure of digital documents ensures that metrological data is not only machine-readable, but also machine-understandable - laying the groundwork for its integration into future digital ecosystems.

C. Proposed Workflow for Co-developing the DX Semantics, Ontology and implementation in XML

To ensure a coherent and effective integration of schema and ontology, we propose the following workflow:

1. **Requirements Gathering**
   Identify the key document types, data structures, and use cases to be supported by the DX schema.
2. **Ontology and XML Schema Co-Design Process**
   - **Develop the DX Ontology**
     Start by building the DX Ontology to model the

key concepts and relationships relevant to the ISO 170XX series. As part of this process, implement a structured class hierarchy that organizes these concepts according to the domains and document types covered by the standards.

- **Derive the DX XML Schema from the Ontology**
  Use the ontology as the semantic foundation for the DX XML Schema. This ensures that the schema is aligned with the ontology's structure and terminology. Use annotations to maintain traceability between schema types and ontology terms, which enables better documentation, validation, and long-term maintainability of both models.

3. **Ontology Alignment and Reuse**
   Where possible, reuse terms and structures from existing ontologies (e.g., SI Reference Point for units, D-SI for measurement values). Align with external vocabularies where needed.

4. **Generation of RDF Representations**
   Use transformation tools or scripts to convert XML documents based on the DX schema into RDF [20] - a graph-based format for representing structured data - in a way that aligns with the DX Ontology.

5. **Validation and Reasoning**
   Develop SHACL shapes based on the ontology to validate RDF data. Use OWL reasoners to check logical consistency and derive implicit knowledge.

6. **Publication and Governance**
   Ensure the ontology is versioned, documented, and published using persistent Universal Ressource Identifiers (URIs). Align with best practices in ontology governance and community involvement.

IV. EXAMPLES FOR DX BASED CERTIFICATES AND REPORTS

Designed as a modular XML schema, the DX does not follow a fixed structure, which can be viewed as an extension of D-SI. Meanwhile, digital certificates and reports primarily consists of a tree structure of types from both the DX and the D-SI. Additionally, custom types can be incorporated into the digital certificate or report when needed.

Every type in the DX stands on its own and is only connected to others when required. However, for a digital certificate or report, it is good practice to follow the ring structure (Figure 2) commonly used in DCC. The ring structure categorizes data based on its usage. The first ring defines administrative data, such as laboratory information, software used, etc. The second ring contains the results of measurements. The third ring defines comments related to the measurements. Finally, the fourth ring can include a document, such as a PDF or HTML file, to provide a human-readable view of the report or certificate.

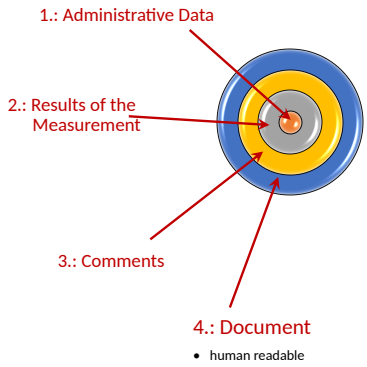


Fig. 2. Ring structure as good practice.

A. *Redefinition of coreData*

Typically, a DX-based schema should include an element called coreData in the first ring (the administrative data section). In the DCC schema, coreData is an element that contains essential information, such as the language of the certificate and a unique identifier.

The DKD (Deutscher Kalibrierdienst) expert report DKD-E 0-2 summarises common requirements for machine-readable certificates and highlights the need for standardised administrative data across the ISO 17000 series [21]. This report recommends to include administrative requirements for reports on conformity assessments, i.e. type of certificate, certificate number, revision number, number of the first certificate issued, name of issuing organisation, address issuing organisation, and date of issue of the certificate.

With the DX, it would be beneficial to redefine coreData as a universal administrative data element on the DKD expert report. This means coreData could be used consistently across all DX-based schemas in different ISO standards, serving as a standardized container for key administrative-level data.

B. *Digital Calibration Certificate version 4*

The DCC version 4 has essentially the same tree structure as DCC version 3. The difference, as shown in Figure 3, is that the type of measurementResultList, comment, document, and signature now follow a different namespace. The types of these four elements are inherited from the DX, as they can be commonly used, while the administrative data is specific to DCC version 4 and thus is customised for this schema.

C. *Digital Calibration Request (DCR)*

The Digital Calibration Request (DCR) was developed as a part of the GEMIMEG-II project, serving as a key dig-

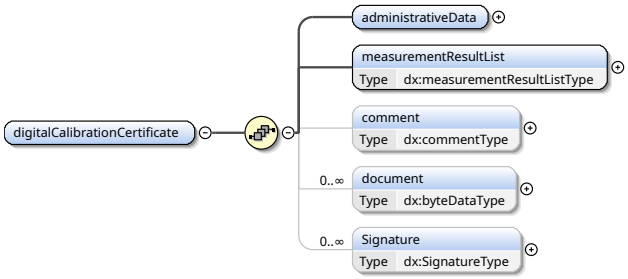


Fig. 3. DCC version 4 schema based on DX.

ital document within a fully digital industrial workflow. It is used to specify and exchange of the calibration requirements between customers and calibration service provider. By receiving a DCR, the calibration service provider can extract the information, including complete and detailed technical specifications for the requested calibration.

The structure of the DCR is slightly different from that of the DCC because, logically, a DCR cannot include the results of a measurement. The DCR provides the measuring points and additional information about the calibration to the calibration laboratory. These details need to be included in the DCR but are not necessarily part of the later DCC. This is accomplished through the element measurementRequest, as shown in Figure 4.

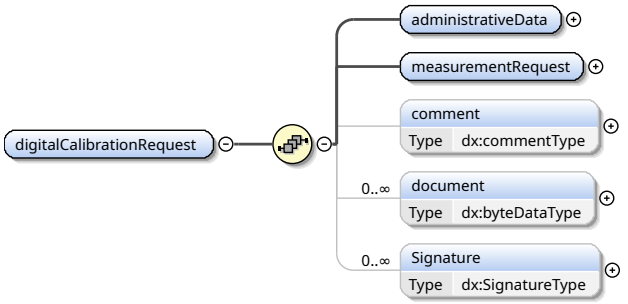


Fig. 4. DCR schema based on DX.

D. Digital Test Certificate (DTC)

ISO/IEC 17025 covers both calibration certificates and test reports (or test certificates), which share common requirements including unique identification and laboratory information (name and address). However, each document type has specific requirements. Calibration certificates must always include measurement uncertainty, while test reports only require this information under certain circumstances. Additionally, test reports contain specific information not found in calibration certificates, such as test conditions and, when applicable, sampling information when the laboratory is responsible for the sampling activity.

Based on the DX schema, PTB is currently developing a Digital Test Certificate (DTC) schema, which follows a similar ring structure as the DCC as shown in Fig. 2. Like the DCC, the DTC includes elements such as measurementResultList, comments, documents, and signature following the types of DX, while its administrativeData element should be contained in its own schema. Compared to the DCC version 4 structure, the key structural difference is in the element dtc:administrativeData, where the child element dtc:sampling is used under the element dtc:item. dtc:sampling is used to describe specific requirements related to reporting sampling, especially the sampling plan and sampling method.

V. ROAD TO FUTURE

In the future, several projects for the DX are in the pipeline.

A. DX Wiki

A Wikipedia [22] for DX is currently under development. The Wikipedia demonstrates how to use DX to develop custom schemas based on it. Additionally, it provides information about commonly used schemas based on DX, such as DCC version 4, DCR, DTR or DTC.

B. DX Library

There is a need for a DX software library, similar to the library for the DCC, which will allow to sign, validate, evaluate formulas, and perform other operations for certificates or reports based on the DX. These functionalities have already been considered during the development of the DX schema but can only be tested and released at a later stage.

C. DCC version 3 to DCC version 4 Converter

Furthermore, there is a need for a converter from DCC version 3 to DCC version 4. This will make it easy to transform a DCC into the upcomming DX standard.

VI. CONCLUSION

The development of DX represents a significant advancement in metrology, providing a flexible and modular XML schema for generating digital certificates and reports that adhere to various ISO standards mainly for the ISO 17000 series. By expanding the DCC, DX offers improved compatibility with ISO standards beyond the scope of the DCC, making it a versatile foundation for a variety of metrological and quality related documents. The modular structure ensures its adaptability to diverse requirements while maintaining consistency in data representation, ensuring compliance with metrological norms and facilitating integration into existing systems.

The versatility of the DX schema to incorporate and support various certificates, such as the DCC, DCR, and

DTC, demonstrates its potential to streamline processes within the industrial supply chain and testing laboratories. Furthermore, the future development of tools like the DX Wiki, a library for DX, and DCC version 3 to DCC version 4 Converter promises to enhance accessibility and ease of use, ensuring that DX continues to evolve in line with emerging needs and standards. The standardized DX schema as a common basis for metrological and other quality related documents is also a key enabler to transfer DX documents into the digital quality document submodel of the asset administration shell. [12] By that, the DX might become a versatile digital foundation for a large number of quality documents for a versatile digital quality infrastructure. to support full digital and automated interoperability along industrial value chain with numerous stakeholders.

ACKNOWLEDGMENTS

Sincere gratitude is extended to the GEMIMEG-II project, funded by the German Federal Ministry for Economic Affairs and Climate Action (BMWK), and the International Standard Accident Number (ISAN) project, sponsored by the Niedersächsische Vorab Initiative of the Volkswagen Foundation (VolkswagenStiftung). These projects have made significant contributions to the development of the DX schema, and their support is appreciated.

REFERENCES

[1] T. Engel. GEMIMEG-II – How metrology can go digital... In: Measurement Science and Technology 34.10 (July 2023), p. 104002. DOI: 10.1088/1361-6501/ace468. URL: https://dx.doi.org/10.1088/1361-6501/ace468.

[2] ISAN project. Accessed: 2025-03-06. URL: https://www.tu-braunschweig.de/informatik/research/projects/international-standard-accident-number.

[3] General requirements for the competence of testing and calibration laboratories. ISO/IEC 17025:2017.

[4] Daniel Hutzschenreuter et al. SmartCom Digital System of Units (D-SI) Guide for the use of the metadata-format used in metrology for the easy-to-use, safe, harmonised and unambiguous digital transfer of metrological data - Second Edition. Version D-SI 1.3.0-2. July 2020. DOI: 10.5281/zenodo.3816686. URL: https://doi.org/10.5281/zenodo.3816686.

[5] Conformity assessment – Vocabulary and general principles. ISO/IEC 17000:2020.

[6] Conformity assessment – Vocabulary and general principles. ISO/IEC 17034:2016.

[7] Conformity assessment – General requirements for the competence of proficiency testing providers. ISO/IEC 17043:2023.

[8] Conformity assessment – Requirements for bodies certifying products, processes and services. ISO/IEC 17065:2012.

[9] The International System of Units (SI). 2014. URL: https://www.bipm.org/en/publications/si-brochure/ (visited on 03/24/2025).

[10] International vocabulary of metrology – Basic and general concepts and associated terms (VIM). JCGM, 2012.

[11] Evaluation of measurement data – Guide to the expression of uncertainty in measurement. JCGM, 2008.

[12] Engel, Thomas. From a digital calibration certificate to a digital quality infrastructure. In: EPJ Web Conf. 323 (2025), p. 01003. DOI: 10.1051/epjconf/202532301003. URL: https://doi.org/10.1051/epjconf/202532301003.

[13] Thomas Engel and Alexandru Popescu. GEMIMEG - Digitalization of calibration processes in metrology. In: Measurement: Sensors (2025), p. 101474. ISSN: 2665-9174. DOI: https://doi.org/10.1016/j.measen.2024.101474. URL: https://www.sciencedirect.com/science/article/pii/S2665917424004501.

[14] SI Reference Point. URL: https://si-digital-framework.org/ (visited on 04/15/2025).

[15] D-SI Ontology. URL: https://gitlab1.ptb.de/d-ptb/d-si/ontology-d-si (visited on 04/15/2025).

[16] QUDT Working Group. QUDT Ontologies: Quantities, Units, Dimensions and Data Types. Accessed: 2025-05-13. 2023. URL: https://qudt.org/.

[17] W3C OWL Working Group. Web Ontology Language (OWL). Accessed: 2025-06-20. 2012. URL: https://www.w3.org/TR/owl2-overview/.

[18] Knublauch, Holger and Kontokostas Dimitris. Shapes Constraint Language (SHACL). Accessed: 2025-06-20. 2017. URL: https://www.w3.org/TR/shacl/.

[19] GO FAIR Initiative. Fair Principles. Accessed: 2025-06-20. 2016. URL: https://www.go-fair.org/fair-principles/.

[20] Cyganiac, Richard and Wood, David and Lanthaler, Markus. Web Ontology Language (OWL). Version 1.1, Accessed: 2025-06-20. 2014. URL: https://www.w3.org/TR/rdf11-concepts/.

[21] T. Krah et al. DKD-E 0-2, Expert report: Common requirements for machine-readable certificates in the ISO 17000 series. 2024.

[22] DX Schema wiki. URL: https://www.dx.wiki (visited on 03/06/2025).

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Metrological AI Reliability Verification

Volker Zeuner, Gulian Couvreur

Federal Institute of Metrology METAS, Lindenweg 50, 3003 Bern-Wabern, Switzerland
Volker.Zeuner@metas.ch, Gulian.Couvreur@metas.ch

***Abstract* – Applications based on Artificial Intelligence (AI) are becoming useful tools in Metrology. The ability of e.g. managing big amounts of devices or analysing Big Data using AI Applications (in the sense of Software) raises speditivity and accuracy of gaining metrological results as well as facilitates metrological analyses on a new scale.**
When AI is involved in measurement, decision making, classification etc., it must be assured that the results are transparent, reproducible, comparable and traceable, corresponding to the established metrological measurement requirements.
Hence, the operation of AI Applications in Metrology should be subject to prior and periodic verification, regarding a classification of the AI types and their use scenarios (e.g. in Legal Metrology), their usage and the expected results thereof.

Keywords:
Metrology, Legal Metrology, Artificial Intelligence, reliable measurement, CRA [1], AIA [2], Cybersecurity, secure operation, conformity, trustworthiness

I. INTRODUCTION

AI Applications open a fundamentally new opportunity on how big amounts of data can be processed in a fast way in order to achieve decisions, verdicts and results regarding a given purpose – even with the possibility to retrieve continually improved ones out of a continuously learning system.

At the same time, the great computational complexity an AI System brings to play calls for a lot of necessary verification steps in order to make sure that such a system keeps its warranted properties during operation.
It is imaginable that for given use cases in metrology the operation of AI Applications looks advantageous, but at the same time lacks the reliability of approved measuring equipment. For specific use cases, the corresponding good practise requirements may be accomplished and verified easily. However, each set of such singleton use cases needs assessment regarding conformity of its elements to be able to assign AI Applications to their appropriate use case classifications.

Resulting from this request, is the subject of classifying AI Applications. The classical instruments of standardisation seem comparably slow and narrow facing a software product with a near indefinite number of possible use cases, product cycles with a fraction of the duration of standardisation phases and implementational degrees of freedom exceeding the coverage by "must, shall and should".

Furthermore, AI Applications are characterised by a certain opaqueness and, consequently, the calls for transparency, explainability or other clarifying actions to be taken in this respect exist. But without an organised approach, how can transparency be achieved?

II. APPROACH

The current state of the art
Currently, no initiatives of qualifying AI Apps for metrological systems are public.

PTB's TraCIM service provides a selection of test data sets to validate evaluation algorithms for metrological software [3]. (cf.: https://tracim.ptb.de/tracim/tenant.xhtml?fragment=quickguide.xhtml)
The verification scheme for Metrological AI Applications for given metrological use cases might profit thereof.

Regarding conformity assessment norms, the status as of January 30, 2025, of corresponding activities of "CEN / CLC / JTC 21 - Artificial Intelligence" for the AI Conformity assessment framework was "Forecasted voting date 2026-07-29".

Under the assumption that the knowledge of a system's properties (regardless of being focused on ethical, metrological, legal or cybersecurity issues) is the paramount requisite to use or to manage it (cf. Lord Kelvin and also Mr. Peter Drucker), the development of appropriate Norms and Standards lags behind the speed of the development cycles of AI Applications as well as in regard of the presumably needed high number of use case specific documents.
At the same time, the approach to applicable specifications governing the AI issues is not "unified".
Standardisation groups and for example the authors of the EU AIA, as a specification for regulation, surely mean the

same, but operate with different semantics.

Still, management guidelines do represent an approach to good practise of managing AI Applications, although it may not be stringently clear what the managed systems' properties really are and whether these stay impaired.

Appreciable examples are for instance

- NIST AI 100-1, Artificial Intelligence Risk Management Framework (AI RMF 1.0) [4]
- ISO/IEC 42001, Information technology — Artificial intelligence — Management system [5]

whereas the first document seems at least to address the shortfall of unknown system properties in its section 5.

A very interesting aspect with well-defined use cases is the usage of AI Applications for medicinal purposes. An obvious advantageous use case lies in supporting imaging technology. Unfortunately, the many degrees of freedom to select and operate AI Applications creates the need for categorisation in order to purchase or develop a suitable solution and also how to qualify the obtained results out of the operation (Jonas Richiardi, Patrick Omoumi1 et. al: "To buy or not to buy—evaluating commercial AI solutions in radiology (the ECLAIR guidelines)" in IMAGING INFORMATICS AND ARTIFICIAL INTELLIGENCE [6]. This approach to a classification of procurable AI Applications bases on a lot of lessons learnt during research and operation of AI Applications. Here, it seems practical to offer a standardised pre-qualification, because this might reduce the number of "not to buy" cases and the spent effort at the same time.

Another valid solution path

"Conformance testing — an element of conformity assessment, and also known as compliance testing, or type testing — is testing or other activities that determine whether a process, product, or service complies with the requirements of a specification, technical Standard, contract, or regulation." [7]
(https://en.wikipedia.org/wiki/Conformance_testing)

Under the assumption that a conformity assessment body (CAB) is qualified to work with an appropriate specification of properties to determine that these properties are present and valid in the test object, it seems appreciable, to make the sponsor of the testing bring a requirement list instead of waiting for a Norm to become applicable.
Likewise, for different phases of an AI Application's life cycle, different specific properties may have to be verified – via adequate sets of specification.
And after all, if the relevant properties are being regarded as "assets" in the meaning of "entities that the owner of a ToE [i.e. Target of Evaluation in the CC] presumably places value upon" [8], there is obviously a way to specify sets of requirements for the properties that matter. Should there be vulnerable assets in the set, the specification of protective functionality and the verification of its correctness and effectiveness would then be part of the conformance testing, ensuring the unimpaired presence of the properties at the time of the testing.

Basic starting condition

A testing scheme for trustworthiness of AI Applications was developed covering the three following areas (respective lists not exhaustive) over all relevant life cycle phases beginning with development specifications and ending with de-commission:

1. 3 Tiers
   a. Operational environment of the AI Application
   b. AI behaviour
   c. Competence of AI User

2. 5 Assets
   a. Transparency
   b. Fairness (non-discrimination)
   c. Beneficence (non-malfeasance)
   d. Accountability
   e. Privacy

3. Assessment of asset handling comprising at least
   a. Contents and structure of output data
   b. Creation of output data
   c. Implementation aspects

On the first view, this classification seems very abstract in regard of a Metrological AI Application. In fact, the categories were defined for a broader scope of possible AI Applications. Still, the classification scheme will prove helpful to characterise AI Applications so that a more focused scope can be defined.

Being elaborate systems based on Software, all AI Applications face justified reservations about their trustworthy functioning – just like all SW.
Combined with generic SW life cycle phases, the three areas above can be projected onto:

i. design of the application and procurement of 3rd party components
ii. development, implementation and training
iii. deployment, commissioning, operation
iv. maintenance, update, upgrade
v. end-of-operation, disposal

During the life cycle, different roles appear, exercised by different legal entities: software developers, system integrators, operators end users etc.

The operational contexts of AI Applications, are characterised by many degrees of freedom. These contexts may span from DIY-developed systems for research purposes over purchased solutions operated for research or service provision over in-house metrological support SW to a marketed Metrological AI Application enhancing an until then customary measurement instrument.
Regardless of the context, all AI Applications will have been examined, whether they grant their intended "warranted properties".
The basic assumption that a purchased product maintains the warranted properties, however, is not valid for AI Applications. They are prone to change their behaviour just after commissioning and during operation.

Consequently, the equivalent of instrumental bias or drift needs to be controlled for a Metrological AI Application and every other one with legal of financial liabilities in case of malfunction. That may be covered sufficiently by in-house solutions for research purposes. For marketable AI Applications a conformity test with a quality seal (or else) may be advisable. Likewise, two different Metrological AI Applications, maybe also with different origin, for the same metrological purpose could be comparable.

III. PROOF-OF-CONCEPT SITUATION

If one would assume a Metrological AI Application similar to a measurement instrument, then each result an AI Application has contributed to should have an equivalent quality as one achieved with a classical measurement instrument alone.
Operating a metrological AI Application in the field, it seems obvious that there is a need for qualifying it via conformity testing.
Operating one in the field of Legal Metrology, it is not imaginable to run it without prior AI Reliability Verification.

Metrological AI Application Use Cases comprise:
- Measurement
- Calibration
- Analyses
- Predictive Maintenance
- Control, Alarming, Decision Making

Goals for AI Applications' Reliability in Metrology are:
- they do what they should
- they keep their (warranted) properties during operation
- they do not do what they shouldn't

Taking into regard the sections 2.44 and 2.45 of BIPM JCGM 200:2012 [9]
- The set of Metrological AI Applications needs to be classified in a way that they can be assigned to their AI class.
- The scope of the requirements for a given AI class can be matched with an AI Application under assessment.
- The fulfilment of the requirements can be verified by applying a defined verification scheme.

Thus, each AI Application can be subject to Conformity Assessment, which should be congruent with the "verification process" according to section 2.44.

A Reliable Metrological AI Application, hence, will have to undergo assessment focusing on:
- measurement requirements,
- safe and secure operation,
- results that are transparent, reproducible, comparable and traceable.

Here, the generic approach as described in section II. needs to be refined.
Purpose, expected benefit and operational environment must be known before the implementation starts.
Depending of these aspects, some "assets" (cf. II.2.) may be excluded, whilst other ones must be added.

After all, if there is a comprehensive list of assets characterising a given AI Application, appropriate measures can be taken to conserve these properties during the AI Application's life cycle.
Conformity testing against such a set of properties, derived from the identified assets seems an appropriate tool to assess, whether the assets are covered and maintained in the phases of design, specification, implementation, training, commissioning, operation and maybe even at the end of operations (cf. previous section).

A. SELECTION OF SPECIFIC REQUIREMENTS

The AI Application should be analysed in regard of the intended warranted properties, for example:
- Measurement: The AI Application shows reproducible output including reliably low uncertainty.
- Data Analytics: The AI Application generates reproducible results.
- Machine Learning; The AI Application's output keeps continuity over time or improves whilst adding new data.

When the specific properties are defined, functional or architectural objectives need to be formulated. Their achievement can be seen as specific assets to be protected and can be added to the list in II.2.

From a metrological point of view, the traceability of the results of an AI Application seems worth a deeper consideration.
The common interpretation of traceability in an AI Application focuses rather on the safeguarding of the prevention of bias or drift in the output results by being able to trace causal interrelations in the AI Applications' functionality.
If the properties of a measurement instrument are well known, it might be possible to model a possible drift of the measurement results, thus being able to operate under known conditions – and even with changing measurement results.
Consequently, given a sufficient transparency regarding an AI Application's functionalities, a traceability of biased or drifting results might also become possible, because the applied assessment scheme would make it explainable.

B. DELINEATION OF AI APPLICATION TYPES

Likewise, the operational specifics must be analysed and specified, for example in regard of:

- generative vs. discriminative results
- predictive capabilities
- decision supporting vs. stand-alone decision-making
- potential liabilities for manufacturers or operators
- in-house vs. public operation

If objectives or assets can be formulated, their achievement can be seen as specific assets to be covered and can also be added to the list in II.2.

C. ASSESSMENT SCHEME PHASES

During the life cycle phases the list of assets will not change. The measures taken surely will differ, but always shall follow the idea to keep all assets unimpaired.
Subject to the conformity testing scheme will always be the list of assets, assumed sources for impairment thereof, processes and functions implemented to protect all given assets against the feared impairment and specific steps taken to ensure the AI Application's producing the desired answers being stimulated by given sets of data.

Non-exhaustive lists of aspects to be covered during different phases:

i. design of the application and procurement of 3rd party components:
   - purpose of the AI Application & AI type
   - input data in training and reality
   - output
   - 3rd party components
   - models and algorithms
   - training data qualification

ii. development, implementation and training:
   - training data
   - output
   - warranted characteristics
   - verification processes
   - acceptance process

iii. deployment, commissioning, operation
   - output vs. training data
   - output vs. reality data

iv. maintenance, update, upgrade
   same as iii.

v. end-of-operation, disposal
   - removal of non-disclosable modules
   - re-use of pooled data

D. ASSESSMENT STEPS

In order to assess the correctness and effectiveness of the implementation as well as the rendering of the warranted characteristics by an AI Application, for each assessment step a qualified and verifiable description of how the (relevant) assets are protected must be documented.
There are three tiers:

- ease-of-use (i.e. safe and secure operational environment)
- regular functioning (the AI Application only does, what it is supposed to do)
- competent user (e.g. like a driver's license)

and five life cycle phases, which defines 15 testing steps.
In the testing steps, the defined set of assets, their freedom from impairment and the taken measures will be analysed. The analysis should at least contain results in regard of how the functionality regarding

- Contents and structure of output data
- Creation of output data
- Implementation aspects

maintains the assets impairment-free.

IV. RESUME AND RATIONALE

This conformity testing scheme for AI Applications is under development. It is characterised by treating the desired properties as assets and testing their presence in the AI Application under assessment and the AI Application's capability to uphold these properties over time in an acceptable way.

The properties are partly non-technical. The manufacturers are free to specify their interpretation of the adequate properties and their protective measures taken.

The testing follows a standardised but use case specific scheme. The result is certifiable by an accredited CAB.

An AI Application as a measurement instrument in the field of metrology would represent a comparably small test object with relatively focused properties and, hence, would represent an ideal proof-of-concept use case for the Metrological AI Reliability Verification. At the same time, a method to qualify AI Applications for use in the field of Legal Metrology will provide room for digital transformation.

Ongoing research regarding this subject will be concluded. The objective is to gain a conceptual approach to be published.

In the meantime, co-operative contributions or partnerships are welcome.

REFERENCES

[1] Cyber Resilience Act, REGULATION (EU) 2024/2847 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

[2] AI Act, REGULATION (EU) 2024/1689 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

[3] PTB's TraCIM service; https://tracim.ptb.de/tracim/tenant.xhtml?fragment=quickguide.xhtml

[4] NIST AI 100-1, Artificial Intelligence Risk Management Framework (AI RMF 1.0) https://doi.org/10.6028/NIST.AI.100-1; January 2023

[5] ISO/IEC 42001, Information technology — Artificial intelligence — Management system First edition 2023-12

[6] Patrick Omoumi, Alexis Ducarouge, Antoine Tournier, Hugh Harvey, Charles E. Kahn Jr, Fanny Louvet-de Verchère, Daniel Pinto Dos Santos, Tobias Kober, Jonas Richiardi; To buy or not to buy—evaluating commercial AI solutions in radiology (the ECLAIR guidelines); European Radiology https://doi.org/10.1007/s00330-020-07684-x

[7] Conformance testing; https://en.wikipedia.org/wiki/Conformance_testing

[8] Common Criteria for Information Technology Security Evaluation; CCMB-2017-04-001 Part 1: Introduction and general model April 2017, Version 3.1, Revision 5

[9] JCGM 200:2012, International vocabulary of metrology – Basic and general concepts and associated terms (VIM) 3rd edition, 2008 version

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Knowledge Management foundations in the Design of the São Paulo State Institute of Weights and Measures (IPEM-SP) Quality Infrastructure Portal: Strategies and Challenges

Robson S. da Silva[1], Roberto A. Filho[2], Marcos Oliveira Junior[3], Eduardo Mario Dias[4], Maria Lídia Rebello Pinho Dias[5]

[1] *Instituto de Pesquisas Tecnológicas de São Paulo, Av. Prof. Almeida Prado, 532 São Paulo- Brasil, robsonss@ipt.br*

[2] *Universidade do Estado do Rio Grande do Norte, Av. Lauro Maia S/N, Patu- Brasil, robertomariano@uern.br*

[3] *Instituto de Pesos e Medidas do Estado de São Paulo, Rua Santa Cruz, 1922 São Paulo- Brasil, mhgojunior@ipem.sp.gov.br*

[4] *Universidade de São Paulo, Av. Prof. Luciano Gualberto, 380 São Paulo- Brasil, emdias@pea.usp.br*

[5] *Universidade de São Paulo, Av. Prof. Luciano Gualberto, 380 São Paulo- Brasil, lidiadias@pea.usp.br*

***Abstract* – This paper presents the conceptual, methodological, and technological foundations for a Knowledge Management (KM) model aimed at the digital transformation of the Quality Infrastructure (QI) ecosystem. As a case study, it examines the Quality Infrastructure Services Portal (PSIQ), currently under development by the Sao Paulo State Institute of Weights and Measures (IPEM-SP), at Technology Readiness Level 2 (TRL2). Using a qualitative, exploratory, and applied approach, the study supports the transition of the portal to TRL3. Classical KM models (SECI, DIKW, APO, and Cynefin) are employed to structure knowledge flows, while enabling technologies—such as Intelligent Virtual Assistants, Semantic Search, Big Data & Analytics, and Robotic Process Automation—are mapped to sustain operationalization. The results indicate that PSIQ can function as a cognitive digital platform for regulatory knowledge, institutional memory, interoperability, and strategic learning. The paper concludes with recommendations to consolidate KM as a strategic function in the digital transformation of QI in Brazil.**

Palavras-chave: Knowledge Management; Quality Infrastructure; Digital Transformation.

I. Introduction

Digitalization is a strategy that holds the potential to foster the efficiency of public policies and the provision of public services [1]. In Brazil, its application within Quality Infrastructure (QI)—a system composed of institutions, policies, legal frameworks, and practices that ensure the quality, safety, and environmental integrity of goods and services—directly impacts components such as metrology, standardization, accreditation, conformity assessment, and market surveillance [2].

In this context, the Institute of Measures and Weights of State of São Paulo develop initiatives focused on the digitalization of Quality Infrastructure (QI) to enhance industrial competitiveness, facilitate access to technical information, and promote integration within the productive ecosystem. Among these initiatives, the Quality Infrastructure Services Portal (QISP) stands out as an intelligent digital platform designed to organize,

systematize, and disseminate normative content associated with the services provided by the institution.

This paper explores the conceptual, methodological, and technological foundations that should underpin Knowledge Management (KM) at QISP. In a technical-regulatory environment such as that of Quality Infrastructure (QI), effective KM [3] requires more than mere storage: it demands the organization and dynamization of various types of institutional knowledge—technical-regulatory (standards, regulations), operational (inspection routines, service instructions), and tacit (expert experiences). The importance of updating routines, emerging technologies (AI, Semantic Search, RPA), complementary tools (ontologies, ECM, visual analytics, digital competencies management), and curation strategies aligned with the institutional mission is also highlighted.

One of the foundational concepts underlying this study is the Technology Readiness Levels (TRLs) framework. Originally developed by NASA to evaluate the maturity of emerging technologies within aerospace programs, TRLs comprise a structured scale of nine incremental stages—from TRL 1, in which basic scientific principles are observed, to TRL 9, where the technology is demonstrated to operate successfully in its target environment [4]. This framework offers a systematic and consistent approach for assessing the developmental stage and readiness of technological solutions. To facilitate understanding of these maturity levels, Figure 1 presents the full TRL structure and its corresponding descriptions.

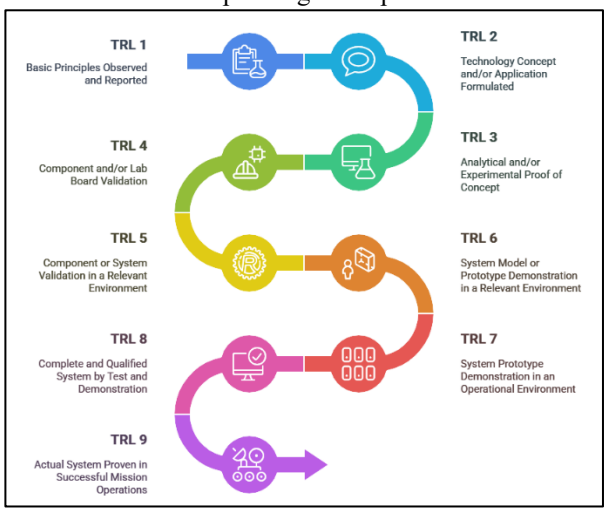


Fig. 1. Stages of Technology Readiness Levels (TRLs)

Currently at TRL 2 (technological formulation), the project requires the definition of structural principles, functional requirements, and enabling technologies. The transition to TRL 3 (experimental validation) depends on a clear methodological approach in designing knowledge flows, curation mechanisms, and interoperability criteria among systems and technical databases.

The central problem lies in structuring, already at TRL 2, the elements that enable the conversion of tacit, explicit, and automated knowledge into a strategic asset. Accordingly, the paper seeks to answer the following research question: how can KM foundations be structured to support the design of QISP and prepare it for TRL 3, ensuring continuous organization, dissemination, and updating technical-regulatory knowledge?

The approach combines classical KM theoretical models with the portal's technological components, modelling knowledge flows, identifying challenges, and proposing recommendations for its future evolution. The research adopts a qualitative, exploratory, and applied approach, focused on technological formulation and the conceptual modelling of the main informational and cognitive elements.

To this end, the paper is structured as follows:

Section II: the articulation between QI and KM, current gaps, and the role of digital platforms.

Section III: the main theoretical models of KM.

Section IV: the technological strategies of QISP and their connection with KM.

Section V: the methodology for analysis and preparation for TRL 3.

Section VI: the challenges and opportunities of KM at QISP.

II. Quality Infrastructure and Knowledge Management: Emerging Intersections

Digital transformation in public administration has driven the adoption of emerging technologies to modernize services, increase transparency, and strengthen institutional capacity. In the context of Quality Infrastructure (QI), this movement promotes the digitization of technical processes, the integration of regulatory systems, and democratic access to technical regulatory knowledge [5].

In this scenario, Knowledge Management (KM) becomes a foundational element in the digital transformation of QI. In sectors such as metrology, standardization, and conformity assessment, KM underpins the capture, systematization, updating, and dissemination of critical information—from technical standards and legal requirements to operational instructions and oversight guidelines—whose continuous curation ensures traceability and regulatory confidence.

An effective KM system integrates people, processes, and technology, fostering alignment between institutional knowledge and organizational objectives [6]. At QISP, developed by IPEM-SP, the goal is to transform this knowledge into a structured digital asset that guides users

in their technical and legal obligations. This requires more than just digitizing documents; it demands information flows oriented toward usability, interoperability, and continuous updating.

However, according to the Final Report of the Public Subsidies Initiative of the National QI Strategy (ENIQ), significant structural challenges persist: deficiencies in governance, a lack of resources for modernizing the national conformity mark, and limitations in the Brazilian Conformity Assessment System (SBAC). These issues are compounded by the low participation of the private sector in standardization, the regional concentration of services, insufficient investments in laboratory infrastructure and technical training, as well as excessive bureaucracy and disjointedness between standardization and regulation [7].

These limitations call for digital solutions to consolidate and streamline regulatory knowledge flows. Platforms such as QISP, by incorporating technologies like Intelligent Virtual Assistance, Semantic Search, Big Data, and Robotic Process Automation (RPA), do not merely provide information but also structure the governance of technical-regulatory knowledge. These technologies enable intelligent screening, contextualized retrieval of norms, automated responses, and data-driven analyses.

By promoting an institutional culture centered on data curation, interoperability, and continuous learning, these platforms reinforce QI as an integrated system. Their adoption fosters responsive regulation and voluntary conformity, enhancing the effectiveness of public actions and the competitiveness of the productive sector.

Thus, the intersections between QI and KM represent a strategic field for institutional modernization. QISP stands as a concrete opportunity to apply KM principles in a digital regulatory environment, functioning as an institutional laboratory for modeling replicable solutions within Brazil's National System of Metrology, Standardization, and Industrial Quality.

III. Knowledge Management Models for Technical-Regulatory Environments

The formulation of a Knowledge Management (KM) model for the Quality Infrastructure Services Portal (QISP) requires a foundation in well-established theoretical frameworks capable of guiding the systematization, curation, and dissemination of technical regulatory content in complex digital environments. This theoretical basis is especially relevant at TRL 2, which focuses on conceptual and technological structuring, ensuring that institutional knowledge is transformed into a reusable, accessible digital asset aligned with the public mission of QI.

The SECI model describes the mechanisms for converting tacit to explicit knowledge through four processes: socialization, externalization, combination, and internalization [8]. At QISP, it contributes to transforming the knowledge of QI experts into structured digital artifacts, such as regulatory flowcharts, FAQs, manuals, and technical protocols. Once systematized, these contents become accessible repositories for both internal and external users, enhancing access to qualified knowledge.

The DIKW hierarchy (Data, Information, Knowledge, and Wisdom) organizes informational complexity into a semantic chain [9]. At QISP, it structures the process from the collection of raw data (e.g., technical records), through the transformation into organized information, to the generation of useful knowledge for normative interpretation and decision-making. This hierarchy supports the definition of logical layers in information architecture.

The methodology of the Asian Productivity Organization (APO), with its five pillars—leadership, people, processes, technology, and results—guides KM in public institutions [10]. When applied to PSIQ, it fosters performance indicators, continuous feedback, and quality control of content, which are essential for the transition to TRL 3, where evidence of viability is required.

Snowden's Cynefin model classifies decision-making contexts as simple, complicated, complex, or chaotic. At PSIQ, it enables distinguishing between demands that can be automated (e.g., standardized norms) and those requiring specialized analysis (e.g., normative exceptions), thereby optimizing the combined efforts of virtual assistants, human curators, and semantic algorithms.

The KM Maturity model identifies an organization's evolutionary stage—from informal practices to a consolidated KM culture—allowing the portal's functionalities to be calibrated according to institutional maturity [11].

The integration of these models results in a complementary theoretical framework. KM Maturity assesses the level of KM maturity; SECI addresses the conversion of knowledge; DIKW structures informational complexity; APO guides governance; and Cynefin defines appropriate cognitive responses. This combined framework supports a robust KM system for QISP, catalysing the digital modernization of QI in Brazil.

IV. Technological Strategies and their Connection with Knowledge

The development of a Knowledge Management (KM) model in Quality Infrastructure (QI) requires technologies capable of structuring and energizing technical-regulatory information flows. In the case of the Quality Infrastructure Services Portal (QISP), KM demands more than a mere digital repository; it requires technological architecture based on accessibility, continuous updating, interoperability, and intelligent curation [3].

Among the essential technologies, Intelligent Virtual Assistance—based on Natural Language Processing (NLP) stands out as a mediator between users and normative content. Well-trained chatbots simulate human interactions and provide precise, contextualized responses, thereby reducing informational bottlenecks and democratizing access to institutional knowledge [12].

Advanced Semantic Search goes beyond keyword-based retrieval by integrating context, user intent, and conceptual synonyms [13]. This approach is critical in QI, where technical terminology varies according to the regulated sector, resulting in a more responsive and inclusive information architecture.

Big Data and Analytics enable the generation of knowledge from large volumes of structured and unstructured data. By identifying patterns, anticipating demands, and feeding back into decision-making processes, these tools reveal frequent inquiries, norms with high non-compliance, and priority topics [14].

Robotic Process Automation (RPA) supports the operationalization of KM by automating repetitive tasks such as document indexing, standardized responses, and metadata updating. This automation liberates human resources for curation activities and reduces errors, promoting scalability and standardization [15].

Additional complementary technologies enrich the KM ecosystem. Technical ontologies foster interoperability between systems and knowledge domains, while Enterprise Content Management (ECM) systems ensure version control, traceability, and document approval processes. Visual analytics tools, including dashboards and dynamic panels, facilitate the monitoring of KM effectiveness [16].

Equally important is the management of digital competencies, which connects institutional knowledge and human development. Solutions that map skills, provide learning pathways, and promote technical-regulatory upskilling reinforce the sustainability of KM as an organizational practice.

The strategic distinction between essential technologies (focused on operations) and complementary ones (aimed at quality and sustainability) consolidates QISP as an adaptive, data-driven platform with embedded institutional intelligence. Consequently, the technological infrastructure of QISP should be understood as a cognitive platform focused on organizational learning, decision traceability, and regulatory transparency. In a dynamic and demanding environment, the integration of these technologies is indispensable for establishing KM as a strategic function in digital QI.

V. Methodology for Structuring and Preparation for TRL3

The nature of this study, centered on the conceptual and technological analysis of Knowledge Management (KM) in Quality Infrastructure (QI), demands a methodological approach that integrates theory, modelling, and application. For exploratory and applied purposes, the research aligns with the logic of projects in the technological formulation phase (TRL 2), aiming to support the transition to TRL 3 (experimental validation).

The approach is qualitative, focused on constructing interpretative and propositional knowledge. It is based on the understanding and systematization of complex phenomena—such as the curation of technical-regulatory knowledge in digital environments—drawing on multiple sources and the integration of theoretical models, institutional practices, and emerging technologies [17].

The envisioned methodology is organized into three integrated axes:

1. Requirements Review and Architecture Analysis In this phase, the requirements of QISP are analysed, with an emphasis on functional flows and the types of knowledge involved: tacit (expert experience), explicit (normative documentation), and automated (algorithmic responses). Traditional QI communication is compared with the flows envisioned for the portal, identifying changes in accessibility, interoperability, and reliability [3]. Conceptual diagrams are also developed using tools such as Napkin AI, graphically representing the transformation of regulatory information into reusable knowledge and identifying the system's points of entry, mediation, and output.

2. Simulation of Cognitive Flows Applied to KM In this stage, use scenarios for QISP are modelled based on four strategic technologies: Intelligent Virtual Assistance, Semantic Search, Big Data/Analytics, and Robotic Process Automation (RPA). For each technology, micro-cycles of KM are designed to demonstrate: the capture of demands and semantic intent (chatbot); contextualized retrieval (semantic search); pattern identification (analytics); and task automation (RPA) [12][14]. These simulations test, in a controlled environment, the integration between technological modules and KM guidelines, particularly regarding the updating and dissemination of regulatory content.

3. Proposition of Criteria for TRL 3 Based on the previous stages, the minimum criteria for transitioning QISP to TRL 3 must be identified. These include: the existence of a conceptually validated digital KM model with defined flows and actors; the implementation of automated routines for capturing, updating, and disseminating normative content; measurement of system responsiveness and assertiveness based on indicators such as response time, semantic adherence, and resolution rate; and the definition of a pilot

testing plan with real users, covering diverse profiles (technicians, entrepreneurs, consumers) and various regulatory themes.

These strategies are essential for consolidating a methodological model capable of supporting the experimental validation of QISP as a KM platform applied to QI, in accordance with TRL 3 requirements.

In this context, considering all the aspects outlined in the study, Figure 2 illustrates the proposed structural foundations for Knowledge Management in QISP—a comprehensive representation of the cognitive operational logic and the digital resources that underpin the portal's proposal.

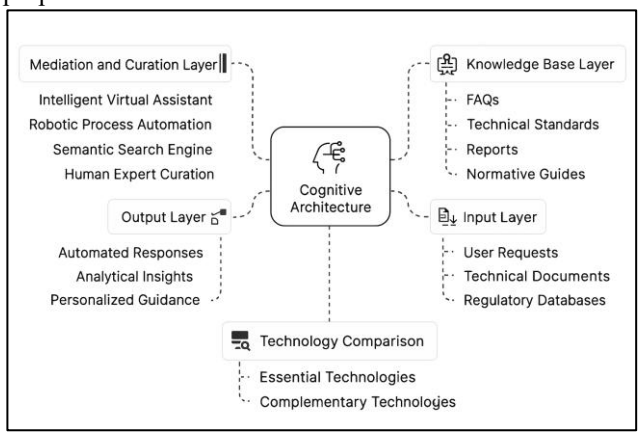


Fig. 2. Cognitive and technological architecture for knowledge Management in PSIQ.

The organization is structured into four interdependent layers: data inputs, which encompass user demands, technical documents, and external normative databases; mediation and curation mechanisms, composed of technologies such as Intelligent Virtual Assistance, RPA, Semantic Search, and specialized curation; a structured knowledge base, where organized normative content, decision flows, and digital repositories are consolidated; and, finally, strategic outputs, corresponding to the delivery of automated responses, predictive analytics, and personalized guidance to users.

Figure 1 also clarifies the distinction between essential and complementary technologies, highlighting how each is positioned and contributes to the governance of technical regulatory knowledge. Thus, it serves as a visual synthesis of the dimensions addressed in Sections III, IV, and V, designed to support the system's conceptual modelling and guide its transition to more advanced levels of technological maturity.

VI. Conclusion and Recommendations

This research explored the conceptual, methodological, and technological foundations required to structure a Knowledge Management (KM) model for the Quality Infrastructure Services Portal (QISP), currently at TRL 2. It demonstrated that the digitization of Quality Infrastructure (QI) requires integrated approaches to transform technical-regulatory knowledge into accessible, updatable, and interoperable assets.

The theoretical analysis revealed that models such as SECI, DIKW, APO, and Cynefin are highly applicable to PSIQ because they enable the conversion of knowledge, the organization of information flows, the alignment of institutional strategies, and the guidance of responses to diverse regulatory contexts. These models underpin KM flows based on human curation, automation, and responsiveness.

In the technological field, four central resources stand out—Intelligent Virtual Assistance, Advanced Semantic Search, Big Data & Analytics, and Robotic Process Automation (RPA)—as fundamental tools for organizing, updating, and disseminating technical knowledge. Complementary technologies such as ontologies, ECM, visual analytics, and digital competencies management further reinforce the sustainability and governance of KM in the digital QI environment.

The proposed methodology demonstrated that preparing for TRL 3 requires more than technical advancement: it is necessary to validate the knowledge architecture, model semantic flows, simulate usage scenarios, and establish evaluation criteria for system maturity. The challenge is to build not only a portal, but also a cognitive platform focused on decision traceability, regulatory transparency, and organizational learning.

Some persistent obstacles include institutional fragmentation, the lack of inter-institutional curation mechanisms, the absence of standardized technical vocabularies, and cultural challenges related to adopting AI-based solutions. However, these barriers also open opportunities for strategic initiatives, such as strengthening IPEM-SP role in digital regulation, expanding access to normative information for micro and small enterprises, and fostering a collaborative KM ecosystem among Sinmetro agencies.

It is recommended that the next project stages include: (i) empirical validation of the simulated flows; (ii) participation of QI specialists in curation and in training the intelligent systems; (iii) use of performance indicators (accuracy, responsiveness, and usability); and (iv) feedback mechanisms based on user interactions.

In conclusion, QISP can consolidate itself as a strategic KM platform in Brazilian QI, provided that its digital architecture is based on scientific foundations, recognized models, and interoperable, adaptive technologies. In advancing in this direction, the portal will not only digitize processes, but also transform the production, mediation, and reuse of regulatory knowledge in the public sector.

References

[1] Reck, J. R., & Hübner, B. H. (2021). “A Transformação Digital do Estado: Digitalização do Governo e dos Serviços Públicos no Brasil”. Revista Eletrônica Direito e Política, 16(3), 1075–1096. https://doi.org/10.14210/rdp.v16n3.p1075-1096.

[2] International Network on Quality Infrastructure, "Quality Infrastructure Definition," Geneva: INetQI, 2018.

[3] Chalmeta, Ricardo & Grangel, Reyes. (2008). Methodology for the Implementation of Knowledge Management Systems. Journal of the American Society for Information Science and Technology.

[4] Mankins, J. C. (1995). Technology Readiness Levels: A White Paper. NASA, Office of Space Access and Technology.

[5] Reck, J. R., & Hübner, B. H. (2021). A Transformação Digital do Estado: Digitalização do Governo e dos Serviços Públicos no Brasil. Revista Eletrônica Direito e Política, 16(3), 1075–1096.

[6] Chalmeta, R., & Grangel, R. (2008). Methodology for the Implementation of Knowledge Management Systems. Journal of the American Society for Information Science and Technology.

[7] Ministério do Desenvolvimento, Indústria, Comércio e Serviços. Relatório Final da Tomada Pública de Subsídios – Estratégia Nacional de Infraestrutura da Qualidade (ENIQ). Brasília: MDIC, 2024.

[8] Nonaka, I., & Takeuchi, H. (1997). The KnowledgeCreating Company. Oxford University Press.

[9] Rowley, J. (2007). The wisdom hierarchy: representations of the DIKW hierarchy. Journal of Information Science, 33(2), 163–180.

[10] Asian Productivity Organization (APO). (2010). Knowledge Management: Facilitators’ Guide. Tokyo: APO.

[11] Pee, L. G. & Kankanhalli, Atreyi. (2009). A Model of Organisational Knowledge Management Maturity. Journal of Information & Knowledge Management.

[12] Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. Machine Learning with Applications, 2, 100006.

[13] López, V., Fernández, M., Motta, E., & Stieler, N. (2017). PowerAqua: Supporting users in querying and exploring the Semantic Web. Semantic Web, 8(3), 403–424.

[14] Chen, M., Mao, S., & Liu, Y. (2014). Big Data: A survey. Mobile Networks and Applications, 19(2), 171–209.

[15] Willcocks, Leslie P., Lacity, Mary and Craig, Andrew, (2015), The IT function and robotic process automation, LSE Research Online Documents on Economics, London School of Economics and Political Science, LSE Library.

[16] Few, S. (2006). Information Dashboard Design: The Effective Visual Communication of Data. O'Reilly Media.

[17] Flick, U. (2009). An Introduction to Qualitative Research. Sage Publications.

Digital Transformation for the United States Air Force Metrology and Calibration Program

Jeremy Latsko [1], Michael Brockway [2], Salvatore Capra [3], Evan Elliott [4]

[1] *Air Force Metrology & Calibration (AFMETCAL), Heath, OH, United States, jeremy.latsko@us.af.mil*
[2] *AFMETCAL, Heath, OH, United States, michael.brockway@us.af.mil*
[3] *AFMETCAL, Heath, OH, United States, salvatore.capra@us.af.mil*
[4] *AFMETCAL, Heath, OH, United States, evan.elliott@us.af.mil*

***Abstract* – The United States Air Force (USAF) Metrology and Calibration program's mission is to ensure USAF readiness through the accuracy, reliability, and traceability (A-R-T) of 700k units of Measurement & Test Equipment (M&TE). While our operations are extensive—500k annual calibrations across 65 global labs—current limitations in data capture, infrastructure, and analytical capabilities prevent us from fully leveraging our data for a modern, data-driven, proactive Quality Infrastructure (QI). Consequently, this paper explains capability gaps between our current and desired end state QIs, the resultant Digital Transformation (DT) strategy, and its technical implementation. This DT strategy will revolutionize our metrology operations, fundamentally improving the quality of A-R-T, and therefore further ensuring USAF readiness.**

I. INTRODUCTION

The Air Force Metrology and Calibration (AFMETCAL) program is the USAF's centralized program office and technical authority for metrology and calibration. Our primary mission is to ensure USAF readiness through A-R-T of 700k units of M&TE [1, 2]. The AFMETCAL mission encompasses 500k annual calibrations across 65 global calibration labs and 35k square meters of lab space. To achieve our mission, our key responsibilities include developing and maintaining 6k USAF calibration procedures, procuring and managing 8k calibration standards, and conducting and funding metrology research and development.

Guided by a focus on "new metrology" and "digital transformation" as identified in the International Committee for Weights and Measures (CIPM) Strategy 2030+, AFMETCAL is pursuing two strategic modernization initiatives [3]. The first invests in "new metrology" by developing quantum-based, intrinsically accurate, and deployable reference standards and M&TE through the NIST-on-a-Chip program [4, 5]. The second, and the focus of this paper, is our DT strategy for a modern, data-driven, proactive QI.

An early gap analysis across A-R-T revealed shortcomings in our current QI, driving the development of our DT strategy based on an enabling robust digital foundation and four technical Lines of Effort: Measurement Assurance, Digital Calibration Chain, Data-Driven Intervals, and AI/ML Analytics. In addition, the adoption of the Digital SI (D-SI) and Digital Calibration Certificates (DCCs) are key components to our strategy [6, 7, 8, 9, 10].

II. DIGITAL TRANSFORMATION

The concept of DT has been described as the total and overall effect of digitalization, which provides a valuable perspective for us [10]. Building on this, we view DT as a strategic organizational change leveraging digitalization to create a data-driven and proactive QI that will improve the quality of A-R-T. Ultimately, improving USAF readiness.

III. THE OBJECTIVE: USAF READINESS

Imagine a fighter jet, armed and fueled, ready to embark on a critical mission. While the jet itself may be fully operational, true military readiness demands more than just a functioning aircraft. If the navigation system is misaligned, the pilot might deviate from the planned flight path or fail to locate critical waypoints. Or if the targeting pod provides inaccurate positioning data, the ability to identify, track, and engage targets effectively could be compromised. In such scenarios, the mission's success—and the pilot's safety—are endangered. This is where metrology and calibration play a critical role. Every sensor, every instrument, every component of that jet relies on the A-R-T of its calibrations to ensure it performs precisely as intended. Calibration, therefore, is not just a technical process but a fundamental aspect of military readiness, demanding the highest levels of A-R-T.

IV. ACCURACY, RELIABILITY, TRACEABILITY: CAPABILITY GAPS

In 2025, AFMETCAL commenced our strategic DT planning. This began with a review of current QI capabilities across A-R-T. These were then compared to newly defined end states based on new concepts, best practices, and capabilities associated with cloud storage, computing, and AI and machine learning algorithms [11, 12, 13]. A subsequent analysis between the current state and desired end state QIs revealed capability gaps outlined in the tables below. This informed us of our technical Lines of Effort discussed later.

A. Accuracy

Table 1. Accuracy capability gaps.

Desired End State	Current Capability	Capability Gap
Integrated MDR* analysis with PFA* thresholds aligned with USAF policy [1]	Reliance on TARs* for risk estimation	No system for TUR calculation or probabilistic risk calculation for global risk with EOPR* and TUR*
Real-time anomaly detection, indicating potential STDs*, M&TE*, or process issues to prevent false acceptance	Real-time analysis not conducted; OOT* only detected at EOP* cal	No system for real-time anomaly detection using cloud computing AI/ML tools
Automated substitution validation ensuring continuity of operations during downtime of STDs	Manual TAR analysis by technicians (error prone)	No systems for TUR calculation or for technicians to select and validate alternate STDs against permitted TUR

* *Defined in the following paragraph.*

Current QI capabilities for accuracy lack advanced risk management and real-time monitoring. Measurement Decision Risk (MDR) analysis with Probability of False Accept (PFA) thresholds, aligned with USAF policy, is absent [1]. Risk estimation relies solely on Test Accuracy Ratios (TAR), excluding End of Period Reliability (EOPR) and Test Uncertainty Ratio (TUR) considerations. Out of tolerance (OOT) detection for standards (STDs) occurs only at the End-of-Period (EOP) calibration, preventing real-time prevention of falsely accepting M&TE. Manual TAR analysis for substitution validation does not represent the full TUR, increasing PFA risk and jeopardizing operational continuity during a standard's downtime.

B. Reliability

Table 2. Reliability capability gaps.

Desired End State	Current Capability	Capability Gap
Interval optimization achieved through data-driven, detailed test point analysis	Analysis based on model groupings	No system for detailed test point analysis at the SN level
Reliable temporary interval extensions through test point data evaluation	Reliance on pass/fail outcomes	Lack of system for using granular test point data for interval extensions
Historical trend and pattern detection, including "bad actor" SNs and indicating potential M&TE, standard, or process issues	Ad hoc manual data collection and analysis	Absence of automated tools to detect trends and patterns based on historical data collection

This analysis for reliability highlights the need for advanced data-driven solutions to optimize calibration intervals and enhance decision-making. Current interval analysis relies on broad model groupings, lacking the granularity of serial number (SN) level test point data. This limits the ability to optimize intervals based on individual equipment performance. Similarly, temporary interval extensions are based on simple pass/fail outcomes, rather than leveraging detailed test point data for more nuanced and reliable extensions. Furthermore, historical pattern and trend detection, including the identification of "bad actor" SNs, relies on ad-hoc manual data collection and analysis. This absence of automated tools for analyzing historical patterns and trends prevents identification of potential issues with M&TE, standards, or processes.

C. Traceability

Table 3. Traceability capability gaps.

Desired End State	Current Capability	Capability Gap
End-to-end digital traceability with DCCs containing test point data	Limited digital storage; primary lab and 65 cal labs operate on separate MIS's*	No integrated MIS and no established framework for implementing DCCs [7, 9]
Automated recall analysis to identify previously calibrated M&TE that exceed PFA requirements	Labor-intensive manual reviews w/o PFA analysis functions	No integrated MIS, no MDR, and no established framework for implementing DCCs [7, 9]

* *Defined in the following paragraph.*

To achieve end-to-end digital traceability, a fully integrated system leveraging DCCs is essential [7, 9]. Current practices rely on limited digital storage and fragmented Management Information Systems (MIS), with primary and calibration labs operating on separate platforms. This lack of integration hinders the implementation of DCCs and prevents seamless data flow across the traceability chain. Consequently, comprehensive test point measurement data is not readily available in a machine-readable, interpretable format. Furthermore, recall analysis is currently a labor-intensive manual process, lacking automated PFA analysis functionality. This hampers the timely identification of calibration points exceeding PFA requirements due to OOT standards.

V. DIGITAL TRANSFORMATION FRAMEWORK

Following our gap analysis, a clear vision for a transformed QI emerged, leading to the development of our DT strategy. We started with our DT framework focused on ensuring USAF readiness through the improved quality of A-R-T. Figure 1 depicts our DT framework with the digital foundation and four pillars representing the four technical Lines of Efforts—or "how

to's"—which are each detailed later in the paper. Figure 2 provides a technical implementation overview.

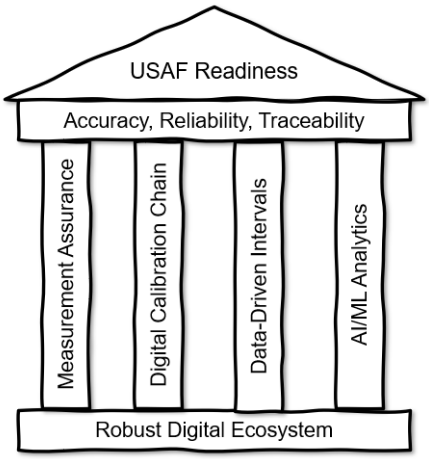


Fig. 1. AFMETCAL's DT framework, built upon a digital foundation and four Lines of Effort. These directly enable USAF readiness through improved quality of A-R-T.

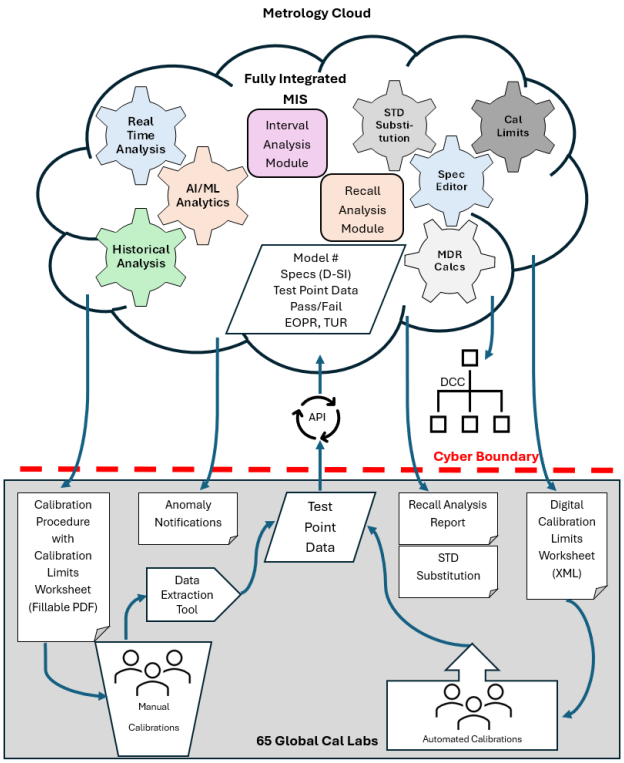


Fig 2. Technical overview of Digital Transformation.

VI. ROBUST DIGITAL ECOSYSTEM

AFMETCAL's DT strategy is built upon a robust digital ecosystem foundation rooted in digital specifications for M&TE and reference standards, comprehensive digital test point measurement data, and cloud storage and computing capabilities. These elements are essential enablers to our Lines of Effort to bridge the capability gaps.

A key component of our DT strategy is the transition to digital specifications. This is achieved through the digitization of existing paper-based specifications and the migration of current digital specifications into the D-SI schema published by Physikalisch-Technische Bundesanstalt (PTB) [6, 7, 8]. The D-SI, underpinned by the BIPM's SI Brochure, the Guide to the Expression of Uncertainty in Measurement (GUM), and fundamental physical constants from CODATA, provides a standardized and increasingly internationally recognized framework for representing quantities and SI units [7]. This approach will ensure compatibility with systems used by the global metrology community, facilitating future data exchange and collaboration.

To create these digital specifications, technical staff will use an organically developed and maintained Specification Editor to input M&TE and reference standard specifications. The specifications will be stored in a centralized, cloud-based database.

Complementing digital specifications, AFMETCAL will leverage the DCC standard, utilizing the XSD schema and XML format developed by PTB [6, 9]. Each calibration will result in a comprehensive DCC, capturing not only high-level aspects of the calibration (e.g., reference standards, environmental conditions), but critically the individual test point measurement data. This granular digital data will be stored within the DCC, which will then be stored in a secure cloud environment.

The standardized nature of the D-SI and the comprehensive data captured in DCCs, readily accessible in the cloud, will unlock the potential of AI/ML algorithms to address A-R-T capability gaps [6, 7, 8, 9].

VII. THE FOUR LINES OF EFFORT

To address the capability gaps identified in Tables 1-3 and achieve our DT, we have established four key Lines of Effort, visually represented as pillars in Figure 1. Each Line of Effort incorporates technical solutions designed to close these gaps, as detailed in the following sections.

Note: In some of the discussion below, artificial intelligence and machine learning algorithms are identified as potential technical solutions. While these approaches are considered viable, this paper does not provide an in-depth analysis of AI/ML applications. Instead, relevant references are included for further research [12, 13, 14, 15]. As we are not AI/ML technical experts, we are collaborating with USAF specialists in this domain to select and deploy the best AI/ML techniques.

A. Line of Effort One: Measurement Assurance

A key objective of our DT strategy is to shift from reactive to proactive measurement assurance, minimizing false acceptance and increasing quality in results. Currently, guard band calculations rely on TARs—lacking the full TUR calculation and the necessary probabilistic calculation for global risk [16]. Furthermore, manual standard substitution processes are error prone. These limitations hinder efficient and reliable measurement assurance. The following table details our DT's technical solutions for measurement assurance desired end states:

Table 4. Technical solutions for measurement assurance.

Desired End State	Technical Solution
Integrated MDR analysis with PFA thresholds aligned with USAF policy [1]	The MDR Calculations Module will allow AFMETCAL technical staff to calculate TURs rooted in specs entered via the Specification Editor. Following ANSI Z540.3 Handbook Method 2 (and eventually Method 1), it will then calculate guard bands at USAF's required 2% PFA using global risk calculations with TUR—when less than 4:1—and EOPR for global risk [1, 16]. These guard bands then populate the requested Calibration Limits Worksheet for automated or manual calibration.
Automated substitution validation ensuring continuity of operations during downtime of STDs	The MDR Calculations Module will allow AFMETCAL technical staff to calculate TURs. Using specs pulled from the Specification Editor, the technician will use the Standard Substitution Module to select candidate substitution STDs. If the TUR is greater than or equal to the permitted TUR in the USAF calibration procedure, new guard bands will be calculated and then populate the requested Calibration Limits Worksheet for automated or manual calibration.

To achieve integrated MDR analysis, the MDR Calculations Module will leverage the Specification Editor and ANSI Z540.3 methodologies to enable complete TUR calculations and guard bands, aligning PFA with USAF policy [1, 16]. These guard bands will automatically populate Calibration Limits Worksheets. For automated substitution validation, the Standard Substitution Module, coupled with MDR Calculations Module, will enable technicians to select acceptable standards (i.e., acceptable TURs and PFA) to ensure operational continuity during a standard's downtime.

B. Line of Effort Two: Digital Calibration Chain

Ensuring robust traceability is essential for confidence in measurement results. A fully realized digital calibration chain links measurements from the flightline to the SI through National Metrology Institutes, ensuring an unbroken, documented record. Currently, AFMETCAL faces challenges such as fragmented MIS's, limited digital data storage, and the absence of DCCs—hindering traceability and alignment [6, 9]. Additionally, recall analysis remains a manual process lacking PFA functionality, delaying error identification and potentially impacting USAF readiness. The following table details our DT's technical solutions for digital calibration chain desired end states:

Table 5. Technical solutions for digital calibration chain.

Desired End State	Technical Solution
End-to-end digital traceability with DCCs containing test point data	Each calibration will have a DCC including calibration test point data following PTB's DCC schema [9]. The Fully Integrated MIS will store all DCC's in machine readable format.The MIS will also link STDs used by unique ID numbers and Job Control Numbers (JCN) to ensure end-to-end traceability.
Automated recall analysis to identify previously calibrated M&TE that exceed PFA requirements	Upon detecting an OOT STD during calibration, the Recall Analysis Module will identify all calibrations performed in its last in-service cycle using the STD's unique ID and JCN. The MDR Calculations Module will update uncertainty, recalculating guard bands per USAF policy [1]. Prior M&TE measurements exceeding new limits will be flagged, generating a Recall Analysis Report for management review and potential recall.

To achieve end-to-end digital traceability, each calibration will have a DCC following PTB's schema, including complete test point measurement data [9]. A Fully Integrated MIS will store DCCs in machine-readable format and link standards using unique ID numbers and Job Control Numbers (JCN) for seamless traceability. When an OOT standard is detected, the Automated Recall Analysis Module will activate, identify impacted calibrations, recalculate guard bands, and flag affected M&TE exceeding PFA requirements [1, 16]. A Recall Analysis report will be generated for lab management review and potential recall actions.

C. Line of Effort Three: Data-Driven Intervals

Optimizing calibration intervals is crucial for minimizing user operational downtime while ensuring measurement reliability. Currently, AFMETCAL's interval analysis is based on model groupings, lacking the granularity of test point data. This generalized approach leads to suboptimal intervals. Furthermore, current temporary interval extensions rely on limited pass/fail outcomes, hindering the ability to provide targeted extensions based on individual M&TE performance. This can negatively impact USAF readiness when equipment is unnecessarily taken out of service. The following table details our DT's technical solutions for data-driven interval desired end states:

Table 6. Technical solutions for data-driven intervals.

Desired End State	Technical Solution
Interval optimization achieved through data-driven, detailed test point analysis	The Interval Analysis Module will calculate calibration intervals using the NCSLI RP-1 Exponential Method [17]. Trained machine learning detection algorithms (e.g., One-Class SVM, Isolation Forest) will analyze test point data by serial number, identifying outliers [12, 13]. An engineer will review the generated report and may intervene by correcting procedures, limiting ranges, or creating new model groupings for interval optimization.
Reliable temporary interval extensions through test point data evaluation	A trained machine learning algorithm (e.g., Gradient Boosted Decision Trees), using factors like historical test point stability, current UUT/STD disagreement, comparable model performance, and existing calibration interval, will generate a reliability score and supporting data [14]. An engineer will review this information to determine—if appropriate—a tailored temporary interval extension.

Calibration intervals are optimized using machine learning (e.g., One-Class SVM, Isolation Forest) to analyze test point data, identify outliers, and inform engineer interventions [12, 13]. Reliable temporary extensions are enabled by a machine learning algorithm (e.g., Gradient Boosted Decision Trees) that considers historical stability, UUT/STD agreement, and other factors to generate a reliability score for engineer review and tailored extension [14].

D. Line of Effort Four: AI/ML Analytics

AI/ML offers a transformative shift in calibration data analysis by moving beyond manual post-calibration reviews to proactive anomaly detection. Current methods rely on labor-intensive evaluations prone to human error, delaying issue identification. Machine learning models can streamline this process by detecting patterns and outliers flagging inconsistencies in Measurement & Test Equipment (M&TE) performance at the unit level—rather than relying solely on model group aggregates. This improvement enables earlier interventions, enhances measurement integrity, and strengthens overall reliability. The following table details our DT's technical solutions for AI/ML analytics desired end states:

Table 7. Technical solutions for AI/ML analytics.

Desired End State	Technical Solution
Real-time anomaly detection, indicating potential STDs*, M&TE*, or process issues to prevent false acceptance	Real-time calibration data, streamed via Amazon Kinesis Data Streams, will be analyzed by tuned machine learning algorithms (e.g., One-Class SVM, Isolation Forest) to detect outliers [12, 13]. Trained machine learning algorithms (e.g., Gradient Boosted Decision Trees) are highly effective for pattern identification in data [14]. Detected anomalies will trigger Anomaly Notifications to technicians for investigation. Any ML algorithm will be tuned to minimize false alarms.
Historical trend and pattern detection, including "bad actor" SNs and indicating potential M&TE, standard, or process issues	Historical calibration data, transported by Amazon Kinesis Data Firehose to Amazon S3, will be analyzed using Isolation Forest and DBSCAN to detect systematic patterns and long-term trends in calibration performance [13, 15]. DBSCAN will cluster serial numbers exhibiting consistent drift or instability, flagging potential M&TE, STD, or process issues for engineers to review [15]. Any ML algorithm will be tuned to miniize false alarms.

Machine learning models analyze real-time calibration data, detecting outliers while minimizing false alarms. Gradient Boosted Decision Trees identify patterns, triggering Anomaly Notifications for technician review [14]. Historical trend analysis, using Isolation Forest and DBSCAN, clusters serial numbers with consistent drift, flagging potential M&TE, standard, or process issues [13, 14]. These automated insights improve measurement reliability and enable proactive issue resolution.

VIII. CONCLUSION

AFMETCAL's DT strategy, driven by the imperative for USAF readiness, addresses critical gaps in our current QI. Limitations in data utilization and analysis hinder our ability to fully ensure the A-R-T of our extensive M&TE inventory. This strategy, built on a robust digital ecosystem incorporating the D-SI and DCCs, enables four key Lines of Effort: Measurement Assurance, Digital Calibration Chain, Data-Driven Intervals, and AI/ML Analytics [6, 7, 8, 9]. Accomplishing these will result in a proactive and data-driven QI, revolutionizing AFMETCAL operations, and fundamentally improving the quality of A-R-T. Most importantly, these efforts will further ensure USAF readiness for safe and effective operations across every mission.

AFLCMC Public Affairs Case Number AFLCMC-2025-0078

REFERENCES

[1] United States Air Force. (2025). Technical Order 00-20-14: Air Force Metrology and Calibration (AFMETCAL) Management. Published under the authority of the Secretary of the Air Force.
[2] United States Air Force. (2020). Air Force Metrology and Calibration Program Management (AFMAN 21-113). Published under the authority of the Secretary of the Air Force.
[3] International Committee for Weights and Measures. (2020). CIPM Strategy 2030+. Sevres, France: Bureau International des Poids et Mesures.
[4] Latsko, J., Fritzsche, R., & Kauffman, R. (2024). Utilizing the SI redefinition and "NIST on a chip" towards ensuring readiness across the United States Air Force, Army, and Navy. XXIV IMEKO World Congress "Think Metrology". Hamburg, Germany.
[5] National Institute of Standards and Technology. (n.d.). NIST on a Chip. Retrieved from https://www.nist.gov/noac
[6] Physikalisch-Technische Bundesanstalt. (2019). D-SI in Short. Retrieved from https://www.ptb.de/empir2018/fileadmin/documents/empir/SmartCom/documents_for_download/D-SI_in_Short_2019-11-04_DE_PTB_NPL.pdf
[7] National Physical Laboratory. (2019). Digital system of units, digital calibration certificate, DCC-XML schema.
[8] Physikalisch-Technische Bundesanstalt. (2022). SI_Format.xsd. Retrieved from https://www.ptb.de/si/v1.3.1/SI_Format.xsd
[9] Physikalisch-Technische Bundesanstalt. (2022). DCC XSD Schema. Retrieved from https://www.ptb.de/dcc/v3.1.0/dcc.xsd
[10] Eichstädt, S., Elster, C., & Schwartz, R. (2021). Brief overview of the future of metrology. Measurement: Sensors, 18, 100269. Retrieved from https://www.sciencedirect.com/science/article/pii/S2665917421002695?via%3Dihub
[11] NCSL International. (2014). Recommended Practice 18: Estimation and evaluation of measurement decision risk. NCSL International. Retrieved from NCSLI.
[12] Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., & Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. Neural computation, 13(7), 1443-1471.
[13] Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008, December). Isolation forest. In 2008 eighth ieee international conference on data mining (pp. 413-422). IEEE.
[14] Machine Learning Plus. (n.d.). An Introduction to Gradient Boosting Decision Trees. Retrieved from: https://www.machinelearningplus.com/machine-learning/an-introduction-to-gradient-boosting-decision-trees/
[15] Sharma, P. (2023, June 26). DBSCAN Clustering Algorithm Explained. Built In. Retrieved from: https://builtin.com/articles/dbscan
[16] NCSL International. (2009). Handbook for the Application of ANSI/NCSL Z540.3-2006: Requirements for the Calibration of Measuring and Test Equipment. NCSL International.
[17] NCSL International. (2010). RP-1: Establishment and Adjustment of Calibration Intervals. Published by NCSL International.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Measurement Data and Information of Non-Automatic Weighing Instruments as Structured Data

Tatyana Sheveleva[1], Gisa Foyer[1]

[1] *Physikalisch-Technische Bundesanstalt, Bundesallee 100, 38116 Braunschweig,*
tatyana.sheveleva@ptb.de; gisa.foyer@ptb.de

***Abstract* – The successful digitization of documents in conformity assessment, e.g. for calibration and test, requires both the compliance with different requirements and the use of a unified machine-readable data structure. A well-known example is the data structure for the Digital Calibration Certificate (DCC). Its use is proposed by the German Calibration Service (DKD) in its Report DKD-E 7-3 for the digitization of calibration certificates for non-automatic weighing instruments (NAWIs). An analysis has been carried out to check whether the proposal can be applied to OIML R 76 test reports. By comparing the measurement data and information in both types of documents, similarities and differences in the content description were identified. The following publication gives an overview of this analysis and its results and provides suggestions for the description of the test reports using the DCC.**

I. INTRODUCTION

The digitization of documents used in conformity assessments of non-automatic weighing instruments (NAWIs), such as calibration and test, poses several challenges: On the one hand, all relevant requirements from different standards and other normative sources have to be considered. On the other hand, a unified machine-readable data structure needs to be applied, that can cover documents for a wide range of products. A good example to this issue is the already established machine-readable data structure for the Digital Calibration Certificate (DCC) [1]. The DCC is intended to replace the paper-based calibration certificates issued until now in accordance with ISO/IEC 17025 [2]. The German Calibration Service (DKD) has published an Expert Report DKD-E 7-3 [3], which, as a result of extensive discussions, includes a detailed guideline for the content description of the NAWI calibration certificates based on the requirements of EURAMET/CG-18/v.02 [4], the Guide on the Calibration of NAWIs.

Since the types of conformity assessment laid down in ISO/IEC 17025 [2] also cover testing, digitizing the corresponding test reports for NAWIs seems to be useful too. NAWIs are tested extensively to support type evaluation of NAWIs, i.e. conformity assessment of the instrument type within legal metrology. The respective requirements, tests and templates for test reports are standardized in the International Recommendations OIML R 76 [5], [6].

Due to the fact that the content of NAWI test reports and NAWI calibration certificates is largely similar, it makes sense to check whether the DCC data structure proposed in DKD-E 7-3 [3] could also be applied to the test reports. For this, a usability analysis was carried out to identify the similarities and the differences between the content of both document types. The detailed description of this analysis and the discussion of its results are therefore the subject of this publication.

The publication is structured as follows: After a brief introduction to the subject, a short overview of overall requirements for the description of the content within a calibration certificate and test reports is given in Chapter II. Chapter III summarizes the findings and gives an outlook of possible future activities.

II. MEASUREMENT DATA AND INFORMATION OF NON-AUTOMATIC WEIGHING INSTRUMENTS

The overall requirements for reporting the results of a calibration and a test are set out in ISO/IEC 17025 [2]. It states that the results shall be provided unambiguously, usually as a report or certificate, and include all data and information necessary to interpret the results correctly [2]. In addition, the standard provides a list of mandatory information, that each calibration and test report must contain. According to them, both types of documents must, among other things, include information on the object of conformity assessment, the measurement methods applied, the influence conditions, and the measurement results which are obtained and used for further interpretation. All of these content elements have been considered for a comparative analysis and represented in detail in separate sub-chapters. For better differentiation the following notation is applied: the DCC elements are in bold, and the quantities are in lower case italics.

A. Object of conformity assessment

ISO/IEC 17025 [2] and EURAMET/CG-18/v.02 [4] require a calibration certificate to provide a detailed description of the object of conformity assessment. Based on this, the DCC structure has been proposed [2], as shown in Fig. 1.

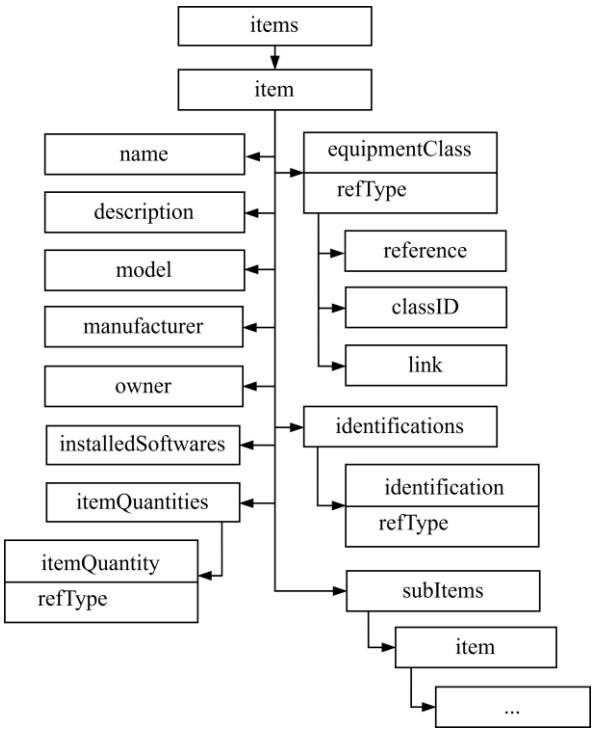


*Fig. 1. DCC structure of **dcc:items** [1].*

An individual calibrated NAWI is described by the element **dcc:item**. Its sub-elements allow to provide a set of further descriptive information. Using the element **dcc:identification**, one or more identification type can be added. Each identification is referenced by a **refType** attribute [3]. The **refTypes** are predefined and can be selected from the centralized "Metrology refType Database" [7]. Here, several types of **refType** are differentiated by their prefixes: The prefix **basic_** relates to the universal methods suitable to different application domains, quantities and instruments [7]. The prefix **NAWI_** refers to the methods used only for the calibration of NAWIs [7]. The prefix **math_** relates to universal mathematical **refTypes**, which are usually used in combination with other **refTypes** to specify them [7]. To identify a NAWI, the **refType basic_serialNumber** [7] ("serial number") is applied, as this is the typical identification type used in calibration certificates [7].

The element **dcc:equipmentClass** and its sub-element **dcc:classID** are used to specify a category of the NAWI. As there is no main specified classification of the NAWIs within the scope of calibration, this element was used to describe the categorization of NAWIs based on one of the metrological characteristics, such as "single range instrument" ("NAWI-SR"), "multiple range instrument" ("NAWI-MR") or "multi-interval instrument" ("NAWI-MI"). The **dcc:reference** and **dcc:link** elements are used in combination with these predefined categories [3].

In cases where the calibrated NAWI is represented as a "complex device" incorporating different "devices" and "parts", they can be presented using **dcc:subItems**. Each of them can then be separately described in detail by the sub-element **dcc:item**. In addition, this sub-element is used to specify several (partial) weighing ranges in case of a multiple range or multi-interval NAWI [3]. The following **refTypes** have been created to refer the item to the ranges: **NAWI_range1**, **NAWI_range2**, **NAWI_range3**, and **NAWI_range4** [7].

All metrological nominal and reference characteristics (**dcc:itemQuantities**) can be separately described by **dcc:itemQuantity** in combination with a **refType.** (Here, the element **dcc:itemQuantity** is one of the variants of the element **dcc:quantity**, which can also be used within other elements such as **dcc:influenceConditions**, **dcc:results**, or **dcc:usedMethods** [3], as shown in Fig. 2.

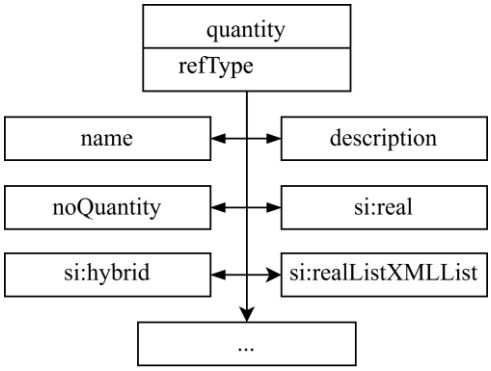


*Fig. 2. DCC structure of **dcc:quantity** [1].*

The following characteristic are already covered by the **refTypes**:

- *actual scale resolution* (*d*) (**NAWI_resolutionOfDisplayingDevice**) [7],
- *verification scale interval* (*e*) (**NAWI_verificationScaleInterval**) [7],
- *nominal value* **(basic_nominalValue**) [7].

For the description of range characteristics, the two different **refTypes** are usually combined to express quantities such as *maximum weighing capacity* (*Max*) (**math_maximum basic_nominalValue** [7]) and *minimum weighing capacity* (*Min*) (**math_minimum basic_nominalValue** [7]) [3].

The requirements for describing a NAWI in the framework of tests according to OIML R 76-2 [6] are similar to those for calibration in several aspects. For example, an unambiguous identification is also mandatory. As the NAWI is tested for a subsequent type evaluation, its "type designation" must be specified [6]. This information

can be covered by **dcc:identification** [1] and its sub-elements, but the corresponding **refType** should be created.

In terms of classification, there is no predefined uniform one. For tests, NAWIs can be classified according to their accuracy class ("special" (I), "high" (II), "medium" (III), and "ordinary" (IIII)), as the accuracy is directly related to further main characteristics such as *verification scale interval* (*e*), *number of verification scale intervals* (*n*), and the *minimum capacity* (*Min*), which are important to describe the NAWIs in a required level of detail [5].

There is also no unified classification for the incorporated "devices" and "parts" of the NAWIs. Instead, different test methods require different classification, each focusing on a specific characteristic. For example, the test report on "weighing performance", "eccentricity" or "repeatability" must include the cate of "tracking device" and/or "setting device", if available. For the "eccentricity test using a rolling load", the "load receptor" has to be specified [6]. This can be described by **dcc:equipmentClass** within **dcc:subItem**.

The main characteristics of NAWI(s), such as *Min*, *Max*, *d*, *e*, and *resolution during the test* [6] can be described by **dcc:itemQuantities**, since these are also partially relevant to a calibration and the required **refTypes** are available [3]. The **refTypes** are still missing for such main characteristics as the *number of verification scale intervals* (*n*) as well as for specific characteristics that are not relevant for calibration and are only important for certain test methods. These are, for example:

- *nominal voltage* (U_{nom}),
- *maximum voltage* (U_{max}),
- *minimum voltage* (U_{min}), and
- *frequency* (*f*)

necessary for the test method "electrical disturbances" [6].

Among the method-specific characteristics, there are also those that carry only qualitative information. As part of an "eccentricity test", for example, it is important to document whether the NAWI is mobile. For a "discrimination test", the characterization in terms of the type of indication is required, as this determine which variant of the "discrimination test" is to be applied [6]. These characteristics can be either described by **refTypes**, which have also to be created, or by the element **dcc:noQuantity**, as shown in Fig. 2.

B. Methods

According to the common requirements of ISO/IEC 17025 for the reporting of results, the identification of the method used is considered as mandatory for each calibration certificate and test report [2]. To cover this requirement, the element **dcc:usedMethods** with its sub-element **dcc:usedMethod** in combination with a **refType** and a further detailed description [3] has been proposed in the DCC [8], as shown in Fig. 3.

In the context of calibration, this requirement is

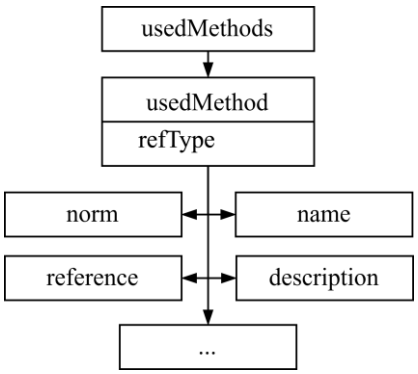


*Fig. 3. DCC structure of **dcc:usedMethods** [1].*

specified for NAWIs more precisely in EURAMET/CG-18/v.02 [4]. Based on this, DKD-E 7-3 [3] defined **refTypes** for the "calibration method" (**NAWI_calibrationMethod**) and for the "measurement uncertainty" (**basic_methodMesasurementUncertainty**) [7].

In contrast to a calibration, a test of NAWI(s) generally includes a larger range of test methods, as these are used for the subsequent type evaluation. A general overview and the detailed description of the test methods is given in OIML R 76-1 [5]. OIML R 76-2 [6] provides a standardized format for the respective test reports. However, all these methods cannot currently be digitized by **dcc:usedMethod**, as the corresponding **refTypes** are missing. Two solutions are possible: (i) to use the general approach as in DKD-E 7-3 [3] and define a new **refType** (**NAWI_testMethod**) or (ii) to define a new **refType** for each test method, e.g. **NAWI_eccentricityTestMethod** or **NAWI_weighingPerformanceTestMethod**. In both cases the DCC is flexible enough to provide further human-readable description.

C. Conditions

According to ISO/IEC 17025 and EURAMET/CG-18/v.02, the conditions must be a mandatory part of each calibration certificate, because they affect the result of the calibration [2], [4]. This requirement has been met in the DCC by use of **dcc:influenceConditions** [8], as shown in Fig. 4.

The following main "influence conditions" have been created so far as **refTypes**: **basic_humidityRelative**, **basic_adjustment**, **mass_airDensity**, **mass_airPressure**, and **basic_temperature** [7].

"Influence conditions" in the DCC are not further specified in categories, such as "operational", "environmental" or "disturbances", but can be expressed in different ways within the element **dcc:data**, as in Fig. 4, among other forms, as text (**dcc:text**) or quantity (**dcc:quantity**), according to the principle described in Sub-chapter A, presented in Fig. 2. So, a *temperature* (**basic_temperature**) [7] can be described by **si:real** with

a single value, unit, and (if applicable) measurement uncertainty. It is also possible to document a *temperature range*. For this, the **refType basic_temperature** [7] is used in combination with the **refTypes math_maximum** or **math_minimum** [7], each described by a **si:real**. In addition, there are also conditions that do not require a numerical value but freely formulated explanations or expressions from a predefined list. For example, the "adjustment status" (**basic_adjustment)** [7] is described by **dcc:status** in combination with one or more explanatory notes (**dcc:text**) [3]. For "status", several values are predefined in DCC ("beforeAdjustment", "afterAdjustment", "beforeRepair", and "afterRepair"), as they are required by ISO/IEC 17025 [2].

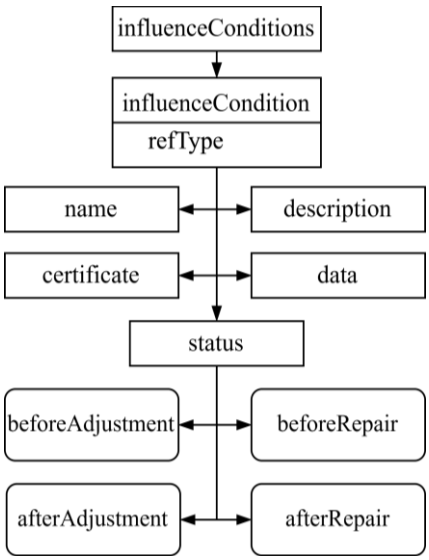


*Fig. 4. DCC structure of **dcc:influenceConditions** [1].*

In the framework of an OIML R 76 test, there are several conditions that must be considered in all test methods required for a type evaluation, as well as those that are method-specific. For instance, an "environmental condition" such as *temperature* is mandatory for each test method. The optional "environmental conditions" *relative humidity* and *barometric pressure* (*Bar. Pres.*) are only documented for specific types of test methods or for testing NAWIs with a specific accuracy class [5]. Each of these conditions can be expressed by one to three individual values that relate to the start, the end of the measurement and/or to the maximum value recorded during the measurement [6].

The DCC structure proposed for the calibration is also partially applicable to a test. Thus, the three "environmental conditions" expressed by a single maximum value, can be fully covered by both the **refTypes** and **si:real**. However, it is not possible to relate these conditions to the start and the end of the measurement, as no **refTypes** are provided. These should be added for the use in test reports. In addition, there is no possibility to express several "operational conditions" which provide the information on how the "devices" and "parts" incorporated in the NAWI operate during the measurement. For example, for some test methods, it is required to specify, whether the "automatic zero-setting device" and "automatic zero-tracking device" are "not-existent", "not in operation", "out of working", or "in operation" [6]. As these expressions are not relevant to calibration and therefore not part of the DCC data structure, they should be given in a harmonized form, similar to the "adjustment status".

D. Measurement results

The description of the measurement results within a calibration according to requirements of [2] and [4] is realized by **dcc:results** [3], as shown in Fig. 5:

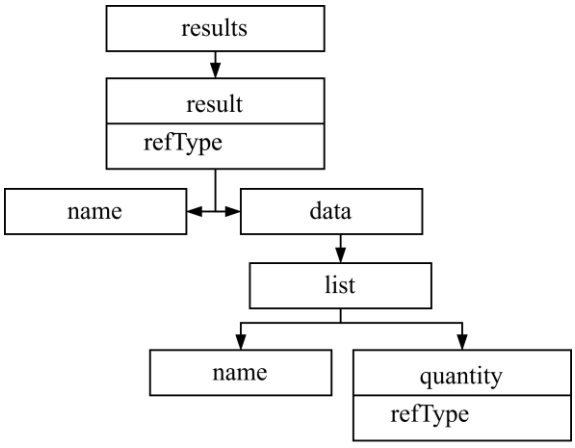


*Fig. 5. DCC structure for **dcc:results** [1].*

Here, an individual measurement can be specified using the sub-element **dcc:result** in combination with one of the following predefined **refTypes**:

- **NAWI_repeatabilityMeasurement** ("repeatability measurement") [7],
- **NAWI_eccentricityMeasurement** ("eccentricity measurement") [7],
- **NAWI_errorOfIndication** ("error of indication") [7], and
- **NAWI_auxiliaryMeasurement (**"auxiliary measurement"**)** [7].

Within **dcc:data**, each quantity (**dcc:quantity**) is represented by a **refType** [3]. The typical **refTypes** used in calibration are:

- **basic_nominalValue** (*nominal value*) [7]
- **basic_referenceValue** (*reference value*) [7],
- **basic_measuredValue** (*measured value*) [7], and
- **basic_measurementError math_maximum** (*maximum deviation*) [7].

The description of the quantities depends on their type and follows the pattern specified in the Digital System of

Units (D-SI) [8]. For example, the *nominal value* (**basic_nominalValue**) [7] of a test load as part of a "repeatability test" is usually documented as a single numerical value with the corresponding unit. In a calibration certificate, it is described using **si:real** [8] as in Table 1:

Table 1. D-SI-based description of a nominal value [3].

Human Readable Equivalent	Sub-elements of si:real	Value Example
value	**si:value**	2
unit	**si:unit**	\kilogram

The same pattern can also be applied to a *reference value* (**mass_referencedValue**) [7].

The *measured values* (**basic_measuredValue**) [7] obtained in one measurement sequence of the "eccentricity test" shall be related to the position of the test load on the load receptor [5]. For this, the following D-SI pattern has been applied, as given in Table 2:

Table 2. D-SI-based description of measured values [3].

Human Readable Equivalent	Sub-elements of si:realListXMLList	Value Example
load position	**si:labelXMLList**	Position1 Position2 Position3 …
measured values	**si:valueXMLList**	0.1004 0.1005 …
SI unit	**si:unitXMLList**	\kilogram

The description of the measurement results in the framework of the OIML R 76-2 test reports [6] have some similarities to the calibration certificate. For example, measurement types such as "repeatability measurement", "eccentricity measurement", and "error of indication" can be covered by the respective **refTypes** used for a calibration. However, OIML R 76-2 also includes a large number of other measurement types for which no **refTypes** have yet been provided, as they are not relevant for calibration, such as "discrimination and sensitivity", "tare", "warm-up time", "voltage variation", and "electrical disturbances" [6].

Most of the main quantities used in a test can be covered by the existing **refTypes** and D-SI quantities. So, the **basic_nominalValue** can be applied to describe the *nominal value* of different types of loads, such as:

- "test load",
- "additional load" (ΔL), which is only required for NAWIs without higher resolution,
- "zero-load" (L_0) as load at or near zero indication, and
- "tare-load" needed, among others, for the test method "stability of equilibrium" [6].

The **refType basic_measuredValue** can be used for an *indicated weight value* (I) [6], which is always an individual value. In cases where several values are obtained depending on the position of the test load, such as in an "eccentricity test", the D-SI quantity pattern **si:realListXMLList** can be reused here. The **refType basic_measurementError** is applicable to an *error of indication* (E) [6], while *mpe* is part of the statement of conformity and serves as a **basic_toleranceLimitUpper** and **basic_toleranceLimitLower** [7]. This is similar to the *tolerance* in [4] but was not considered in [3].

Some test-specific quantities do not have **refTypes** because they are not applicable within a calibration, such as the *corrected error of indication* (E_c) as the difference between the *error of indication* and the *error calculated at or near zero indication* (E - E_0). Additionally, the **refType** for E_0 (**zeroReduced** [7]) actually exists only in relation to the force-related quantity and could be adopted for NAWIs within the framework of calibration or tests. Some measurements, such as those carried out within "repeatability test", requires the difference between the *maximum* and *minimum error of indication* ($E_{max} - E_{min}$) in order to make the result statement about the success or failure if the test. Since the "repeatability test" within a calibration does not require such a difference but a *standard deviation* of the measured value, a **refType** for ($E_{max} - E_{min}$) shall be created. Furthermore, the **refType basic_referenceValue** from the calibration is not applicable, as tests based on OIML R 76-2 are only carried out with the *nominal values* of the loads [6].

III. CONCLUSIONS

The comparison showed that the DCC data structure for a NAWI calibration certificate can largely be applied to the test reports according to OIML R 76 [5], [6]. Most of the information/data that occur in the test reports can be captured by the DCC data structure, e.g. the object of conformity assessment, the method, the conditions, and the results.

In general, it is not possible to describe test-specific information that does not appear in a calibration certificate and is therefore not reproducible in machine-readable form using DCC. In many cases, this is caused by missing **refTypes**. With regard to the description of the object of conformity, these are, for example, the identification as "type designation" as well as some specific metrological and functional characteristics that can be treated as "influence conditions". As most of these conditions form the basis of specific test methods as part of a type evaluation, they do not occur in this form in a calibration. This also concerns the **refTypes** for the test methods as such.

Another important aspect is the difference in how the classification of the NAWIs is treated. For a calibration, the categorization was fixed in terms of measurement ranges and measurement intervals [3], as these are the focus of each calibration. The categorization for a test is more complex. However, each classification used within

different tests can be described by the DCC due to its generalized structure.

Despite the existing differences between the two types of documents presented here, it is extremely important to adhere to existing structures and procedures during digitization to maintain consistency in such a wide area as conformity assessment. In addition, many other areas related to conformity assessment, such as manufacturing, standardization or market surveillance, may also benefit from the harmonization, especially those, where the data and information from both calibration certificates and test reports can be reused, as e.g. Asset Administration Shell (AAS) or Digital Product Passport (DPP).

IV. ACKNOWLEDGMENTS

This publication was written within the project framework German QI-Digital Initiative, funded by the Federal Ministry for Economic Affairs and Energy.

REFERENCES

[1] Physikalisch-Technische Bundesanstalt, “DCC – Digital Calibration Certificate.” Accessed: Apr. 25, 2025. [Online]. Available: https://www.ptb.de/dcc/v3.3.0-rc.1/autogenerated-docs/Dokumentation.html

[2] ISO/IEC, “ISO/IEC 17025:2017: General requirements for the competence of testing and calibration laboratories,” 2017.

[3] Deutscher Kalibrierdienst, “DKD-E-7-3: Instruction in how to use the DCC schema to create a digital calibration certificate for non-automatic weighing instruments,” 2025, *DKD, Braunschweig*.

[4] European Association of national Metrology, “Calibration Guide EURAMET/cg-18/v.02 – Englisch Version,” 2009, *EURAMET e.V, Braunschweig*.

[5] International Organization of Legal Metrology, “OIML R 76-1: Non-automatic weighing instruments. Part 1: Metrological and technical requirements – Tests. 2006 (E),” 2006, *Bureau International de Métrologie Légale, Paris*. Accessed: Jun. 22, 2025. [Online]. Available: https://circabc.europa.eu/webdav/CircaBC/nbg/NoBoMet%20(Notified%20Bodies%20in%20Legal%20Metrology)/Infor mation/news_5ba7ba0a-bf8b-4421-8078-2f757d68a80b/NoBoMet_Documentation_D-CoC_V1

[6] International Organization of Legal Metrology, “OIML R 76-2: Non-automatic weighing instruments. Part 2: Test report format,” 2007, *Bureau International de Métrologie Légal, Paris*.

[7] “Metrology refType Database.” Accessed: Apr. 02, 2025. [Online]. Available: https://digilab.ptb.de/dkd/refType/vocab/index.php

[8] D. Hutzschenreuter *et al.*, “SmartCom Digital System of Units (D-SI) Guide for the use of the metadata-format used in metrology for the easy-to-use, safe, harmonised and unambiguous digital transfer of metrological data - Second Edition (D-SI 1.3.0-2). Zenodo,” 2020. doi: https://doi.org/10.5281/zenodo.3816686.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Digitalized Third-Party Validation for Calibration Service: a System Design Example

Hiroshi Watanabe[1], Yoshitaka Shimizu[1], Katsuhiro Shirono[1], Toshiyuki Fujimoto[2]

[1]*National Institute of Advanced Industrial Science and Technology (AIST), National Metrology Institute of Japan (NMIJ), Tsukuba 305-8563, Japan, hiroshi-watanabe@aist.go.jp, y-shimizu@aist.go.jp, k.shirono@aist.go.jp*
[2]*National Institute of Advanced Industrial Science and Technology (AIST), Planning Headquarters, Tsukuba 305-8561, Japan, t.fujimoto@aist.go.jp*

***Abstract* – We present a new case study of digitalized third-party validation for a calibration service of purity. The third-party validation is an add on to the calibration service, conducted by a third party independent from laboratory and customers, aiming at assuring quality of results. The case features frequent validations, and information leakage-free by checking data within the laboratory rather than externally, that have not been addressed in the previous cases. We describe the design and implementation of the system.**

I. INTRODUCTION

Accredited Testing and Calibration services are carried out in accordance with a predetermined procedure in laboratories. The procedure is reviewed by an accreditation body at some regular intervals, such as 3 years. As a result, most customers may assume that the results of the services are sufficiently reliable for their purposes. However, our concern is that the laboratories cannot clearly demonstrate how they implemented the procedures to obtain the individual results. They have no specific means of assuring customers beyond simply reminding them of their accreditation status.

A. Digitalized third-party validation for Calibration service

As a means to complement the system of accreditation, the authors have proposed integrating validation [1, 2, 3] in laboratory calibration processes, meeting the following conditions: conducted by a third party independent of both the customer and the laboratory; performed promptly each time a specific work is executed or data is generated; controlled or automated by software; checks the work log or data to confirm the process is on track, and records proof of the checks or the confirmation in the third party.

In integrating the third-party validation into calibration process, laboratory can convey quality of activities, results, and relevant data to customers by using the result of validation. In addition to the the proposal of the approach [1], we have developed two practical cases that implement the approach as proof of concept: vernier-caliper calibration [2] and calibration of reference photovoltaic devices [3].

These two cases of integrating third-party validation gave clear insight into the approach and showed feasibility of the approach. On the other hand, they revealed remaining two concrete challenges in implementing the approach as follows:

B. Investigation on timing and content of a validation

If a calibration service takes longer period, such as over a few weeks, the more frequent the validations, the more reliance the customer may have in quality of results in laboratory. However, too frequent validations may cause unnecessarily increasing the complexity of the control and cost while decreasing the thruput. In the previous cases, we did not implement frequent validations: one validation in vernier-caliper case [2], and two validations in reference photovoltaic devices case [3]. It is important to design timing and content of a validation to increase reliability as much as possible without increasing costs. To discuss an appropriate frequency, demonstrating cases along with necessity of implementing frequent validations is also indispensable.

C. Laboratory and customer information leakage-free

The data to be checked in a validation may include information of laboratory or customer. If the data is transferred outside of the laboratory and checked at there, any leakage of laboratory or customer information from the transferred data becomes our great concern. In order to decrease the risk of information leakage from the transferred data, we can consider in two directions as follows:

- developing secure and confidential checking system, even in outside, that can handle the transferred data without any information leakage, or
- checking the data within the laboratory rather than externally, and eliminating, originally, any customer or laboratory information from the data transferred outside.

The previous cases [2, 3] addressed the first direction and elaborated the checking system at a third party organization by employing smart contract on a blockchain. The variety of techniques, whether blockchain or not, that realize the first direction needs to be developed further. On the other hand, it has not been addressed yet whether we can develop validation that implements the second direction. In practice, the second direction might be preferred by the laboratories that do not want to be bothered by the transferred data outside.

The present study develops a new case of the third-party validation for calibration service of purity. In developing, we address the above challenges in order to wider the applications of the approach. The main contribution of the present paper is to show a new case of integrating validations with the following both features:

- frequent validations as described in Section IV.B., and
- information leakage-free, following the second direction different from the previous cases, by checking data within the laboratory rather than externally as described in Sections III.D. and IV.C..

The case broadens the scope of the approach, since these features have not been addressed in past cases. Today, there are many people in metrology community who are looking for ways to address digital transformation. The system design of the case is of interest to them and offers many valuable insights.

The paper is organized as follows: Section II introduces the procedure of calibration service of purity which we integrate third-party validations. Section III describes a design of third-party validation. Then, Section IV introduces concrete implementation of a few validations. Section V mentions related work. Finally, we discuss and conclude in Section V.

II. CALIBRATION SERVICE OF PURITY

A calibration service of purity used for integration of third-party validation is introduced. The service is part of an accredited calibration service that receives a request with a sample compound from customers, conducts both NMR and GC experiments in parallel, and issues calibration certificate that includes the calibration result (purity). Fig.1 shows a simplified procedure of the calibration service. Any experiment comprises three activities: sample preparation, measurement, and data analysis. Of these, the activities measurement and sample preparation are repeated until sufficient measurement data is available for the data analysis. Once both experiments are completed and both results of data analysis are obtained, a calibration result is determined and a calibration record file is compiled from these results. The calibration certificate is issued based on the calibration record file. There are multiple personnel in charge of the service, with roles divided as follows: Calibration Authority, NMR staff, and GC staff. The entire service, from receiving a request to issuing certificates, takes 2 weeks to 2 months, at most.

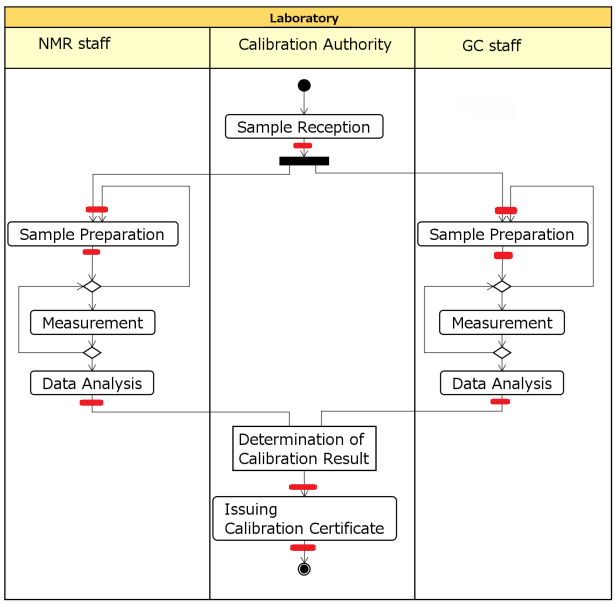


Fig. 1. Procedure of calibration service of purity, and timings of validations. The small red signs indicate the timings for validations described in Sections iii. and iv.

Keeping environment (temperature and humidity) in the laboratory stable during the activities of sample preparation is crucial for later valid calibration results. Some sample compounds absorb moisture from the air, altering their weight and purity. Therefore, acceptable ranges of environment is specified in a quality manual as environmental conditions for each sample compound. We ensure that the laboratory environment meets the environmental conditions during the activities.

III. DESIGNING VALIDATIONS

A. Basic Design of Validations

Table 1 shows a list of validations we considered, whose timings are also indicated on the procedure in Fig. 1. We integrate validations after each of the following activities: sample reception, sample preparation, data analysis, and issuing calibration certificates. Among these activities, before each of the following activities, we also integrate validations: sample preparation and issuing calibration certificate. We call the validation before issuing calibration certificate ConformityAssessment, since, in this validation, we check the calibration record file with the calibration result that is to be transcribed to the certificate.

Table 1. List of Integrated Validations

Name	Timing	Operator	Data to be checked and compiled into data package	Log
SampleReception	after **SR**	CA	reception information sample information	Y
pre-SamplePreparation	before **SP**	MS	environment (temperature and humidity), personnel and equipment	N
post-SamplePreparation	after **SP**	MS	sample preparation record file for NMR or GC, environment (temperature and humidity), personnel and equipment	Y
DataAnalysis	after **DA**	MS	summary report for NMR or GC, measurement data, sample information	Y
ConformityAssessment	before **ICC**	CA	calibration record file	Y
RegistrationOfCertificates	after **ICC**	CA	calibration certificates and their IDs	Y

SR= Sample Reception, **SP**= Sample Preparation, **DA**= Data Analysis, **ICC**= Issuing Calibration Certificate, CA=Calibration Authority, MS=NMR staff or GC staff

B. Local Application and Local Storage

Let us introduce some terms before we explain concrete design. A *local application* is software that is designed and developed by the third party but is deployed and operated in the laboratory. A *local storage* is a storage installed in laboratory to allow sharing information between local applications. A *data package* for activity is a text data in JSON format generated by local applications, which stores a collection of data produced or generated during the activity. A *completion log* of an activity is a text message generated and sent by local applications that indicates the completion of the activity. It consists of a control ID (which identifies the activity within a series of activities in a calibration service), a hash of the activity's data package and a time stamp.

For each validation in Table 1, we design a local application. Fig.2 shows a basic design for local applications prepared to align the designs for individual applications. The applications generally behaves as follows: when the operator uploads data, the local application retrieves from the local storage the relevant past data packages, called "context"; it checks the data by using the context; after the check, it compiles the data with the check results into a new data package; this new data package is then saved back to the local storage; finally, the application sends a completion log to the third party.

Note that the local storage must be properly access controlled, since it affects the later validation results. So, we encrypt the local storage using a third-party key. The third party ensures that only local applications have access to the storage using the key, making the contents directly inaccessible to laboratory members.

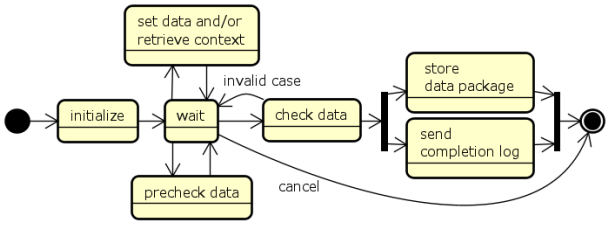


Fig. 2. Basic design for local applications

C. Certifier Server in third-party organization

In the third party, we prepare a certifier server with the following two functions: logging completion and integrity checking. The function logging completion stores the completion logs sent from local applications. The function integrity checking returns the result of comparing the incoming list of pairs of control ID and hash from ConformityAssessment and stored completion logs in the server.

D. Data Transfer from Local applications to Certifier Server

A general data flow diagram for a local application is illustrated in Fig.3. The message from local application to certifier server does not contain any laboratory or customer information. The application does not send data packages to the server, but sends messages such as completion logs, lists of pairs of control ID and completion log, or signals of beginning or ending calibration, at most. These messages do not contain any laboratory or customer information, except control IDs.

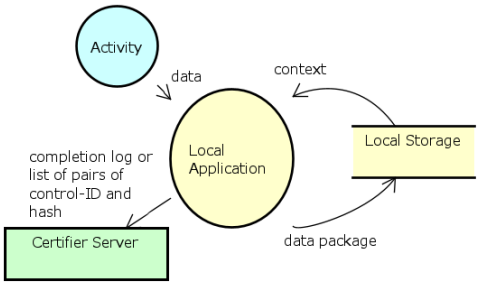


Fig. 3. Dataflow for local application

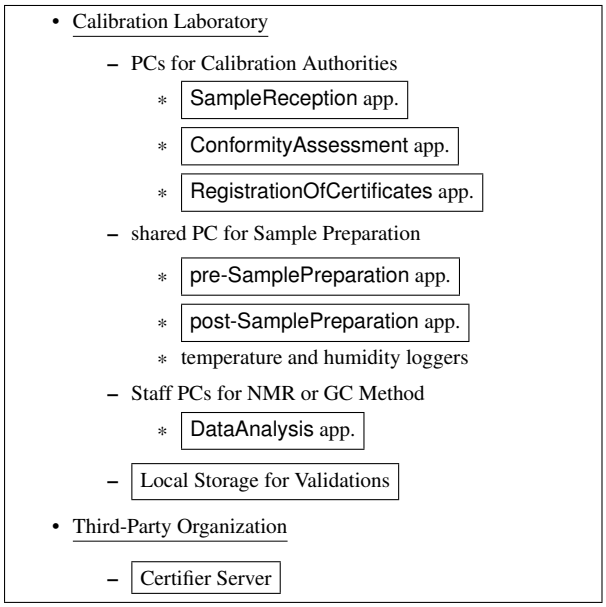


Fig. 4. System Configuration of entire system.

IV. IMPLEMENTING VALIDATIONS

We developed a system for validations based on the design described in the previous Section. Here, we describe an overview of the entire system, and a few concrete validations implemented as local applications, as shown in Fig.2.

A. System Configuration

Fig 4 shows the configuration of the entire system. In the laboratory, the local applications are installed in dedicated PCs for personnels or a shared PC for sample preparation with connected temperature and humidity loggers; the local storage is also prepared inside the laboratory. In the thrid-party organization, a certifier server is prepared.

B. Validations for Sample Preparation

In the calibration service (Fig.1), the activity of sample preparation is repeated during both NMR and GC experiments. For each sample preparation, the NMR or GC staff uses local applications to initiate validation pre-SamplePreparation before the activity and validation post-SamplePreparation after the activity.

The pre-SamplePreparation verifies specific environmental conditions of the sample compound before moving on to the sample preparation. In the "check data" state shown in Fig.2, the local application checks both

- the qualifications of personnel and equipment, and
- the environmental conditions by using the latest temperature and humidity data directly extracted from the logger connected to the shared PC.

If both succeed, the application displays with the message "Move on to Sample Preparation"; otherwise, the application halts the operation and prevents the staff from beginning the activity with deviations from quality manual; in this case, some corrective action is required to continue. We omit implementing "send completion log" in the local application.

The post-SamplePreparation leaves completion log in the third party immediately after completing the sample preparation. The completion-log sent by the local application includes the hash of the data package, which contains both the sample preparation record noted by the staff and data from the previous pre-SamplePreparation.

The completion logs stored in the certifier server in the third party cannot be falsified later by the laboratory member. So, they provide evidence to show integrity for the data packages stored in the laboratory, which we show in Section IV.C..

For every sample preparation, the present case repeats the pair of validations above. The combination of validations not only prevents deviations from quality manual at real time but also promptly leaves completion logs in third party.

C. ConformityAssessment

The calibration certificate with calibration result is issued based on the calibration record file in the activity "Issuing Calibration Certificate" shown in Fig.1. Just before the activity, the calibration authority uses the local application to start the validation ConformityAssessment.

The ConformityAssessment, conducted through the collaboration of the local application and the certifier server, verifies the calibration record file as follows. In the "set data and/or retrieve context" state shown in Fig.2, the calibration authority uploads calibration record file, then the local application retrieves all relevant data packages from the local storage, including sample reception, data analysis, and sample preparation. In the "check data" state, the application checks that the calibration record file does not contain any of the following errors:

- data omissions or inconsistencies, such as inconsistency of data between sheets inside Excel file,
- copying errors from data packages, and
- miscalculations, such as for calibration result, uncertainty and compensation.

Meanwhile, the certifier server checks the integrity of the retrieved data packages as follows: The local application recalculates hashes from the data packages, pairs the hashes with control IDs, creates a list of pairs of control ID and hash, and sends this list to the certifier server. The server then uses the integrity checking function (stated in Section III.C.) to compare the received list with the stored logs.

If both succeed, the local application indicates that the result is valid. If either check fails, it indicates that the result is invalid and displays the errors found.

Note that the validation provides a conformity assessment of calibration record file by combining the local application and the certifier server, including a consistency check with data from past data packages. Also, note that the local application does not send laboratory or customer information outside the laboratory, except for control IDs used to manage the service.

V. RELATED WORK

The present work follows the previous works [1, 2, 3] that explored integrating (third-party) validation in laboratory calibration processes. Specifically, these studies addressed micropipette calibration [1], vernier-caliper calibration [2], and reference photovoltaic devices calibration [3]. These focus on the third-party validation for calibration services more strongly than other related studies in the rest of this section. A key aspect of these works is implementation using smart contracts in blockchain technology. In contrast, the present study addressed a design that does not assume blockchain technology. Further details on the differences between the present study and the previous works are elaborated in the Introduction.

In addition, the line of work was heavily influenced by research on blockchain technology applications in metrology, especially in conformity assessment. This includes blockchain applications for legal metrology [4], measuring system implementation to enhance security and simplify regulatory management [5], and PKI implementation for smart meters [6].

In the digitalization of calibration services, Digital Calibration Certificates (DCC) [7] is currently a hot topic in the metrology research community. Beyond DCC, there are also studies on lab data management studies, such as [8], which utilize semantic web technologies and ontologies to enhance data interoperability and reusability in accordance with FAIR principles [9]. Both DCC and FAIR principles do not directly relate to the present topic of third-party validation.

The challenge of frequent validations (in Section I.B.) is inseparable from a principle of 'contemporaneous data' that means the data should be recorded at the time it is generated. The principle is known as the C of ALCOA principles, which gives a set of guidelines to ensure Data integrity (see [10, 11, 12]) in GxP practices, such as GMP, GLP, and GCP. Although the challenges are similar, we have not found any relevant studies in the field of data integrity that correspond to our third-party validation. Still, we have seen similar system development. For instance, a blockchain-based monitoring system for pharmaceutical manufacturing lines [13] performs integrity inspections of reports from various parts of the lines in real-time and conducts monitoring to reduce risks.

VI. DISCUSSION AND CONCLUSION

We add some remarks and future works, and then conclude.

1. The design choice of implementing local application on PCs was made to connect additional equipment and extract data from them, not just from temperature and humidity loggers.

2. We recall a point of validation that the third party implements checks in the validation based on the information gathered from the laboratory. We also remind the laboratory to consider and decide on corrective actions regarding the validation results in advance when implementing the system.

3. The development of the case is completed, but it has not been deployed and put into actual operation with the calibration service. The actual operation will be our future works.

4. The present case was provided with minimal security measures since we did not anticipate the presence of malicious members in the laboratory, such as those who report false activities or falsify results. Security measures for them will be future works too.

5. For an application of the third-party validation results, another futuer work will be to design a service of the third-party organization to endorse the calibration results to the customer.

In the present study, we have developed a new case of integrating third-party validation into calibration service for purity. The case addressed the identified challenges of both frequent validations and information leakage-free. The information leakage-free was instantiated by checking the data inside the laboratory and eliminating, originally,

customer and laboratory information from the data transferred to outside. The case broadens the applications of our developing approach.

REFERENCES

[1] K.Shirono, N.Takegawa, M.Moni, D.Peters, “DESIGN OF THE DIGITALIZED CONFORMITY ASSESSMENT FOR LABORATORY ACTIVITIES IN INDIVIDUAL CERTIFICATIONS”, Proc. of M4Dconf2022., September 2022, doi:10.21014/tc6-2022.009.

[2] K.Shirono, A.Hirai, G.Matsui, O.Sato, H.Watanabe, T.Takatsuji, T.Fujimoto, “Proof of concept for the digitalized certification of measurement data: An example of vernier-caliper calibration”, AMCTM XIII, World Scientific, 2025, pp. 260-267, doi:10.1142/9789819800674_0024.

[3] S.Igari, H.Watanabe, T.Fujimoto, K.Shirono, “Quality-compliance validation system for individual calibration and testing events: case study of primary calibration of reference photovoltaic devices”, Meas. Sci. Technol., vol. 36, March 2025, doi:10.1088/1361-6501/adb643.

[4] D. Peters, J. Wetzlich, F. Thiel and J. -P. Seifert, "Blockchain applications for legal metrology," 2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Houston, TX, USA, 2018, pp. 1-6, doi:10.1109/I2MTC.2018.8409668.

[5] W. S. Melo, A. Bessani, N. Neves, A. O. Santin and L. F. R. C. Carmo, "Using Blockchains to Implement Distributed Measuring Systems," in IEEE Transactions on Instrumentation and Measurement, vol. 68, no. 5, pp. 1503-1514, May 2019, doi:10.1109/TIM.2019.2898013.

[6] W. Melo, R. C. S. Machado, D. Peters and M. Moni, "Public-Key Infrastructure for Smart Meters using Blockchains," 2020 IEEE International Workshop on Metrology for Industry 4.0 & IoT, Roma, Italy, 2020, pp. 429-434, doi:10.1109/MetroInd4.0IoT48571.2020.9138246.

[7] The Physikalisch-Technische Bundesanstalt, Digital Calibration Certificate (DCC) - Wiki, https://wiki.dcc.ptb.de/.

[8] Schilling, M., Bruns, S., Bayerlein, B., Kryeziu, J., Schaarschmidt, J., Waitelonis, J., Dolabella Portella, P. and Durst, K. (2025), Seamless Science: Lifting Experimental Mechanical Testing Lab Data to an Interoperable Semantic Representation. Adv. Eng. Mater., 27: 2401527. doi:10.1002/adem.202401527.

[9] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). doi:10.1038/sdata.2016.18.

[10] U.S. Department of Health and Human Services Food and Drug Administration, Data Integrity and Compliance With Drug CGMP Questions and Answers Guidance for Industry, 2018, https://www.fda.gov/regulatory-information/search-fda-guidance-documents/data-integrity-and-compliance-drug-cgmp-questions-and-answers.

[11] Medicines and Healthcare products Regulatory Agency, Guidance on GxP data integrity, 2018, https://www.gov.uk/government/publications/guidance-on-gxp-data-integrity.

[12] World Health Organization, TRS 1033 - Annex 4: WHO Guideline on data integrity, 2021 https://www.who.int/publications/m/item/annex-4-trs-1033/.

[13] Durá, M., Leal, F., Sánchez-GarcÃa, Á. et al. Blockchain for Data Originality in Pharma Manufacturing. J Pharm Innov 18, 1745-1763 (2023). doi:10.1007/s12247-023-09748-z.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Digitalization of Calibration Workflow at National Institute of Metrology Thailand (NIMT)

Sunantiya Parana[1,*], Nititorn Kenyota[1], Narueson Nanna[1], Praiya Thongluang[1]
Narin Chanthawong[2] and Jariya Buajarern[2]

[1]*Digital Transformation Centre, National Institute of Metrology (Thailand), Pathumthani, Thailand*
[2]*Dimensional Metrology Department, National Institute of Metrology (Thailand), Pathumthani, Thailand*
Corresponding author: sunantiyap@nimt.or.th

***Abstract* – This paper presents the strategic digital transformation of workflows at the National Institute of Metrology Thailand (NIMT). The implementation of digital technologies has significantly reduced human errors, enhanced accuracy, and increased transparency in data handling. NIMT has proactively developed and deployed digital platforms across the organization to support this transformation. A key outcome of this initiative is the Equipment Management and Tracking System (EMTs), developed in alignment with ISO/IEC 17025, which enables efficient electronic management of measurement instruments. The integration of technologies such as Optical Character Recognition (OCR) and Digital Calibration Certificates (DCC) has further improved data accuracy and reduced processing time. In support of quality assurance, NIMT is also planning to develop a digital system to enhance its audit processes and ensure continued compliance with quality standards. This digital shift not only strengthens NIMT's operational capabilities but also aligns Thailand's measurement infrastructure with international metrological practices.**

I. INTRODUCTION AND METROLOGICAL CHALLENGES

The digitalization of calibration workflows at NIMT represents a strategic shift in how measurement data is managed, processed, and utilized. However, this transition is not without its challenges. As with many national metrology institutes worldwide, NIMT faces a complex set of obstacles that must be addressed to achieve an efficient and sustainable digital transformation. These challenges can be categorized into five primary areas [1]:

1. Infrastructure and Budget Limitations

A fundamental requirement for digital transformation is the establishment of a robust and secure IT infrastructure. This includes cloud-based data centers, scalable storage systems, high-speed networks, and cybersecurity protocols, all of which are essential to support real-time calibration workflows and digital certificate issuance. However, as a public agency, NIMT often faces budgetary constraints, limited flexibility in procurement, and lengthy approval processes. These limitations can hinder the timely acquisition and deployment of critical digital technologies, delaying initiatives such as the integration of real-time data acquisition from calibration instruments or the implementation of secure metrology cloud services [2].

2. Complexity of Metrology Systems and Data Integration

Metrology systems inherently involve the handling of highly sensitive and precise data from calibration values and measurement uncertainty to traceability chains and international equivalence. Migrating such data to a digital system requires absolute precision and consistency. One major challenge lies in integrating data from diverse instruments and legacy systems into a centralized digital platform. These older systems were often developed in isolation and were not designed for digital interoperability, resulting in fragmented databases and inconsistent data formats. Harmonizing such data across platforms without compromising accuracy or traceability presents a significant technical hurdle in the digitalization of calibration workflows

3. Development of Digital Skills for Personnel

While NIMT staff possess deep expertise in scientific measurement and engineering, the shift toward digital workflows requires an expanded skillset particularly in areas such as data analytics, software integration, automation systems, and artificial intelligence. Many personnel need upskilling to adapt effectively to new tools and systems used in digital calibration processes. This skill development requires structured training programs, ongoing knowledge transfer, and a cultural shift toward embracing technological innovation. Without these efforts, the full benefits of digital transformation may not be realized, and the risk of system underutilization or operational errors may increase

4. Data Security and Regulatory Compliance

In the digital calibration context, data integrity and security are non-negotiable. Metrological data not only supports critical industrial applications but also serves as a

legal reference in regulated sectors. As such, digital calibration systems must comply with national cybersecurity laws, such as the Personal Data Protection Act (PDPA), and adhere to international standards for information security NIMT must ensure that its digital platforms are designed to prevent unauthorized access, data breaches, and service disruptions. At the same time, the integration of new technologies must be carefully aligned with regulatory requirements to avoid legal or operational risks.

5. Policy Changes and Interagency Coordination

Digital transformation initiatives are often influenced by national policy directions and interagency collaboration. At NIMT, policy shifts, especially those related to digital governance or public sector reform, can have a direct impact on resource allocation, strategic focus, and project continuity. Furthermore, digital calibration workflows require coordinated data exchange and system integration with external stakeholders, including government agencies, industry clients, and international metrology networks. Any misalignment in policy, priorities, or technical standards may result in delays or inconsistencies in implementation.

Therefore, the digitalization of calibration workflows at NIMT entails navigating multiple layers of complexity ranging from infrastructure and budget limitations to human resource development, data security, and policy alignment. Each of these challenges must be strategically addressed to ensure that the transition from traditional to digital calibration not only improves efficiency and accuracy but also meets international expectations and national regulatory standards [1, 3]

II. STRATEGIC GOALS FOR DIGITALIZING THE CALIBRATION WORKFLOW

The transition from analog processes to digital systems has become an essential strategy for national metrology institutes, including the NIMT, to respond effectively to the accelerating pace of technological advancement and the growing complexity of industrial demands [3, 4]. The adoption of digital technology is not merely an operational upgrade it serves as a fundamental enabler for enhancing Thailand's measurement infrastructure to meet internationally recognized standards. Despite the clear benefits, the path toward full digitalization of calibration workflows presents various technical, organizational, and policy-related challenges. To overcome these barriers, it is vital for NIMT to establish well-defined strategies and objectives that provide direction and coherence for the transformation process. The following strategic goals have been identified as central to the success of this digital transition:

1. To enhance operational efficiency: Streamlining calibration workflows through digital platforms reduces manual errors, minimizes redundant processes, and accelerates service delivery turnaround times. Automation and real-time data exchange support higher productivity and greater responsiveness to customer needs.

2. To elevate the precision and reliability of measurement standards: Digital systems facilitate accurate storage, processing, and analysis of calibration data, ensuring consistency and reducing uncertainty. These capabilities are essential for maintaining traceability and reinforcing confidence in national measurement standards.

3. To promote transparency, traceability, and accountability: The use of centralized digital platforms enhances the traceability of calibration records and strengthens document security. Digital audit trails and secure access controls ensure compliance with legal, industrial, and international standards.

4. To facilitate data and workflow integration across agencies: By harmonizing systems and protocols with both national and international stakeholders, NIMT can improve collaboration and data interoperability. This ensures that calibration results align with global frameworks, such as those promoted by the BIPM and regional metrology organizations.

To foster digital competency and readiness among personnel: A critical success factor lies in developing the digital capabilities of NIMT's workforce. This includes training in digital tools, data analytics, and emerging technologies empowering staff to adapt to new systems and drive innovation from within [4].

III. METHODOLOGY AND EXPECTED RESULTS

1. To handle the challenges of infrastructure and budget limitations

The implementation of digital systems at NIMT will bring several positive outcomes, particularly in enhancing the accuracy and speed of operations. One of the key advantages is the reduction of errors associated with manual data entry. For instance, Optical Character Recognition (OCR) technology, which converts images into editable and searchable text, has been applied in research conducted by Naruesorn Nanna et al. [5]. Their study, titled A Two-Layer Deep Learning-Powered Optical Character Recognition for Enhanced Calibration Processes, was part of the DX research group at NIMT. The study found that OCR technology significantly reduced calibration times, comparing the traditional method (03:05:58 hours) with OCR software (02:20:02 hours), resulting in a time reduction of up to 24.70%. Furthermore, OCR technology improved the accuracy of data processing. An example of the OCR software in use is shown in Figure 1.

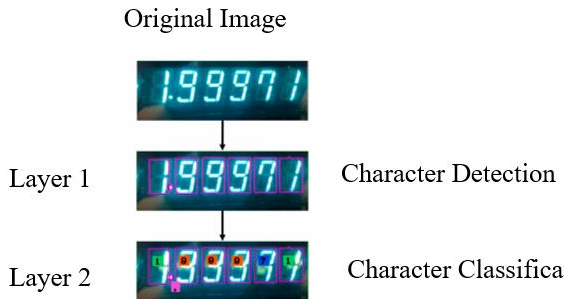


Figure 1. presents the steps for character detection and classification in measurement results using OCR software

The OCR process also helps mitigate budget constraints by reducing the need to invest in full-scale systems to receive or read real-time data from machines and laboratory equipment [5, 6].

2. Complexity of Metrology Systems and Data Integration

To address the challenge, NIMT has established a robust process for maintaining the quality and accuracy of its calibration services. Yearly, the calibration process is thoroughly reviewed to ensure compliance with established standards. This review process, known as the audit, assesses whether the procedures and outcomes meet the required criteria [7]. If any part of the process is found to be non-compliant, the auditor issues a Non-Conformance (N/C) report. Should the N/C be confirmed, the Quality Management (QM) department issues a Corrective Action Request (CAR), directing the responsible personnel to address and resolve the issue.

However, there is currently no system in place to efficiently track and manage the resolution of CARs in a timely manner. To address this limitation and enhance operational efficiency, NIMT plans to implement a digital system for tracking CAR resolutions. This system will facilitate more streamlined, accurate, and faster responses to corrective actions, supporting the broader goal of digitalization and significantly improving operational effectiveness and modernization. In addition, NIMT is currently building a database that captures the scope and capabilities of each calibration service. This system will be connected to the relevant standard body part in Thailand, enabling real-time verification of measurement capabilities and ensuring that NIMT's services remain aligned with national standards.

3. Development of Digital Skills for Personnel

Traditionally, lab staff used Microsoft Excel to record data, calculate results, and print calibration certificates in paper format for customers. Digital Calibration Certificates (DCC) [8, 9] are a new concept for them. To support this shift, NIMT has developed a web-based platform at dcc.nimt.or.th. It allows NIMT staff to generate DCC in PDF/A3 with attached XML files, integrated with internal systems for accuracy and security. General users can generate XML-only DCC. The platform includes schema and Schematron validation and uses the Electronic Transactions Development Agency (ETDA) verification services that is a government agency under the Ministry of Digital Economy and Society to verify digital signatures, ensuring compliance with digital standards. This approach not only streamlines the DCC generation process but also supports the development of new skills among NIMT staff.

4. Data Security and Regulatory Compliance

NIMT has enhanced its data management processes by making them more systematic and accessible. Electronic documents are now stored in the Equipment Management and Tracking System (EMTs) [8, 10]. Figures 2. shows an overview of the duty structures in EMTs and Figures 3. illustrates the data storage and management system. When equipment is received along with its certificate, the owner stores the certificate and related data in electronic form in the EMTs. Authorized users can then log in to search for the equipment data within the system.

The key benefits of this system include easier document retrieval, improved data privacy, and enhanced security.

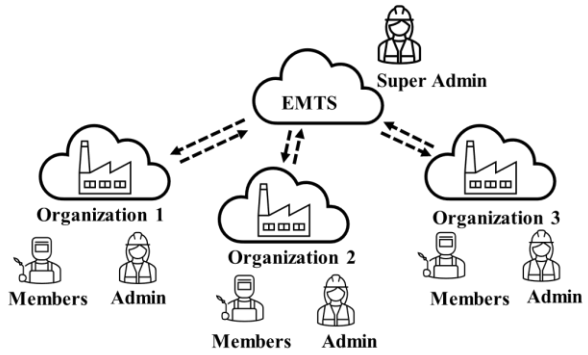


Figure 2. Diagrams of the duty structures in EMTs system

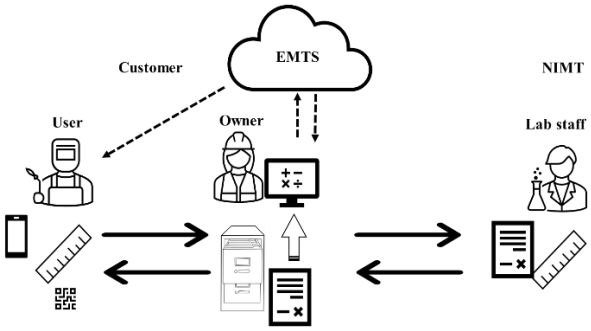


Figure 3. illustrates the data storage and management system – EMTs

5. Policy Changes and Interagency Coordination

National policies are often difficult to predict and cannot be directly changed [11]. However, NIMT can prepare for

digital transformation by improving its internal structures and processes to effectively support these changes. By fostering cooperation both within the organization and with external partners, leveraging modern technologies, and continuously monitoring progress, NIMT can strengthen its readiness and enhance the efficiency of its calibration services at the international level.

IV. CONCLUSION

The digitalization of calibration workflows at the National Institute of Metrology Thailand (NIMT) represents a significant advancement in aligning national metrology practices with international standards. By implementing systems such as the EMTs, OCR technologies, and DCC. NIMT has improved data accuracy, reduced processing time, and enhanced transparency across its services. These developments address key challenges, including legacy system integration, infrastructure limitations, regulatory compliance, and workforce digital readiness. The planned introduction of a digital system to support audit and quality management processes further reinforces NIMT's commitment to continuous improvement and ISO/IEC 17025 compliance. This transformation has not only increased operational efficiency but also fostered a culture of innovation and collaboration within the organization. Staff have gained essential digital skills, and interdepartmental coordination has improved, enabling more responsive and reliable calibration services. NIMT's experience illustrates that successful digital transformation requires strategic planning, cross-functional engagement, and sustained effort. It provides a practical model for other metrology institutes seeking to modernize their systems in response to evolving technological and industrial demands.

REFERENCES

[1] E. Henriette, M. Feki, I. Boughzala, "The shape of digital transformation: a systematic literature review". In: Information Systems in a Changing Economy and Society, vol. 431, 2015.

[2] V. Ramaswamy, K. Ozcan, "Brand value co-creation in a digitalized world: an integrative framework and research implications," Int. J. Res. Market, vol. 33(1), 2016, pp. 93-106.

[3] C.Matt, T.Hess, A.Benlian, "Digital transformation strategies", Bus. Inf. Syst. Eng., vol.57(5), 2015, pp.339–343.

[4] T.Hess, C.Matt, A.Benlian, F.Wiesböck, "Options for formulating a digital transformation strategy," MIS Quarterly Executive., vol.15(2), 2016, pp.123-139.

[5] N. Nanna, N. Chanthawong, J. Buajarern, "Design and development of OCR software for remote measurement and calibration". J. Phys.: Conf. Ser, vol. 2653: 012012, 2023.

[6] G. Grzeczka and M. Klebba, "Automated calibration system for digital multimeters not equipped with a communication interface". Sensors, vol. 20: 3650, 2020.

[7] C.R.H. Barbosa, M.C. Sousa, M.F.L. Almeida, and R.F. Calili, "Smart manufacturing and digitalization of metrology: A systematic literature review and a research agenda," *Sensors*, vol. 22: 6114, 2022.

[8] S. Parana, P. Thongluang, N. Chanthawong, J. Buajarern, "Analyzing the conformance of DCC prototype architecture to calibration laboratory expectations report". Proceeding of 3rd International DCC Conference, 2023, pp. 471-479.

[9] L.H.J.M. van der Aa, R.J.M. van der Heijden, and J.J.M. van der Meer, "Digital skills at work – conceptual development and empirical exploration". *Technol. Forecast. Soc. Change*, vol. 202: 122015, 2024.

[10] E. Henriette, M. Feki, and I. Boughzala, "Information systems in a changing economy and society". In: Information Systems in a Changing Economy and Society, p. 431, 2015.

[11] National Institute of Metrology (Thailand)., (2023). Five-year action plan (2023–2027) of the National Institute of Metrology, National Institute of Metrology Press, Thailand.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

"SASO Proficiency Test Machine" - Advanced Pythonic AI Algorithms for Automating and Validating ISO 13528 & ISO 5725-2 at Saudi Standards, Metrology, and Quality Organization-SASO-KSA

Fahad A. AlMuhlaki[1], Saad A. Bin Qoud [1], Rayan A. AlYousefi [1], I. AlFaleh[1], N. Qahtani[1], Khaled AlEnizi[1], AbdulRahman AlMrhom[1], A. El-Matarawey[2]

1. NMCC-SASO- Saudi Arabia
2. National Institute of Standards (NIS-Egypt)

***Abstract* – In the field of laboratory proficiency testing and method validation, linking to international standards such as ISO 13528, ISO 17043 and ISO 5725-2 is essential for ensuring data accuracy, consistency, and inter-laboratory comparability. However, simple implementation these standards can be led to insufficient error-handling and difficult to scale across large datasets. This study introduces a novel Python-based AI framework automating ISO 13528 and ISO 5725-2 compliance, enhancing data integrity and decision-making confidence. We introduce a robust software framework built on modern Python libraries and machine learning techniques that streamline key statistical computations required by these standards—including outlier detection, repeatability and reproducibility assessment, and performance scoring. The system not only automates routine calculations but also introduces intelligent validation checks that flag anomalies, inconsistencies, or deviations from expected patterns, thereby enhancing data integrity and decision-making confidence. By integrating AI-enhanced analytics with standardized evaluation protocols, this work bridges the gap between traditional statistical methods and modern computational capabilities.**

***Keywords*:** Python, AI algorithms, ISO 13528, ISO 5725-2, automation, statistical validation, proficiency testing, data quality, SASO, NMCC

I. INTRODUCTION

Proficiency testing (PT) and method validation are critical components of laboratory quality assurance programs, ensuring that measurement results are accurate, reliable, and comparable across different facilities.

The objectives of PT are:

- Assess the performance of the participating laboratories in the scope of the PT scheme.
- Provide technical support to the participating laboratories.
- Identify the technical issues in the performance of the participating laboratories and recommend potential solutions.
- Provision of additional confidence in services.
- Assess and validate the uncertainty claimed by the participants.
- Identify the sources of error in their performance.
- Support the participating laboratories with technical/statistical recommendation to improve their performance.

Standards such as ISO 13528 [1], ISO 5725-2[2] and ISO 17043[3] provide detailed guidelines for evaluating laboratory performance and validating measurement methods. However, traditional approaches to implementing these standards often rely on manual calculations or spreadsheet-based tools, which can be labor-intensive, error-prone, and difficult to scale. The advent of modern programming languages like Python offers an opportunity to streamline and automate the application of these standards. By applying robust statistical techniques, visualizations and efficient computational tools, laboratories can achieve faster, more consistent results while minimizing human error. Automation helps ensure consistent interpretation of laboratory performance and measurement accuracy across multiple datasets. This paper explores the development of a Python-based AI automation framework for implementing key aspects of ISO 13528 and ISO 5725-2, with a focus on robust statistical methods such as median-based estimators and outlier-resistant calculations. Through this approach, we aim to demonstrate how automation enhances the efficiency, accuracy, and transparency of proficiency testing and method

validation processes. In many real-world scenarios, laboratory data may contain outliers due to equipment malfunctions, human error, or environmental variability. Classical statistical measures like the mean and standard deviation are highly sensitive to such outliers. To address this, ISO 13528 recommends the use of robust estimators. These provide more reliable central tendency and dispersion estimates, ensuring that performance evaluations remain accurate even when data is not perfectly normal. Python [4], with its rich ecosystem of scientific computing packages, provides an ideal platform for implementing the statistical requirements of ISO 13528 and ISO 5725-2. Libraries such as NumPy [5] and SciPy [6] offer built-in functions for calculating z-scores, zeta score. Normalized En value, repeatability limits, and outlier detection criteria. Pandas [7] simplifies data manipulation, while Matplotlib [8] and Seaborn [9] allow for clear visual reporting of results. Together, these tools enable the development of fully automated pipelines for evaluating laboratory performance and validating measurement methods.

II. MECHANISM OF STATISTICAL MODEL

Figure 1. illustrates a structured process flow for the development, testing, and evaluation of calibration/test materials, as outlined by relevant ISO [1,3] standards. This flowchart provides a comprehensive overview of the steps involved in ensuring the quality and reliability of such materials, which are critical for accurate measurements and laboratory performance assessments. All figures that demonstrated in this work are coming from different examples.

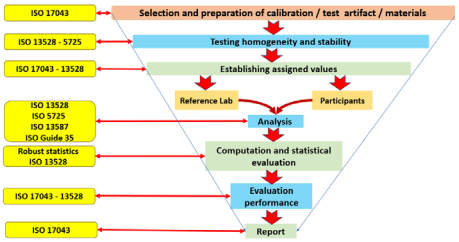


Fig. 1. Statistical Model Flowchart.

II.I. SELECTION / PREPARATION OF CALIBRATION/TEST MATERIALS

The process begins with the selection and preparation of calibration or test materials according to [3]. This step ensures that the equipment, device, artifact or materials chosen are suitable for their intended purpose and meet the necessary specifications. Proper preparation is crucial to maintain consistency and accuracy throughout subsequent stages.

II.II. TESTING HOMOGENEITY AND STABILITY

Once the materials are prepared, they undergo rigorous testing according to [1, 2] to evaluate their homogeneity (uniformity across samples) and stability (consistency over time). These tests are essential to confirm that the materials will provide reliable results during use.

We validate the homogeneity using both models of ISO 13528, ISO 5725-2 and analysis of variances (ANOVA).

II.II.A. OUTLIER

The algorithm checks the outlier by series of tests includes Cochran and Grubbs tests as shown in figure 2.

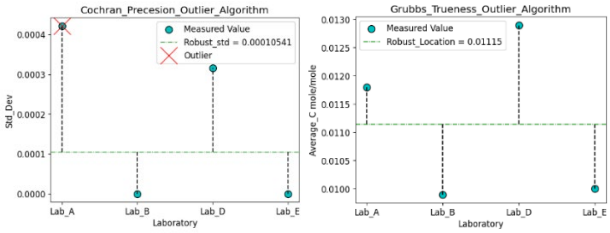


Fig. 2. Outlier detection algorithm

II.II.B. Homogeneity

The algorithm checks the homogeneity by investigate the between and within laboratories standard deviations, figure 3. The criterion of homogeneity is $s_s \leq 0.3\sigma_{pt}$ by compare the between-sample standard deviation s_s with the standard deviation for proficiency assessment σ_{pt}. The criteria $s_s \leq 0.3\sigma_{pt}$ or $s_s \leq 0.1\delta_E$ are used to determine whether the proficiency test items are adequately homogeneous, δ_E represents the target value difference or the minimum detectable difference between samples that is considered significant for the proficiency test. It is often based on the performance requirements or the intended use of the test. These criteria ensure that the variability among samples does not significantly affect the performance evaluation.

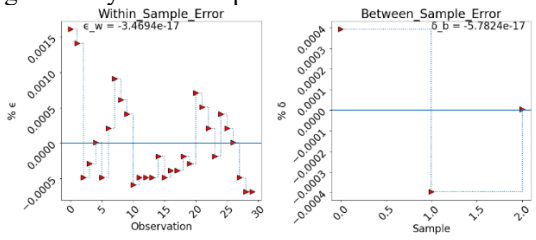


Fig. 3. Between-Within labs standard deviation algorism.

If the homogeneity criterion is satisfied, algorithm used directly σ_{pt} but if not, an extension parameter has been added to compensate if z-score was selected.

II.II.C. STABILITY

The algorithm checks the stability as:

a. If the protocol is used z score (test, samples, chemicals, concentration, etc..), algorithm investigate the difference between averages based on the distributions as $|\overline{y_1} - \overline{y_2}| \leq 0.3\ \sigma_{pt}$ and its expanding form $|\overline{y_1} - \overline{y_2}| \leq 0.3\ \sigma_{pt} + 2\sqrt{u_{\overline{y_1}}^2 - u_{\overline{y_2}}^2}$.

b. If the protocol used En or zetta score (calibration, equipment, device, etc..), algorithm investigate the stability by three parallel techniques.

- *Allowance error*: A suitable statistical criterion was established in order to ensure that the changes in the artifact's results are within the uncertainty and have no

impact on the evaluation of performance for the participants. $\delta_s = |Y_2 - Y_1| \leq U_{Ref}$

- *Trend analysis*: linear regressing fitting, standard error of fit is used as a compensate factor for reference lab uncertainty, figure 4.

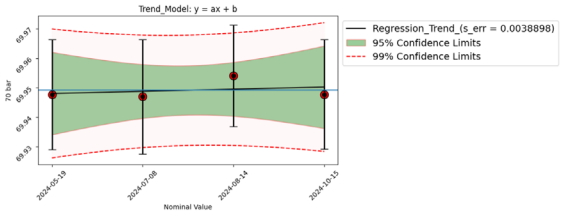


Fig. 4. Trend model.

- *Cumulative sum control chart (CUSUM QC Chart)*. The Algorithm sets the limits to $\pm 2\sigma$ to detect the cumulative error that exceeds the 95% confidence level. The QC was applied to the results reported by the reference laboratory at each nominal point, figure 5.

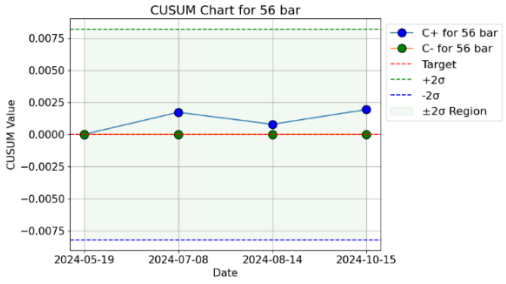


Fig. 5. CUSUM QC Chart.

II.III. ESTABLISHING ASSIGNED VALUES

After confirming homogeneity and stability, assigned values are established for the materials. This involves determining the true or reference values that the materials should exhibit. This step is critical for ensuring that the materials can serve as reliable references in calibration and testing processes. The algorithm treated this part as "if statement coding".

a. If assigned values are coming from reference lab, kernel density plot is used to measure the tendency to center location, it provides a smooth curve describing the general shape of the distribution of a data set (figure 6). The idea underlying the kernel estimate is that each data point is replaced by a specified distribution (typically normal), centered on the point and with a standard deviation bandwidth σ_k. In general and examination inspection, we used $\sigma_k = 0.9\frac{s^*}{p^{0.2}}$, $\sigma_k = 0.25\delta_\varepsilon$ respectively. uncertainty associated with the reference values was evaluated considering the uncertainty of the reference laboratory and the precision of the measurements. The combined standard uncertainty was evaluated as $u_c = \sqrt{u_s^2 + u_{NMCC}^2}$

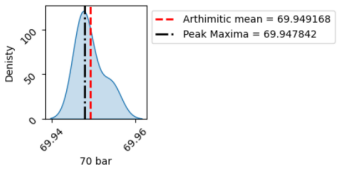


Fig. 6. Kernel density plot.

b. If assigned values are coming from robust statistics.

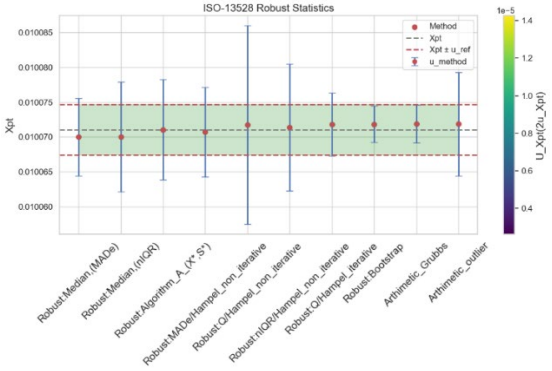


Fig. 7. ISO-13528 robust statistics techniques.

II.IV. REFERENCE LABORATORY AND PARTICIPANT ANALYSIS

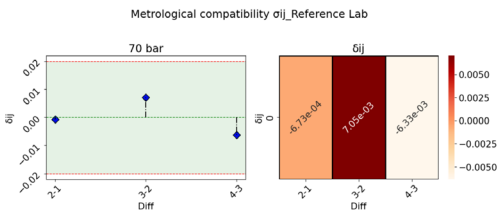


Fig. 8. Metrological compatibility criterion.

Participants: Other laboratories or participants also analyze the same materials independently. Their results are compared against those from the reference laboratory to assess consistency and identify any discrepancies.

II.IV.A. TRUENESS ASSESSMENT

In this section, the error in the measurements results reported by the participants (i.e., the deviation from the reference value) was assessed regardless of the uncertainty reported by the participants. The aim of this section is to investigate whether the measurements conducted by the participants require further correction, figure 9. The error in the results reported by the participant was compared with the reference value. As the results reported by the participants are within the tolerance limits, it indicates a good trueness for the results reported by the participant.

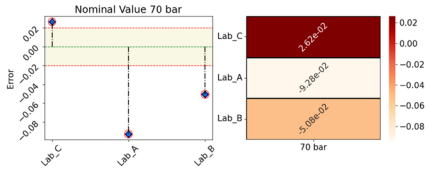


Fig. 8. Trueness Assessment.

II.IV.B. ACCURACY ASSESSMENT

In this section, the deviations of the results reported by the participants are assessed considering the uncertainty reported by the participants. The results reported by the participants were graphically represented, while the expanded uncertainties of the reference values were represented as tolerance lines. Figure 9 illustrates the accuracy assessment for the results reported by the participants. The uncertainty line should be intersected with red line otherwise the accuracy of this measurement is considered to be poor. The laboratory shall consider the comments and recommendations given in this and previous sections and set the appropriate correction/corrective actions for these issues.

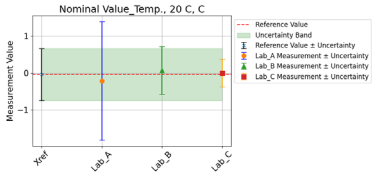


Fig. 9. Accuracy Assessment.

II.IV.C. UNCERTAINTY ASSESSMENT

The uncertainty values reported by the participants have been assessed using the repeatability plot. In this statistical analysis, confidence intervals are established based on the reference value and its associated standard uncertainty. The results of laboratories that lie outside these confidence intervals are considered significantly different from the reference values and corrective/correction actions shall be considered by the corresponding laboratories, figure 10.

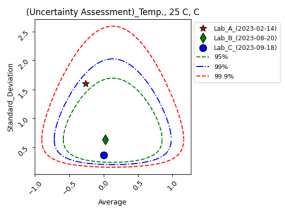


Fig. 10. Uncertainty Assessment.

II.V. COMPUTATION AND STATISTICAL EVALUATION

To check how consistent the errors were across different participants, algorithm used a method developed by Mandel, which is outlined in [2]. This method helps us understand how reliable and repeatable the measurements are when different labs use the same technique. The goal is to see how much the results vary from one lab to another and whether these variations are within an acceptable range. For each lab, we calculate two key numbers: one that shows how precise their measurements are (how close repeated measurements are to each other), and another that shows how accurate they are (how close their measurements are to the true value). This allows us to identify any labs that might have issues with their measurement methods or equipment, figure 11.

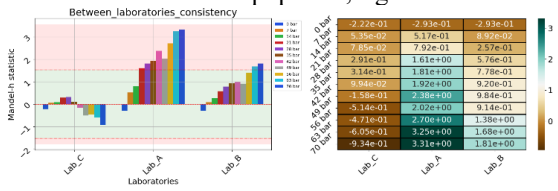


Fig. 11. Mandel's statistic.

II.VI. EVALUATION OF PERFORMANCE

Based on the statistical evaluations of [1, 3], the performance of the materials and the participating laboratories is assessed. This includes evaluating how well the materials meet the required specifications and how accurately the laboratories perform their analyses. Feedback from this evaluation helps identify areas for improvement and ensures that the materials, equipment, device or artifact are fit for purpose, figures 12, 13 for En Score evaluation and Z-score evaluation respectively.

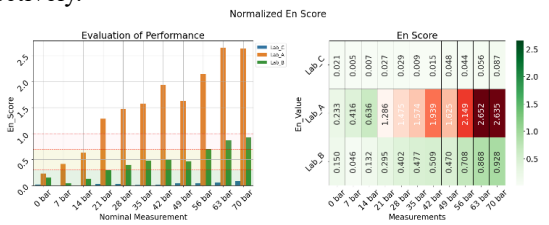


Fig. 12. Normalized En Score Evaluation.

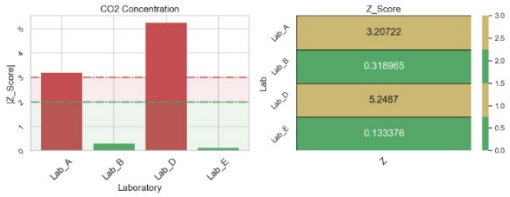


Fig. 13. Z-Score Evaluation.

II.VII. REPORTING

The final step involves generating a comprehensive report summarizing the entire process according to the requirements of ISO 17043. This report documents the results of the homogeneity and stability tests, assigned values, statistical evaluations, and performance assessments. Each step is designed to enhance the reliability and accuracy of the

artifacts, ultimately supporting high-quality measurement practices in various scientific and industrial applications. The systematic approach not only ensures compliance with international standards but also fosters transparency and reproducibility in laboratory operations.

III. PYTHON SCRIPT CODE

Python [4] has gained significant traction as one of the most popular programming languages, largely due to its simplicity, adaptability, and comprehensive ecosystem. The language's syntax is intentionally designed to prioritize clarity and brevity, enabling users to write code that is both intuitive and efficient, which is particularly advantageous for beginners and experts [10]. Python's extensive library support, including tools like NumPy, Pandas, and TensorFlow, offers powerful capabilities for tasks such as data manipulation, artificial intelligence, and automation [11, 12]. Our code consists of more than 1200, 584 and 682 lines of code for calculations, training the model and validation respectively. We introduced here some examples and welcome for any request, help and support to our colleagues or reader of this article around the world.

III.I. STABILITY / HOMOGENEITY CODE

To assess homogeneity, we build a homogeneity check function (*homogeneity_check*) on our dataset, which included multiple samples from two groups (e.g., Sample 1 and Sample 2). The homogeneity test evaluates whether the variance across groups is consistent, using a specified standard deviation (*sigma_pt*) as a threshold for proficiency assessment, figure 14. The results of this test were printed to confirm whether the data met the criteria for homogeneity.

```
# Example data for homogeneity check
homogeneity_data = [
    [10.1, 10.2, 10.3],  # Sample 1 replicates
    [10.4, 10.5, 10.6],  # Sample 2 replicates
    # ... add more samples
]
sigma_pt = 0.5  # Proficiency assessment SD

# Run homogeneity check
hom_result = homogeneity_check(homogeneity_data, sigma_pt)
print(f"Homogeneity Check: s_s={hom_result['s_s']:.4f}, s_w={hom_result['s_w']:.4f}")
print(f"Homogeneous? {'Yes' if hom_result['homogeneous'] else 'No'}")

# Example data for stability check
y1 = 10.2  # Average before
y2 = 10.3  # Average after
u_y1 = 0.1  # Uncertainty in y1
u_y2 = 0.1  # Uncertainty in y2

# Run stability check
stab_result = stability_check(y1, y2, sigma_pt, u_y1, u_y2, use_expanded=True)
print(f"Stability Check: Difference={stab_result['difference']:.4f}, Criterion={stab_result['criterion']:.4f}")
print(f"Stable? {'Yes' if stab_result['stable'] else 'No'}")

# Cochran's test for outlier detection
cochran_result = cochran_test(homogeneity_data)
print(f"Cochran's Test: Stat={cochran_result['cochran_stat']:.4f}, Critical={cochran_result['critical_value']:.4f}")
print(f"Outlier detected? {'Yes' if cochran_result['outlier_detected'] else 'No'}")
```

```
Homogeneity Check: s_s=0.2041, s_w=0.1000
Homogeneous? No
Stability Check: Difference=0.1000, Criterion=0.4328
Stable? Yes
Cochran's Test: Stat=0.5000, Critical=7.7086
Outlier detected? No
```

Fig. 14 Stability / homogeneity code.

Next, we conducted a stability check to evaluate the consistency of measurements over time or under varying conditions. This involved comparing two sets of measurements (y_1 and y_2) with their respective uncertainties (u_y_1 and u_y_2). The stability check function (*stability_check*) calculated the difference between the means of the two datasets and determined whether the observed variation was within acceptable limits. The output indicated whether the data was deemed "Stable" based on predefined criteria. Finally, to identify any potential outliers in our dataset, we applied Cochran's test for outlier detection (*cochran_test*). This test evaluates whether any single observation significantly deviates from the rest of the dataset. The critical value obtained from the test was compared against a pre-defined threshold to determine whether an outlier was detected. The results of these analyses ensured that our data met the necessary statistical standards for further processing and interpretation.

III.II. DECISION TREE TRAINING CODE

The decision tree [13] is a supervised machine learning algorithm [14-19] used for both classification and regression tasks. It works by recursively partitioning the dataset into subsets based on feature values, creating a tree-like structure where each internal node represents a decision rule (e.g., "Is feature X > threshold?"), figure 15.

```
data = data
X = data.data  # Features
y = data.target  # Labels

# Split the dataset into training and testing sets (80% training, 20% testing)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Initialize the Decision Tree Classifier
clf = DecisionTreeClassifier(max_depth=3, random_state=42)  # Limiting depth to prevent o

# Train the model on the training data
clf.fit(X_train, y_train)

# Make predictions on the test data
y_pred = clf.predict(X_test)

# Evaluate the model's performance
accuracy = accuracy_score(y_test, y_pred)
print(f"Accuracy: {accuracy:.2f}")
print("\nClassification Report:")
print(classification_report(y_test, y_pred))
```

Fig. 15 Decision Tree training code.

III.III. Z-SCORE CODE

The code snippet and accompanying plot illustrate the process of analyzing a dataset using statistical methods to assess its distribution and identify key quantiles. The first step involves calculating z-scores for a sample dataset (*sample_props_20*). Z-scores are computed using the formula $z = \frac{lab_i - \mu}{\sigma}$ where μ is the assigned, reference, consensus value based on method used and σ is the standard deviation, figure 16.

```
zscores = [(value - mu) / std for value in sample_props_20]
quantile_z=np.quantile(zscores, [0.025,0.15, 0.5,0.85 , 0.975])

# Plot the histogram.
plt.hist(zscores, bins=100, density=True, alpha=0.6, color='g')
# Plot the PDF.
xmin, xmax = plt.xlim()
x = np.linspace(xmin, xmax)
p = norm.pdf(x, 0, 1)
plt.plot(x, p, 'r', linewidth=2)
plt.axvline(x=quantile_z[0], color='b', linewidth=1.5, dashes=[6, 2])
plt.axvline(x=quantile_z[1], color='b', linewidth=1.5, dashes=[6, 2])
plt.axvline(x=quantile_z[2], color='b', linewidth=1.5, dashes=[6, 2])
plt.axvline(x=quantile_z[3], color='b', linewidth=1.5, dashes=[6, 2])
plt.axvline(x=quantile_z[4], color='b', linewidth=1.5, dashes=[6, 2])
```

```
<matplotlib.lines.Line2D at 0x1d20e82b490>
```

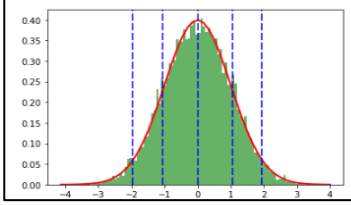


Fig. 16 Z-Score Code.

IV. MODEL VALIDATION

Annex E (Informative) in [1] provides a set of examples designed to validate the procedures outlined in the main document. We have successfully validated all 15 examples and compared the results from our model with those mentioned in the standard. Our results fully match the standard values, with even greater precision in the number of significant digits. For simplicity, we present only two of these validated examples here.

E.3 Comprehensive example of atrazine in drinking water

A proficiency testing scheme for Atrazine in drinking water has 34 participants. All coming section will match and compare the values that are calculated by ISO 13528 [1] and SASO model. According to the table E4 in [1], figures 17 and 18 are described the sorted data increment in standard and in SASO model.

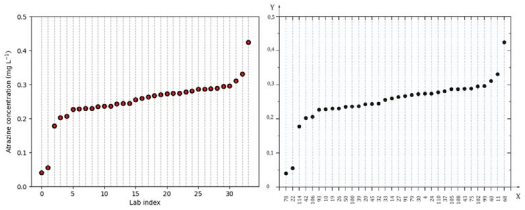


Fig. 17 SASO Model. *Fig. 18 ISO 13528 Model.*

The Kernel density plot for participant results is shown in figures 19, 20.

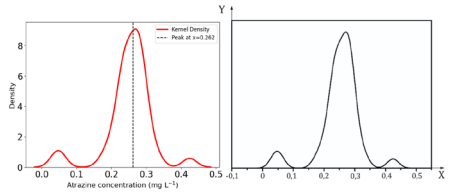


Fig. 19 SASO Model. Fig. 20 ISO 13528 Model.

ISO 13258 model provide the calculated values of location, standard deviation and uncertainty, SASO model match those numbers with performance 99.98% as in table 1.

Table 1. Summary Statistics of SASO Model

Robust: Median, nIQR	[52]: Median_ [52]: 0.262	[57]: nIQR [57]: 0.04023405749999	[58]: u_nIQR [58]: 0.00862510491
Robust: Median, MADe	[52]: Median_ [52]: 0.262	[53]: MADe [53]: 0.038557999999999	[58]: u_nIQR [58]: 0.00862510491
Robust: Algorithm A (x*, s*)	[70]: itr_alga [70]: 0.257016	[69]: itr_alga(df.xi,s [69]: 0.039495557070355	[71]: u_alg_A [71]: 0.0084832722
Robust: Q/Hampel	[103]: Results.Xpt [103]: 0.259633141	[107]: Results.opt.iloc [107]: 0.04257499999997	[111]: Results['u_Xpt [111]: 0.00919740426
Bootstrap (for mean)	[104]: Results.X [104]: 0.2503278	[113]: Results.opt.iloc[4 [113]: 0.066768844	[115]: Results['u_Xpt [115]: 0.011583478251
Arithmetic, outliers removed	[41]: np.mean(Gru [41]: 0.258761290	[117]: Results.opt.iloc[5 [117]: 0.033683012118	[118]: Results['u_Xpt'] [118]: 0.006147504708
Arithmetic, outliers included	[105]: Arth_out[8 [105]: 0.25121176	[120]: Results.opt.iloc[6 [120]: 0.0672108160699707	[118]: Results['u_Xp [118]: 0.0115347

CONCLUSION

A novel smart algorithm has been developed to automate the implementation of key standards, including ISO 13528, ISO 5725-2, and ISO 17034, leveraging machine learning techniques implemented in Python. This algorithm demonstrates exceptional performance by successfully validating all 15 examples provided in ISO 13528, achieving nearly complete alignment with the standard's results or, in some cases, even surpassing their accuracy. The robustness and precision of the algorithm underscore its potential as a reliable tool for automating complex statistical procedures in proficiency testing and interlaboratory comparisons. Looking ahead, the next steps involve enhancing its accessibility and scalability by integrating the algorithm into a FastAPI framework and reprogramming the codebase to function as a web-based application.

REFERENCES

1. ISO 13528:2022 Statistical methods for use in proficiency testing by interlaboratory comparison.
2. ISO 5725-2:2019 Accuracy (trueness and precision) of measurement methods and results - Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method.
3. ISO/IEC 17043:2023 Conformity assessment - General requirements for the competence of proficiency testing providers.
4. Python Software Foundation, *Python.* Available: https://www.python.org.
5. NumPy Developers, *NumPy.* Available: https://numpy.org.
6. SciPy Developers, *SciPy.* Available: https://scipy.org.
7. Pandas Development Team, *Pandas.* Available: https://pandas.pydata.org.
8. Matplotlib Development Team, *Matplotlib.* Available: https://matplotlib.org.
9. Seaborn Development Team, *Seaborn.* Available: https://seaborn.pydata.org.
10. Python Software Foundation, *Python 3 Reference Manual.* Scotts Valley, CA: CreateSpace Independent Publishing Platform, 2009. ISBN: 978-1-4414-1269-0.
11. T. E. Oliphant, "Python for Scientific Computing," *Computing in Science & Engineering,* vol. 9, no. 3, pp. 10-20, 2007. DOI: 10.1109/MCSE.2007.58.
12. W. McKinney, Python for Data Analysis: Data Wrangling with Pandas, NumPy, and I Python. Sebastopol, CA: O'Reilly Media, 2010.
13. Scikit-learn Developers, *Decision Trees in Scikit-learn.* Available: https://scikit-learn.org/stable/modules/tree.html.
14. J. R. Quinlan, "Induction of Decision Trees," *Machine Learning,* vol. 1, no. 1, pp. 81-106, 1986, doi: 10.1007/BF00116251.
15. T. M. Mitchell, *Machine Learning* . New York, NY, USA: McGraw-Hill, 1997.
16. C. M. Bishop, *Pattern Recognition and Machine Learning* . New York, NY, USA: Springer, 2006.
17. S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach,* 3rd ed. Upper Saddle River, NJ, USA: Pearson Education, 2010.
18. L. Breiman, "Statistical Modeling: The Two Cultures," *Statistical Science* , vol. 16, no. 3, pp. 199-231, 2001, doi: 10.1214/ss/1009213726.
19. V. N. Vapnik, *The Nature of Statistical Learning Theory.*

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

A Timing Accuracy Assessment System Prototype for Multiple NTP Servers

Deepak Sharma[1,2], Divya Singh Yadav[1,2], Preeti Kandpal[1], Bharath Vattikonda[1,2], Ashish Agarwal[1,2]

[1]*CSIR National Physical Laboratory, New Delhi, India*
[2] *Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, India*
deepak.sharma@nplindia.res.in

***Abstract* – Time synchronization is a critical aspect of modern IT systems. The Network Time Protocol (NTP) plays a crucial role in maintaining consistent and accurate time across digital infrastructure worldwide. NTP is aligned with digital metrology vision, enabling precise timekeeping through the remote correction of clocks. National Metrology Institutes (NMIs) around the world, the custodians of national time standards, play a pivotal role in disseminating accurate time information through NTP services. Multiple NTP servers are generally considered for redundancy and load distribution. To maintain this ecosystem, it is crucial to ensure the timing accuracy and reliability of all the NTP servers. This work presents a system prototype for ensuring timing accuracy of multiple NTP severs. The prototype has been developed considering self-developed Python program for collecting and processing NTP data along with a Grafana based visualization dashboard for timing accuracy assessment. The system may be useful in detecting and isolating faulty NTP servers. The system may be enhanced with Machine Learning based early anomaly detection and alerting mechanisms. The work may be useful for developing application specific customized dashboards for different metrological systems.**

I. INTRODUCTION

With the rise of time-sensitive and innovative applications, precise timing has become more critical than ever before. Clocks are perhaps among the early instruments to corrected remotely with the exchange of timing messages rather than physical artifacts. Remote correction allows a clock to stay on-site and in operation while its performance is monitored in parallel and any deviations from nominal behavior are detected and corrected. In coordination with the International Bureau of Weights and Measures (BIPM), National Metrology Institutes (NMIs) across the globe plays a pivotal role in global timekeeping and time synchronization [1]. Various modern domains such as IT systems, Cyber-Physical Systems (including the Internet of Things), and Industry 4.0 require timing accuracy ranging from milliseconds to microseconds [1-3]. Multiple methods exist to achieve time synchronization and ensure traceability to a recognized time reference, e.g., Network Time Protocol (NTP) [4]; Precision Time Protocol (PTP); White rabbit; All in view global navigational satellite system (AVGNSS); Common view global navigational satellite system (CVGNSS) and two-way satellite time and frequency transfer (TWSTFT) [1-3].

Network Time Protocol (NTP) is one of the most widely used technologies for global time synchronization, which maintains system clocks on various devices on network/Internet with an accuracy within 100 milliseconds [1]. NTP servers help in coordinating/maintaining the system time of networked computers/devices in agreement with UTC. NTP service providers, like NMIs across the globe, consider multiple NTP servers for redundancy and NTP load distribution. The behavior of the internal clocks on NTP servers can be dynamic in comparison to the highly stable clocks preferred in metrology. Monitoring is a crucial aspect of IT systems for maintaining the reliability, performance, and security. It leads to improved efficiency and reduces downtime/costs. Monitoring timing accuracy is a highly specialized area to ensure all time providing systems within a network maintain precise and synchronized time, often to the level of milliseconds or microseconds. Unlike general IT monitoring, which focuses on performance indicators like CPU usage, memory consumption, etc., the timing accuracy monitoring targets specific metrics such as time offset and delay.

This paper presents an open source technologies based timing accuracy assessment system prototype for multiple NTP servers. It may be very useful for NMIs and other entities hosting multiple NTP servers. The system considers open-source, platform-independent technologies, which makes it easy to replicate and customize for an application specific environment. The concept may be helpful for developing similar systems for other metrological domains.

The paper is organized as follows. Section II introduces timing accuracy assessment aspects of Network Time Protocol. This section also covers relevant literature. Section III presents the methodology and architecture of the system. The functional system prototype and future

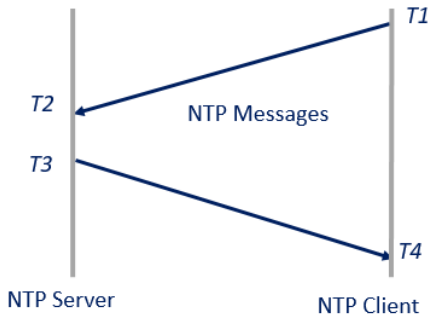


Fig. 1. NTP message exchange.

plans are discussed in Section IV. Section V concludes the paper.

II. TIMING ACCURACY ASSESSMENT FOR NETWORK TIME SYNCHRONIZATION

The time synchronization in NTP is based on four timestamps (*Fig.1*). At time *T1* (Originate Time) a client sends an NTP request to an NTP server, which is received by the NTP server at *T2* (Receive Time). The server sends a reply at *T3* (Transmit Time), which is received by the client at time *T4*. The time offset (θ) and round trip delay (Δ) are calculated as

$$\theta = 1/2 * [(T2-T1) + (T3-T4)] \quad (1)$$

$$\Delta = (T4-T1) - (T3-T2) \quad (2)$$

A client device can correct its clock based on the offset value (θ) with respect to the reference NTP server. NTP operates through a hierarchical structure, with Stratum-1 servers at the top. These servers obtain their time from highly accurate time sources, such as atomic clocks or Global Navigation Satellite Systems (GNSS) like Global Positioning System (GPS) receivers, which are typically traceable to Coordinated Universal Time (UTC). This traceability enables consistent synchronization of system clocks across networked devices. Lower-level NTP servers (with higher stratum values) receive accurate time from upper-level NTP servers (with lower stratum values). The stratum-1 NTP servers, which are hosted at NMIs, like NPL India (NPLI), take time references (generally 1 pulse per second signal) directly from their atomic time scale and hence provide the UTC traceable time over the network/Internet within 100 milliseconds accuracy.

NMIs across the globe are working on timing accuracy assessment for NTP services for end customers. National Research Council (NRC), Canada monitors NTP servers and document their traceability for traceability related client requests [5]. The Real Instituto y Observatorio de la Armada (ROA) considers NTP service calibration assessment over a Local Area Network based on data acquired from NTP server files [6]. A study considers python based program for offline performance analysis of six public NTP services based on manually collected NTP data from two distant locations [7]. Lombardi et al. [8] presented a system under SIM Time and Frequency Metrology Working Group for comparisons of NTP servers located in different regions. The system considers

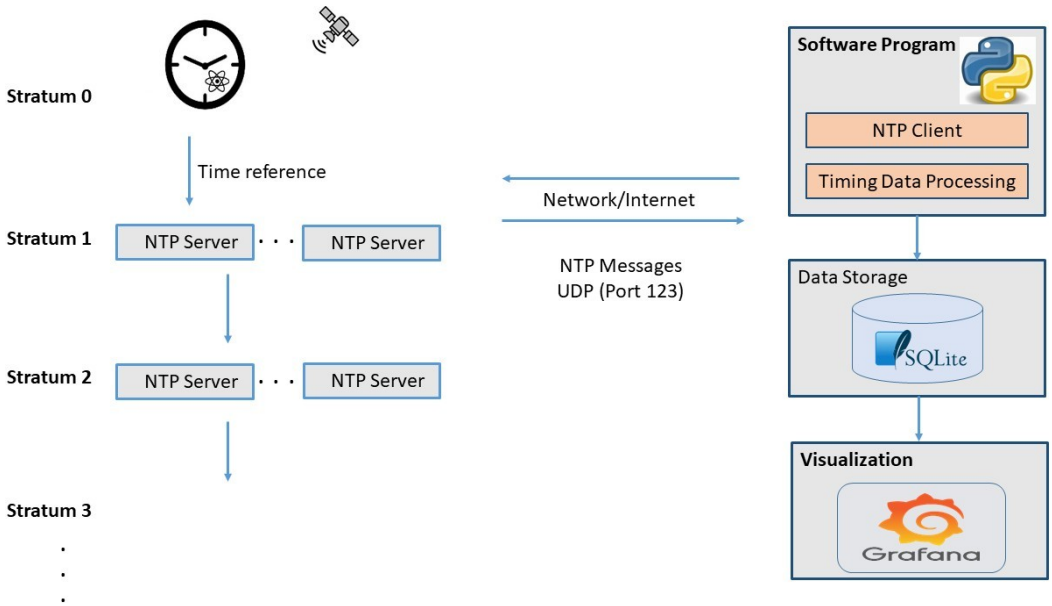


Fig. 2. System Architecture.

a 100 ns resolution clock board, synchronized to time reference via a 1 PPS signal with compensated delay. It compares NTP timestamps using both direct and common-view methods, and publishes the accuracy results of NTP servers on a website. Although proprietary tools are also available for NTP monitoring and other time and frequency metrology applications, open-source solutions offer greater transparency and customization flexibility, making them well-suited for NMIs. For instance, the Open Traceable Time Platform is an open-source timing system based on GNSS time transfer [9].The timing accuracy assessment of NTP servers is an important requirement and open source technologies based solution may be explored in this area.

III. METHODOLOGY

The system architecture of presented system prototype is shown in Fig.2. The developed system considers a python based program, which uses ntplib library for NTP messages exchange with multiple NTP servers. The key information is extracted from the NTP packets, processed and stored in a database. The NTP message exchange process is repeated periodically in rounds after few minutes interval to get the latest information. The NTP messages received from all the NTP servers under observation in a particular round are processed. The mean value of time offset values of all the servers for a particular round is calculated and relative offset values for each server is calculated with respect to the calculated mean value for the round. Median value may also be considered instead of mean value for each round. It is assumed that the client device clock remains stable and does not deviate/drift significantly within a round due to its internal clock variation. The data is currently stored in SQLite database and the functionality has been tested with PostgreSQL database also.

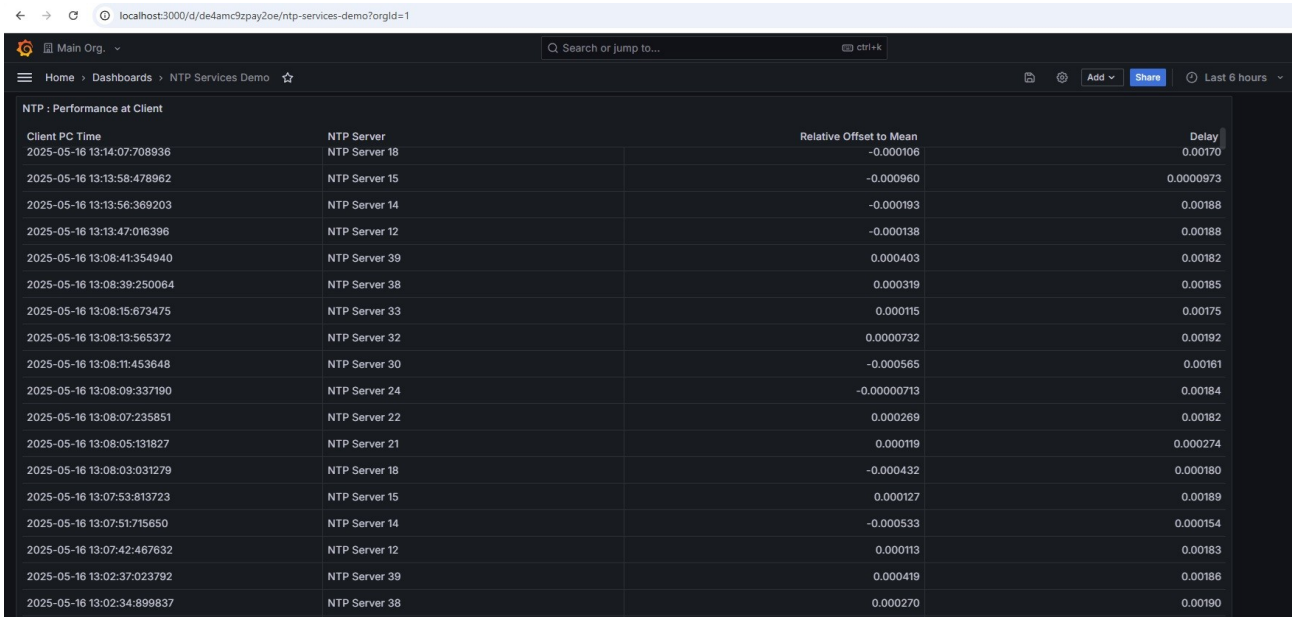
Grafana Open Source Software [10] is considered for data visualization. Grafana Open Source Software is a versatile visualization and monitoring platform with multi-platform compatibility and support for variety of data sources like SQL, NoSQL and time series databases like Prometheus and CloudWatch. It helps in creating real-time dashboards to showcase data through insightful graphs and visualizations helpful in monitoring, alerting and system behavior analysis. Grafana Enterprise is a commercial edition with additional features and support. Multiple dashboards are created based on customized SQL queries on database connectivity.

IV. RESULTS AND DISCUSSION

The system prototype including the python program, the database and the Grafana dashboard (Fig. 2) are deployed on a computer on Local Area Network. The NTP client requests are generated from this computer to the NTP servers hosted in a secured zone. The NTP information is processed and stored in the database, where the Grafana software can fetch the latest information from the database. Grafana dashboards can be accessed using any web browser from any computer on the network with the IP address of the hosting computer.



Fig. 3. Latest relative Offset and Delay values based on NTP message exchange



Client PC Time	NTP Server	Relative Offset to Mean	Delay
2025-05-16 13:14:07:708936	NTP Server 18	-0.000106	0.00170
2025-05-16 13:13:58:478962	NTP Server 15	-0.000960	0.0000973
2025-05-16 13:13:56:369203	NTP Server 14	-0.000193	0.00188
2025-05-16 13:13:47:016396	NTP Server 12	-0.000138	0.00188
2025-05-16 13:08:41:354940	NTP Server 39	0.000403	0.00182
2025-05-16 13:08:39:250064	NTP Server 38	0.000319	0.00185
2025-05-16 13:08:15:673475	NTP Server 33	0.000115	0.00175
2025-05-16 13:08:13:565372	NTP Server 32	0.0000732	0.00192
2025-05-16 13:08:11:453648	NTP Server 30	-0.000565	0.00161
2025-05-16 13:08:09:337190	NTP Server 24	-0.00000713	0.00184
2025-05-16 13:08:07:235851	NTP Server 22	0.000269	0.00182
2025-05-16 13:08:05:131827	NTP Server 21	0.000119	0.000274
2025-05-16 13:08:03:031279	NTP Server 18	-0.000432	0.000180
2025-05-16 13:07:53:813723	NTP Server 15	0.000127	0.00189
2025-05-16 13:07:51:715650	NTP Server 14	-0.000533	0.000154
2025-05-16 13:07:42:467632	NTP Server 12	0.000113	0.00183
2025-05-16 13:02:37:023792	NTP Server 39	0.000419	0.00186
2025-05-16 13:02:34:899837	NTP Server 38	0.000270	0.00190

Fig. 4. Offset and Delay values based on NTP message exchange.

The dashboard shown in Fig.3 visualizes the latest values of relative offset of each NTP server with respect to the mean offset value of all the servers as observed by the NTP client based on NTP message exchange. The NTP message exchange round trip delay for each server is also shown, as a higher round trip delay may lead to higher measurement uncertainty. The image shows that the relative offsets values for all the servers are within 1 millisecond and the round trip delay values are within 2 millisecond. It shows that all the servers are available for NTP packet exchange and providing correct time to the end client. The dashboard shown in Fig. 4 visualizes server specific timestamped relative offset to mean and the delay values based on NTP message exchange in a tabular form. This allows behavior analysis and finding anomalies considering the information for longer time duration. The delay variation gives insights about the network performance also.

The work is planned to be enhanced with customized dashboards, alerting mechanisms and machine learning based insights.

V. CONCLUSION

Network Time Protocol is one of the most widely used technologies for time synchronization over networks/Internet. NMIs and other entities generally consider multiples NTP servers for redundancy and handling larger number of NTP requests. Monitoring of such systems requires specialized tools for ensuring timing accuracy integrity and early detection and removal of any malfunctioning NTP server. An open-source customizable solution is presented based on a python software and Grafana based dashboards. The python based software generates NTP requests to the multiple NTP servers, collects the responses, processes the data and stores in a database. Grafana dashboards are created, which fetch the data based on queries for requirement-specific visualization. The performance of system prototype is found satisfactory. The system may be further enhanced with additional customized dashboards and alert generation. Machine Learning techniques can be considered for finding anomalies and discovering hidden patterns and trends in the collected data. The system may be customized for verity of other metrological applications.

ACKNOWLEDGEMENT

Authors are thankful to CSIR India, CDOT India and CSIR-National Physical Laboratory (NPLI), India for supporting this work.

REFERENCES

[1] D. K. Aswal, “Metrology for inclusive growth of India” Springer Nature, 2020, doi: 10.1007/978-981-15-8872-3.

[2] Sharma, D., Yadav, D., Vattikonda, B., Agarwal, A. (2024). Timing Accuracy for Internet of Things and Industry 4.0 Applications: Technologies and Research Trends. In: Yadav, S., Garg, N., Aggarwal, S.G., Jaiswal, S.K., Kumar, H., Achanta, V.G. (eds) Recent Advances in Metrology. AdMet 2022. Lecture Notes in Mechanical Engineering. Springer, Singapore.

[3] M. Lévesque and D. Tipper, "A survey of clock synchronization over packet-switched networks,"

IEEE Communications Surveys & Tutorials, vol. 18, pp. 2926-2947, 2016.

[4] D. Mills, J. Martin, J. Burbank, and W. Kasch. 2010. RFC 5905: Network Time Protocol Version 4: Protocol and Algorithms Specification. RFC Editor, USA.

[5] Douglas, Rob, André Charbonneau, and Marina Gertsvolf. "Remote calibration of time from NTP servers." Metrologia 55, no. 6 (2018): 855.

[6] Vélez, Carmen, Javier Díaz, Alfonso Osuna, Héctor Álvarez-Martínez, and Héctor Esteban. "Assessment of an NTP service calibration over a Local Area Network." Measurement Science and Technology 35, no. 6 (2024): 065013.

[7] F. F. A, A. Agarwal, D. Sharma, D. S. Yadav and S. M. R, "Time Synchronization Performance Analysis based on NTP Data Captured from Two Distant Locations," 2025 International Conference on Innovation in Computing and Engineering (ICE), Greater Noida, India, 2025, pp. 1-6, doi: 10.1109/ICE63309.2025.10984309.

[8] Lombardi, Michael A., Judah Levine, J. Mauricio Lopez, Francisco Jimenez, John Bernard, Marina Gertsvolf, Harold Sanchez et al. "International comparisons of network time protocol servers." In Proceedings of the 46th annual precise time and time interval systems and applications meeting, pp. 57-66. 2014.

[9] M. J. Wouters, E. L. Marais, A. S. Gupta, A. S. bin Omar and P. Phoonthong, "The Open Traceable Time Platform : Open Source Software and Hardware for Dissemination of Traceable Time and Frequency," 2019 Joint Conference of the IEEE International Frequency Control Symposium and European Frequency and Time Forum (EFTF/IFC), Orlando, FL, USA, 2019, pp. 1-2, doi: 10.1109/FCS.2019.8856026.

[10] Grafana OSS and Enterprise, Available: https://grafana.com/docs/grafana/latest/?pg=oss-graf&plcmt=hero-btn-2#grafana-oss-and-enterprise

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

The Regional Metrology Organisations Coordination Working Group of the CIPM FORUM on Metrology and Digitalization. Helping Emerging NMIs on Their Path to Digital Transformation

Nikita D Zviagin[1]

[1] *D.I.Mendeleyev Institute for Metrology, Russia, n.d.zviagin@vniim.ru, zviaginn@gmail.com*

***Abstract* – On the first Meeting of the CIPM Forum on Metrology and Digitalization (FORUM MD) in the March 2024 several task and working groups were established. Amongst them is the Working Group on RMO Coordination (FORUM MD WG RMO) which includes the chairpersons of working or task groups responsible for the digitalization projects inside 6 RMOs (AFRIMETS, APMP, COOMET, EURAMET, GULFMET and SIM). One of the main tasks of the group is to provide a link through RMOs between the Forum and every NMI (and their stakeholders), including NMIs that are not participating in the work of the Forum and are not yet the Signatories of the CIPM MRA and/or the Metre Convention, thus providing emerging NMIs with the possibility to get recent information on the topic and observe best practices in the field that have been already implemented by developed NMIs. The paper gives information about the FORUM MD WG RMO, its structure and participants, the work already done, current activities and future plans.**

I. INTRODUCTION

Regional Metrological Organisations (RMOs) play significant role in providing link between CIPM CCs and Forums, which due to their missions and rules include NMIs with highest expertise in corresponding areas and small, young and emerging NMIs. The task of digital transformation of metrology that in many cases require use of high-end informational technologies, expensive equipment, involving of experienced IT-specialists and/or additional education of inner metrological staff makes the gap in possibilities and capabilities between big and smaller NMIs even bigger. In that case the way how RMO organises its work, distributes information, provides learning courses and coordinates the cooperation between NMIs in the region becomes crucial. Moreover the situation persists on the higher level. Different RMOs supported by their NMIs demonstrate different capabilities in the specific issues of digital transformation. In order to resolve that problem and facilitate the process of interregional cooperation on digital transformation matters the CIPM FORUM on Metrology and Digitalisation (the FORUM MD) on its first meeting in the March 2024 has established the Working Group on Coordination between Regional Metrology Organisations (the WG RMO). The group consists of six members – the heads of corresponding RMO working/focus or task groups on digitalisation and a number of interested observers.

This paper talks about the mission of the WG RMO, its objectives, the work done through the first year of activity, plans and perspectives for the future.

II. MISSION AND OBJECTIVES

Following the initial recommendations received during the First Meeting of the FORUM MD, discussions inside the WG RMO throughout the year and the recommendations of FORUM MD Working Group on Strategy the terms of references for the WG RMO we formulated as follows:

Area:

The purpose of this group is to harmonize procedures on different digitalization topics between the RMOs, interacting with the FORUM MD Task Groups and Working Groups.

Activities:

Coordinate the work between RMOs in the field of Digital Transformation in Metrology.

Exchange information on activities in regions to share knowledge transfer and explore synergies.

Communicate regional needs, relating to digital transformation, to the Forum, the CIPM, the BIPM.

Provide a link through RMOs between the Forum and every NMI (and their stakeholders), including NMIs that are not participating in the work of the Forum and are not yet the Signatories of the CIPM MRA and/or the Metre Convention, and

Discuss the needs of the emerging NMIs.

These terms of references were approved on the Second Meeting of the FORUM MD that took place at the BIPM in the February 2025[1].

III. STRUCTURE

The Working Group consists of six Members representing six RMOs . Each of the members heads the corresponding working/focus or task group in his region, namely:

For AFRIMETS: Digitization Working Group
For APMP: Focus Group on Digital Transformation (DXFG) [2]
For COOMET: Task Group on the issues of digital transformation [3]
For EURAMET: Working Group on Metrology for Digital Transformation [4]
For GULFMET: Digital Transformation Working Group [6]
For SIM: Metrology for Digital Transformation Working Group [7]

Also there is a number of observers from the NMIs interested in the activities of the group. The participation in the group in status of observer is open for any representative of the NMIs which are either Members or Observers of the FORUM MD. The group is headed by the Chair, who is elected for the two years term from the members (although it's not the case for the present moment with current interim chair who has been appointed by the FORUM MD).



Fig. 1. The 3rd Meeting of WG RMO

IV. WHAT HAS BEEN DONE SO FAR

There have been 4 meetings of the working group during the period between the first and the second meetings of the FORUM MD that were organized by the BIPM through the ZOOM platform. The screenshot from one of the meeting is shown on the Fig. 1.

The 5th meeting of the WG RMO took place in hybrid format at the BIPM on 18th February 2025 during the FORUM MD week.

The two main topics were in the focus during those meetings:

- the exchange of the information on current and planned activities in the digital transformation area in the RMOs

- the analysis of current situation on the set of issues inside the RMOs. The analysis has been organized as follows:

A. The Analysis

An analysis table has been prepared by the Chair. It consisted of several sections each dedicated to one of the following issues from the area of digital transformation which were chosen on the base of existing FORUM MD Working groups list and some additional topics identified by the members of the WG RMO:

- Digital Calibration Certificates (DCCs)
- FAIR Data
- Artificial Intelligence (AI)
- BIPM APIs
- SI reference point (SIrefPoint)
- Semantics
- Authentication of documents
- Sensor networks
- Automation of processes, robotization
- Legal metrology matters: digital type approval certificates, measuring instruments passports

Each section had the following list of questions:

- Are there any works implemented in the RMO? The level of implementation
- Main problems? What help is needed from other RMOs and the FORUM?
- What an RMO could suggest? (Workshops, guidelines, webresources)
- What are the main stakeholders and end users of the sevice in the RMO.

Every member of the WG RMO has filled in the cells of the table, presenting that way the summary information on achievements and lack of knowledge on different issues in their RMOs.

The main aim of the analysis was in

- Identifying common problems and immediate actions needed and
- Developing of a scope of recommendations to RMOs and FORUM MD on organizing joint workshops and preparing joint guidelines on different issues

The Table of Analysis has been distributed between the FORUM MD TGs. Feedback was discussed on the meeting of the WG RMO during the FORUM MD Week in February 2025.

The zoomed-out part of the table of Analysis is presented on the Fig 2.



Fig. 2. The table of Analysis

V. THE RESULTS

After the conducted analysis the WG RMO has came to the following conclusions:

- On most of the issues of Digitalisation the works in RMOs are only on the beginning stage;
- DCC is probably the most covered issue so far;
- Other topics are either covered by one or two RMOs or not covered at all;
- On the current stage there is not too much work for the WG RMO. The Group has to wait for the task groups to provide some results of their work and then help in distribution and implementation of it.

The following recommendations were prepared:

- FAIR Data is chosen as a «hot topic» and WG RMO suggests to have and propose organising a webinar (or series of webinars) to the FORUM MD
- Use of e-learning platfrom for the need of the FORUM is reccomended

VI. FUTURE PLANS

The idea on the series of Fair Data webinars has been approved by the FORUM MD and supported by the members of the Task Group on FAIR data. So the RMO WG plans to conduct together with the TG FAIR data webinars on the topic for all the RMOs. The main aim of these webinars is to introduce the principles of FAIR data to the global metrological community mostly focusing on the emerging and smaller NMIs which are not aware about the topic. This will be one of the main activities for this year.

The WG RMO will continue actively meeting throughout the year, discussing the activities, taking place inside the RMOs and sharing the information with each other.

More active collaboration with other task groups of the FORUM MD is planned.

VII. CONCLUSIONS

The active discussions inside the WG RMO during the meetings together with the positive feedback from other task groups, working group on strategy and the FORUM MD as a whole have shown that FORUM MD decision on creation of the WG RMO has been the right and rational thing. The coordination between different NMIs inside one RMO or globally still requires more attention. Emerging NMIs need help from the CIPM, BiPM, its CCs and FORUMS especially in such a difficult and complex issue as the digital transformation for metrology.

REFERENCES

[1] FORUM-MD Working Group on coordination between Regional Metrology Organizations (FORUM-MD-WG-RMO) webpage on the BIPM

website: https://www.bipm.org/en/committees/fo/forum-md/wg/forum-md-wg-rmo

[2] APMP Digital Transformation webpage: https://apmpweb.org/portal/list/index/id/52.html

[3] COOMET Digital transformation in metrology webpage: https://www.coomet.net/activities/digital-trasformation-in-metrology-tg-digtr?type=94

[4] EURAMET Metrology for digital transformation webpage https://www.euramet.org/technical-committees/tc-im/metrology-for-digital-transformation

[5] GULFMET Digital Transformation Working Group webpage: https://www.gso.org.sa/gd/committees/GENERAL/committee/DT-WG

[6] SIM Metrology for Digital Transformation Working Group webpage: https://sim-metrologia.org/

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Data Quality Characteristics for Improved Metrology in Sensor Networks

Mads Johansen[1*], Anupam Prasad Vedurmudi[2], Martha Arbayani Zaidan[3], Milos Davidovic[4], Gertjan Kok[5], Maitane Iturrate-Garcia[6], Shahin Tabandeh[7]

[1]*FORCE Technology, Hørsholm, Denmark*
[2]*Physikalisch-Technische Bundesanstalt (PTB), Braunschweig, Germany*
[3]*University of Helsinki, Helsinki, Finland*
[4]*Vinca Institute of Nuclear Sciences (VINS), Belgrade, Serbia*
[5]*Van Swinden Laboratorium (VSL), Delft, Netherlands*
[6]*Federal Institute of Metrology (METAS), Bern-Wabern, Switzerland*
[7]*Technical Research Centre of Finland (VTT), Helsinki, Finland*
[*]*Corresponding author. E-mail address: majh@forcetechnology.com*

***Abstract* – Sensor networks are becoming increasingly practical to deploy in largely varying settings, which combined with the growing availability of low-cost sensors and the increasing scale of sensor networks, makes it highly challenging to ensure the trustworthiness and reliability of measurements and data. Factors such as physical inaccessibility and cost constraints make it infeasible to use established methods for calibration, further increasing the difficulty of assessing measurement uncertainty and ensuring traceability in sensor networks. In addition, the large volume of data generated makes the assessment of data quality in sensor networks infeasible without automated, efficient, and reliable methods. This paper explores how well-known data quality characteristics can be applied in a metrologically sound manner, enabling quality assessments even when reference data or traditional calibration data are unavailable.**

Keywords: Data quality, sensor network, metrology

I. INTRODUCTION

Sensor networks are a valuable asset in many settings, such as industrial and smart city applications, as they enable a data-driven approach to decision making by collecting real-time data from various entities. Sensor networks are becoming increasingly practical to deploy, in part because of the availability of low-cost sensors and open-source tools. However, low-cost sensors often have lower measurement quality. Combining this with sensor networks becoming ever-larger and the vast amounts of data they can collect, makes it particularly challenging to ensure trustworthy and reliable data. [1]

Established methods for sensor calibration are used to ensure traceability of measurements and estimate measurement uncertainty. However, these methods involve removing a sensor from its deployment location, transporting it to a laboratory for calibration and redeploying it afterwards. For many sensor networks, this is infeasible due to several factors, such as physical inaccessibility and cost constraints, further adding to the difficulty of assessing data quality. In itself, the huge amounts of data make the assessment of data quality by conventional means impractical without the availability of automated, efficient, and reliable methods, due to the significant amounts of resources required [1]. In cases where such methods are not established, discovering issues relies on the manual inspection of data, which results in substantial risks of basing conclusions and decisions on poor quality data. Thus, from a metrological perspective, new methods must be developed to measure data quality in sensor networks.

Real-world sensor networks vary greatly in many aspects, including associated legal and financial aspects [2, 3]. Some sensor networks use highly specialized and costly sensors for normative applications (e.g. network of gas flow meters in gas distribution pipelines), which will need to be calibrated regularly to avoid high financial and legal implications in case of incorrect measurements. Other sensor networks are less specialized, possibly using battery-powered sensors, and have lower associated costs (e.g. for maintenance, sensor replacement, etc.). These kinds of sensor networks are becoming quite common, for example, in smart cities and citizen science applications (e.g. for monitoring of local air quality and comparison with reference air quality networks).

This paper focuses on less specialized sensor networks, with limited legal or financial implications where data quality is somewhat unknown due to the mentioned challenges and where a strict metrological procedure is not mandatory.

The paper is structured as follows. In Section II. alternative calibration methods for sensor networks are intro-

duced, highlighting methods that do not require additional reference sensors. Section III. highlights metadata as a tool to address different data quality characteristics. Section IV. describes different data quality characteristics relevant to sensor data quality from a metrological perspective and discusses how they can be applied in the described context. Lastly, Section V. concludes the paper.

II. CALIBRATION METHODS

There are a few different approaches to performing calibration in sensor networks in scenarios where no ground-truth reference sensor is available. Such calibrations are likely to be performed as online calibrations, which refer to calibrations performed during normal operation of sensors [4], i.e. without removing sensors or turning off (parts of) the sensor network. In-situ calibration is used in the case of online calibration, as it refers to the calibration of a sensor performed at the location of its deployment without having to disassemble and transport it to a calibration laboratory [5]. It is possible to perform both with and without a reference sensor. Co-calibration can be considered as a special case of in-situ calibration where nearby sensors already present in the network are used as reference devices [6]. In the case of co-calibration, the reference value itself may have to be estimated at the position of the device under test using appropriate interpolation and sensor fusion techniques. This is also related to blind calibration, which refers to calibration without controlled stimuli or ground-truth data [4].

From a sensor perspective self-calibration refers to the sensor being calibrated against its own measurements based on physical self-evident principles or similar. From a sensor network perspective it is similar to co-calibration where calibration is performed using other sensors in the network. [4]

Pre-calibration referring to calibration performed before deployment is also possible for sensors in sensor networks both in laboratory and in-situ. With this approach, the challenge lies in performing recalibration of sensors at a later stage and in potential differences between controlled and uncontrolled environments. If sensors are calibrated using pre-calibration, they can be used to calibrate other sensors through distributed calibration after deployment. [4]

Assuming pre-calibration is performed, another cost-reducing form of calibration is global calibration, where a calibration model is derived using the data from a subset of sensors [7]. The resulting model can then be transferred to similar sensors operating under comparable conditions, thereby reducing the complexity of future recalibrations.

Lastly, a sensor might have been pre-calibrated from the factory, or there might be a specification of its uncertainty in its data sheet. If no other method is possible, this data should be provided as metadata for the sensor in the sensor network system.

III. METADATA

Metadata, i.e. structured data that provides information about another piece of data, can vastly improve the trustworthiness, reliability, and (re)usability of sensor network data. In the case of sensor networks, especially sensor networks without access to reference sensor data, metadata can provide useful information about data structure, measurement units, measurement variables, placement of sensors, measurement method, uncertainty, etc. Metadata should be stored and accessible in the system. Having no metadata can complicate the analysis and meaningful use of the generated measurement data, as it will be ambiguous and difficult to use. Improving the metadata can not only help understanding and use of data, it can also help improve the quality, addressing the below-mentioned data quality characteristics. Examples of how are given in the respective sections for the data quality characteristics.

IV. DATA QUALITY CHARACTERISTICS

If none of the above calibration methods are possible or feasible for a given sensor network it is even more challenging to describe the uncertainty and traceability of measurements and to measure the data quality in general.

There exist several well-known data quality characteristics and these are being used within different domains to measure and describe data quality. The number and definitions of data quality characteristics vary from source to source. Not all are relevant from a metrological perspective and depending on the use case data quality requirements for the different characteristics can vary.

Table 1 lists data-quality characteristics drawn from two standards: ISO/IEC 5259-2 [8], which is based on ISO/IEC 25012 [9], and ISO 8000-210 [10], which is part of the ISO 8000 series on data management. Table 1 also lists the data quality characteristics selected for this paper due to their relevance from a metrological perspective.

ISO/IEC 5259-2 includes the data quality characteristics from ISO/IEC 25012 and uses the same categorization: inherent data quality and system-dependent data quality. Compared to ISO/IEC 25012, ISO/IEC 5259-2 extends with an additional category called "additional data quality characteristics". Only the inherent data quality characteristics of ISO/IEC 5259-2 are listed in Table 1. This is because the inherent characteristics are relevant for the data itself and not just for the system around it.

ISO 8000-210 focuses on data quality characteristics relevant to sensor data and describes only five characteristics.

A subset of characteristics (accuracy, completeness, consistency, precision, timeliness, traceability) was selected as most relevant from a metrological perspective and proposed as a basis for describing data quality in a metrologically sound manner. The selected characteristics are shown in Table 1. All have relatively consistent definitions,

Table 1. Data quality characteristics from two standards and selected characteristics for this paper.

ISO/IEC 5259-2	ISO 8000-210	Selected
Accuracy	Accuracy	Accuracy
Completeness	Completeness	Completeness
Consistency	Consistency	Consistency
Credibility	Precision	Precision
Currentness	Timeliness	Timeliness
Accessibility		Traceability
Compliance		
Confidentiality		
Efficiency		
Precision		
Traceability		
Understandability		

and four of them (accuracy, completeness, consistency, timeliness) are among those mentioned most frequently in literature with traceability and precision not mentioned as often [11], but they are essential to metrology.

It is necessary to make a note on the difference between timeliness and currentness. Timeliness was selected as a characteristic instead of currentness to keep more aligned with ISO 8000-210 because of its focus on sensor data and because of the umbrella aspect of timeliness, which also includes currentness. Timeliness is described in ISO/IEC 5259-2, but not included under it in Table 1, as in that standard timeliness is defined as the latency between when data is recorded and when data is available for use. This definition is in line with DAMA DMBOK [12], which gives an overview of different characteristics from different sources. It describes timeliness as "expectation of availability" and currentness as "current with the real world", but it also describes timeliness as an umbrella term of different aspects including currency and latency. It is the currentness part which is most interesting from a metrological point of view, i.e. not the latency but with emphasis on having the correct timestamp of the measurement.

ISO 8000-210 [10] includes another aspect which is interesting from a metrology point of view, where the data quality characteristics are categorized using the three categories for data quality described in ISO 8000-8 [13]: Syntactic, Semantic, and Pragmatic. Even if metrology focuses mainly on the semantic quality part, the syntactic category, encompassing the representational precision characteristic (i.e. representation of the measurement values with the correct amount of digits matching the measurement precision) is also important. The pragmatic category, which comprises timeliness, is likewise crucial for data quality because accurate timestamps strongly affect subsequent analyses. This is shown in Table 2.

Table 2. ISO 8000-210 data quality characteristics categorized according to ISO 8000-8

Category	Characteristic
Syntactic	Representational precision
Semantic	Accuracy
	Completeness
	Measurement precision
	Consistency
Pragmatic	Timeliness

A. *Data Quality Measurement*

When it comes to describing the quality of a dataset the data quality characteristics are usually applied using different data quality rules with associated metrics and measures for each characteristic.

In this paper a measure is "measuring" how many data points adhere to a given criterion, a metric is a measurement function calculating the result for a rule, and the rule itself is a threshold defining when the result is acceptable or not. The metrics are usually of the following form:

$$Metric = \frac{A}{B} \cdot 100\% \tag{1}$$

Where *A* is a measure, usually a count of how many data points adhere to a given criterion, and *B* is the total amount of data points considered for the given rule. The fraction can be multiplied by 100 % to express the result as a percentage.

In this way a rule could, for example, state that the result for completeness should be 95 % or above, i.e. the rule defines a target. This would be measured by having *A* be the counted number of data points and *B* be the expected number of data points. If the result is below the target, the data quality for that characteristic is not acceptable.

B. *Accuracy*

Accuracy is one of the most important data quality characteristics from a metrological perspective. Accuracy expresses how closely the data reflect the real world. Traditionally, accuracy is assessed with a reference sensor, which has been calibrated using reference standards that are traceable to the International System of Units (SI), that quantifies the calibrated sensor's uncertainty. This method also ensures the traceability of the sensor. Without a reference sensor, other methods must be used.

One method is to make multiple measurements over a short period of time. The difference between the measurements, which is a measure of repeatability, can be used as a lower limit for the accuracy of the sensor.

Monitoring the rate of change is another simple technique. Looking at the change in measurement value be-

tween two consecutive measurements can help indicate issues or accuracy of a sensor, e.g. if it is too large.

Another alternative method is to use sensor fusion, which can be used in two different ways. One is to have multiple sensors measuring the same variable. In that case, variations in the measurements between the different sensors can be used to indicate accuracy. Another way is to use multiple sensor signals to infer another measurement result and comparing the inferred measurement with an actual measurement of a given variable. This is a common practice in air quality applications. For example, if we have a single device including a temperature, humidity, CO_2 and pressure sensor, we could infer the data for the humidity sensor based on the temperature and pressure sensor outputs. Differences between the inferred value and the actual measurement may indicate accuracy and sensor issues.

The correlation between sensors can also be used to estimate accuracy. E.g. if sensors in a network are measuring the same variable, but at different locations, the correlation between sensors can be utilized to indicate when sensors deviate from their normal operation or when they drift.

All methods mentioned above can provide further insight into the measurement accuracy of sensors. Although they do not address measurement uncertainty in a metrologically traceable manner, they can still be used to assess the trustworthiness of sensor network data from a metrological perspective.

Metadata regarding accuracy can be provided by including sensor specifications such as the uncertainty or measurement range from the data sheet of a sensor. Furthermore, a description of measurement principle and data processing methods can provide insight into potential quality issues.

C. Traceability

Metrological traceability [14] of measurements is essential from a metrological standpoint and impossible to provide without reference sensors. The calibration methods mentioned above can provide some form of traceability, in having a well-defined approach for estimating a substitute for the full measurement uncertainty of the sensors. But they are only "traced" back to other sensors in the network, or models of the network, and does not provide the same traceability as the traditional approach.

In cases without reference sensors, metadata can mitigate the lack of traceability by providing sensor-specific information about sensor type, uncertainty, measurement principle, measurement range, etc.

Another aspect of traceability is data provenance, which is concerned with providing metadata about the origin and modifications of data as it moves through different processes and transformations. This can be important when correcting data issues, such as outliers and imputation of data where data are otherwise missing, because changes to the data are "traced" as they occur.

D. Precision

Precision is split in two as shown in Table 2 above; measurement precision, which is similar to the definition in VIM [15], and representational precision, which refers to the number of decimals or significant digits.

It is important to be aware of both. Measurement precision is directly related to the precision of the sensor and, therefore, the reliability of the data. Representational precision is important because different representations might have different functionalities or applications associated with them. Rounding errors or loss of information can occur if values are rounded in different parts of a system. Furthermore, floating point numbers in computers have limited precision which in some cases can lead to errors.

Metadata defining the number of decimals for sensor measurements and for stored measurement data will directly address the precision characteristic of data quality.

E. Timeliness

Timeliness is a measure of the availability of up-to-date data for a given application and depends on the correct timestamping of data to make sure it is known when a measurement was made. Incorrect timestamps render the data non-current; this, in turn, reduces usability because the true measurement time is unknown, especially where analyses are based on measurement times, such as time of day.

It is also useful to store metadata on the frequency of sensor measurements and the frequency of sending data to the cloud. This will make it easier to determine if data is being collected at the correct frequency.

F. Completeness

Completeness has a high impact on quality since missing data severely reduces the usefulness of data. As an example, sensor networks produce time series data where development of forecasting models will be greatly impacted by large gaps in the data. Missing data imputation is a common practice where gaps are filled using different methods. These can range from simple methods such as last observation carried forward (LOCF) , mean imputation, or linear interpolation, to more sophisticated methods, for example using machine learning [1, 16]. Poor completeness severely affects reliability and applicability as data can either be deemed unacceptable or artificial data is needed to fill in the gaps. In the latter case, data provenance is important to keep track of the inserted data.

Metadata can help improve completeness by providing information about placement of sensors, applied communi-

cation technologies, etc. which can be used to investigate the source of potential issues.

G. Consistency

DAMA DMBoK [12] defines consistency as the absence of difference, when comparing two or more representations of a thing against a definition. This definition fits well with a definition from a metrological perspective, where consistency can relate to sensor measurements being consistent over time, i.e. under the same conditions the sensor will output the same measurement result. Another aspect is sensors being consistent with each other, i.e. sensors are consistent with each other if they, under the same conditions, produce the same outputs.

When considering sensor data, one important challenge is the mitigation of sensor drift, where a sensor over time will produce different measurement results under the same conditions.

Another challenge related to consistency is detection of a stuck sensor, which repeatedly outputs an identical value. Depending on the number of significant digits of the sensor output this might be the result of a sensor issue.

Metadata can address some issues regarding consistency. It is preferable for later data analysis that format, unit, measurement method, etc. are consistent throughout a sensor network. This improves understandability and (re)usability of data. Consistent use of units also limits the number of conversions, each of which can impact the accuracy and precision of measurements.

V. CONCLUSION

For cases where a traditional metrological approach to sensor calibration cannot be used, an alternative approach to data quality within metrology has been presented. A selection of data quality characteristics was selected from literature, representing the primary focus areas of metrology. It is proposed to actively use these to address data quality challenges in a metrologically sound manner, because while different alternative calibration methods exist, if no reference sensor is available, there is no traceability, and data quality needs to be addressed through other means.

Metadata plays a big role in addressing data quality issues in sensor networks from a metrological point of view, as this can help inform how the system works and where to improve it.

Guidelines could be made to describe what information needs to be available in a system to make quality assessment manageable from a metrological perspective. Guidelines could define minimum requirements for metadata, alternative calibration methods where applicable, or other processes to make sure the system and its application is metrologically acceptable and fit-for-purpose in view of the goal pursued by the sensor network.

VI. ACKNOWLEDGMENTS

The project (22DIT02 FunSNM) has received funding from the European Partnership on Metrology, co-financed from the European Union's Horizon Europe Research and Innovation Programme and by the Participating States.

REFERENCES

[1] H. Teh et al. "Sensor data quality: a systematic review". In: *Journal of Big Data* 7 (Feb. 2020). DOI: 10.1186/s40537-020-0285-1.

[2] S. Tabandeh et al. "Sensor network metrology: Current state and future directions". In: *Measurement: Sensors* (2025), p. 101798. ISSN: 2665-9174. DOI: 10.1016/j.measen.2024.101798.

[3] J. Pearce et al. "End-user case studies in sensor network metrology". In: *Measurement: Sensors* (2025), p. 101819. ISSN: 2665-9174. DOI: 10.1016/j.measen.2025.101819.

[4] J. M. Barcelo-Ordinas et al. "Self-calibration methods for uncontrolled environments in sensor networks: A reference survey". In: *Ad Hoc Networks* 88 (May 2019), 142–159. ISSN: 1570-8705. DOI: 10.1016/j.adhoc.2019.01.008.

[5] F. Delaine et al. "In Situ Calibration Algorithms for Environmental Sensor Networks: A Review". In: *IEEE Sensors Journal* 19.15 (2019), pp. 5968–5978. DOI: 10.1109/JSEN.2019.2910317.

[6] M. Gruber et al. "A3.2 - Co-Calibration in Distributed Homogeneous Sensor Networks". In: *SMSI 2023*. Jan. 2023, pp. 47–48. DOI: 10.5162/SMSI2023/A3.2.

[7] S. De Vito et al. "A global multiunit calibration as a method for large-scale IoT particulate matter monitoring systems deployments". In: *IEEE Transactions on Instrumentation and Measurement* 73 (2023), pp. 1–16. DOI: 10.1109/TIM.2023.3331428.

[8] ISO/IEC. *Artificial intelligence – Data quality for analytics and machine learning (ML) – Part 2: Data quality measures*. Standard 5259-2:2024. Geneva, Switzerland, 2024.

[9] ISO/IEC. *Software engineering – Software product Quality Requirements and Evaluation (SQuaRE) – Data quality model*. Standard 25012:2008. Geneva, Switzerland, 2008.

[10] ISO. *Data quality – Part 210: Sensor data: Data quality characteristics*. Standard 8000-210:2024. Geneva, Switzerland, 2024.

[11] L. Zhang et al. "Data Quality Management in the Internet of Things". In: *Sensors* 21.17 (2021). ISSN: 1424-8220. DOI: 10.3390/s21175834.

[12] D. International. *DAMA-DMBOK: Data Management Body of Knowledge (2nd Edition)*. Denville, NJ, USA: Technics Publications, LLC, 2017. ISBN: 1634622340.

[13] ISO. *Data quality – Part 8: Information and data quality: Concepts and measuring*. Standard 8000-8:2015. Geneva, Switzerland, 2015.

[14] P. J. Brewer et al. "SI traceability and scales for underpinning atmospheric monitoring of greenhouse gases". In: *Metrologia* 55.5 (2018), S174. DOI: 10.1088/1681-7575/aad830.

[15] BIPM et al. *International Vocabulary of Metrology — Basic and general concepts and associated terms (VIM)*. Joint Committee for Guides in Metrology, JCGM 200:2012. (3rd edition). DOI: 10.59161/JCGM200-2012.

[16] T. Emmanuel et al. "A survey on missing data in machine learning". In: *Journal of Big Data* 8 (Oct. 2021). DOI: 10.1186/s40537-021-00516-9.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - M4DConf 2025
Benevento, Italy, September 3-5, 2025

Research on Digital Methods for Testing and Calibration Schemes

QI Yulin, CUI Weiqun, LIU Xinmeng*

National Institute of Metrology, Beijing, China, liuxm@nim.ac.cn

Abstract – To adapt to the digitalization of product design, a digital model and establishment method for testing and calibration schemes are proposed, which are used for the digital modeling and simulation verification of testing and calibration schemes. This method uses the System Modeling Language (SysML) to construct a digital model of the testing and calibration scheme. Taking the calibration scheme of a typical optical measuring instrument as an example, a digital model is built based on MBSE software, and simulation verification and processing analysis are carried out in combination with measured data. The empirical results show that the proposed digital method provides a unified model architecture for the testing and calibration schemes in metrological design, which can facilitate the rapid establishment of digital models and virtual simulation verification. The proposed method is adapted to the digitalization of product design based on MBSE, filling the gap in the digital method for testing and calibration schemes in metrological digitalization.

Keywords: Metrology, Product design, Digital model, MBSE, Virtual simulation, SysML

I. INTRODUCTION

Metrology digitization represents the global development trend in metrology. Driven by research institutes such as the International Bureau of Weights and Measures (BIPM) and Physikalisch-Technischen Bundesanstalt (PTB), the digitization of metrological results has reached a mature stage, with the establishment of specifications like Digital International System of Units (D-SI) and Digital Calibration Certificates (DCC) [1]. The National Institute of Metrology, China (NIM) has developed formatting protocols and generation algorithms for DCC, which are now under trial in domains such as time-frequency metrology [2]. Notably, the utilization of digital testing and calibration result data hinges on the availability of corresponding digital scheme information. Current frameworks like D-SI exclude calibration scheme descriptions, while DCCs can only document calibration methods in textual form. These limitations render existing metrological digitization tools inadequate for systematic representation of testing and calibration schemes, and they lack mechanisms for simulative verification.

One of the keys to integrating testing and calibration into product digital design, digital manufacturing, and digital maintenance lies in establishing a scientific and reasonable digital model for testing and calibration schemes. It should involve digitizing static information such as metrological characteristics, calibration conditions, and calibration items, as well as dynamic information including calibration methods, calibration result processing methods, and uncertainty evaluation methods.

Product design paradigm is shifting from traditional Document-Based Systems Engineering (DBSE) to Model-Based Systems Engineering (MBSE), aimed at enabling clear, precise description and management of product architecture, interfaces, and interaction dynamics while meeting demands for information integration and visual analysis [3-6]. In MBSE, digital models serve as irreplaceable cornerstones [7-8]. This integration spans requirements, functional specifications, structural designs, behavioral patterns, and other dimensions.

This paper establishes a digital model for testing and calibration schemes using SysML based on the MBSE design methodology. Through the simulation and analysis functions of the model, designers can verify the schemes via simulation during the design phase, predict potential issues in testing and calibration processes in advance, and optimize measurement methods and workflows.

II. DIGITAL MODELING METHODS FOR TESTING AND CALIBRATION

A. Product Testing and Calibration Scheme Design Process

The testing and calibration process of products is an interactive process involving multiple elements such as personnel, metrological equipment, measured objects, metrological methods, metrological environments, and data [9]. The design process of product testing and calibration schemes generally includes two parts: product metrological

requirement analysis and testing & calibration scheme design [10]. Effective metrological requirement analysis is the prerequisite for carrying out testing and calibration scheme design. Aimed at product metrological requirements, the testing and calibration scheme design gradually clarifies key elements such as testing requirements, methods, interfaces, and testing cycles, completes the design of testing schemes for each parameter, then integrates the testing schemes of each parameter to form a product testing scheme and confirm each testing device. For each parameter of the testing device, according to the parameter calibration requirements, the calibration method, traceability chain, calibration interface, and calibration cycle are determined to form the calibration scheme for the corresponding parameter. The design process of product testing and calibration schemes is shown in Fig. 1.

B. Digital Methods and Models for testing and Calibration Schemes

As shown in Fig. 1, the design processes and elements of testing schemes and calibration schemes are similar. To avoid repetition, this paper only presents the digital method for calibration schemes. For testing equipment, there are generally multiple parameters to be calibrated, and corresponding parameter calibration schemes need to be designed for each parameter, including parameter calibration requirements, parameter calibration methods, parameter traceability chains, parameter calibration interfaces, and parameter calibration cycles. By integrating the calibration schemes for different parameters of the testing equipment, the equipment's calibration scheme is formed, including calibration requirements, calibration methods, traceability chains, calibration equipment, calibration interfaces, and calibration cycles.

During the transformation and development of product design processes toward MBSE, SysML , an extension of the Unified Modeling Language (UML), which has emerged as a pivotal MBSE modeling language, are widely adopted for digital modeling and analysis in systems engineering [11-12]. SysML provides system developers with standardized graphical and textual representation methods to clearly depict a system's key elements, such as structure, behavior, requirements and their interrelationships [13].

Testing and calibration methods are essentially measurement processes, generally comprising four parts: measurement principles, environmental conditions & control measures, out-of-tolerance judgment methods, and data processing & uncertainty evaluation methods. Thus, the digital model for calibration schemes in this paper mainly includes these four modules. Kinds of modules are modeled using elements and diagrams in SysML to finally form a complete calibration scheme model as follows:

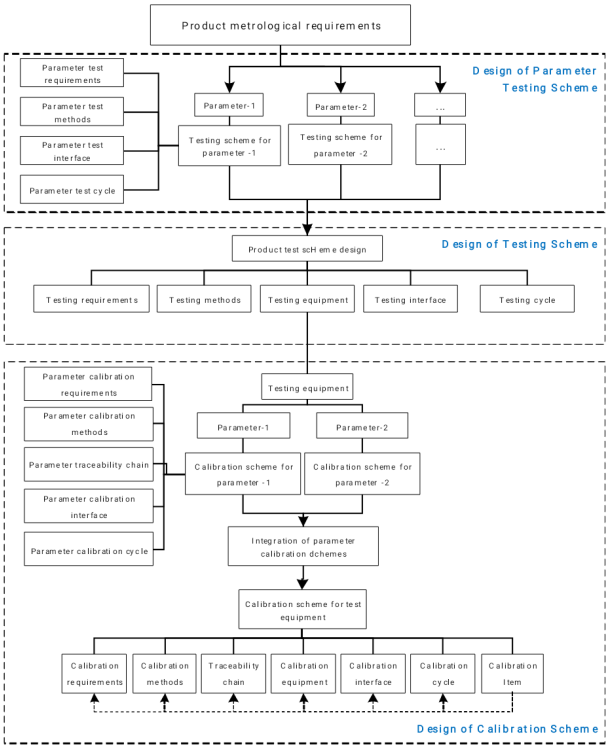


Fig. 1. Testing and Calibration Scheme Design Process.

(1) Calibration Scheme Architecture: Package element is utilized to describe constituent documents and Package Diagram (PKG) is employed to illustrate the scheme's composition, organize and manage model elements. By grouping related elements into distinct packages, the structure of complex system models becomes clearer, more intuitive, and easier to maintain.

(2) Model Composition:Block element is employed to depict model components and Block Definition Diagram (BDD) is used to illustrate model composition, present system modules and their interrelationships in a structured manner.

(3) Internal Structure and Signal Transmission of the Model: Block element is used to describe model components and Port is used to define interfaces, then Block Definition Diagram (BDD) with an Internal Block Diagram (IBD) is utilized to refine the system structure. This diagram primarily illustrates interconnections and interaction relationships

among internal components of a block.

(4) Calibration Activities and Process Description: Activity Diagrams (ACT), State Machine Diagrams (STM), and Sequence Diagrams (SD) are utilized to appropriately model calibration workflows corresponding to calibration methods. Simulation can be employed to verify process correctness.

(5) Calibration Item Simulation and Uncertainty Evaluation: through binding parameter values to calibration method models in a Parametric Diagram (PAR), the correctness and rationality of calibration method models are validated.

III. CASE STUDY ON THE IMPLEMENTATION OF DIGITALIZATION APPROACHES

A. Typical Calibration Scheme for Optical Measuring Instruments

The hemispherical emissometer, which is used to measure the hemispherical emissivity of materials and evaluate their absorption and radiation characteristics, serves as a typical optical measuring instrument. It plays an indispensable role in fields such as material science, aerospace, and building energy conservation [16]. In the product design and manufacturing of related fields, the hemispherical emissometer is an important testing equipment for the thermal physical properties of materials [17]. Taking this as an example, this paper introduces the digital method of the calibration scheme. The MBSE design tool used complies with the SysML specification, and the model-related graphics in the following text are exported from the finally completed digital model of the calibration scheme.

B. Digital Model of Calibration Scheme

Overall architecture of the calibration scheme for the hemispherical emissometer is designed using Package elements and a Package Diagram (PKG), as shown in Fig. 2. The overall structure follows the design process illustrated in Fig. 1. First, the calibration object is correctly described as the starting point for designing the calibration scheme. Then, parameter calibration schemes are designed for the main calibration parameters of the hemispherical emissometer. Finally, the calibration schemes for each parameter are summarized, deleted and aggregated to form the equipment calibration scheme, which mainly includes equipment calibration requirements, calibration methods, traceability chains, calibration interfaces, calibration cycles, etc. Meanwhile, calibration items and calibration equipment for the hemispherical emissometer are established.

(1) Digital Model of Hemispherical Emissometer for Materials

The model of the calibrated object primarily includes the composition of the equipment under calibration, such as internal interfaces and signal transmission, functional descriptions, and usage processes.

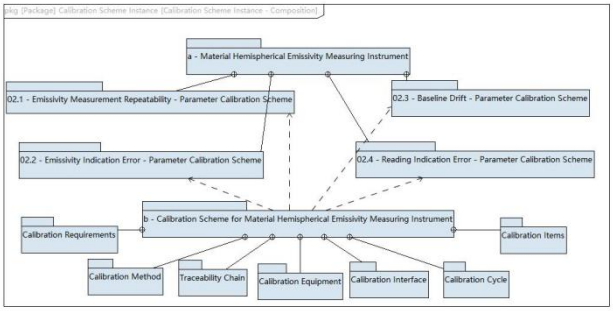


Fig. 2. Composition of Calibration Schema.

As shown in Fig. 3, the Block element is used to establish the model of the calibrated object, i.e., the hemispherical emissometer, and the Block Definition Diagram (BDD) is employed to describe its model composition. To facilitate the subsequent design of the Internal Block Diagram (IBD) and simulation of the calibration process, the model is constructed as a system comprising five components based on the actual functions and composition of the measuring instrument: light source, spectroscopy system, detection system, integrating sphere sample chamber, and control system.

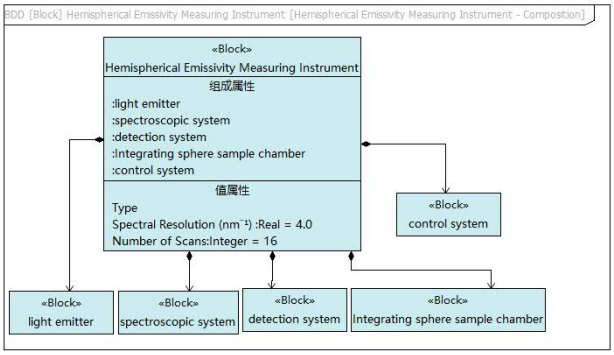


Fig. 3. Composition of Hemispherical Emissometer Model.

Furthermore, the Internal Block Diagram (IBD) is utilized to describe the internal composition of the hemispherical emissometer, including its interfaces and signal transmission, as shown in Fig. 4.

The Activity Diagram (ACT) can be used to depict the relevant measurement process as shown in Fig. 5.

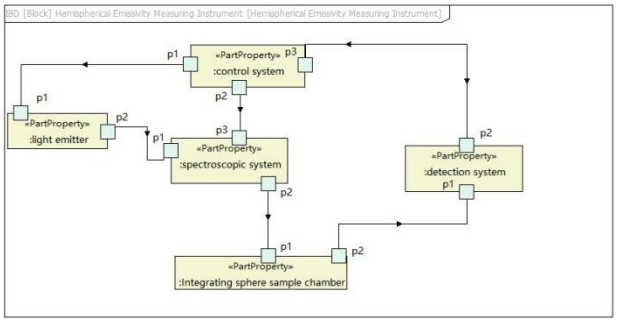


Fig. 4. Internal Signal Flow Diagram of Hemispherical Emissometer.

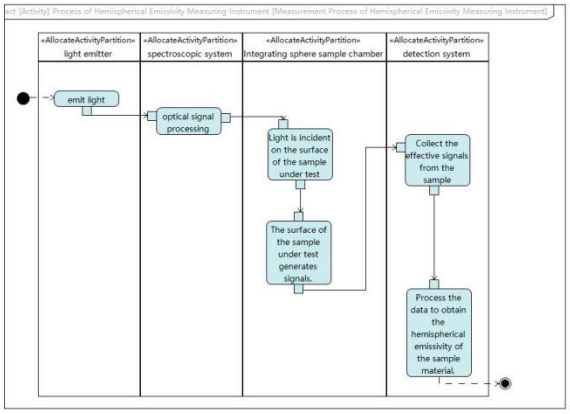


Fig. 5. Measurement Process of Hemispherical Emissometer.

(2) Digital Model of Baseline Drift Parameter Calibration Scheme

The main calibration parameters of the hemispherical emissometer include baseline drift, emissivity measurement repeatability, wavenumber indication error, and emissivity indication error. The composition of baseline drift parameter calibration scheme can be represented by a Block Definition Diagram (BDD), as shown in Fig. 6.

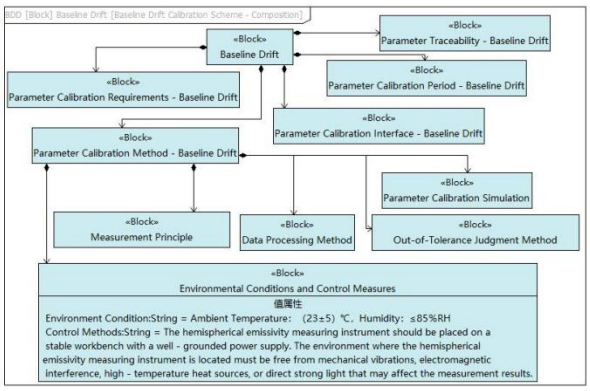


Fig. 6. Baseline Drift Calibration Schema.

The parameter calibration method module further specifies the calibration methods for specific parameters. For example, the baseline drift calibration method is to select an emissivity standard piece with a nominal value of 0.9 to calibrate the emissivity measurement capability of the hemispherical emissometer, take 1 minute as the measurement time interval, and continuously measure 10 emissivity data. The data processing method is to calculate the relative drift according to Eq. (1).

$$\Delta_{100} = \frac{\varepsilon_{max} - \varepsilon_{min}}{\varepsilon_{max} + \varepsilon_{min}} \times 100\% \quad (1)$$

Where subscript max is the maximum value among the 10 emissivity data points, and subscript min is the minimum value among the 10 emissivity data points.

The block definition diagram (BDD) is used to establish the baseline drift data processing method model, as shown in Fig. 7.

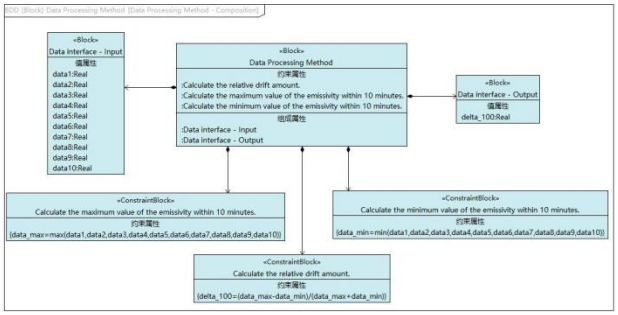


Fig. 7. Composition of Baseline Drift Data Processing Model.

Furthermore, in the Parametric Diagram (PAR), specific parameter values are bound to the data processing model, as shown in Fig. 8, to complete the construction of the baseline drift data processing model calculation simulation program.

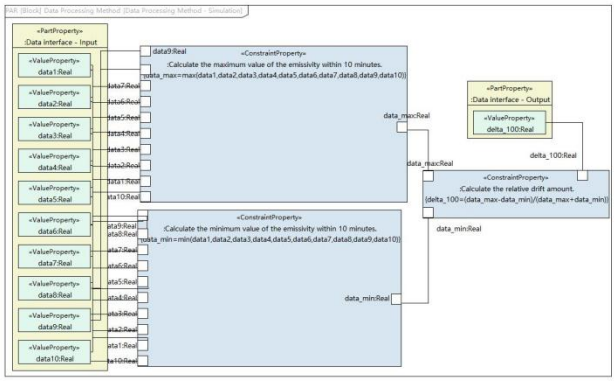


Fig. 8. Baseline Drift Data Processing and Calculation.

For the calculated baseline drift, the out-of-tolerance judgment method in the calibration scheme can set limits for the calibration results. For

example, the baseline (100% point) drift of the hemispherical emissometer is required to not exceed 0.02, based on which out-of-tolerance judgment can be made. Similarly, the Block Definition Diagram (BDD) can be used to establish the baseline drift out-of-tolerance judgment method model, as shown in Fig. 9. In addition to the data input, data output, and out-of-tolerance judgment modules, the data processing module calls the baseline drift data processing module. In the Parametric Diagram (PAR), the parameter values are bound to the out-of-tolerance judgment data processing model, as shown in Fig. 10, to complete the construction of the baseline drift out-of-tolerance judgment model calculation simulation program.

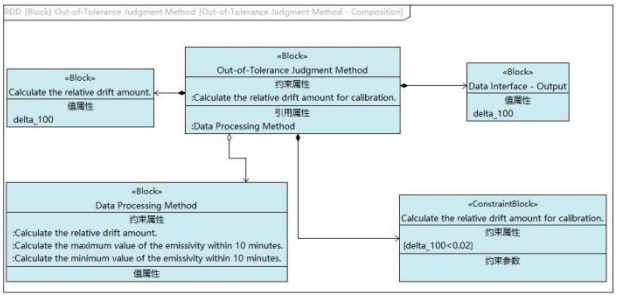


Fig. 9. Composition of the Baseline Drift Out-of-Tolerance Judgment and Processing Model.

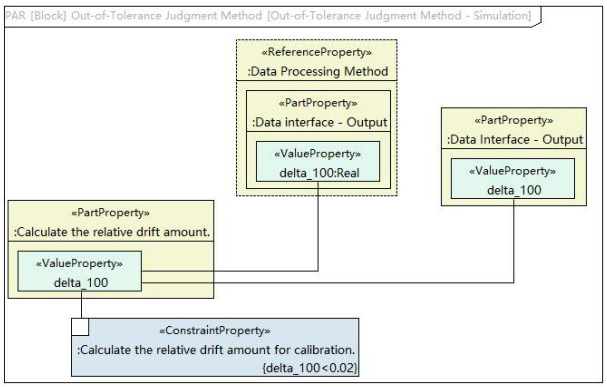


Fig. 10. Calculation of Baseline Drift Out-of-Tolerance Judgment and Processing.

After establishing the data processing model, the construction of a specific simulation model can be further completed. Assuming 10 emissivity data points are collected, they are input into the input module of the calibration method simulation model, and the baseline drift out-of-tolerance judgment method is selected as the data processing module to construct a baseline drift out-of-tolerance judgment simulation model.

After establishing the simulation model, Parametric Diagram (PAR) is used to bind the emissivity data to the input interface of the out-of-tolerance judgment method model. Then, model simulation can be performed. The emissivity data can be processed using the model established earlier. The out-of-tolerance judgment result is shown in Fig. 11. The simulation shows that the baseline drift of this hemispherical emissometer is 0.0326, exceeding the tolerance of 0.02. The judgment result is displayed in red, indicating that it does not meet the drift requirement.

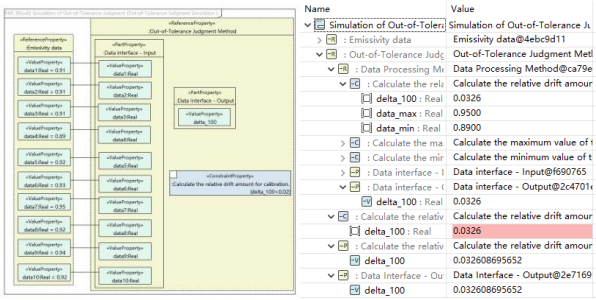


Fig. 11. Baseline Drift Out-of-Tolerance Judgment Simulation Calculation and Results.

(3)Digital Model of Emissivity Indication Error Calibration Scheme

The emissivity indication error also includes the components shown in Fig. 7. The calibration method for the emissivity indication error parameter involves calibrating the hemispherical emissometer using a hemispherical emissivity measurement standard. The measurement model is:

$$\Delta\varepsilon = \varepsilon - \varepsilon_m \qquad (2)$$

Where ε_m is the calibration value at a specific wavenumber, ε is the indication value of the measuring equipment, and $\Delta\varepsilon$ is the indication error of the measuring equipment. The measurement uncertainty of emissivity mainly originates from measurement repeatability and the emissivity measurement standard.

According to the GUM standard, an uncertainty evaluation model can be established to carry out uncertainty verification. Considering the versatility of the model, Monte Carlo method is used. As am example, it is assumed that the standard uncertainty introduced by measurement repeatability follows a normal distribution with a mean of 0.910 and a standard deviation of 0.007. The standard uncertainty of the emissivity measurement standard follows a uniform distribution with a mean of 0 and a half-width of 0.015. With a confidence probability of 0.95, 10^6 Monte Carlo simulations are performed. The

expanded uncertainty of the measurement result for the indication error is:

$$U_{95} = [-0.021, 0.021] \quad (3)$$

(4)Digitalization of Calibration Scheme for Hemispherical Emissometer

After designing calibration schemes for each parameter, it is necessary to integrate the calibration requirements, methods, interfaces, traceability chains, and calibration cycles of all parameters to form an equipment calibration scheme. Meanwhile, suitable standards should be selected as the calibration equipment for the testing equipment by synthesizing the calibration schemes of all parameters. For example, in the calibration scheme of the hemispherical emissometer, calibration equipment such as emissivity standards with nominal values of 0.1, 0.3, 0.5, 0.7, and 0.9, as well as polystyrene infrared wavelength standard materials, are selected by integrating calibration requirements such as baseline drift and emissivity indication error.

IV. RESULTS

Empirical results in this paper show that the digital method proposed in this paper can provide a unified and effective model architecture for product inspection and calibration schemes. In the follow-up research, the control and data interaction methods with the digital twin models of testing and calibration equipment will be further studied.

V. ACKNOWLEDGMENTS

This research was supported by Basic Research Funding for Central Public Research Institution of China (No. AKYZZ2420) and National Metrology Data Center of China(No. APT2501-6).

REFERENCES

[1] Hu Bo. “Review of Research and Practical Progress in metrology Digitization Development”, Metrology Science and Technology, vol.67, No.5, 2023, pp.39-44.

[2] Wang Tingting, Cui Weiqun. “An Overview of Algorithm Traceability”, Metrology Science and Technology, vol.67, No.5, 2023, pp.23-30.

[3] Hou Xiong, Li Dongfang, “MBSE Technology and Its Application in Aerospace Control”, Aerospace Control, vol.41, No.5, 2023, pp.3-13.

[4] Zhao Ziyue, Gong Siyuan, Qu Jiansu, et al. “Research on Key Common Metrology Technologies for Intelligent Manufacturing of Large-scale Equipment”, Metrology Science and Technology, vol.68, No.2, 2024, pp.60-75.

[5] Chen Chuan, Chen Yuefei, Zeng Lin, et al. “Application and Research Progress of Digital Twin in the Field of Intelligent Manufacturing”, Metrology Science and Technology, vol.12, 2020, pp.20-25.

[6] Xiao Shunan, Guo Xiaojian, “Hot Topics and Evolution Trends of Digital Twin Research at Home and Abroad — — Knowledge Graph Analysis Based on CiteSpace”, Metrology Science and Technology, vol.67, No.8, 2023, pp.16-28+74.

[7] Dou Chen, Zhou Zili, “Analysis of the Importance of Metrology in the Digital World and the Development Direction of Metrology”, Metrology Science and Technology, vol.66, No.9, 2022, pp.70-73.

[8] Zhao Bin. “Cloud Metrology: A New Model of the metrology Ecosystem under China's Version of Industry 4.0”, Metrology Technology, vol.11, 2016, pp.58-60+66.

[9] Lv Qi, Wang Junbo, Zhang Ying, et al. “Research on Measurable Design of Low-frequency Standard ShakerResearch on Measurable Design of Low-frequency Standard Shaker”, Metrology Science and Technology, vol.66, No.4, 2022, pp.101-107+88.

[10] Sun Ruoduan, Hao YingWei, Liu Xinmeng. “Measurability Research of Grating Monochromator Wavelength Calibration equipment”, Metrology Science and Technology, vol.67, No.6, 2023, pp.3-8.

[11] GB/T 28174.1-2011, Unified Modeling Language (UML) — Part 1: Infrastructure [S]

[12] Wu Juan, Wang Mingzhe, Fang Huajing. “System Architecture Product Design Based on SysML”, Systems Engineering and Electronics, No.4, 2006, pp.594-598.

[13] Zhou Shuhua, Cao Yue, Zhang Zheng, et al. “System Design and Simulation Integration of Complex Electromechanical Products Based on SysML and Modelica”, Journal of Computer-Aided Design & Computer Graphics, vol.30, No.4, 2018, pp.728-738.

[14] Wang Ruiqin, Han Ying, Gao Yuan, “Research on the Measurement Methods of Spectral Emissivity and Hemispherical Emissivity of Infrared Materials”, Journal of Astronautic Metrology and Measurement, vol.44, No.4, 2024, pp.40-43+81.

[15] Gan Haiyong, Yu Tianlai, Kuang Bo, et al. “Development Status and Trends of the Metrology and Testing Technology System for Optical Properties of Materials”, Measurement Technology, vol.44, No.6, 2014, pp.1-19.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

A Modular Windows-Based Intelligent API for Traceable Drone Positioning Using UWB-OptiTrack Fusion and AI-Based Residual Learning

Ihtisham Ul Haq[1], Luigi D'Alfonso[1], Giuseppe Fedele[1], Francesco Lamonaca[1, 2]

[1]*Department of Computer Science, Modelling, Electronics and Systems (DIMES), University of Calabria, 87036 Rende, CS, Italy*
[2]*National Research Council Institute of Nanotechnology (CNR-NANOTEC), Rende, 87036, Italy*

***Abstract* – Accurate and traceable drone positioning is crucial for autonomous aerial navigation, especially in GNSS-denied environments. Traditional approaches using Ultra-Wideband (UWB) sensors or Kalman Filters (KF) struggle with multipath interference, non-line-of-sight, and environmental uncertainties, and are often limited to Linux-based Application Programming Interfaces (APIs). This work presents an innovative framework based on: a novel modular Windows-based API for real-time drone positioning, the integration of Kalman filter for optimal multisensor data fusion and trajectory smoothing, AI-driven residual learning to correct systematic estimation errors, and metrology-compliant uncertainty modeling. The system enables real-time swarm deployment and pose-aware feedback using an auxiliary vision based positiong system (OptiTrack) and UWB data. A feedforward neural network compensates for residual errors in Kalman-filtered trajectories, while Monte Carlo simulations establish traceable 95% confidence intervals. Experimental tests show that the proposed framework reduces RMSE by over 40% across axes, with strong regression accuracy greater than 94%.**

I. INTRODUCTION

Accurate drone positioning is essential for autonomous navigation, industrial inspection, swarm coordination, and environmental monitoring. While GPS-based methods are widely used, their effectiveness declines in cluttered or indoor environments. Ultra-Wideband (UWB) technology provides high temporal resolution and robustness to multipath interference, making it suitable for such settings [1, 2, 3]. However, UWB systems may still experience performance degradation in non-line-of-sight (NLoS) conditions and are sensitive to calibration biases, leading to trajectory distortions and time-varying errors [4, 19, 6].

Probabilistic filtering approaches such as the Kalman Filter (KF) and its nonlinear extensions (EKF, UKF) are often applied to mitigate these issues through multi-sensor data fusion [7, 8]. Yet, their accuracy can be limited by Gaussian noise and linearity assumptions that are not always met in real-world scenarios [9, 8]. Recent work leverages Artificial Neural Networks (ANNs) to model and correct complex error patterns in UWB-based positioning, using metrological ground truth from systems such as OptiTrack [10, 19, 11]. Embedding metrological principles, bias quantification, and uncertainty modeling has further advanced measurement accuracy and reliability [8, 19, 3].

For drone swarms, real-time synchronization and scalable multi-agent fusion are crucial. Modular APIs support decentralized UWB-OptiTrack integration and enable robust swarm operations [11, 12]. Sensor fusion strategiesâcombining UWB, IMU, and barometric data with AI models such as multilayer perceptrons offer resilience against sensor drift and dynamic conditions [9, 13]. However, a fully integrated, standards-oriented framework for filtering, AI correction, and metrological uncertainty remains lacking.

This work introduces a unified, modular positioning framework that combines Kalman filtering, ANN-based residual correction, and rigorous metrological uncertainty modeling for traceable, real-time positioning of both individual drones and swarms. UWB and OptiTrack data are fused via ANNs and Monte Carlo-based assessment. The remainder of this paper is structured as follows: Section ii. presents the methodology, Section iii. reports experimental results, and Section iv. discusses conclusions and future work.

II. DRONE TRAJECTORY REFINEMENT AND API INTEGRATION

It is proposed a robust framework that integrates OptiTrack (OT) motion capture [15] and Ultra Wideband (UWB) modules for accurate drone trajectory estimation and benchmarking via a modular, decentralized API. The pipeline consists of intelligent data acquisition, Kalman-filtered state estimation, uncertainty quantification, AI-based residual learning, and ensemble fusion. Fig. 1 out-

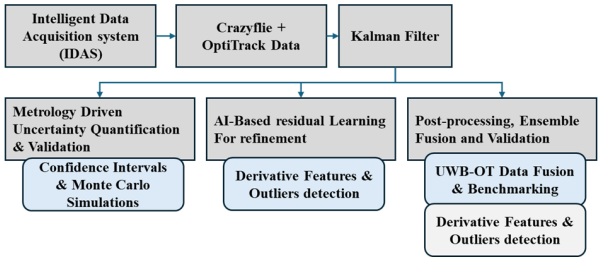


Fig. 1. Block scheme of the proposed framework for drone trajectory refinement.

lines the complete process.

The Intelligent Data Acquisition System (iDAS) module is used to synchronize and record positional data from Crazyflie drones [16] with both UWB and OptiTrack sensors. Real-time data streaming is enabled through a MATLAB-based API, interfacing with the NatNet SDK [17] and UWB radio modules [18].

A. Kalman Filter for Position Estimation

To fuse noisy measurements from multiple sources, a linear Kalman Filter (KF) is used with the standard discrete-time state-space model:

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{B}\mathbf{u}_{k-1} + \mathbf{w}_{k-1}, \tag{1}$$

$$\mathbf{z}_k = \mathbf{H}\mathbf{x}_k + \mathbf{v}_k, \tag{2}$$

where $\mathbf{x}_k$ is the 6-DOF state vector defined as $\mathbf{x}_k = [x_k,\, y_k,\, z_k,\, \dot{x}_k,\, \dot{y}_k,\, \dot{z}_k]^\top \in R^6$, representing the three-dimensional position and velocity components (positions and speeds in R^3). The control input $\mathbf{u}_{k-1} = [u_x,\, u_y,\, u_z]^\top \in R^3$ typically denotes commanded accelerations or velocity increments. The state transition matrix $\mathbf{A} \in R^{6\times6}$ and control input matrix $\mathbf{B} \in R^{6\times3}$ are structured for a constant-velocity motion model, while the measurement matrix $\mathbf{H} \in R^{3\times6}$ extracts the position components from the state vector. The process noise $\mathbf{w}_{k-1} \sim \mathcal{N}(0,\, \mathbf{Q})$ and measurement noise $\mathbf{v}_k \sim \mathcal{N}(0,\, \mathbf{R})$ are assumed to be zero-mean, mutually uncorrelated Gaussian random vectors, with covariances $\mathbf{Q}$ and $\mathbf{R}$, respectively. The matrix sizes are: $\mathbf{x}_k \in R^6$, $\mathbf{u}_{k-1} \in R^3$, $\mathbf{A} \in R^{6\times6}$, $\mathbf{B} \in R^{6\times3}$, $\mathbf{z}_k \in R^3$, and $\mathbf{H} \in R^{3\times6}$.

$$\mathbf{w}_{k-1} \sim \mathcal{N}(0, \mathbf{Q}),\ \mathbf{v}_k \sim \mathcal{N}(0, \mathbf{R}),\ E\left[\mathbf{w}_{k-1}\mathbf{v}_j^\top\right] = 0\ \forall k, j. \tag{3}$$

B. AI-Based Refinement, Anomaly Detection, and Fusion Benchmarking

Residual errors from Kalman-filtered trajectories are adaptively corrected using a feed-forward neural network that models the residual as

$$\mathbf{e}_k = \hat{\mathbf{x}}_k^{KF} - \mathbf{x}_k^{GT}, \quad \hat{\mathbf{e}}_k = \mathcal{F}_\theta(\mathbf{z}_k), \tag{4}$$

with the refined estimate given by

$$\mathbf{x}_k^{ref} = \hat{\mathbf{x}}_k^{KF} - \hat{\mathbf{e}}_k. \tag{5}$$

To further enhance robustness, time-aware profiling is conducted by computing velocity and acceleration,

$$v_k = \frac{x_k - x_{k-1}}{\Delta t}, \quad a_k = \frac{v_k - v_{k-1}}{\Delta t}, \tag{6}$$

and applying z-score-based outlier rejection ($|z| > 3$) to eliminate noisy or delayed measurements. Refined position estimates from UWB and OptiTrack (OT) are then fused via weighted averaging,

$$\mathbf{x}_f = w_{UWB} \cdot \mathbf{x}_{UWB} + w_{OT} \cdot \mathbf{x}_{OT}, \tag{7}$$

followed by an ensemble mean fusion to yield the final trajectory:

$$\mathbf{x}_{final} = \frac{1}{M}\sum_{i=1}^{M} \mathbf{x}_i^{(ref)}. \tag{8}$$

Here, $\mathbf{x}_{final}$ denotes the final fused trajectory estimate, M is the total number of estimation models or fusion sources, and $\mathbf{x}_i^{(ref)}$ represents the i-th refined trajectory estimate contributed by each model or fusion branch. Performance is validated using Root Mean Squared Error (RMSE), Standard Deviation (SD), and 95% confidence coverage, ensuring the methodology provides traceable, adaptive, and high-fidelity trajectory estimation through heterogeneous sensing and metrologically rigorous fusion.

Algorithm 1 Unified Framework for Trajectory Refinement and Swarm Execution

Require: Drone set $\mathcal{D}$, OptiTrack and UWB modules, model parameters

Initialize: Synchronize clocks; activate sensor streams; assign unique IDs for $d \in \mathcal{D}$
Register pose acquisition callbacks and prepare logging
Synchronize: Wait for all drones to reach ready state
while mission is active **do**
  for all drone $d \in \mathcal{D}$ **(in parallel) do**
    Obtain sensor measurements $\mathbf{z}_k^{(d)}$ (UWB, OT)
    State Estimation: Fuse via Kalman filter *(Eqs. 1–2)*
    Uncertainty Quantification: Monte Carlo + CI *(Eqs. 3–4)*
    Outlier Removal: Velocity/acceleration profile, z-score threshold
    Residual Correction: AI-based learning *(Eq. 5)*
    Multi-Sensor Fusion: Weighted average *(Eq. 6)*, ensemble mean *(Eq. 7)*
    Issue control setpoints; log data for validation
  end for
end while
Finalize: Stop logging; disconnect; evaluate RMSE, SD, CI coverage =0

III. RESULTS AND DISCUSSION

The proposed framework was experimentally validated for drone trajectory estimation using UWB and OptiTrack systems, with further refinement via artificial neural networks (ANNs). The following analysis presents quantitative and visual results, including error metrics, uncertainty quantification, and ANN-based correction, to assess positioning fidelity and system robustness.

A. *UWB and OptiTrack-Based Trajectory Evaluation*

Figure 2 illustrates the Crazyflieâs UWB-based trajectory across X, Y, and Z axes, revealing minor noise and Y-axis drift, as well as step changes in altitude. Despite these variances, the overall movement trend is captured well, supporting the operational viability of the UWB-based estimator. Figure 3 presents the OptiTrack-estimated trajectory in X, Y, and Z, exhibiting smooth and continuous position traces with well-defined altitude steps and stable lateral motion. This reflects the systemâs stable tracking and strong temporal consistency throughout the experiment.

B. *Error Characterization Between UWB and OT*

Figure 4 presents the positional error (OT - CF) along all three axes. The X-axis shows moderate error oscillations peaking at ± 0.2 m, while the Y-axis reveals a pronounced negative bias of approximately -0.35 m, suggesting consistent underestimation by the UWB system. The Z-axis error remains within ± 0.05 m, highlighting better alignment during vertical maneuvers.

Figure 5 overlays OT and CF trajectories, clearly contrasting the precise OT path with the slightly lagging UWB estimates. This overlay further substantiates the systematic deviation in the Y-axis and confirms the capacity for correction via residual learning. Figure 6 displays a 3D trajectory overlay, where the CF (dashed red) and OT (solid blue) paths are juxtaposed. Discrepancies are particularly evident in lateral movements, highlighting the need for

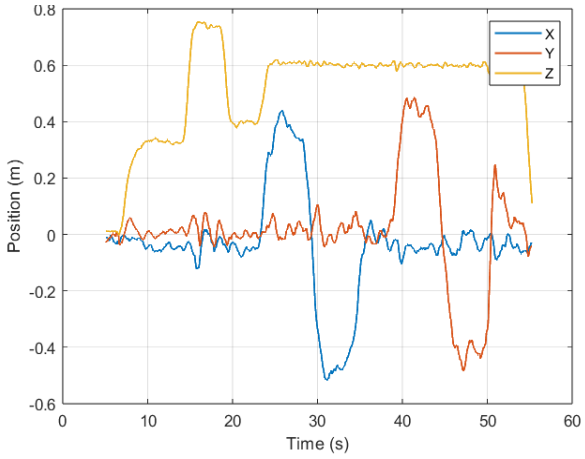


Fig. 2. UWB Based Crazyflie Estimated Trajectories

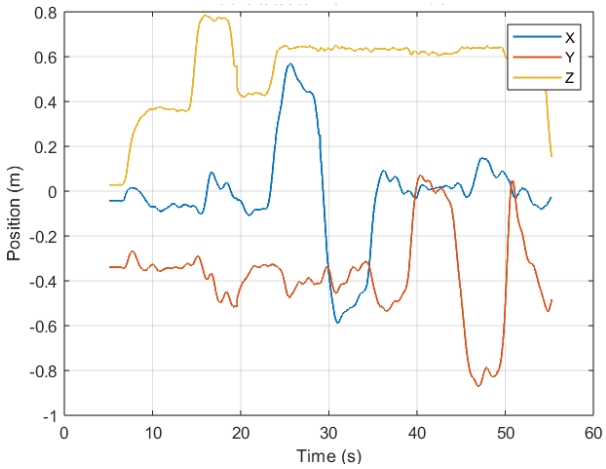


Fig. 3. OptiTrack Estimated Trajectories

trajectory correction to reduce spatial offset and maintain consistent spatial alignment.

C. *Statistical Error Distribution and Uncertainty Estimation*

Figure 7 offers histograms of the positional errors. The X-axis error is centered around 0.025 m, whereas the Y-axis exhibits a negatively skewed distribution peaking at -0.35 m, consistent with prior observations. Z-axis errors remain narrowly centered, indicating relatively accurate height estimation by the UWB module. Figure 8 demonstrates the outcome of a Monte Carlo-based uncertainty quantification, wherein 1000 random trials computed the standard deviation of error vectors. The mean uncertainties are approximately 0.045 m (X), 0.065 m (Y), and 0.012 m (Z). The Y-axis again shows the largest variability, reinforcing its susceptibility to error due to lateral drift or antenna positioning misalignment.

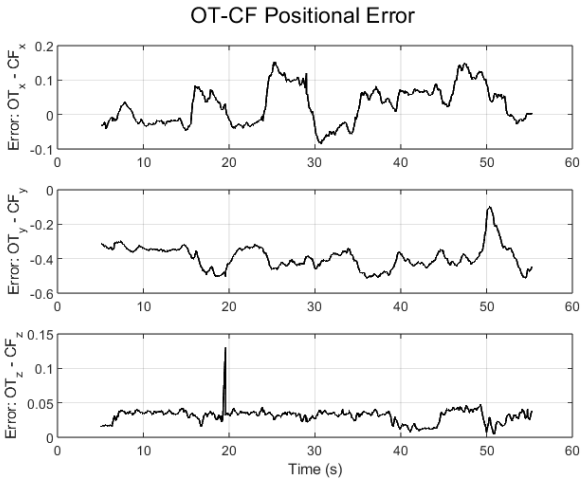


Fig. 4. OT - CF Positional Errors in X, Y, Z Dimensions

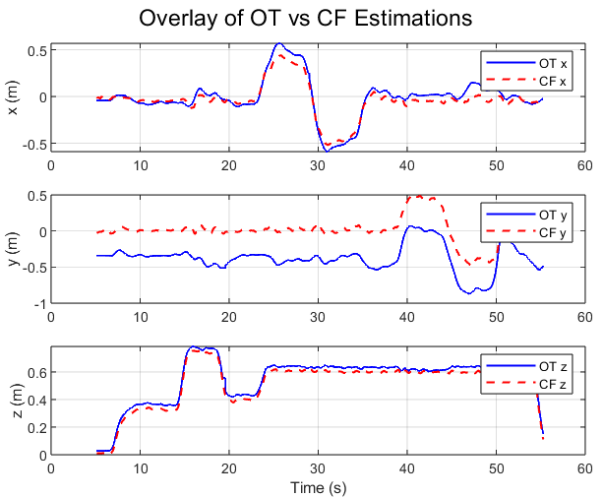


Fig. 5. Overlay of OT vs. CF Trajectory Estimations

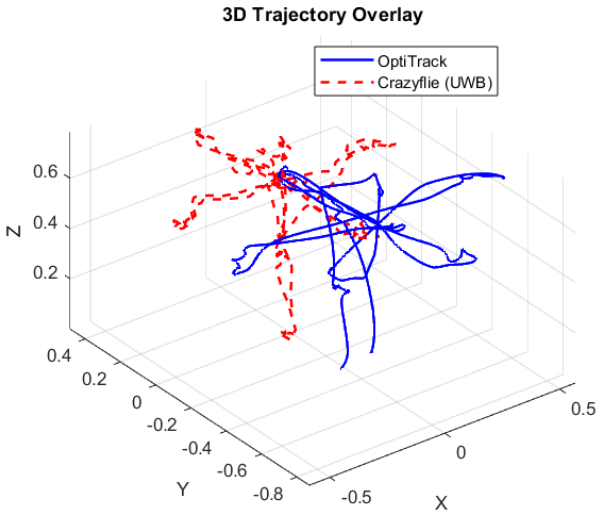


Fig. 6. 3D Trajectory Overlay of UWB and OptiTrack Estimations

D. AI-Based Residual Learning for Estimation Refinement

Figure 9 introduces the AI-enhanced refinement using a shallow ANN, where the green line shows the corrected CF estimates. The ANN-learned residuals successfully align the CF trajectory with the OT ground truth across all axes. The Y-axis correction is especially notable, reducing the systematic bias and enhancing fidelity.

E. ANN Training Performance and Regression Validation

Figure 10 shows the mean squared error (MSE) curves for training, validation, and test sets over 122 epochs, with the minimum validation MSE (0.0053) attained at epoch 116, reflecting strong convergence and negligible overfitting. The log-scale Y-axis highlights a two-order magnitude error reduction, evidencing the effectiveness of the

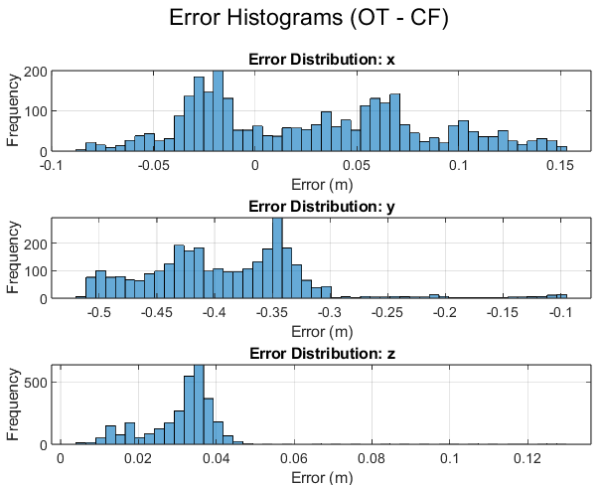


Fig. 7. Error Histogram Distribution Between OT and CF

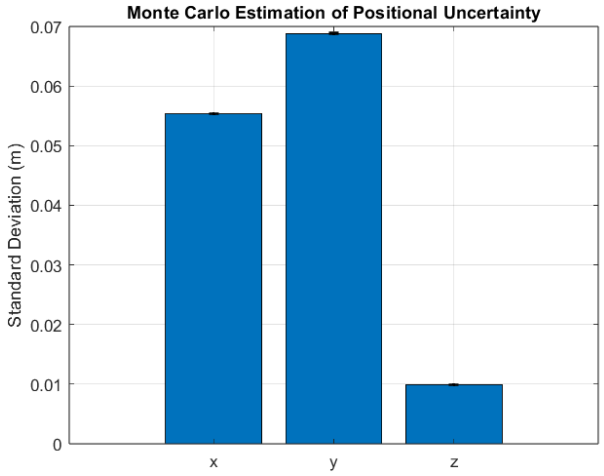


Fig. 8. Monte Carlo-Based Positional Uncertainty Quantification

shallow ANN for residual learning. Following the approach in [19], the network employs three fully connected layers with ReLU activations and a linear output to transform the three-dimensional residual input into a corrected $[x, y, z]$ position. Figure 11 illustrates the internal training dynamics, where the gradient plot confirms stable optimization, and the adaptive μ plot demonstrates efficient convergence via the Levenberg-Marquardt algorithm. The early stopping criterion, triggered after six consecutive validation check failures, effectively prevents overfitting.

Figure 12 plots the distribution of residual errors (Target - Output) for training, validation, and test sets. The concentration of errors around zero suggests the model effectively learned the mapping from UWB positions to correction residuals. No significant bias or skew is observed, indicating balanced learning across all splits. Figure 13 displays correlation plots between predicted residuals and ground truth across training ($R = 0.947$), validation ($R = 0.933$), test ($R = 0.943$), and aggregated

datasets ($R = 0.945$), where R is the Pearson correlation coefficient. The consistently high R-values and close alignment along the diagonal confirm the ANN robust generalization and predictive performance.

The proposed methodology outperforms the one without the refinement across all metrics, including RMSE, MAE and CI coverage (Table 1). The proposed AI-driven framework achieves over 56% reduction in Y-axis RMSE and tighter 95% confidence intervals, underscoring enhanced positional accuracy. These results substantiate the effectiveness of combining metrological validation with AI-based correction for robust, high-precision swarm localization, establishing the MATLAB API as a practical tool for traceable positioning in metrology-driven and multi-agent systems.

Table 1. Comparison of Positional Accuracy Metrics Before and After AI Refinement

Axis	Before Refinement (KF)			After Refinement (KF + AI)		
	RMSE (m)	MAE (m)	CI95 (m)	RMSE (m)	MAE (m)	CI95 (m)
X	0.0462	0.0364	±0.057	0.0241	0.0189	±0.038
Y	0.0678	0.0511	±0.068	0.0295	0.0237	±0.046
Z	0.0254	0.0203	±0.017	0.0106	0.0079	±0.009

F. Comparison with Prior ANN-Based Compensation Methods

Recent advances in neural network-based compensation for UWB and sensor fusion systems have improved positioning accuracy, but key limitations remain in the literature. For example, Chen and Kia [19] achieved a 35% RMSE reduction using a feedforward ANN for UWB-OptiTrack bias correction, yet did not address metrological uncertainty or multi-agent scenarios. Almassri et al. [10] and Bao et al. [9] reported accuracy improvements via ANN and MLP-based approaches; however, their studies lacked formal uncertainty quantification and deployment in traceable, real-world settings. Salimpour et al. [11] proposed autoencoder-based UWB anomaly detection, but did not consider metrological traceability or coordinated swarm control. In contrast, the present work integrates feedforward ANN residual learning with Monte Carlo-driven uncertainty analysis in a modular Windows/MATLAB API, achieving over 56% RMSE reduction on the Y-axis, reporting confidence intervals, and enabling traceable, real-time multi-agent localization. This represents a significant advancement in both technical and operational applicability over previous methods.

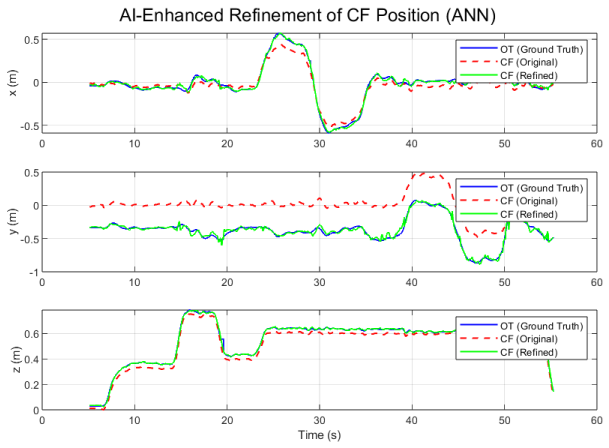


Fig. 9. ANN-Based Residual Learning for Refined Estimation

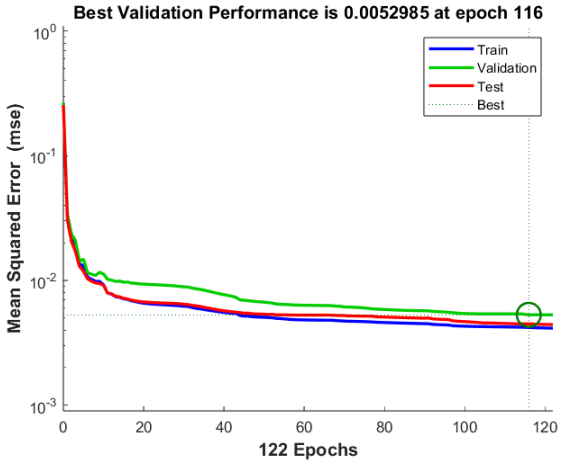


Fig. 10. ANN Training Performance Curve

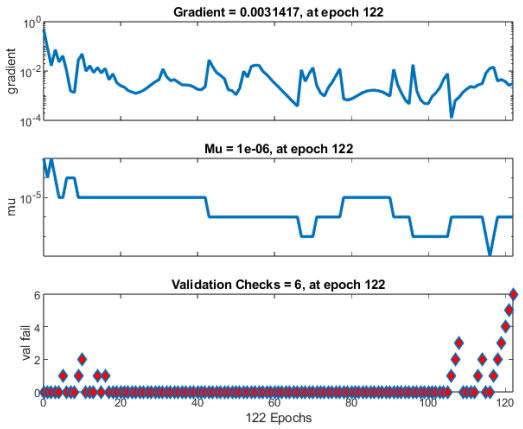


Fig. 11. ANN Training State Metrics (Gradient, Mu, Fail Checks)

IV. CONCLUSION

This work introduced a modular, Windows-based framework for traceable drone positioning, unifying Kalman filtering, AI-driven residual correction, and uncertainty quantification. By leveraging UWB and OptiTrack data

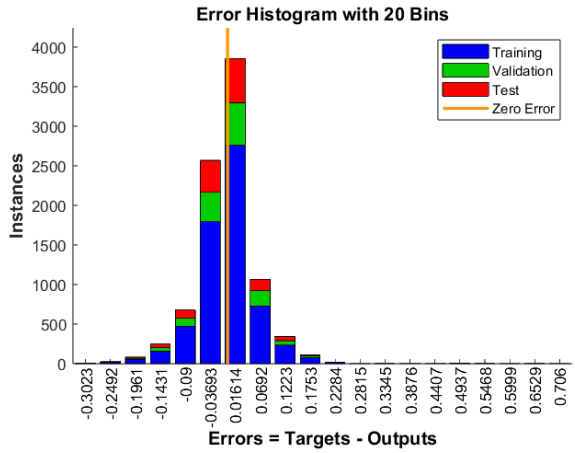


Fig. 12. ANN Residual Error Histogram

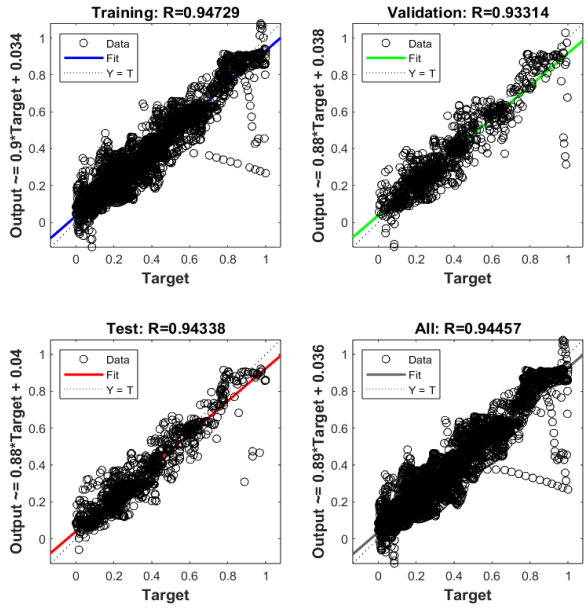


Fig. 13. ANN Regression Analysis Across All Data Splits

within a real-time MATLAB API, the approach enables robust, scalable localization for both individual and swarm drones. Experimental results confirm significant accuracy gainsâover 56% Y-axis RMSE reduction and tighter confidence intervals, demonstrating the effectiveness of integrating AI with metrological approach. The systemâs modular design, comprehensive uncertainty modeling, and realtime swarm execution advance the state-of-the-art for accurate drone positioning in complex environments. Future directions include extending sensor fusion to vision-based and inertial modalities, enhancing adaptive filtering, and further aligning with ISO metrology standards to enable broader certification and deployment in autonomous robotics.

REFERENCES

[1] L. KramariÄ, N. JeluÅ¡iÄ, T. RadiÅ¡iÄ, M. MuÅ¡tra, "A Comprehensive Survey on Short-Distance Localization of UAVs," *Drones*, vol. 9, no. 3, p. 188, Mar. 2025.

[2] W. Shule, C.M. Almansa, J.P. Queralta, Z. Zou, T. Westerlund, "UWB-based localization for multi-UAV systems and collaborative heterogeneous multi-robot systems," *Procedia Comput. Sci.*, vol. 175, pp. 357â364, Jan. 2020.

[3] A. Krummenauer, V.E. Gomes, V.C. Nardelli, "Estimation of Measurement Uncertainty of the Real-Time Location System (RTLS) with Ultra-Wideband (UWB) Technology," *Metrology*, vol. 3, no. 2, pp. 113â130, Mar. 2023.

[4] M.A. Shalaby, C.C. Cossette, J.R. Forbes, J. Le Ny, "Calibration and uncertainty characterization for ultra-wideband two-way-ranging measurements," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2023, pp. 4128â4134.

[5] C. Chen, S.S. Kia, "OptiTrack-Aided Supervised Learning for Neural Network-Based Ultra-Wideband Ranging Bias Correction," *IEEE Sensors J.*, vol. 24, no. 6, pp. 8484â8492, Feb. 2024.

[6] F. Ma, J. He, X. Zhang, "Robust Kalman filter algorithm based on generalized correntropy for ultra-wideband ranging in industrial environment," *IEEE Access*, vol. 7, pp. 27490â27500, Feb. 2019.

[7] L. MarkoviÄ, M. KovaÄ, R. Milijas, M. Car, S. Bogdan, "Error state extended Kalman filter multi-sensor fusion for unmanned aerial vehicle localization in GPS and magnetometer denied indoor environments," in *Proc. Int. Conf. Unmanned Aircraft Syst. (ICUAS)*, Jun. 2022, pp. 184â190.

[8] J. Jia, K. Guo, W. Li, X. Yu, L. Guo, "Composite filtering for UWB-based localization of quadrotor UAV with skewed measurements and uncertain dynamics," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1â3, Feb. 2022.

[9] B. Bao, C. Luo, Y. Hong, Z. Chen, X. Fan, "High-Precision UAV Positioning Method Based on MLP Integrating UWB and IMU," *Tsinghua Sci. Technol.*, vol. 30, no. 3, pp. 1315â1328, Dec. 2024.

[10] A.M. Almassri, N. Shirasawa, A. Purev, K. Uehara, W. Oshiumi, S. Mishima, H. Wagatsuma, "Artificial neural network approach to guarantee the positioning accuracy of moving robots by using the integration of IMU/UWB with motion capture system data fusion," *Sensors*, vol. 22, no. 15, p. 5737, Jul. 2022.

[11] S. Salimpour, P.T. Morón, X. Yu, T. Westerlund, J. Peña-Queralta, "Exploiting redundancy for UWB anomaly detection in infrastructure-free multi-robot relative localization," *Front. Robot. AI*, vol. 10, p. 1190296, Dec. 2023.

[12] M. Nattke, R. Rotta, R. Natarov, O. Archila, P. Mykytyn, "Precise sensors for localization in the drone swarm," in *Proc. iCCCâiCampus Cottbus Conf.*, 2024.

[13] W. Zhao, A. Goudar, M. Tang, X. Qiao, A.P. Schoellig, "Uncertainty-aware Gaussian mixture model for UWB time difference of arrival localization in cluttered environments," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2023, pp. 5266â5273.

[14] F.B. Sorbelli, C.M. Pinotti, S. Silvestri, S.K. Das, "Measurement errors in range-based localization algorithms for UAVs: Analysis and experimentation," *IEEE Trans. Mobile Comput.*, vol. 21, no. 4, pp. 1291â1304, Sep. 2020.

[15] OptiTrack, "PrimeX 13 Motion Capture Camera," OptiTrack Website, `https://optitrack.com/cameras/primex-13/`, accessed July 14, 2025.

[16] Bitcraze, "Crazyflie 2.1 Open Source Drone," Bitcraze Website, `https://www.bitcraze.io/products/old-products/crazyflie-2-1/`, (July 2025).

[17] OptiTrack, "NatNet SDK," OptiTrack Website, `https://optitrack.com/software/natnet-sdk/`, (July 2025).

[18] Bitcraze, "Crazyradio 2.0," Bitcraze Website, `https://www.bitcraze.io/products/crazyradio-2-0/`, (July 2025).

[19] C. Chen and S. S. Kia, "OptiTrack-Aided Supervised Learning for Neural Network-Based Ultra-Wideband Ranging Bias Correction," *IEEE Sensors Journal*, vol. 24, no. 6, pp. 8484–8492, 2024.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Swiss Quality Infrastructure in Transition

Peter Blattner[1], Oscar De Feo[2], Fabiano Assi[3]

[1] METAS, Lindenweg 50, CH-3003 Bern-Wabern, peter.blattner@metas.ch
[2] METAS, now at: NAGRAVISION SÀRL, KUDELSKI GROUP, Route de Genève 22-24 CH-1033 Cheseaux-sur-Lausanne, oscar.defeo@oxanne.net
[3] METAS, Lindenweg 50, CH-3003 Bern-Wabern, fabiano.assi@metas.ch

***Abstract* – In response to the accelerating digitalization of society, the Swiss Quality Infrastructure (QI) is undergoing a strategic transformation. To support this, METAS organized a two-part workshop series in 2024 and 2025. The first workshop mapped the complex stakeholder landscape and explored key QI concepts like competence and trust, highlighting both formal and informal relationships. The second workshop moved from analysis to action, identifying pain points and generating digital solution ideas, including AI, and machine-readable standards. Participants identified possible initiatives such as a QI Data Space, Digital Calibration Certificate, and quality-IoT systems, while also addressing internal readiness challenges. Key outcomes included the need for shared infrastructure, international alignment, and seed funding to support implementation. The workshops provided a starting point for shaping a resilient, digitally enabled QI system tailored to Switzerland's strengths and future needs.**

I. INTRODUCTION

The accelerating digitalization of society is catalysing profound transformations across all sectors—and the Swiss Quality Infrastructure (QI) is no exception, quite the opposite. In response to these shifts, a two-part workshop series was launched in 2024 and 2025 to explore and co-create the future of the QI system in Switzerland. The workshops were initiated by the chair of the Technical Committee INB NK 195 Conformity Assessment, which is the Swiss mirror committee of the ISO CASCO, and they were hosted by the Federal Institute of Metrology (METAS). METAS is mandated by the Swiss Federal Council to coordinate the Swiss Quality Infrastructure [1]. It convened a diverse range of stakeholders, including accreditation bodies, regulators, certification and inspection organizations, industry leaders, standardization experts, and consumer protection representatives.

This initiative acknowledges the dual challenge that QI faces today: leveraging digital technologies to improve performance, meet stakeholder requirements, and ensure trust and transparency; while simultaneously navigating the disruptions these technologies can bring to long-standing organizational structures, concepts, and roles. Particular attention has been paid to the concept of competence, a cornerstone of accreditation and conformity assessment, which is increasingly being tested by technologies such as ML/AI-powered systems

II. WORKSHOP 1: MAPPING STAKEHOLDERS AND EXPLORING CORE CONCEPTS

The first workshop focused on understanding the architecture of the Swiss QI system, specifically by identifying its key stakeholders and mapping their functional relationships.

Accreditation Bodies	NGOs
Accreditation Organizations	NMIs
Accredited Labs	Notified Bodies
Calibration Laboratories	Regional & International Organizations
Certification Bodies	Regulators Legislators
Consumers	Standardisation Bodies
Customs	Technical Committees
Financial Service Providers	Technical Experts
Industrial Value Chain	Testing Laboratories
Inspection Bodies	Tribunal Court
Market Surveillance	

Fig. 1. List of QI Stakeholder.

Through co-constructive and gamified group exercises, participants collaboratively:

- Drafted a map of QI stakeholders, identifying about 20 classes, including accreditation bodies, certification bodies, testing and calibration laboratories, and more (see Fig 1).
- Distinguished types of relationships between stakeholders (e.g., legal, normative, utilitarian, hierarchical, cooperative, and informal). An example of relationship in the QI is shown in Fig. 2.

- Selected and analysed pillar QI concepts, such as competence, trustworthiness, standard, and metrological traceability, and examined how these are "used" across the network—explicitly (e.g., documented competence requirements in accreditation) and implicitly (e.g., assumed trustworthiness between certification bodies and regulators).

Key findings from Workshop 1 included:

- The diversity of stakeholder relationships—ranging from cooperative to competitive—highlights the systemic complexity that digital solutions must address.
- The most dominant and strongest relationships are of type "financial" and "legal", followed by "normative".
- The concept of competence is deeply embedded throughout the network, appearing not only in formal accreditation paths but also in informal exchanges of expertise and judgment.
- Participants recognized the need to distinguish between internal digital transformation (within organizations) and external transformation (across the stakeholder network), setting the stage for deeper exploration in the second workshop.

Subsection heading should be in 10pt italic letters and numbered using capital alphabet letters.

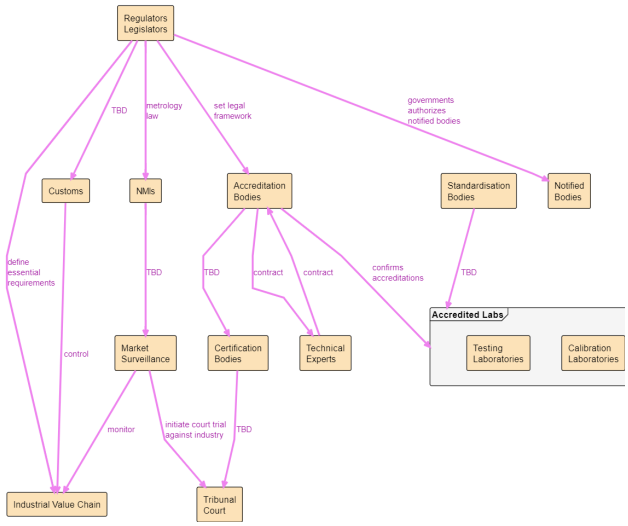


Fig. 2. Example of outcome of workshop 1: relationships of the QI (focus on legal)

III. WORKSHOP 2: FROM CONCEPTS TO ACTION – DESIGNING DIGITAL INITIATIVES

Building on the conceptual groundwork of the first workshop, the second session shifted from mapping the system to designing practical and visionary digital transformation initiatives. It followed a structured, stepwise approach designed to transform abstract challenges into actionable digital solutions for the Swiss Quality Infrastructure (QI). The day unfolded in several interlinked phases, combining systemic analysis, collaborative ideation, and organizational reflection. The group worked in teams, which intermediately exchanged their views and ideas.

A. Identifying Pain Points – Market Pull in the Swiss QI Network

The workshop began by revisiting the functional network of the Swiss QI, as established in the first session. Participants worked in groups to identify concrete pain points and systemic challenges within this network—areas where current processes, interactions, or concepts fall short or face increasing pressure due to digitalization. More than 10 specific pain points or market-driven needs were identified across different stakeholder relationships.

B. Generating Techno-Push Ideas

In the next step, each team co-ideated at least three digital technology-based initiatives—so-called techno-push solutions—which could potentially enhance or transform current QI operations. These ideas were inspired by emerging or underutilized technologies, such as AI, blockchain, digital identities, and data spaces, and were not necessarily tied to previously identified pain points.

C. Designing Systemic Digital Solutions

The market-pull and techno-push results were then brought together. For each pain point or opportunity, participants identified one or more digital solutions, selecting relevant technologies or business models, and reusing techno-push ideas where applicable.

D. Categorizing External Initiatives by Domain of Action

Before turning to internal implications, participants categorized their external digital initiatives based on their domain of action:

- Optimization, referring to solutions that improve or streamline existing processes (e.g., high-speed communication infrastructure).
- Disruption or Revolution, referring to transformative interventions that fundamentally alter stakeholder interaction or compliance mechanisms (e.g., secure data spaces, fully interoperable IoT ecosystems covering the full QI, or machine-actionable standards that only permit device operation if (potentially evolving) standard requirements are met).

This categorization helped assess both the ambition level and implementation maturity of each proposed initiative

and served as a transition point to consider internal readiness.

E. Focusing on Internal Organizational Gaps

Following the external categorization, each team reflected on their institution's ability to engage with the proposed digital solutions. Participants reflected on internal impediments, capability gaps, or structural limitations that could hinder effective participation—ranging from missing harmonization of in-house and international standards (including terms and definitions), outdated IT infrastructure, and lack of in-house digital competencies, up to unclear governance for innovation.

F. Ideating Internal Digital Readiness Initiatives

In response to these internal challenges, teams co-created targeted digital initiatives aimed at enhancing organizational readiness. These were similarly categorized as either:

- Optimization-oriented (e.g., process automation, internal data management systems, more active contributions to international standardization), or
- Disruptive (e.g., redefinition of internal roles, adoption of agile regulatory tech approaches).

This step ensured that institutions not only envision future participation in digital QI ecosystems but also define concrete pathways to get there.

G. Connecting the Dots: Synergies and Conflicts

To conclude the workshop, participants reviewed the landscape of developed initiatives—external and internal. They were asked to assess:

- Whether the use cases were interconnected, and if so, how.
- Where synergies or potential conflicts existed.
- Which initiatives could benefit from shared infrastructure, mutual learning, or coordinated funding.

It was interesting to observe that high similarities in the ideas of the different teams emerged.

IV. OUTCOME

The initiatives that emerged from the collaborative design sessions included:

- QI Data Spaces: Participants emphasized the importance of creating secure, interoperable data spaces to facilitate trusted information exchange across stakeholders, including aspects like identity management. The idea aligns with existing national initiatives [2]—five data space projects have already been identified by the Swiss Federal Chancellery —and a dedicated QI Data Space could be a promising avenue for further exploration.
- Digital Calibration Certificate (DCC): The DCC and its enabling infrastructure were recognized as a crucial element for digital metrology, with attention given not only to the certificate itself but also to the entire end-to-end process leading to its issuance and use within industry and their ERP systems.
- IoT and Big Data: The potential of networked infrastructure and large-scale data analytics to support real-time conformity, monitoring, and predictive maintenance was discussed. These technologies open up transformative opportunities for market surveillance, certification, and traceability.
- Smart Standards: Participants identified the need for machine-interpretable and self-updating standards that could support automation in conformity assessment and regulatory compliance, reducing interpretation ambiguity and manual workload.

In addition to these thematic initiatives, several cross-cutting structural needs were articulated to support implementation:

- The development of business models and value propositions tailored to different stakeholder perspectives, to ensure long-term viability of digital initiatives.
- The establishment of seed funding mechanisms to support prototyping and piloting of digital QI tools.
- The importance of exchange and alignment with international initiatives, particularly through collaboration and learning from the European and global harmonization efforts.

These findings point to the necessity of not only designing technical solutions, but also establishing the financial, legal, and collaborative frameworks that can sustain and scale them.

V. LOOKING FORWARD

Together, the two workshops laid a foundational architecture for the digital transformation of the Swiss Quality Infrastructure (QI). Switzerland's strengths—its compact stakeholder ecosystem, culture of pragmatism, and excellence in precision—position it well to lead by example in QI digitalization. Building on this momentum, we will begin developing a concept for a QI data space to enable trusted, interoperable, and secure data exchange across stakeholders. Additional workshops will focus on identifying and addressing the need for harmonization in digital calibration certificates, a critical element for cross-sector and cross-border interoperability. Informal collaborations have also emerged within the DACH region (Germany, Austria, and Switzerland), pointing to a growing regional interest in aligning efforts and fostering mutual learning. These workshops mark the beginning of a coordinated and forward-looking journey toward a digitally enabled QI system.

REFERENCES

[1] Strategic objectives of the Federal Council for the Swiss Federal Institute of Metrology for the period 2025–2028 , https://www.fedlex.admin.ch/eli/fga/2024/3019/fr (accessed on 16.05.2025)

[2] Swiss Data Ecosystem principles, https://www.bk.admin.ch/bk/en/home/digitale-transformation-ikt-lenkung/datenoekosystem_schweiz/grundlagen-fuer-datenraeume.html (accessed on 16.05.2025)

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Advancing Digital Quality Infrastructure: Transforming Laboratory Processes for Enhanced Efficiency and Reliability

Anna-Maria Elert, Lena Meyer, Nanine Brunner, Michael Melzer, Claudia Koch

[1] *Bundesanstalt für Materialforschung und -prüfung (BAM),*
Unter den Eichen 87, 12205 Berlin, Germany, {firstname.lastname}@bam.de

Abstract – **The digital Quality Infrastructure (QI) holds significant potential for ensuring and enhancing the accuracy, reliability, and efficiency of laboratory processes. Establishing digital QI tools and processes and the integration into the larger digital QI ecosystem however comes with many technical and organizational challenges at the laboratory but also larger system level. This paper presents solutions of a digital QI tool set that are being developed within the German initiative QI-Digital, our own experiences in the implementation of Digital Calibration Certificates (DCCs) as eAttestation in our laboratory, as well as the structured process we have been establishing to engage with the laboratory community to support adoption of the digital QI**.

I. INTRODUCTION

The digital transformation of laboratories is a critical component of the broader shift towards a digital Quality Infrastructure (QI). As key components of the QI system, laboratories need to be integrated into the evolving digital QI ecosystem.

The initiative QI-Digital seeks to develop solutions for a modern, digital QI that allows for smarter and automated quality assurance [1]. This paper introduces some of the QI-Digital tools that are most relevant for laboratories — such as SMART standards and Digital Quality Documents—and shares insights from implementing the Digital Calibration Certificate (DCC) in BAM's accredited calibration laboratory. As intensive stakeholder engagement is crucial to successfully establish these tools in practice, the paper continues to describe our structured stakeholder approach. Finally, the paper synthesizes key learnings and provides recommendations.

II. BACKGROUND

A. Digital Transformation in Laboratories

An international study by BAM across 15 countries worldwide [2] revealed that most conformity assessment bodies (CABs), including laboratories, are still in early stages of digital maturity, with limited adoption of digital technologies.

While there are many projects and initiatives under keywords like Lab of the Future or Lab 4.0 [3], laboratories are not just some type of organization, but a central corner stone within the QI system – a perspective that needs specific consideration. Both, the digitalization of QI and laboratories are closely interconnected and must therefore be driven in close coordination.

To address this, BAM, supported by the German Accreditation Body DAkkS and the VUP (the German Association of Independent Laboratories), hosted the workshop series "*Digital Quality Infrastructure in Testing and Calibration Laboratories – From Trend to Daily Business"* with over 100 participants. The aim was to identify specific needs and challenges and elaborate a vision of digital QI in laboratories [4, 5]. Section IV provides deeper insights and recommendations derived from this stakeholder dialogue. In the following we will present the concept of a digital QI, its building blocks and their implications for laboratories.

B. Digital Quality Infrastructure

The traditional system of quality assurance and infrastructure is still largely document- and paper-based and involves a great deal of manual, bureaucratic effort. The QI-Digital initiative, led by key German QI institutions, seeks to shift towards a data-driven, automated and networked quality assurance system that meets the needs of an increasingly digital economy [1].

The transformation goes far beyond the mere digitalization of existing processes −the initiative strives towards a genuine paradigm shift in the QI to a smart, interconnected QI ecosystem that encompasses modern methods as well as tools embedded in standardized digital platforms and technologies. The following section presents core tools and components of this digital QI ecosystem and how laboratories can benefit from it.

III. DIGITAL QI-ECOSYSTEM FOR LABORATORIES

A. Digital QI-Toolbox

The digital QI offers significant potential for improved

efficiency and error reduction in daily laboratory routines. Specifically machine-readable versions of standards and quality documents, integrated in digitally connected processes, can play an important role.

SMART Standards

Standards define the requirements that products and processes need to fulfill. SMART Standards convert them from a human-only- into a machine-readable form allowing to feed them directly into laboratory systems and workflows. This way, limit and tolerance values can be imported automatically, and test or measurement parameters are loaded without manual re-typing, and automatically cross-checked with the latest version of the standard. Every result is tagged with that version—greatly enhancing traceability and efficiency and reducing errors [6].

Quality documents

Quality documents such as test reports, certificates of conformity, calibration certificates, or reference material certificates, are part of the daily routine in laboratories. Today, they are still issued as paper forms or static PDFs, which makes handling them a tedious effort. Media breaks create duplicate work, increase the risk of typos and version mix-ups, and make every subsequent review time-consuming. Thus, a range of quality related documents are currently being transformed in machine-readable format.:

- The Digital Calibration Certificate (DCC) [7] is an XML-based calibration certificate that consolidates all technical and administrative calibration data and meets ISO / IEC 17025 requirements.
- A Digital Certificate of Conformity (D-CoC) [8] is the machine-readable, structured counterpart to the paper-based Certificate of Conformity under the EU's New Legislative Framework. It consolidates every conformity assessment detail—modules, revisions, regulatory notes—into a unified data model that supports automated processing and plausibility checks.
- Digital Reference Material Documents (DRMD) [9] are machine-readable certificates based on ISO 17034 that encapsulate all key reference-material information—administrative data, certified and uncertified properties with uncertainties, and statements like handling and use, safety information or metrological traceability—within a unified XML schema.

Seamless integration of digital QI documents into laboratory, production and quality assurance processes requires standardized semantics and ontologies, and solutions for trustworthy verification [10]. The envisioned **DX Schema** [11] provides a unified, modular framework, from which many digital QI documents can be derived. Core fields in the schema — such as issuer, scope, applicable standards, validity period, and cryptographic signature—are harmonized across all document types, yet each schema still allows domain-specific extensions [10].

Digital certificates can feed directly into a company's data-management or ERP system, allowing automatic compliance checks and true end-to-end traceability—essential e.g. for the Digital Product Passport (DPP). Benefits include lower administrative burden, seamless data exchange across global value chains, and a single, trustworthy source for various stakeholders.

Digital Accreditation Symbol and eAttestation

To verify the authenticity of quality documents from accredited conformity assessment bodies, the German accreditation body DAkkS has introduced the **digital accreditation symbol** [12]. Embedded in digital certificates, this machine-readable, advanced electronic seal confirms the issuing body's accreditation and enables real-time validation of integrity and authenticity of the entire certificate data. This turns digital certificates into tamper-proof "**eAttestations**" that integrate seamlessly into digital workflows.

B. Digital Ecosystem for QI

For digital QI tools to deliver their full value, they must operate within a cohesive ecosystem. This requires harmonized semantics, standardized data management and sharing, and integration across platforms and stakeholders. The latest developments towards industrial international data spaces and the Asset Administration Shell (AAS), make such a digital QI ecosystem possible.

- **Data Spaces** offer a federated, open infrastructure for sovereign data exchange based on common agreements, rules and standards. They form a data ecosystem, on the basis of which data and services can be made available, networked, shared, validated and used in a trustworthy manner. This increases security, data protection, transparency, and trust, but also efficiency, scalability and adaptability [13] – making it a valuable option for laboratories and their stakeholders.
- The **AAS** (in accordance with the IEC 63278 series) as an administration shell for the digital twin represents each asset digitally and offers great potential for standardized data collection and documentation. It plays a key role in connecting QI systems to Industry 4.0 environments. The standardized AAS sub-models are being developed in the Industrial Digital Twin Association (IDTA) that also has a working group (No. 02065) for digital quality documents [14].

C. Use Case Scenarios: Digital QI for Calibration Laboratories

The following section illustrates practical experiences in the implementation of the DCC in the accredited calibration laboratory of BAM. We discuss the specific steps taken and the benefits realized in daily routines [15].

The accredited calibration lab specializes in three main areas: force, temperature, and electrical quantities. With a staff of three, approximately 200 certificates are issued per year within the accredited scope. Until recently, every Pt100 resistance-thermometer calibration resulted in a PDF certificate that technicians had to create, layout, export, and e-mail to customers—followed by manual data re-entry into the laboratory's quality-management system (QMS) and the customer's asset database. With the introduction of the DCC in our laboratory, the certificate generation can be automated and no longer end with paperwork.

Implementing a DCC for temperature quantities, specifically resistance thermometers calibrated according to the DKD-R 5-1 rule, the laboratory has further been a pilot laboratory for the digital accreditation symbol, allowing for piloting the eAttestation.

Digital calibration workflow

To establish a fully digital workflow in our laboratory, we followed several steps [15]:

1. *DCC template creation*

The first step was the creation of a DCC template based on the DCC scheme provided by PTB. This XML-schema was developed to be suitable for miscellaneous calibrations (*e.g.* from simple weight pieces to complex network analyzers) and is by now accepted as a world-wide standard for machine-readable calibration certificates. [7,16] The DCC template is specific to our laboratory and the calibration method, including placeholders for metadata and measurement data, making the middleware's task of gathering and matching information straightforward.

2. *Creation of a standardized data structure / interface between DCC and any existing calibration software*

A standardized data structure must be created for the interaction between the software with which the calibration is performed and the software with which the calibration data is to be written to the XML sample (middleware). LabVIEW and Python were used in our case. Harmonization between the DCC template (XML), the user interface (e.g. LabView) and the middleware (e.g. Python) are important.

3. *Write data to the XML template*

Measured calibration values, together with the metadata captured in LabVIEW (identifiers, dates, item descriptions, customer information, and more) are transferred to the middleware, which then automatically fills the corresponding placeholders in the XML template. This fully automated workflow ensures that DCCs are generated with both high efficiency and accuracy.

It may be necessary to make slight changes to the DCC structure for specific options within a procedure. In the case of temperature calibrations, a crucial feature is the toggling of the display item, distinguishing between resistance thermometers with a display and bare temperature probes like Pt100, which require a conversion of resistance measurement values into temperature – and accordant information in the DCC. Our DCC software is able to handle both and also supports multiple item calibrations with the automated creation of individual DCC files. All generated DCCs can be readily validated against the DCC schema in the authorization process, *e.g.* with an XML editor, to ensure that the architecture defined there is adhered to.

4. *Signature and eAttestation*

To ensure that the DCC is securely protected against subsequent changes and that the origin of the data can be authenticated, the finalized and approved XML is provided with a digital signature or an electronic seal.

As an accredited laboratory, we apply a digital accreditation symbol to every DCC, enabling recipients to verify our current accreditation status electronically. The German accreditation body DAkkS supplies a certificate file that cryptographically seals each DCC, much like an electronic signature. When the document is validated, this seal instantly shows whether any data have been altered, authenticates the issuing laboratory, and signals if the sealing certificate has been revoked or was expired.

In our simple yet powerful DCC demonstrator for thermal process monitoring, the software automatically validates every incoming certificate's accreditation symbol by checking three criteria: data integrity (no alterations), issuer authenticity, and the laboratory's current accreditation status. Only when all three checks pass is the DCC admitted to the process-control system.

Once accepted, the application extracts the metrological data, performs a linear regression on the temperature calibration points, and applies the resulting calibration function to correct the raw sensor readings. It then expands the certificate's uncertainty budget with additional contributions—such as those arising from the correction itself, long term stability and sensor hysteresis—and recalculates the process-specific acceptance limits accordingly, all in real time.

The result is a one-click workflow that turns a DCC into an instant source of metrological traceability and delivers a quality-assured process conformity assessment.

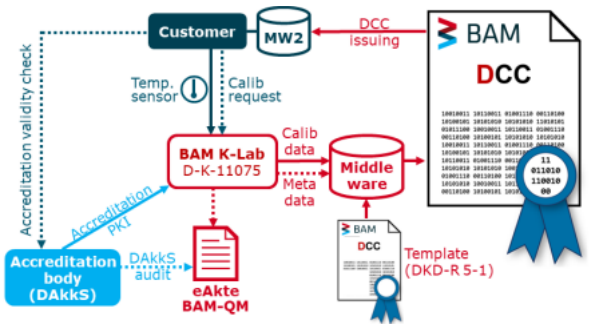


Figure 1 – Process for the automated generation of accredited DCCs (eAttestations) in the BAM calibration

laboratory.

Customer perspective: use at a technical facility

In order to demonstrate the added value of DCCs also on the receiver side, a simple DCC demonstrator for the use case of thermal process monitoring was developed. [17] Upon DCC reception, the software validates the XML-structure and the electronic seal, verifies the major administrative data (*e.g.* performance date and calibrated range), reads out its metrological data, performs a linear regression of the temperature calibration values, uses the obtained calibration function to correct the raw sensor readings, extends the DCC measurement uncertainty values with further contributions (e.g. from the correction and the hysteresis) and adjusts the process-defined acceptance limits to account for the measurement uncertainty, all in real-time. Hence, it automatically utilizes a DCC to obtain metrological traceability and a quality-assured process conformity assessment, practically at the push of a button.

Our next step is to extend the use case to the laboratory's customer. Specifically, we plan to issue DCCs for selected temperature sensors that are then used at our hydrogen refueling station test platform in designated sample processes. In that scenario, the DCC data are automatically fed into the process control technology (PCT) of the refueling station, evaluated and the measurement data from the sensors are linked in real time with the metrological information from the respective DCC.

This enables the corrections required for metrological traceability to be carried out and the measurement data to be quality-assured. The PCT thus carries out a continuous conformity assessment of the processes that are running, which is also quality-assured and therefore reliable thanks to the inclusion and correct further processing of the DCC information, including the measurement uncertainty.

The entire process of transferring, verifying and correctly integrating the metrological information from the calibration of the sensors is therefore fully automated and no longer ties up highly qualified personnel and other resources for the operator of a technical facility. Thus, automation enhances efficiency and accuracy, with significant time and cost savings for companies by accelerating processes and conserving resources. This, together with the increased safety due to the avoidance of errors from manual data transfer and processing, is the decisive added value for users of digital metrology and the DCC.

IV. DIGITAL QI FOR LABORATORIES: STAKEHOLDER DIALOGUE AND ACTIONS

A. Requirements from laboratories' perspectives and recommendations

To better understand the requirements and obstacles of a digital QI from laboratories' perspective, BAM, DAkkS and VUP conducted a series of four stakeholder workshops (see Section II A). The discussions led to specific recommendations to support the digital transformation across the laboratory ecosystem [4, 5].

Overall, participants identified two core **benefits**: improved efficiency (through faster, error-reduced processes), and increased appeal to skilled professionals (through automation allowing focus on demanding tasks). This frees up resources for further optimization, innovation, and the creation of new (digital) business models.

Yet, laboratories also face **hurdles in four areas** along which distinct **recommendations** have been deducted meant to support laboratories [4, 5]:

- **Knowledge and Networks**: The complexity of digital transformation makes orientation difficult. Laboratories need accessible guidance, best-practice examples, and cross-sector networks that connect them with IT providers, regulators, and researchers. Practical training is essential to build confidence in using new tools.
- **Tools and IT Architecture**: Digital tools must be user-friendly for non-IT staff and compatible with existing systems. Open-source solutions with well-documented APIs and strong data protection standards are crucial for trust and adoption.
- **Regulatory Frameworks**: Existing standards and regulations must be updated to accommodate digital processes. Simplified procedures, legal clarity, and standardized interfaces will improve interoperability and acceptance.
- **Business Development**: Adoption will accelerate when leadership sees clear returns. Practical implementation blueprints and collaborative spaces for co-developing business models can help translate digital potential into measurable value.

The workshop results reveal that the rollout of a digital QI will not succeed through technology development alone but needs practical support, dialogue and transfer efforts. Laboratories need strong partner networks, ready access to information for sound decisions, and hands-on help with implementation—and they are eager to collaborate. Addressing those needs, we have taken next steps by launching the "LabHub".

C. Building and strengthening the laboratory community: LabHub

Building on the insights from the dialogue process, BAM, DAkkS and VUP jointly drafted a concept for a "LabHub" addressing the identified needs and recommendations. Meant to support the digital transformation in laboratories, it represents a one-stop shop for **knowledge, training, and collaboration** on

digital topics, including but not limited to QI. Implementing the digital QI in laboratories will be one important aspect of the LabHub. Yet, to ensure its adaptability and integration in the manifold other areas of digital transformation of the laboratory ecosystem, the scope needs to go beyond QI.

Making key resources for the digital transition at large accessible to laboratories is at the core. To this end, a website shall be established with curated information at a glance, including how-to notes and demonstrators; a wiki of best practices and experiential know-how; an academy offering online and on-site courses and workshops; information on funding schemes; a "toolbox" of ready-to-use software kits, guidelines; as well as a steady stream of push-and-pull updates delivered via the website, a newsletter, and social media.

LabHub is also meant to orchestrate events and networking sessions, such as within the annual QI-Digital Forum, industry gatherings like the VUP annual meeting, other conferences by associations, and specialized formats (for example, Accreditation Conferences).

It could also help advancing technological development and transfer them into deployable solutions by hosting collaborative development environments—business camps and innovation labs, plus "Co-Development Circles" in which developers and users refine tools together, as recommended by recent studies.

There are many organizations, initiatives and projects working on different but oftentimes also very similar aspects around Lab4.0, Lab of the Future, Smart Labs, etc. To create transparency, overcome silos, identify interfaces between all those initiatives, BAM, DAkkS and VUP have invited such stakeholders to join forces in the LabHub. In a Kick-off Workshop in April 2025 participants have pledged their support in turning the idea of a LabHub into action. The next steps involve mapping relevant initiatives and projects, exploring their synergies and intersections, mapping their contributions along the laboratory's process chain, and aligning data formats.

In addition, communication with the target group is actively promoted. Alongside the development of the web platform, this also includes talks and presentations as well as further co-operation with other potential partners, particularly at an international level (e.g. with Eurolab).

V. CONCLUSION AND OUTLOOK

The transition to digital Quality Infrastructure (QI) marks a major step forward for laboratories, offering enhanced efficiency, accuracy, reliability and safety. As both beneficiaries and key contributors, laboratories must be actively involved in shaping and implementing digital QI solutions. This paper has outlined the core components of a digital QI toolbox—such as SMART standards, and machine-readable quality documents—as well as the respective ecosystem and has demonstrated their practical application in an accredited laboratory setting.

However, technology alone is not enough. Real progress requires actual implementation and seamless integration across the entire ecosystem of laboratories. To support this, a structured stakeholder engagement process was presented that helped to better understand laboratory needs and promote adoption. The newly launched LabHub aims at connecting stakeholders, foster collaboration, and accelerate the uptake of digital tools—transforming isolated efforts into a coherent, interoperable, and future-ready QI landscape. This will help transfer digital QI solutions into everyday value for laboratories, paving the way for a fully interoperable, future-proof QI.

VI. CITATIONS AND REFERENCES

[1] QI-Digital (2025). Ensuring Quality Smarter. With a digital Quality Infrastructure. Website of the initiative QI-Digital. www.qi-digital.de

[2] L. Ladu, Koch, C., Ashari, P. A., Blind, K., & Castka, P. (2024). Technology adoption and digital maturity in the conformity assessment industry: Empirical evidence from an international study. Technology in Society, 77, 102564. urn:nbn:de:kobv:b43-601930.

[3] Zupancic, Pavlek & Erjavec (eds.) (2021). Digital Transformation of the Laboratory: A Practical Guide to the Connected Lab. Wiley-VCH.

[4] BAM (2024). Digitale Qualitätsinfrastruktur in Prüf- und Kalibrierlaboren - Vom Trend zum Tagesgeschäft - Ergebnisbericht zum Dialogprozess. https://nbn-resolving.org/urn:nbn:de:kobv:b43-612645

[5] BAM (2024). Dialogue process 'Digital QI in Testing and Calibration Laboratories - From Trend to Daily Business'. Summary. https://www.qi-digital.de/fileadmin/user_upload/website/veranstaltungen/Begleitstudien_2024/241206_QID_Ergebnisbericht_Kurzfassung_EN.pdf

[6] DIN/DKE (2024). The business oriented benefit of Smart Standards in standard application processes. Whitepaper.https://www.dke.de/resource/blob/2349926/07fde5efcf4e2b83d6d4c82816a0cef3/din-dke-a4-whitepaper-iii-en-data.pdf

[7] PTB. Digital Calibration Certificate Website. https://www.ptb.de/dcc/(accessed 2025/07/08)

[8] European Coordination Group for Notified Bodies in Legal Metrology, Documentation: Digital Certificate of Conformity (D-CoC). Document 1, 10 May 2023. Online, https://circabc.europa.eu/ui/group/5eaf2706-b5f1-48cc-8b1e-ecde6e4b68fe/library/3500b071-1bc5-4f29-9c8d-eea2825c3ecf/details

[9] M. Melzer, M. Monavari, S. Richter, J. van de Kreeke, C. Koch (2024). The Digital RM Certificate Derived from the DCC. CCQM Online Workshop on Digital and FAIR Chemical and Biological Reference Data and Certificates, Online meeting, 09.09.2024.

[10] T. Engel (2025). Digital Quality Documents and Certificates. In: Meyendorf, N., Ida, N., Singh, R., Vrana, J. (eds) Handbook of Nondestructive

Evaluation 4.0. Springer, Cham. https://doi.org/10.1007/978-3-030-48200-8_86-1
[11] T. Engel and Popescu (2025) “GEMIMEG – Digitalization of Calibration Processes in Metorology,” in Measurement: Sensors. Elsevier, doi: https://doi.org/10.1016/j.measen.2024.101474
[12] DAkkS. Das digitale Akkreditierungssymbol. Website. https://www.dakks.de/de/digitales-akkreditierungssymbol.html (accessed 2025/05/20)
[13] QI-Digital (2023). Quality-X: A Federated Digital Ecosystem for the Future Quality Infrastructure. Whitepaper. https://www.qi-digital.de/fileadmin/user_upload/website/publikationen/1022_Brosch%C3%BCre_Quality-X_v4.pdf
[14] IDTA (2025). AAS Submodel Templates (Website): https://industrialdigitaltwin.org/en/content-hub/submodels
[15] N. Brunner & M. Melzer (2024). Wie funktioniert ein digitaler Kalibrierschein? Report. https://netzwerke.bam.de/Netzwerke/Content/DE/Downloads/qi-digital-dcc.pdf?__blob=publicationFile
[16] PTB. 5th International DCC Conference Website: https://www.dcc-conference-2025.ptb.de/
[17] Youtube. M. Melzer, N. Brunner, K. Nattuveettil, M. Thomas, C. Tiebe (2024). Automated Generation and Utilization of accredited temperature DCCs Video: https://www.youtube.com/watch?si=ra1itRHZZM_Zz6UM&v=j40LJPGZjtk&feature=youtu.be (accessed 2025/07/08)

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Towards an inclusive and agile implementation roadmap for a digital quality infrastructure

Jens Niederhausen[1]

[1] *Physikalisch-Technische Bundesanstalt, Abbestr. 2-12, 10587 Berlin, jens.niederhausen@ptb.de*

***Abstract* – How do we best transition from an outdated QI with many manual processes and analogue *quality documents* to an efficient and interconnected QI that fully leverages digital and automatized processes and granular *quality data*? This paper aims to identify digital transition strategies that can account for the realistic boundary conditions that exist in today's diverse QI landscape, particularly the different digital maturity levels across institutions and nations as well as elaborate regulatory frameworks. Informed by related initiatives, the paper presents promising approaches to map out transition roadmaps for individual starting conditions and transition velocities.**

I. INTRODUCTION

The digital transformation of the quality infrastructure (QI) holds great promise for a complete revolution of how quality is assured. However, like other cases of digital transformation, this disruptive change requires a re-assessment of established processes and additionally includes changes to and implications for products, services, and business models [1]. Notably, it has been found that 70% of all digital transformation initiatives do not reach their stated goals, which was attributed to a lack of coordination and systematic reporting [2]. Thus, for an established system that works productively and effectively, it should be carefully considered how and when to initiate change, even when the disadvantages of the status quo are already obvious. Successful initiatives have been correlated with the combination of digital proactiveness, change commitment, and organizational flexibility [3]. In addition, a *performance infrastructure* that consists of suitable people, processes, and tools working in concert, was found to "help[s] create effective executive-level alignment, communication, and coordination during a transformation." [4]

The QI transformation is different from that of a single organization, since it relies on the coordinated change of several elements of a larger system that is not under the control of a central body (see Fig. 1). The paper explores two specific aspects that seem particularly relevant for the timely and efficient adaptation of digital solutions in the distributed and de-central system that is the QI:

i) inclusiveness: If new solutions are incompatible with legacy processes or emerging solutions employed by other actors, they pose technical barriers and interrupt ongoing activities. This must be avoided as much as possible. Technologically, backward compatible and interoperable technological implementations of digital QI tools and processes allow individual actors to initiate change independent of their stakeholder network. Organizationally, a *performance infrastructure* can ensure that all relevant actors are up to speed, guaranteeing a coordinated digitalization roadmap where no stakeholder is left behind or slows the other ones down.

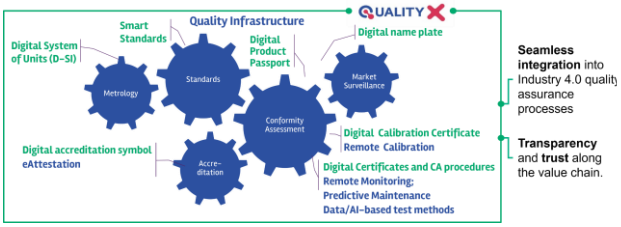


Fig. 1. The Digital QI Ecosystem Quality-X *as envisioned by the initiative* QI-Digital [15][16]

ii) agility: the transition of a complex system like the QI is not completely deterministic. The solution that seems most promising at one point in time will not necessarily prevail in the marketplace of ideas. Agile approaches are most suitable to deal with such a scenario. This is related to *organizational flexibility* and can be supported by a suitable *performance infrastructure*.

II. METHODS

This paper proposes a three-step approach. The first step is to analyze the *current QI* with respect to its essential elements and their digital maturity (see Fig. 2 and Ref. [5]), focusing on the following QI tools and processes:

- (Meta) data formats
- Norms and standards
- Result reports of conformity assessments
- Proof of accreditation and other endorsements
- Reports and documents relevant for customs and market surveillance authorities
- Corresponding processes and procedures

The second step consists in compiling a vision of the *future QI*. This is done by collecting i) the political and technological goals that the QI must address based on its mission as well as expectations from industry and society

ii) technical solutions that are being pursued or considered for the QI's further development.

The final step consists in mapping pathways between the current and future QI, prioritizing those pathways that allow reaching the optimal digital QI ecosystem with minimal disruptions and detours during the transition period (i.e., the time during which several different digital maturity levels co-exist). This entails choosing and timing the integration of the available technical solutions in a way that maximizes the degree to which the QI goals are achieved at all times.

Note that all three steps are rather substantial tasks by themselves in view of the diverse and complex QI stakeholder network. It is thus not within the scope of this article to provide a comprehensive assessment of either. Rather, sections III A to C present a non-exhaustive compilation based on the author's own experience and network. In addition, Section III D describes the *innovation ecosystem* of the QI-Digital initiative as an example for a *performance infrastructure* for the QI's multi-stakeholder network

III. RESULTS AND DISCUSSION

A. Current Status of the QI

Currently the QI is still completely document based. These documents include reports and certifications (from accreditation, calibration, testing, inspection, certification, market surveillance, customs, reverification) as well as norms and standards. On the one hand, the format and terminology of these documents as well as the corresponding QI processes differ significantly between different jurisdictions (municipalities, federal states, countries, regions) as well as domains (e.g., different measurement categories or certification systems), severely hindering their efficient integration into a wider digital transformation. On the other hand, laws, multilateral certification systems, as well as mutual recognition agreements ensure broad validity of these documents. This combination translates into an urgent need for harmonized formats, terminologies, and processes within rather rigid legal and administrative frameworks.

Despite these complications, the digital transformation of QI formats and documents has been ongoing for almost a decade and digital solutions for the representation of units [6], calibration certificates [7][8], and the accreditation symbol of the German Accreditation body DAkkS [9] are already available, while SMART Standards [10], a digital certificate of conformity [11], and the digital product passport (DPP) [12] are in the pilot phase. Notably, the DPP is the only element that consequently enforces the digital format. The other cases by and large allow analogue options for the foreseeable future.

Solutions for the digital mapping of processes have also been forthcoming for several years now. Mostly, these focus on user portals that allow registered QI actors to

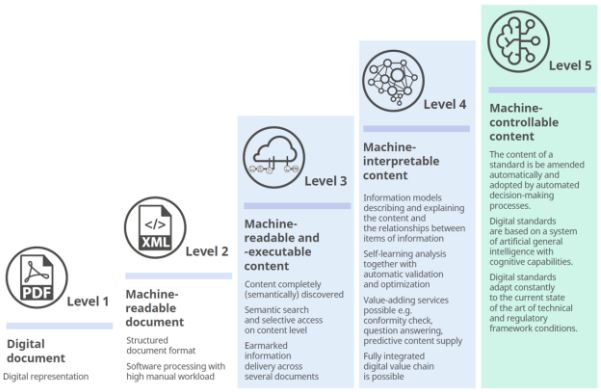


Fig. 2. Extended utility model developed by IDiS [5] that allows assessing digital maturity.

manage their profile, purchase norms and standards, as well as submit requests and retrieve answers for conformity assessments and accreditations. In terms of knowledge sources, certification scheme owners (e.g. IECEx) and accreditation bodies (e.g. DAkkS) provide information on the nature and status of their issued certificates via graphical user interfaces that are connected to their central databases. Market surveillance and customs share information on investigated products (test results, product identification data, economic operator information, accident information, information on measures taken by surveillance authorities etc.) between authorities via central systems (e.g. ICSMS in the EU).

API-based services are much rawer, and often limited to free content like concepts, definitions, and terminologies (e.g., the BIPM SI Digital Framework [13]).

B. Features of the future QI

The International QI Network (INetQI) defines the QI as "The system comprising the organizations (public and private) together with the policies, relevant legal and regulatory framework, and practices needed to support and enhance the quality, safety and environmental soundness of goods, services and processes.

The quality infrastructure is required for the effective operation of domestic markets, and its international recognition is important to enable access to foreign markets. It is a critical element in promoting and sustaining economic development, as well as environmental and social wellbeing." [14]

From this, the following societal, economic, and environmental requirements for the QI can be derived

- Trustworthy and transparent
- Inclusive and internationally recognized
- Supporting innovation and competitiveness

Due to its role as key element of a functioning economy

and facilitator of international market access, the QI system must also meet high standards regarding

- Safety, security, and data sovereignty
- Reliability, resilience, and adaptability
- Efficiency, sustainability, and cost-effectiveness

The German initiative QI-Digital has formulated its vision as "[a] quality infrastructure for the digitalized and sustainable future that enables modern and efficient quality assurance right through to global value creation networks," echoing some of the aspects listed above.

To meet the above goals to the highest degree possible and to enable modern QI procedures like co-calibration, virtual measurement instruments [17], remote monitoring, predictive maintenance, and AI-based methods, the following technological characteristics of the future QI's digital ecosystem (see Fig. 1) are conducive:

- interconnected and decentralized
- data-based (vs. document-based)
- FAIR (i.e. findable, accessible, interoperable, reusable)

These features enable additional data-driven services and thus enhance the overall value of the digital QI ecosystem for the economy, science, and society.

The future digital QI ecosystem can build on technologies that constitute the foundation of modern data ecosystems. The adaptation of the following concepts can thus significantly contribute to an optimized achievement of the above-stated QI requirements.

- Digital twins
- Dataspaces
- (Self-sovereign) digital IDs
- Digital wallets
- Electronic attestation of attributes
- Semantic-web approaches [18]

The large spectrum of measurement procedures, certification systems, product categories, and product life cycle stages result in diverse QI processes and tools. Mapping the requirements to technical solutions already available or under development and precisely identifying possible gaps is a task for the respective domain experts and beyond the scope of this work, which instead aims to highlight the necessary support structure for trans-domain-harmonization and -coordination discussed in Section D.

C. Inclusive and agile approaches

Fig. 2 shows the extended utility model [5] that can help analyze and steer the digital transformation of the QI elements. While the scale goes all the way up to level 5, achieving an end-to-end level 1 QI process is already a major improvement in terms of efficiency, since it prevents media-discontinuities and enables mapping complete process chains to the digital realm. Unified data formats

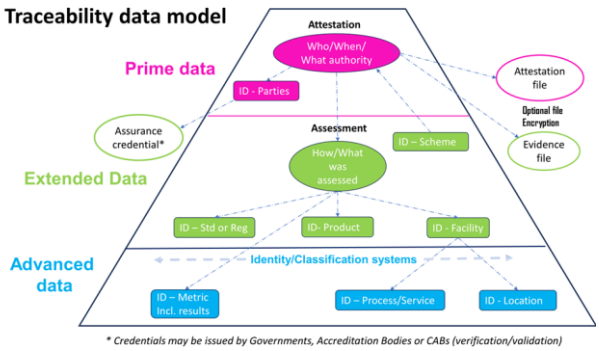


Fig. 3. Conceptual model developed by UN/CEFACT [22].

and open application programming interfaces (APIs) are paramount for reducing manual workload and maximizing efficiency gains. In terms of guaranteeing inclusiveness, handling documents in several different ways should be facilitated by the chosen file type. For example, a PDF file can be read and processed by a human while the machine-interpretable version of that document (e.g., as XML) can be contained as metadata of the PDF file. Extensible documents like XML allow adding metrological and semantic metadata without compromising backward compatibility. This increases the degree of machine-interpretability and enables automatization and the integration of artificial intelligence. E.g. the digital calibration certificate (DCC) offers a long-time stable version 3.0 and more recent versions that enable advanced features [8].

Currently, paper documents are still pervasive and, in many cases, required to meet legal requirements on identity authentication and accessibility. Electronic identities and signatures are increasingly adapted (e.g. enabled by the eIDAS regulation in the EU), however, given that almost three million people have never used the internet in Germany alone, processes will still have to account for communication via traditional mail and paper documents for all acts that include the public.

When exclusively dealing with economic actors, purely digital solutions can be enforced more readily. One prominent example is the EU's digital product passport (DPP), which will become mandatory for the first product category in 2027.

The CIRPASS 2 consortium developed "a 'basic' DPP system that has been stripped down to its most fundamental elements and that is easily applicable to extremely low-cost products with mandatory 'simplest DPP' requirements and focuses on short-term DPP deployments for fast adoption of DPPs by industry"[19]. This approach ensures long-term inclusiveness and prompt uptake of the DPP. Advanced users can still benefit from more advanced implementations that make use of dataspaces to greatly simplify the automatic integration of data from down the value chain as well as the electronic attestation of attributes, for example, by means of



Fig. 4. The QI-Digital Mission (above) and Innovation Ecosystem (below) with its fields of action and a list of relevant technologies and measures as well as the stated vision of a data-based quality assurance enabled by a digitally transformed QI. [15]

verifiable credentials [20].

A related example concerns product conformity certificates. A UN/CEFACT project group has specified the high-level process for the digital exchange of product conformity certificates along the value chain based on 8 principles that address authentication, unique IDs, standard vocabularies, verifiable endorsements, as well as certificate status and validity. The business requirements specification proposes a *traceability data model* (see Fig. 3) and points out that "the approach should be suitable for parties operating at various levels of digital maturity".[21][22].

Digital certificates based on data models that encode the full certificate content in a machine-interpretable way like the D-CoC [11] are a natural extension of the traceability data model. So are the use of verifiable credentials and the implementation in DPP initiatives like the EU DPP and the UNTP [23].

D. The QI-Digital Innovation Ecosystem and Roadmap

The proposed three-step approach consists in matching requirements with technical solutions as well as configuring and timing the implementation such that the QI disruption is minimized. Given the complexity of the QI, each step is a major undertaking that must account for a large stakeholder network and a plethora of processes and tools.

Within the last four years or so, QI-Digital has developed and tested its *Innovation Ecosystem*, effectively a *performance infrastructure* for consultation, alignment, and coordinated implementation with the QI community (see Fig. 4). It consists of four major elements:

1) The digital tools and procedures and their integration into an interconnected and interoperable data ecosystem.
2) Pilot projects that provide infrastructure and technical platforms for R&D in realistic test environments
3) An analysis of the regulatory framework and administrative practices to identify and overcome obstacles for the implementation of the digital QI ecosystem.
4) Measures for sustainable transfer of developed solutions and proactive incorporation of industry needs via comprehensive stakeholder engagement.

While elements 1) – 3) are necessary for a full transformation, it is element 4) that decides about its ultimate success. Only if all stakeholder needs are accounted for and solutions are rolled out effectively, broad acceptance and coordinated and wide-spread adoption can be ensured. It is through these transfer and coordination measures that stakeholder engagement and agile development on a system level is operationalized.

QI-Digital consists of the central actors of the German QI from metrology, standardization and accreditation. Conformity assessment bodies and market surveillance authorities are represented via the advisory council, which also features representatives from industry, academia, and society as members, as well as representatives from relevant ministries as permanent guests.

The advisory council gives all QI stakeholders a say in the transformation process and thus legitimizes the initiative. Sub-groups give strategic advice and contribute to position papers and the analysis of the regulatory framework. In addition, the council has proven indispensable as a steppingstone for the comprehensive engagement of several communities, most notably calibration and testing labs and market surveillance bodies.

One important finding is that one-shot activities by and far do not create significant impact. It usually takes multiple steps to arrive at a meaningful result. For example, big events like the yearly QI-Digital Forum create awareness for an upcoming stakeholder consultation. The respective community then meets several times to assess their status and prioritize their needs. The final step consists of dedicated trainings and workshops, which close knowledge gaps and set up implementation projects.

IV. CONCLUSIONS

The goal of the presented approach is to assist QI actors in their individual digitalization journey, increasing their confidence and motivating them to adapt digital solutions more readily. While the sketched approach in principle

allows systematic analysis and optimization of digital transformation undertakings with respect to inclusiveness and agility, this is beyond the scope of this paper. Instead, the presented examples consider some key strategies as a first step towards a more comprehensive derivation of best practices and guiding principles that are yet to follow.

PTB is part of QI-Digital [15], which has compiled and maintains a *Roadmap for the Digital Transformation of the Quality Infrastructure*, where more details on the development status of the digital QI tools and proposes steps for their further development and scalable roll-out can be found. The interested reader is invited to engage in a dialogue with QI-Digital to jointly shape the QI digital transformation process.

REFERENCES

[1] C. Matt, et al., Digital Transformation Strategies, Bus. Inf. Syst. Eng. 2015, 57(5):339–343

[2] S. ZoBell, Why digital transformations fail: Closing the $900 billion hole in enterprise strategy. Forbes. 2018,

[3] D. K. Nguyen, et al., Leveraging synergy to drive digital transformation: A systems-theoretic perspective, Information & Management 2023, 60(7):103836, DOI: 10.1016/j.im.2023.103836.

[4] M. Bucy, et al., The 'how' of transformation, McKinsey&Company 2016.

[5] DIN/DKE, "IdiS.Whitepaper: Scenarios for digitizing standardization and standards", 2021.

[6] D. Hutzschenreuter, et al.: "SmartCom Digital System of Units (D-SI) Guide for the use of the metadata-format used in metrology for the easy-to-use, safe, harmonised and unambiguous digital transfer of metrological data - Second Edition", Zenodo, DOI: 10.5281/zenodo.3816686.

[7] S. Hackel, et al., "The Digital Calibration Certificate (DCC) for an End-to-End Digital Quality Infrastructure for Industry 4.0.", Sci. 2023, 5(1):11, DOI: 10.3390/sci5010011.

[8] PTB, "Digital Calibration Certificate", URL: https://www.ptb.de/dcc/, accessed May 2025

[9] DAkkS, The digital accreditation symbol, URL: https://www.dakks.de/en/the-digital-accreditation-symbol.html, accessed May 2025

[10] ISO, "IEC/ISO SMART", URL: https://www.iso.org/smart, accessed March 2025

[11] The D-CoC Team, Digital Certificate of Conformity - Version 1.0, 2023, URL: https://www.ptb.de/dcoc/.

[12] CIRPASS 2 Project, URL: https://cirpass2.eu/, accessed May 2025.

[13] Bureau International des Poids et Masures, "Si Digital Framework, Digital References for FAIR measurement data, beta testing version", https://si-digital-framework.org/, accessed March 2025.

[14] International Network on Quality Infrastructure, "Quality Infrastructure Definition", URL: https://www.inetqi.net/documentation/quality-infrastructure-definition/ accessed May 2025

[15] QI-Digital Initiative, URL: https://www.qi-digital.de, accessed March 2025.

[16] QI-Digital Initiative, "Quality-X: A federated digital Ecosystem for the future quality infrastructure", 2023. URL: https://www.qi-digital.de/en/publications, accessed May 2025.

[17] PTB, "Competence Centre VirtMet", digital.ptb.de/en/virtmet, accessed May 2025.

[18] C. Brown, et al., "An Introduction to Linked Data and the Semantic Web", First International IMEKO TC6 Conference on Metrology and Digital Transformation, Berlin, Germany, 19-21 September 2022, DOI: 10.21014/tc6-2022.037.

[19] T. Tammet, et al., „The simplest DPP System Proposal – A thought experiment". CIRPASS-2 Consortium, May 2025. DOI: 10.5281/zenodo.15367064.

[20] H. van Nieuwenhuijze, et al., "D4.1 Reference Architecture", CIRPASS-2 Consortium. Mai 2025. DOI: 10.5281/zenodo.15467720.

[21] UN/CEFACT, "White Paper on Digital Product Conformity Certificate Exchange", 2023, https://unece.org/trade/documents/2023/10/white-paper-digital-product-conformity-certificate-exchange, accessed May 2025.

[22] UN/CEFACT, "Business Requirements Specification on Digital Product Conformity Certificate Exchange - High Level Process", https://unece.org/trade/documents/2024/07/session-documents/brs-digital-product-conformity-certificate-exchange-high, accessed May 2025.

[23] UNECE, "UN Transparency Protocol" URL: https://uncefact.github.io/spec-untp/, accessed May 2025.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

National Metrology Laboratory of the Philippines Digital Calibration Certificate: DigiCert

Ahdrian Camilo C. Gernale[1], Nathaniel Ken A. Aquino[2], Roj Gian D. Gorospe[3], Mark Joseph C. Nicasio[4]

[1]*Industrial Technology Development Institute, Taguig City, accgernale@itdi.dost.gov.ph*
[1]*Asian Institute of Management, agernale.PhDinDS2028@aim.edu*

[2] *University of Santo Tomas, Manila City, nka.aquino02@gmail.com*

[3] *University of Santo Tomas, Manila City, gorosperojgian@gmail.com*

[4] *University of Santo Tomas, Manila City, nicasio.markjoseph641@gmail.com*

Abstract— This paper presents the design and implementation of a system application named *DigiCert*, a Digital Calibration Certificate (DCC) application developed by the National Metrology Laboratory of the Philippines (NMLPhil) of the Industrial Technology Development Institute to modernize calibration workflows and enhance certificate management through standardized, machine-readable documentation. DigiCert improves the process of changing old formats like scanned images and PDF files into organized XML using the Physikalisch-Technische Bundesanstalt (PTB) DCC XML schema. It also allows for manual data entry with a real-time XML preview and can convert back and forth between XML and an easy-to-read PDF. The system employs advanced techniques to enhance images, detect text arranged in a grid pattern, and recognize characters in order to extract both tabular and free-form data from those images. For PDF, it uses Python libraries like pdfplumber and regex parsing to ensure high accuracy. A user-friendly graphical user interface guides the operator through import, edit, and export functions across five dedicated windows, enforcing DCC schema compliance and reducing administrative overhead. DigiCert's flexible design and compliance with ISO/IEC 17025:2018 standards guarantee that the application can work well with other systems, keep track of changes, and maintain security through cryptographic signatures, making it a suitable choice for modern digital measurement and certified calibration labs.

I. INTRODUCTION

A. What is a Digital Calibration Certificate?

A DCC is an electronic version of a traditional calibration certificate that provides standardized, machine-readable documentation of a device's calibration. It is part of the digital transformation in metrology (the science of measurement), aiming to improve accuracy, traceability, interoperability, and efficiency in calibration processes [1].

The DCC uses the widely accepted XML format, making it machine-readable and easily integrated into digital processes. It uses trusted digital signatures, which are created and checked with established security methods, to ensure that a DCC is genuine, has not been altered, and cannot be denied by the sender. Various government systems already rely on these security methods. Additionally, the DCC complies with the DIN EN ISO/IEC 17025:2018-03 standard, making it suitable for use in accredited laboratories [1].

Currently, the PTB, Germany's National Metrology Institute, has developed a web application called the GEMIMEG-Tool, which allows users to generate DCCs by inputting the required data. This tool is publicly accessible through PTB's website.

The goal of this project is for NMLPhil to develop its own DCC application and integrate it into the organization's routine operations. The application, named **DigiCert**, aims to streamline and accelerate the calibration process while aligning with advancements in digital technology and improving certificate management by enabling digital storage.

DigiCert comes equipped with several key features:

1. Convert existing file formats (e.g., .jpg, .pdf) into XML to minimize manual data entry.
2. Supports manual input and editing of certificate data for flexibility.
3. The system has the ability to transform XML files into a format that is easily readable by humans or PDF, thereby guaranteeing accessibility and ease of use.

The design of these capabilities aims to offer a more efficient, user-friendly, and future-ready solution for managing calibration certificates.

B. Benefits of DCC

DCC allows for calibration certificates to be **machine-readable**, allowing for seamless integration with various systems that can operate on the data gathered from the calibrations [3]. The DCC follows a tree-like structure, organizing the data hierarchically, which makes it easier for machines to piece together information.

It also enhances the **traceability** and ensures better tracking of the measurements and standards used during the calibration. This digital traceability not only supports compliance with the standards to meet quality requirements but also increases the validity and reliability of the results gathered in the calibration process [2].

An XML format for the calibration certificates also increases the efficiency of gathering data and speeds up the process of uploading, validating, and using the data. Additionally, it reduces the administrative overhead and minimizes the amount of paperwork required for processing.

An XML format also increases the interoperability and compatibility of calibration certificates. Its structure enables the data to be placed in a hierarchical structure, which allows for better data placement and could be used for data sharing across laboratories, customers, and regulatory agencies [2].

C. Samples of Valid Certificates

One of the current limitations of the application is that it is primarily optimized to read and recognize calibration certificates issued by NMLPhil, particularly certificates from the Force Laboratory. As a result, its compatibility with certificates from other laboratories or formats may be limited at this stage.

D. Scope and Limitations of the Application

Given that the application is in its initial development phase, it comes with certain defined scopes and inherent limitations.

1. The application is currently optimized to recognize NMD's Force Calibration Certificates. Other NMD laboratories can still use it, but it may not fully support all inputs at this stage.
2. PDF inputs will have better results in reading data than image inputs because of OCR limitations.
3. The XML-to-PDF feature of the application currently does not support XML calibration certificates from other laboratories, such as international institutions, as of this moment. It can only process XML files generated through the application's own XML export feature.
4. The XML-to-PDF conversion is not a one-to-one replica of the original certificate; its primary purpose is simply to make the content human-readable.

II. DCC Schema

The DCC follows a schema developed by PTB, which is available on their website. The development of the application primarily used this schema as a guideline, although it served as a reference. Calibration certificates don't have a strict format; they tend to be similar or closely aligned. The schema used in the application was based on PTB's standard but was adapted and structured to better align with NMLPhil's calibration certificate format.

The project DigiCert was developed in compliance with this schema to ensure standardization and compatibility with digital calibration frameworks.

At the top level is the digitalCalibrationCertificate element, which is of type dcc:digitalCalibrationCertificateType. This serves as the root element, containing all the main sections (also referred to as the "four rings" of the DCC) [1]:

1. Administrative Data (administrativeData) – Contains important metadata about the calibration process such as customer information, certificate number, and laboratory details.
2. Measurement Results (measurementResults) – Holds the measurement data, including values recorded during the calibration and their associated uncertainties.
3. Comment (comment) – An optional field that allows for human-readable notes or remarks.
4. Document (document) – Stores supporting documents or binary data relevant to the calibration, represented by the byteDataType.
5. Additionally, the schemaVersion attribute provides versioning for the DCC structure, allowing systems to interpret the document format correctly.

By following this structure, the DigiCert application ensures that every XML output it generates is aligned with international standards for digital calibration certificates.

III. Methodology

The entire DigiCert application was developed using the Python programming language, utilizing various libraries to ensure full functionality and seamless integration of its features. The following sections describe how the various features of the application were developed and implemented.

A. Image to XML

A primary function of the application for converting calibration certificates to an XML (DCC) format is to convert an image of a certificate to the DCC format. By passing an image as input to the application, the system can read the data from it, convert it to the structure of the XML schema, and allow the user to add or edit more details before exporting the calibration certificate as an XML file.

1. Approach used for Optical Character Recognition (OCR)

There are three main steps in reading the calibration data from images to turn them into texts for the system to be able to process and convert them to the XML format. The steps taken include image preprocessing, text detection, and text recognition. These steps were followed sequentially to ensure that most of the calibration details are able to be properly read and be ready for conversion to XML.

Preprocessing the input images comes first before going through the next steps. This utilizes several image-enhancing algorithms to improve the quality and clarity of the images before the main steps of optical character recognition are applied. The preprocessing of the image includes several changes, such as removing noise, combining color channels for better clarity, adjusting contrast with the Contrast Limited Adaptive Histogram Equalization algorithm (CLAHE), and using Gaussian blurring. CLAHE is an image processing technique that enhances the contrast of images, particularly in areas where contrast is low. Denoising reduces the color noise while preserving edges, which makes the texts more distinguishable from the background. Using Lab color space, also known as CIELAB, separates color information into lightness (L) and two color channels (a and b). The L channel represents the lightness of a color, ranging from 0 (black) to 100 (white). Additionally, the image is changed to the Lab color space, which separates lightness (L) from color (A and B). This process helps improve contrast, especially in the lightness channel, making faded or uneven text easier to see. Finally, Gaussian blur and sharpening are also used, wherein a blurred version of the image is subtracted from the original (using weighted difference) to enhance edges and other fine details in the document, resulting in clearer character strokes for the OCR to recognize.

The next step is the text detection, wherein the goal is to identify the regions of interest in the image, particularly tabular or grid-based layouts that are likely to contain text. To find these areas, the image is first changed to grayscale and then inverted using adaptive thresholding to help separate the text and lines (foreground) from the background. Following this the horizontal and vertical line structures are then extracted using erosion and dilation with shaped kernels to help in isolating table lines or cell boundaries. Using the detected lines, a grid-like mask is then constructed by combining the horizontal and vertical lines. Contours of the rectangular regions are then extracted. The detected cells are then filtered based on their respective area and calculated median threshold to remove noise and to avoid detecting the cell's table as another "cell," as it would duplicate the OCR on the table. The bounding boxes are sorted top-to-bottom and left-to-right to preserve the natural reading order.

Finally, for the OCR, we used a machine OCR algorithm called Tesseract. This employs a two-stage adaptive recognition process. It begins by identifying character patterns, then uses context to refine its understanding of words and sentences, and then converts image regions into actual text. We used several Tesseract configurations, such as *--oem 3*, *--psm 12*, and *--l eng*. The configuration *--oem 3* utilizes the LSTM-based OCR engine, which provides high accuracy for both printed and handwritten texts that are the target input of the DCC system. Meanwhile, *--psm 12* specifies the type of page segmentation, treating the image as sparse text with possible orientation and script detection. As the calibration certificates are a combination of descriptions, tables for results, and paragraphs, *--psm 12* was chosen as the best configuration. Finally, *--l eng* tells Tesseract that the expected language is set to only English. If the DCC system expands to other languages, this configuration can be deleted.

2. Using OCR on a scanned document

After running the OCR on the system, here are the generated results saved on a temporary PDF file to be converted into an XML format:

For the scanned images, the OCR system can better detect texts as there is less environmental noise that is happening in the image. Additionally, the foreground is already quite differentiable from the background.

Images captured with a camera could result in a less accurate OCR, as there are many outside factors that could add noise to the image. Even a blurred or low-quality image can affect the quality of the text recognition system. However, the image still retains most of its details.

B. PDF To XML

The conversion of calibration certificates from PDF to XML ensures accuracy and reliability in extracting data from text-based PDF documents since the application leverages the pdfplumber Python library.

The Python library pdfplumber is specifically designed for extracting text, tables, and metadata from text-based PDF documents. It is particularly effective for digital PDFs that contain selectable text, rather than scanned image-based documents. Using this Python library, the application is able to accurately retrieve

paragraph-level and line-by-line text from PDF pages and seamlessly process PDFs with multiple pages.

The process begins when the user selects a PDF file through a file dialog interface. Once the file is chosen, it is opened using the pdfplumber.open() function, which loads all the pages of the PDF. The application then loops through each page, using the extract_text() method to retrieve the textual content. All the extracted text from the pages is combined into one string called raw_text, which serves as the foundation for further data extraction.

After obtaining the text, the application extracts tables from each page using the extract_tables() method. Each table is processed by assuming that the first row contains the headers, while the following rows represent the data entries. These tables are stored in a dictionary format where each header is used as a key, and the corresponding values are the rows of data beneath it. This structure allows for easier access and organization of tabular information.

Specific pieces of information needed for calibration records are extracted from the raw text using regular expressions. This includes details such as the certificate number, calibration date, customer name, and address. For example, a pattern like CERT-\d+ is used to find the certificate number. The same technique is used to extract other important fields, and additional logic is applied when the formatting of the certificate varies across files.

Once all relevant text and table data are extracted, they are organized into a structured Python dictionary called calibration_info. This dictionary uses clear and consistent key names such as "certificate_number," "calibration_date," and "customer_name," with each key assigned its corresponding value. This structure prepares the data for easy conversion into other formats.

The dictionary is then saved as a JSON file using Python's built-in json.dump() function. Saving the data as JSON ensures that it remains structured, readable, and ready for further processing. This step also helps separate the data extraction phase from the XML conversion phase, allowing greater modularity in the system.

Finally, the JSON file is sent to a separate script responsible for generating the XML output. This script reads the JSON data using json.load() and constructs an XML structure. Each field in the JSON dictionary is turned into a corresponding XML element. The resulting XML file is then saved and formatted according to the required schema, making it suitable for submission, integration, or storage.

C. Creating New DCC

The creation of New DCC is designed to simplify the manual creation of XML files for calibration records through an interactive graphical interface. The process begins with the initialization of the application, where static configuration details, such as laboratory information and software metadata, are loaded from a JSON file.

When the application starts, it builds a simple and user-friendly interface. On the left, users enter details like certificate number, date, equipment, and customer info using text boxes and drop-downs. On the right, a live preview shows the XML output as data is entered, helping users check their inputs right away.

As users enter values into the input fields, the script collects all relevant data and compiles it into a structured dictionary. This dictionary is then used in the XML generation phase, where a predefined XML template is loaded and populated with the input data. Proper XML namespaces are handled to maintain schema compatibility.

The real-time preview updates dynamically as users edit the form, providing an accurate representation of what the final XML will look like. Once all information is complete, the user can export the XML file by choosing a save location through a file dialog. A confirmation message appears after a successful export. Users are then navigated back to the main menu, allowing them to start a new entry or perform other tasks.

This structured and intuitive process ensures that users can generate consistent, accurate XML files for calibration data without needing to manually code the XML structure themselves.

D. Conversion process of XML to PDF

This is the overall process of converting the XML-generated DCC of DigiCert back to PDF. The process starts by setting up the script, which includes defining XML namespaces and creating a helper function to safely extract text from XML elements. The XML file is then loaded and parsed. Using XPath queries, important data is pulled from specific parts of the XML, such as software details, administrative information, equipment used, lab details, responsible individuals, customer data, and measurement results. XPath stands for XML Path Language, which is a non-XML syntax to provide a flexible way of addressing (pointing to) different parts of an XML document [4].

Next, the content for the PDF is organized into sections. Each section is added to a list called the "story," which holds everything that will go into the PDF. A special table is created for the measurement results to make the data easier to read.

Once all the content is ready, the PDF is generated using ReportLab's SimpleDocTemplate. Finally, the finished PDF file is saved to the user's chosen location. This ensures that all XML data is clearly and correctly converted into a well-structured PDF document.

IV. The Application Result

Since the output of the project was an application, its graphical user interface (GUI) was designed to be interactive and user-friendly to ensure optimal usability, as shown in Figure 1. The application consists of five main functional windows. This section presents and describes each window in detail.

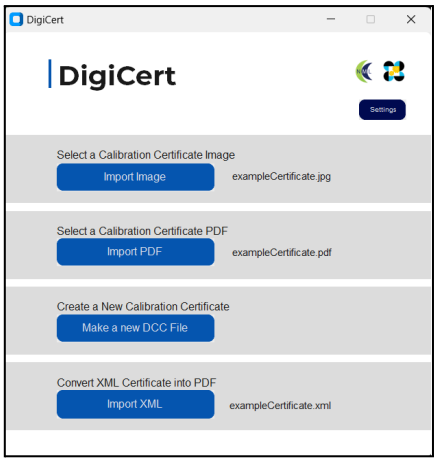


Figure 1. MAIN WINDOW

The main window greets the user upon launching the DigiCert application. The main window functions as the central hub, providing access to all the application's core features and functionalities. As previously mentioned, the app allows users to scan and convert image files to XML, convert PDF files to XML, create a new XML file from scratch or manually input data, and finally, convert an XML certificate into a human-readable format (PDF). All of these features will be discussed in further detail in the succeeding sections of this paper.

The "Settings" button is located on the upper-right portion of the main window. When clicked, it redirects the user to the settings window, where they can edit or modify certain information within the application.

This is the layout of the settings window. It contains static values such as country, language, and laboratory information. These static values are typically constant and do not require frequent changes. However, users still have the option to modify them when necessary. By default, the fields are pre-filled with standard values as shown above. After making any adjustments, users can click the “Save” button. The changes will then be applied and reflected accordingly upon returning to the main window.

Each text box provides visual feedback: it will appear red if a required field is empty and green once valid input is entered. This helps guide users to complete all necessary fields. The form is scrollable, allowing users to review and verify the entire certificate content.

It is strongly recommended that users follow the format indicated by the placeholder text in each field. This ensures that the resulting XML file adheres to the DCC standard and maintains consistency.

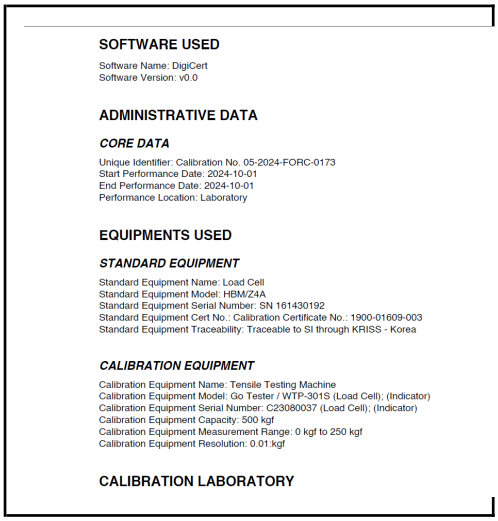
Once all information has been reviewed and finalized, users can click the "Export" button to generate the XML version of the certificate in Figure 2.

The New XML window shares the same layout and appearance as the other windows in the application. However, the key difference is that users must manually input all the required information from scratch. Unlike the image or PDF import features, no data is automatically extracted in this window, so users are expected to fill in each field completely.



Figure 2. EXPORT

Figure 3 shows what the export notification results are when the XML file is exported. Once the export is complete, the user will be redirected back to the main window.



SOFTWARE USED

Software Name: DigiCert
Software Version: v0.0

ADMINISTRATIVE DATA

CORE DATA

Unique Identifier: Calibration No. 05-2024-FORC-0173
Start Performance Date: 2024-10-01
End Performance Date: 2024-10-01
Performance Location: Laboratory

EQUIPMENTS USED

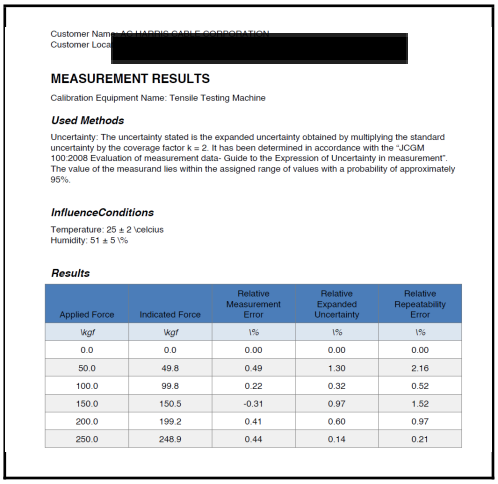
STANDARD EQUIPMENT

Standard Equipment Name: Load Cell
Standard Equipment Model: HBM/Z4A
Standard Equipment Serial Number: SN 161430192
Standard Equipment Cert No.: Calibration Certificate No.: 1900-01609-003
Standard Equipment Traceability: Traceable to SI through KRISS - Korea

CALIBRATION EQUIPMENT

Calibration Equipment Name: Tensile Testing Machine
Calibration Equipment Model: Go Tester / WTP-301S (Load Cell); (Indicator)
Calibration Equipment Serial Number: C23080037 (Load Cell); (Indicator)
Calibration Equipment Capacity: 500 kgf
Calibration Equipment Measurement Range: 0 kgf to 250 kgf
Calibration Equipment Resolution: 0.01 kgf

CALIBRATION LABORATORY



Customer Name: [redacted]
Customer Location: [redacted]

MEASUREMENT RESULTS

Calibration Equipment Name: Tensile Testing Machine

Used Methods

Uncertainty: The uncertainty stated is the expanded uncertainty obtained by multiplying the standard uncertainty by the coverage factor k = 2. It has been determined in accordance with the "JCGM 100:2008 Evaluation of measurement data- Guide to the Expression of Uncertainty in measurement". The value of the measurand lies within the assigned range of values with a probability of approximately 95%.

InfluenceConditions

Temperature: 25 ± 2 \celcius
Humidity: 51 ± 5 \%

Results

Applied Force	Indicated Force	Relative Measurement Error	Relative Expanded Uncertainty	Relative Repeatability Error
\kgf	\kgf	\%	\%	\%
0.0	0.0	0.00	0.00	0.00
50.0	49.8	0.49	1.30	2.16
100.0	99.8	0.22	0.32	0.52
150.0	150.5	-0.31	0.97	1.52
200.0	199.2	0.41	0.60	0.97
250.0	248.9	0.44	0.14	0.21

Figure 3. SAMPLE PDF FILE

Finally, users can also use the XML to PDF feature to transform the generated XML into a human-readable format. A PDF file will be exported containing information in the XML.

V. Conclusion

To improve certificate management and modernize the reporting of calibration through standardized, machine-readable documentation, NMLPhil created DigiCert, a powerful DCC application, which was presented in this paper. The primary goals and accomplishments of the system include

1. DigiCert efficiently operationalizes the conversion of conventional scanned images and PDF report files into PTB-compliant DCC XML.
2. Using smart OCR technology for images and accurate PDF reading with Python tools, DigiCert has demonstrated a dependable way to convert between XML and easy-to-read formats while ensuring both data accuracy and user-friendliness.
3. The application features a five-window GUI that guides users through every stage of import, editing, and export, and instant XML previews that enforce schema compliance

REFERENCES

[1] Digital Calibration Certificate - DCC - PTB.de. (2024, March 26). https://www.ptb.de/cms/en/research-development/ptbs-innovation-clusters/innovation-cluster-for-digitalization/kernziel1einheitlichkeitim/digital-calibration-certificate-dcc.html

[2] Koval, M., Tesař, J., Schönhals, S., Sharma, A., Balslev‑Harder, D., Galvan‑Hernandez, C., Baumann, H., & Mahovský, D. (2025). Digital Calibration Certificates (DCC): Benefits and Challenges to overcome for implementation [Conference paper]. EPJ Web of Conferences, 323, Article 01001. https://doi.org/10.1051/epjconf/202532301001

[3] Bernien, M., Straka, M., Bock, T., & Niepraschk, R. (2025). Digital calibration certificates and automated calibration workflow in the Vacuum Metrology Group of PTB. Measurement: Sensors, 38 (Supplement), 101490. https://doi.org/10.1016/j.measen.2024.101490

[4] https://developer.mozilla.org/en-US/docs/Web/XML/XPath

2025 IMEKO Joint Conference
TC8 - TC11 - TC24
Torino, Italy, September 14-17, 2025

Development of Frontend Interface for Digital Calibration Certificate for AC High Current Source Parameters

Paramita Guha[1,2,*], Arun Ram Prasath R T[1,2], Manish Kumar Tamrakar[1], Shrikrishan[1], Priyanka Jain[1,2]

[1] *CSIR-National Physical Laboratory, Dr KS Krishnan Marg, Pusa, Delhi, India*
[2] *Academy of Scientific & Innovative Research, Ghaziabad, India*
**Corresponding Author paramita.guha@nplindia.res.in*

***Abstract* – CSIR-National Physical Laboratory, India an apex level body which provides calibration services to various secondary, tertiary calibration and testing labs and industries across the country. In general, calibration reports are commonly created in paper or word format to produce PDFs, which are time-consuming, tedious, and susceptible to human errors. Due to this several issues, BIPM- France is taking suitable goals to overcome this issue through the development of DCC. As globally leading NMIs like PTB, Germany, etc were working towards DCC developments, CSIR-NPL (NMI of India) takes part in DCC working group which are in-line with the leading NMIs. The main aim of this work is to create and implement a web application based on react.js to create digital calibration certificates. By deploying this solution, there will be substantial gains in data traceability, security, reliability, and accessibility. The project encompasses the creation of a Digital Calibration Certificate (DCC) platform that substitutes the conventional manual procedures with an automated, scalable system with the ability to produce structured and verifiable digital certificates. To further verify the authenticity and ease of validation, a model based on QR codes has been implemented that produces a unique, scannable code corresponding to a cloud-hosted copy of the certificate, permitting secure and instantaneous access. Also, advanced cryptographic algorithms like cryptographic hashing and PDF encryption are applied to secure the digital certificates and to ensure that the certificates remain tamper-proof and intact over their lifecycle. This end-to-end digital solution not only updates the laboratory's calibration report process but also establishes a new standard for secure, transparent, and effective calibration services.**

I. INTRODUCTION

CSIR-NPL, the NMI of India has a central role in ensuring the accuracy and consistency of measurements by hosting six of the seven SI base units: meter (m), kilogram (kg), kelvin (K), second (s), ampere (A), and candela (cd). It maintains traceability of measurements to the SI units through its comprehensive calibration services and thereby provides critical assistance to various industries, calibration and testing laboratories, and research institutions all over India and other SAARC nations. In spite of their critical significance to the business process, the standard practice of preparing calibration reports remains antiquated today—using paper-based records and manually generated PDF files. Not only are the conventional methods wasteful of precious time and efforts, but also prone to fallibility due to human errors. Such methods therefore risk compromising on the accuracy as well as credibility of calibration records. Moreover, these methods cannot meet the modern requirements of secured handling of data, traceability, and interfacing data between systems.

To overcome these limitations, this work proposes the design and development of a web-based application using React.js—a robust and modern JavaScript framework. The proposed platform aims to automate the creation of calibration certificates, thereby improving the real-time digital record management. By embracing digital transformation, this initiative aligns with the broader industry trend of automation and enhanced data-driven processes. For security and authenticity purposes, the system employs cryptographic hashing and QPDF to shield PDF certificates against unauthorized access and tampering. Also, a QR code generation system has also been developed with *Streamlit* and Python. After the calibration certificate is created, it is stored in a secure cloud storage platform (such as Google Drive), and a downloadable share link is generated. The link is then converted into a QR code, which is placed directly on the certificate. Users can scan the QR code to access or validate the certificate, allowing for immediate, secure, and traceable access from anywhere. This work creates a new platform in the field of calibration services, providing a scalable, secure, reliable and robust solution

that enhances trust in digital calibration information and assists with the metrological traceability towards SI Units.

II. LITERATURE SURVEY

Researchers in [1] have developed a bilingual NIM-DCC [1, 2] metamodel based on the PTB-DCC metamodel [3]. They also have developed software for generating and validating NIM-DCC. The system had been deployed on a microservices architecture, supporting service scalability and coordinating development and operation. The system has different layers, multiple fields, clear interfaces and standardized transparency. The proposed methodology has been tested and applied in more than 10 laboratories of NIM and local metrology institutes

In [4], the problem of secure and authentic signing of official e-government documents using XML have been discussed. It suggests "authentic PDF" can be used as a solution. The authentic PDF technique has been used to satisfy key requirements like, classic visual looks, signature representation with authenticity, recreation of the electronic version from a printout etc. The authors in [5] explore the potential of using Convolutional Neural Networks (CNNs) for symmetric encryption. The authors implemented a CNN-based symmetric encryption scheme and compared its performance against the Advanced Encryption Standard (AES) using various file sizes and content. The comparison focused on execution time, memory usage, and processor load during encryption and decryption. While the paper acknowledges that the robustness of the CNN encryption against cryptanalytic attacks was not evaluated, the experimental results suggest that CNNs could be a viable option for symmetric cryptographic applications under the specific conditions tested.

This research in [6] addresses the increasing necessity of verifying authenticity, non-repudiation, and integrity of electronic documents, especially PDF documents. PHP has been used to create a web-based application that incorporates the RSA digital signature scheme with the SHA-3 hash value to securely sign and authenticate PDF documents. RSA, one of the most common public-key cryptographic methods, delivers digital signing functionality, and SHA-3 guarantees the integrity of the document by creating a distinctive digital fingerprint. The system has tried on PDF files of 6 KB and 23 MB sizes, and tested across principal web browsers—Google Chrome, Microsoft Edge, and Mozilla Firefox. The signing and verification times are average in all browsers but efficient and comparable, taking approximately 1.3309 seconds on Chrome, 1.2565 seconds on Edge, and 1.2667 seconds on Firefox.

The researchers in [7] present a novel image encryption model that combines a Convolutional Neural Network (CNN) with an intertwining logistic map to generate robust secret keys. This approach leverages the chaotic nature of the logistic map, influenced by initial conditions, control parameters, and CNN-derived keys, to produce diverse sequences for pixel scrambling and manipulation through permutation, DNA encoding, diffusion, and bit reversion. The model further enhances security by employing different subkeys, private keys, and public keys generated by the CNN, leading to a larger keyspace and improved confusion and diffusion. Extensive analysis against various attacks (cropping, noise, differential, statistical) and evaluations of key space, sensitivity, histogram, information entropy, and correlation coefficient demonstrate that the proposed method outperforms existing image encryption techniques. The results highlight significant improvements in information entropy, randomness, resistance to attacks, and overall efficiency, establishing the model as a more stable, secure, and reliable solution for image encryption. [7]

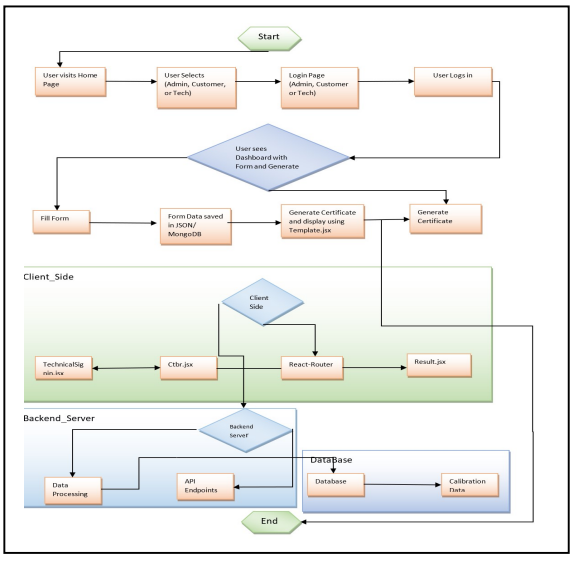
III. METHODOLOGY

The methodology outlines the structured approach used to design, develop, and implement the digital calibration certificate generation system. It follows a modular and web-based architecture, ensuring scalability, ease of access, and data security. The system is designed using React.js for the frontend and integrates secure mechanisms to validate and protect certificate data.

A. Overall Architecture: DCC

The Digital Calibration Certificate (DCC) system streamlines the process of generating and managing calibration certificates through a user-friendly web interface. The system begins with a clear and straightforward homepage featuring three primary navigation buttons: "Administration," "Technical," and "Customer." When users select the "Technical" button, they are directed to a secure sign-in page where they must enter their credentials to gain access to the technical functions. The overall architecture is shown in Fig. 1.

Fig. 1: Overall architecture of DCC



Upon successful login, users are redirected to the TechnicalCalibrationForm.jsx page. This form is crucial for submitting detailed technical information required for calibration. In addition to filling out the form, users must upload a file containing relevant data. Once the form and file are submitted, the system processes the information and navigates the user to the Result.jsx page.

The Result.jsx page presents all the entered data and the contents of the uploaded file in a comprehensive and organized manner. Users can review this information and download the results as a file, ensuring they have a complete and accessible record of the calibration details. This structured workflow not only enhances efficiency but also ensures accuracy and convenience in the calibration certification process.

The architecture of the Digital Calibration Certificate generation platform is designed to automate the certificate creation process by leveraging both manual form entry and bulk CSV uploads. The system is built with modular components, covering form initialization, data input, CSV parsing, form validation, PDF generation, and secure file handling. The flowchart of the work is given in Fig. 2.

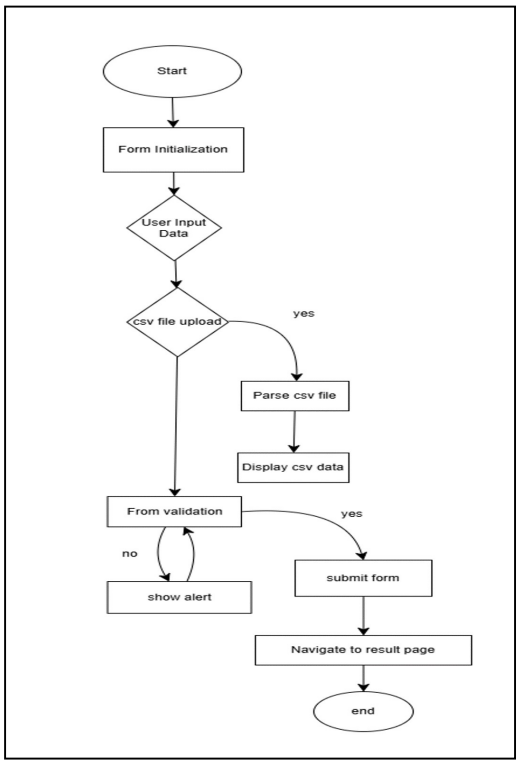


Fig. 2. Generation of DCC from user input to result

B. Architecture of Cryptography

Fig. 3 illustrates the overall workflow of the secure PDF generation and download process. It begins when the user initiates the download. The frontend captures the webpage, generates the PDF, embeds a unique hash for security, and sends it to the backend. The backend then signs and encrypts the PDF before returning it. Once received, a secure download link is generated for the user, completing the download process smoothly and securely.

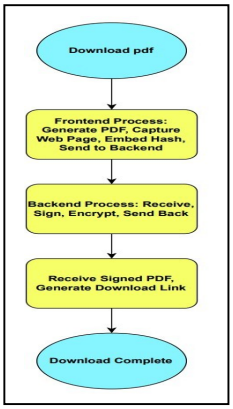


Fig. 3. Flowchart for secure PDF generation.

C. Architecture for QR code generation

Figure 4 depicts the flowchart of QR code-based system used for secure digital access to calibration certificates. PDF generation and upload to a cloud storage solution like Google Drive is the starting point. A shareable link is generated and encoded as a QR code. This QR code is placed on the certificate so that users can scan and download or verify the certificate securely from any device. The flow provides both accessibility and document authenticity.

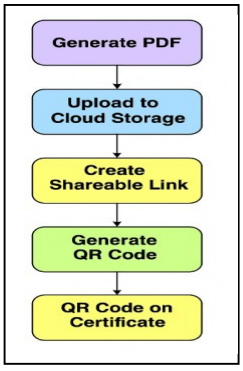


Fig. 4. Flow chart for QR code generation.

IV. DATA COLLECTION AND TECHNIQUES USED

For the generation of DCC, calibrations are performed on a Current Injection Test Set with inputs as 415Volts, A.C., 1 Phase, 50 Hz, and outputs are 0-3 Volts/5000Amperes, 0-6 Volts/2500 amperes, 0-12 Volts/1250Amperes and 0 - 415 Volts/60Amperes [8]. The test set was manufactured by Automatic Electric Ltd., Mumbai. For the measurements, standard environmental conditions, viz., temperature: $(25 \pm 2)^0$C and relative humidity: (50 ± 10) %, are maintained. The calibrations are carried out by a comparison method using a standard Current Transformer (with associated uncertainty as ±0.0025%) along with a standard Digital Multimeter (with associated uncertainty as ±0.002% - ±0.07%). The results obtained as shown in Table 1.

Table 1: Calibration results of current injection test set

Indicated Value (A)	Measured Value (A)	Coverage Factor (k)	Expanded Uncertainty (%)
100	103.2	2	0.70
500	508.4	2	0.16
1000	1013.3	2.21	0.07
1500	1517.4	2.25	0.09
2000	2021.5	2	0.07
2500	2522.9	2	0.06
2500	2521.7	2	0.08
3000	3023.6	2.52	0.28
3500	3525.4	2.21	0.13
4000	4031.0	2.37	0.16
4500	4528.5	2.52	0.23

This data is used for the generation of DCC models. Data is initially received in .csv format then internally converted into JSON objects, for seamless processing across the frontend and backend systems. Before certificate generation, the system performs preprocessing steps such as parsing and validating the CSV structure, detecting missing or malformed values, standardizing formats (like dates and units), and preparing clean, consistent data for output [9]. To maintain data integrity, the system includes real-time validation rules during input, automatic error detection for uploaded files, and checks for duplicates or inconsistencies. Additionally, cryptographic hashes are embedded into the generated PDFs to ensure authenticity and prevent tampering.

For programming, JavaScript has been used as a primary language for both frontend and backend development due to its wide ecosystem and flexibility. React.js is used to build a responsive, component- based frontend that handles user input and displays the generated certificate technically similar to paper format. The vite.js is used to develop as a frontend build tool which enables fast development and real- time updates during coding. Node.js Facilitates backend logic, including request handling, data processing, and certificate generation. Also, express.js and MongoDB are used to store calibration data, user records, and metadata related to certificate generation. For the PDF security, QPDF and Hashings are used to generate a unique digital fingerprint for each certificate, verifying the integrity and detecting any changes made after generation [10-12]. The development environment includes Visual Studio Code, Node.js, npm, and Git on Windows/Linux systems.

V. RESULTS AND DISCUSSION

For the generation of digital certificates, different form pages are generated. The snapshots of the forms are given in Figs. 5 and 6. Finally the QR code based digital certificate for the data of Table 1 is given in Fig. 7.

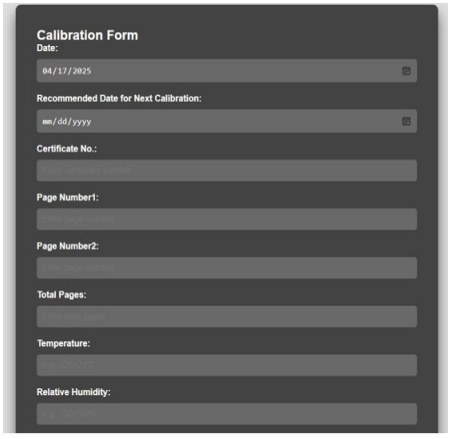


Fig. 5(a): Snapshot of the administration form

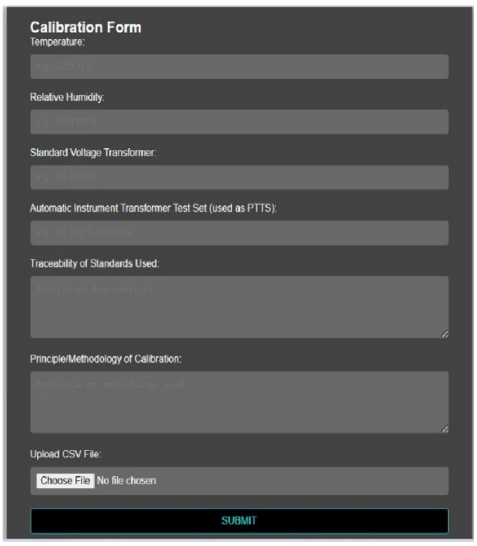
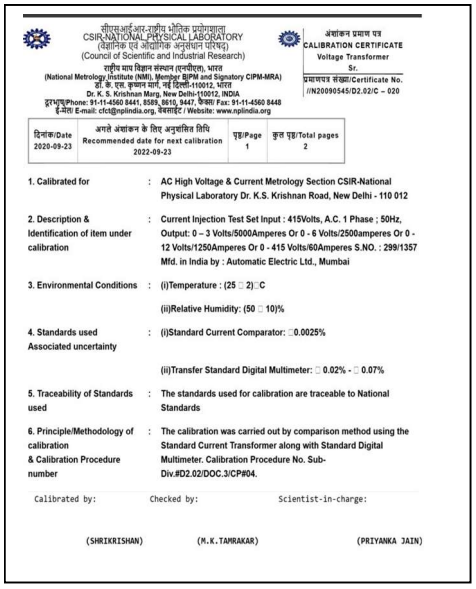


Fig. 5(b). Snapshot of technical form



CSIR-NATIONAL PHYSICAL LABORATORY
(वैज्ञानिक एवं औद्योगिक अनुसंधान परिषद्)
(Council of Scientific and Industrial Research)
राष्ट्रीय माप विज्ञान संस्थान (एनपीएल), भारत
(National Metrology Institute (NMI), Member BIPM and Signatory CIPM-MRA)
डॉ. के. एस. कृष्णन मार्ग, नई दिल्ली-110012, भारत
Dr. K. S. Krishnan Marg, New Delhi-110012, INDIA
दूरभाष/Phone: 91-11-4560 8441, 8589, 8610, 9447, फैक्स/ Fax: 91-11-4560 8448
ई-मेल/ E-mail: cfcl@nplindia.org, वेबसाईट / Website: www.nplindia.org

अंशांकन प्रमाण पत्र
CALIBRATION CERTIFICATE
Voltage Transformer
Sr.
प्रमाणपत्र संख्या/Certificate No.
//N20090545/D2.02/C – 020

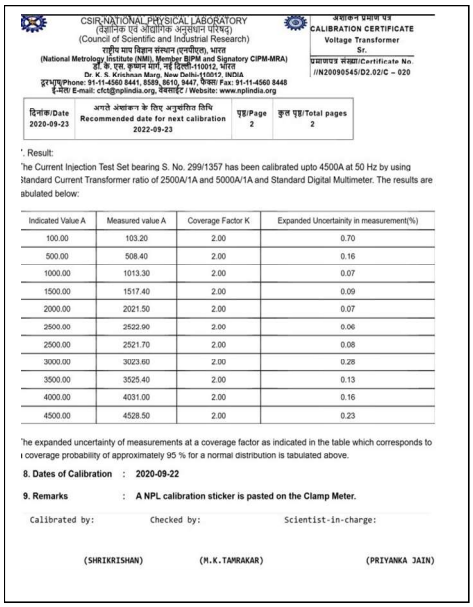
दिनांक/Date	अगले अंशांकन के लिए अनुशंसित तिथि Recommended date for next calibration	पृष्ठ/Page	कुल पृष्ठ/Total pages
2020-09-23	2022-09-23	1	2

1. Calibrated for	:	AC High Voltage & Current Metrology Section CSIR-National Physical Laboratory Dr. K.S. Krishnan Road, New Delhi - 110 012
2. Description & Identification of item under calibration	:	Current Injection Test Set Input : 415Volts, A.C. 1 Phase ; 50Hz, Output: 0 – 3 Volts/5000Amperes Or 0 - 6 Volts/2500amperes Or 0 - 12 Volts/1250Amperes Or 0 - 415 Volts/60Amperes S.NO. : 299/1357 Mfd. in India by : Automatic Electric Ltd., Mumbai
3. Environmental Conditions	:	(i)Temperature : (25 □ 2)□C (ii)Relative Humidity: (50 □ 10)%
4. Standards used Associated uncertainty	:	(i)Standard Current Comparator: □0.0025% (ii)Transfer Standard Digital Multimeter: □ 0.02% - □ 0.07%
5. Traceability of Standards used	:	The standards used for calibration are traceable to National Standards
6. Principle/Methodology of calibration & Calibration Procedure number	:	The calibration was carried out by comparison method using the Standard Current Transformer along with Standard Digital Multimeter. Calibration Procedure No. Sub-Div.#D2.02/DOC.3/CP#04.

Calibrated by: Checked by: Scientist-in-charge:

(SHRIKRISHAN) (M.K.TAMRAKAR) (PRIYANKA JAIN)

Fig. 6(a). First page of DCC



CSIR-NATIONAL PHYSICAL LABORATORY
(वैज्ञानिक एवं औद्योगिक अनुसंधान परिषद्)
(Council of Scientific and Industrial Research)
राष्ट्रीय माप विज्ञान संस्थान (एनपीएल), भारत
(National Metrology Institute (NMI), Member BIPM and Signatory CIPM-MRA)
डॉ. के. एस. कृष्णन मार्ग, नई दिल्ली-110012, भारत
Dr. K. S. Krishnan Marg, New Delhi-110012, INDIA
दूरभाष/Phone: 91-11-4560 8441, 8589, 8610, 9447, फैक्स/ Fax: 91-11-4560 8448
ई-मेल/ E-mail: cfcl@nplindia.org, वेबसाईट / Website: www.nplindia.org

अंशांकन प्रमाण पत्र
CALIBRATION CERTIFICATE
Voltage Transformer
Sr.
प्रमाणपत्र संख्या/Certificate No.
//N20090545/D2.02/C – 020

दिनांक/Date	अगले अंशांकन के लिए अनुशंसित तिथि Recommended date for next calibration	पृष्ठ/Page	कुल पृष्ठ/Total pages
2020-09-23	2022-09-23	2	2

', Result:
'he Current Injection Test Set bearing S. No. 299/1357 has been calibrated upto 4500A at 50 Hz by using Standard Current Transformer ratio of 2500A/1A and 5000A/1A and Standard Digital Multimeter. The results are abulated below:

Indicated Value A	Measured value A	Coverage Factor K	Expanded Uncertainty in measurement(%)
100.00	103.20	2.00	0.70
500.00	508.40	2.00	0.16
1000.00	1013.30	2.00	0.07
1500.00	1517.40	2.00	0.09
2000.00	2021.50	2.00	0.07
2500.00	2522.90	2.00	0.06
2500.00	2521.70	2.00	0.08
3000.00	3023.60	2.00	0.28
3500.00	3525.40	2.00	0.13
4000.00	4031.00	2.00	0.16
4500.00	4528.50	2.00	0.23

'he expanded uncertainty of measurements at a coverage factor as indicated in the table which corresponds to coverage probability of approximately 95 % for a normal distribution is tabulated above.

8. Dates of Calibration : 2020-09-22

9. Remarks : A NPL calibration sticker is pasted on the Clamp Meter.

Calibrated by: Checked by: Scientist-in-charge:

(SHRIKRISHAN) (M.K.TAMRAKAR) (PRIYANKA JAIN)

Fig. 6(b). Second page of the DCC



Fig. 6(c): QR code for the DCC

VI. CONCLUSION

The Digital Calibration Certificate (DCC) platform is a milestone in the process of modernizing calibration reporting at CSIR-NPL. The transition away from manual, paper-based documentation to a secure, automated web-based platform is a major improvement in obtaining more reliable digitized data, as well as improving efficiency and accessibility. Implementing cryptographic hashing and PDF protection on behalf of the DCC, users significantly enhance the integrity and security of the documents. Furthermore, leverages QR code technology that provides end-users with an efficient and expedient mechanism for authenticating and verifying certificates.

VII. ACKNOWLEDGMENT

The authors are thankful to CSIR NPL for providing all the support and PTB Germany for the technical support under Capacity Building Using regional exchange of knowledge (CABUREK) program.

REFERENCES

[1] Hackel, S. Gustav, Härtig, Frank, Hornig, Julia, and W. Thomas, “The digital calibration certificate”, PTB-Mitteilungen Vol. 127 2017, vol. Issue 4, p. 7, 2017, https://doi.org/10.7795/310.20170403

[2] “Digital Calibration Certificate – DCC”, Apr. 14, 2021. https://www.ptb.de/cms/en/researchdevelopment/into-the-future-with-metrology/the-challengesof-digital-transformation/kernziel1einheitlichkeitim/digitalcalibration-certificate-dcc.html

[3] Xingchuang Xiong, Zilong Liu, Kan Kan, Yiwei Zhu, Wei Zhang, Xiang Fang, “Design and implementation of a digital calibration certificate web service system based on microservice architecture”, Measurement: Sensors, vol. 38, 2025, https://doi.org/10.1016/j.measen.2024.101487

[4] T. Neubauer, E. Weippl and S. Biffl, "Digital signatures with familiar appearance for e-government documents: authentic PDF," First International Conference on Availability, Reliability and Security (ARES'06), Vienna, Austria, 2006, pp. 8 pp.-731, https://doi.org/10.1109/ARES.2006.54

[5] R. Forgáč and M. Očkay, "Contribution to Symmetric Cryptography by Convolutional Neural Networks," 2019 Communication and Information Technologies (KIT), Vysoke Tatry, Slovakia, 2019, pp. 1-6, https://doi.org/10.23919/KIT.2019.8883490

[6] A. Sharif, D. S. Ginting and A. D. Dias, "Securing the Integrity of PDF Files using RSA Digital Signature and SHA-3 Hash Function," 2021 International Conference on Data Science, Artificial Intelligence, and Business Analytics (DATABIA), Medan, Indonesia, 2021, pp. 154-159, https://doi.org/10.1109/DATABIA53375.2021.9650121

[7] K. Raghuvanshi, S. Kumar, S. Kumar, S. Kumar, "Image encryption algorithm based on DNA encoding and CNN, Expert Systems with Applications", vol. 252, Part B, 2024, https://doi.org/10.1016/j.eswa.2024.124287

[8] H. Brom, L. Jol, G. Rietveld, E. So, "High-Current AC Current Transformer Calibration Using an Automated Sampling System", Conference on Precision Electromagnetic Measurements, July 2012, https://doi.org/10.1109/CPEM.2012.6250697

[9] R. Campos, J. Carlos, Rico-Chagollán, Marianab, Chacón-Olivares, María del Carmenc and Guzmán-Hernández, Manuel Alejandro, "Frontend and Backend: The new approach to the development of a Web platform for automating the control and administration of degree processes at ITESI", Journal of Computational Technologies, vol. 8, no. 21, pp. 1-9, 2024, https://doi.org/10.35429/JOCT.2024.8.21.2.9.

[10] A. Castiglione, A. Santis, C. Soriente, "Security and privacy issues in the Portable Document Format", *Journal of Systems and Software,* vol. 83, no. 10, 2010, pp. 1813-1822, doi: https://doi.org/10.1016/j.jss.2010.04.062

[11] B. Khan, R. Olanrewaju, M. Morshidi, R. Mir, Laiha, B. Kiah, "Evolution and analysis of secured hash algorithm (SHA) family", Malaysian Journal of Computer Science, Vol. 35 (3), 2022, pp. 179-200.

[12] S. Patra, M. Rani, "Evaluation and categorization of hashing algorithms based on their applications", vol. 55, no. 3, pp. 540-552, 2025.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

The Digital Reference Material Document: From Paper Certificates to Interoperable Data Objects in Digital Quality Infrastructure

Matthias Prellwitz, Claudia Koch, Silke Richter, Johannes van de Kreeke, Michael Melzer, Mehran Monavari*

Bundesanstalt für Materialforschung und -prüfung (BAM), Unter den Eichen 87, 12205 Berlin, Germany, {firstname}.{lastname}@bam.de

***Abstract* – Reference materials (RMs) are essential for traceable and reliable measurements in science and industry, yet their certificate/document remain largely paper-based. The Digital Reference Material Document (DRMD) project at BAM introduces a transformative approach by converting traditional RM certificates/documents into machine-interpretable, XML-based digital assets. Building on the Digital Calibration Certificate framework, DRMDs encode ISO 33401 requirements and integrate semantic standards like D-SI and material identifiers. These digital documents support automated data exchange, integration into laboratory systems, and interoperability e.g. via asset administration shells and data spaces. The paper presents the DRMD schema concept, and outlines the path toward international harmonization and large-scale deployment, positioning DRMDs as a cornerstone of a digital quality infrastructure.**

I. INTRODUCTION

Reference materials (RM) are indispensable when it comes to ensuring the accuracy and reliability of measurement results and thus generating trust in analytical and other quality-assuring services. Inaccurate measurement values or those that are misinterpreted or incorrectly processed or transferred can lead to wrong decisions and thus to considerable follow-up costs or even hazards. Certified RMs ensure the traceability of measurement results to the SI [3], national measurement standards or recognized reference values and are frequently used for calibration and method validation, and the determination of measurement uncertainties. For accredited testing and calibration laboratories, the use of (certified) RMs for ensuring traceability of the results is considered essential as per ISO 17025.

Certified RMs come with a certificate that provides the value of the specified property, its associated uncertainty, and a statement of metrological traceability, and represent a central element of the daily routine of many laboratories for quality assurance. Non-certified RMs come with a so-called RM document, which does not necessarily provide a traceability statement [16]. Both RM certificates and documents come in paper or PDF format, the handling of the information is causing manual efforts, e.g. for data entry and processing. Providing these certificates and documents in a machine-readable way can contribute to significantly optimized process efficiency and accuracy.

The Digital Reference Material Document (DRMD) [18] transforms the traditional certificate and document defined in ISO 33401 into machine-interpretable XML objects. DRMDs function as digital assets that may integrate into laboratory information systems and populate digital twins via asset administration shell [2].

This paper shows how the Digital Reference Material Document fits within the emerging digital quality infrastructure, and supplements human-readable PDF certificates with a machine-interpretable asset. It presents the v0.2.0 XML schema, and open-source tooling and plans for integration into RM databases such as COMAR [17], then shows how DRMDs plug into digital twins and dataspace exchanges. We close by outlining the path to international harmonisation and large-scale deployment.

II. BACKGROUND

A. *Digital Quality Infrastructure*

The system of quality assurance and infrastructure that has been practiced for decades is still largely document- and paper-based and involves a great deal of manual bureaucratic effort. The central actors of the quality infrastructure (QI) in Germany have established the QI-Digital initiative [21] to jointly develop solutions to transform the system holistically into a data-based quality assurance system that meets the requirements of an increasingly digital economy. Among the variety of quality related documents that will be transformed into an appropriate digital representation are e.g. calibration or conformity certificates, test reports – and the RM certificate/document. Their efficient integration into calibration and quality assurance processes requires standardization regarding semantics and ontologies, as well as solutions for trustworthy verification [9].

*Corresponding author

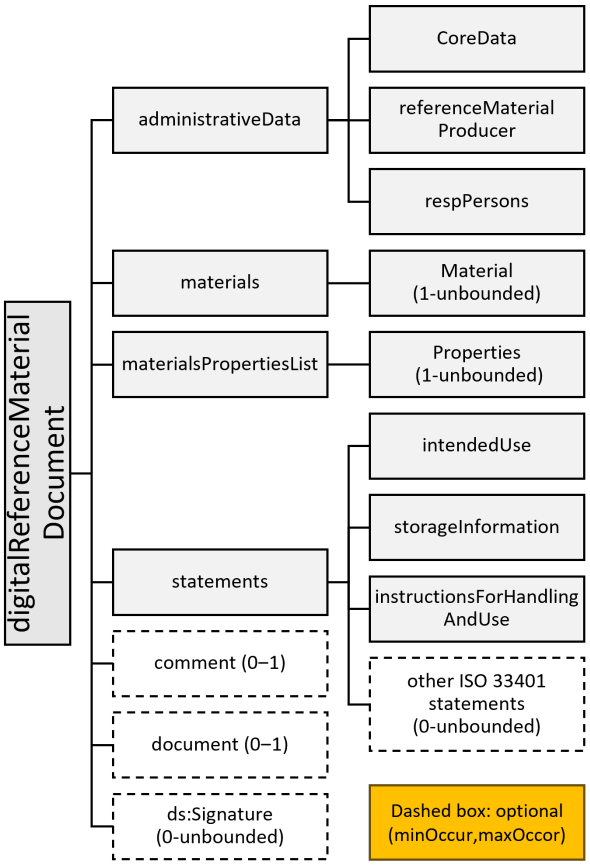


Fig. 1. Snapshot of the DRMD schema structure [18]

B. (Digital) Reference Material Documents

Today, (analogue) RM certificates and documents are usually created manually by the manufacturers, and then published or distributed with the reference material as a PDF copy to the customer, who then enters the relevant data manually into the instrument software.

In the context of the digital transformation of reference materials, the National Research Council Canada (NRC) as well as the National Institute of Standards and Technology (NIST, U.S.) are modelling databases that store the information contained in the RM certificates and documents of their institute [5, 19]. The resulting data architecture accommodates a broad spectrum of information – including quantified property values, multipart and serialised materials, various dependencies, and explicit traceability statements – thereby providing a machine-interpretable foundation for reference material metadata [4]. In Germany, the Federal Institute of Material Research and Testing (Bundesanstalt für Materialforschung und -prüfung, BAM) is driving the development of the Digital Reference Material Document (DRMD).

The DRMD encodes all ISO 33401 certificate requirement – property values, uncertainties, and metrological-traceability statements – in a machine-readable structure that can be ingested directly by laboratory software and other digital quality infrastructure services, eliminating manual re-keying and transcription errors. DRMDs offer significant process improvements. They will allow a laboratory that buys a RM to validate a procedure or to calibrate its device with it – or to calibrate many devices in a routine laboratory – to load the DRMD into the device software, making all RM information (analytes, content, minimum sample size, existing accreditation, etc.) available in machine-readable form. This way, the calibration/analysis can be carried out automatically - without a technician having to type all the contents of the analytes in the sample into the software by hand.

Thus, standardized data formats simplify the exchange of information between systems, make quality control more robust lowers administrative costs, and minimises error prone procedures.

C. Technical Implementation

The DRMD will be based on an XML schema built on the successful Digital Calibration Certificate (DCC) [9, 13, 18] as developed by PTB and others. The current DRMD schema introduces six clearly separated containers (cf. Fig. 1):

- **administrativeData**: includes core metadata: unique identifiers, period of validity and RM producers, etc.
- **materials**: list of material objects; each material embeds description, minimum sample size and *materialClass* tags that may reference external taxonomies.
- **materialsPropertiesList**: includes measurement results, which uses the incorporated D-SI XML scheme (Digital System of Units) to express property values and their uncertainties and *unitsProperties* can be linked to existing taxonomies.
- **statements**: intended use, storage instructions, commutability and other ISO 33401 clauses.
- **comment**: optional free remarks
- **document**: optional binary documents
- **signature**: multiple XML-DSig blocks for integrity and accreditation proofs.

The DRMD toolkit contains (i) the XSD, (ii) XSLT style sheets that render human-readable HTML, (iii) validated XML examples for best practices. The following tools are currently in development to ease development and usage of a proposed DRMD scheme:

- To accelerate market diffusion and adoption of DRMDs, BAM has developed a prototype for an online tool that allows producers of RMs to create their own schema conforming DRMDs.

- An AI powered generator tool to accelerate the transformation of existing PDF certificates into DRMDs is under development. The prototype for this LLM-based generation is showing initial success and will help generating DRMDs for inclusion into RM databases .

BAM hosts an international database that RM producers can use to list their materials (currently more than 2000). Several DRMDs have already been uploaded to COMAR to demonstrate their feasibility. Integrating DRMDs broadly in this database, supported by the abovementioned tools, will accelerate their diffusion.

III. DIGITAL ECOSYSTEM FOR DRMD

The transformation from analogue to digital processes in QI goes far beyond the mere digitization of existing processes. While manual work, paper-based documentation and siloed data storage are still commonplace today, QI needs to evolve into a networked system with automated processes and seamless digital documentation. This integration not only enables more efficient and secure data management, but also creates the basis for anticipatory, use-oriented quality assurance processes.

To leverage the potential of a digital QI, its digital tools, and procedures as well as relevant actors need to be seamlessly integrated into an interoperable, cross-system and coherent digital ecosystem. As the complexity of distributed systems is determined by its interfaces, latest developments towards industrial international data spaces [6] as well as the standardized documentation and access of data, e.g. with the help of the Asset Administration Shell [2], enable such an ecosystem.

For quality documents such as the DRMD these developments represent technical foundations for an efficient and trustworthy information modelling and sharing.

A. DRMD and Asset Administration Shell

The Asset Administration Shell (AAS) was developed as part of the German *Industrie 4.0* [12] initiative to create a standardized digital representation of physical assets. The motivation behind AAS is to enable interoperability among various components in industrial systems, facilitating seamless integration and communication between physical and digital entities.

The AAS acts as a digital twin of physical assets, providing a comprehensive digital model that includes all relevant information and functionalities as well as unified access to it. Thereby, it supports the entire lifecycle of an asset, from creation to disposal, by continuously updating information. Its modular structure consists of submodels that describe specific aspects of an asset, such as energy consumption, maintenance schedules, and documentation. Implementing the AAS offers several advantages: It enhances efficiency by providing precise insights for monitoring and control, and thus enables well-informed decision-making. The AAS optimizes processes by replacing heterogeneous data landscapes with unified, interoperable data structures, increasing transparency and operational efficiency. The International Digital Twin Association (IDTA) provides comprehensive guidelines, submodel templates, and specifications that ensure interoperability and consistency across digital twins. By promoting standardized data structures and semantic definitions, IDTA facilitates seamless integration and communication between various industrial systems, enhancing the efficiency and effectiveness of digital transformation initiatives.

As data schemes display partial overlap of specific data elements, a modular approach is selected for the development of AAS submodel templates. Generic data points are subject to harmonization and transferred from namespaces, with the objective of facilitating their modular application across diverse instances. Synergies may evolve from e.g., general and administrative data, the representation of measurement results, and SI units. Existing AAS submodel templates [14, 15] are also evaluated for the purpose of reusing parts. The DCC data schema is the most mature submodel template available and therefore serves as foundation for further digital QI developments, such as the DRMD and the D-CoC [22]. These will be progressively integrated with the aim of establishing a modular D-X scheme that accommodates multiple QI certificates and documents.

Ultimately, these AAS sub-modeled quality documents could easily be used and managed as information basis for digital product passports or other documents along product life cycles [8].

B. DRMD and Data Spaces

Data spaces [20] are collaborative ecosystems designed to facilitate secure and controlled data sharing among various stakeholders. Originating as an alternative to restrictive bilateral data sharing agreements, data spaces aim to retain data with their owner while granting controlled access to authorized parties. Key features include a secure infrastructure that protects data privacy and complies with regulations such as general data protection regulation (GDPR) [23]. They enable the integration of data from multiple sources based on agreed standards and technology building blocks, enabling interoperability. Robust governance processes are in place to ensure data accuracy, reliability, and trustworthiness. Additionally, data spaces allow data providers to set clear rules about how their data can be used, ensuring data sovereignty and trust. Data spaces are well-suited to foster cooperation among partners and unlock new data sources. Governance frameworks improve the quality and reliability of shared resources. Data spaces align with ethical principles by ensuring data sovereignty,

trust, and fair value exchange. They also support AI applications by providing the massive amounts of high-quality data needed, addressing immediate data needs, and supporting long-term AI development.

The incorporation of foundational components such as digital identities and credentials to conformity assessment entities, in conjunction with SMART standards in data formats that facilitate machine-executable comparison of process parameters with established norms, has the potential to enhance the efficiency of digital QI.

C. DRMD and Digital Trust

Ensuring the integrity of digital documents and data structures is essential to build trust equivalent to traditional paper-based formats. Electronic signatures achieve this by applying a hash function to the content, which is then signed using a verified digital identity. Asymmetric cryptographic techniques are key to safeguarding documents from tampering or unauthorized access, securing information during both transmission and storage. Legal recognition of electronic signatures in many jurisdictions grants them the same validity as handwritten signatures, supporting regulatory compliance and legal processes. Additionally, electronic signatures improve workflow efficiency by reducing the need for physical documents, saving time and lowering costs related to printing, shipping, and storage. They also support remote work and global collaboration through easy document access and signing. Various technical solutions have been developed to support digitally signed certificates.

Public Key Infrastructure (PKI) is a system that uses public and private cryptographic keys to create and verify digital signatures. Certificate Authorities (CAs) issue digital certificates to authenticate signers' identities. PKI ensures document integrity, authenticity, and non-repudiation. In Germany, the national accreditation body *DAkkS* offers a taylored PKI as digital accreditation symbol, i.e., an electronic company seal issued through D-Trust, providing a verified digital identity, exclusively for accredited bodies [7]. These electronic seals can be applied to PDF or XML files with the unique feature of digitally verifying the respective accreditation status of the issuing body on the receiver side of the document.

Verifiable credentials (VCs) [1] are open standards by the World Wide Web Consortium (W3C) which allow third-party verification of digital certificates using decentralized identifiers (DIDs) and cryptographic proofs. VCs are particularly useful for identity verification and access control. Within digital signature frameworks, verifiable presentations enable secure and private sharing of VCs. Privacy-preserving techniques, such as zero-knowledge proof, facilitate the selective disclosure of data, thereby ensuring that only the necessary attributes are revealed for verification purposes.

Trust anchors are foundational in both PKI and VCs, ensuring the credibility of certificates and credentials. In PKI, trust anchors are typically root certificates from trusted CAs. These validate the certificate chain but present challenges in management, particularly in issuance and revocation. In VCs, trust anchors are often DIDs, offering decentralized trust with stronger privacy—but also introducing complexity in management. Secure and recognized DID management is essential for broad VC adoption. The European Blockchain Services Infrastructure (EBSI) [11], an EU initiative, uses blockchain to strengthen digital trust and identity. It addresses challenges in trust anchors via verifiable data registries (VDR) and identity management through a decentralized, transparent, open-source framework that supports VCs and DIDs.

The European Business Wallet (EUBW) aims to simplify regulatory compliance and strengthens digital trust in business-to-business (B2B) and business-to-government (B2G) interactions by providing a unified identity and trust framework. It helps businesses manage identity verification across borders, reducing fraud and ensuring compliance with EU regulations. The implementation of trustworthy identity anchors and the credential issuing within public organisations is of particular importance, especially in the safety certifying testing, inspection and certification (TIC) domain.

In summary, it is imperative to employ digital trust in the context of conformity assessment documentation to ensure the secure exchange of documents. As is currently being discussed in industry and politics at the European level, open standards with flexible trust levels accredited by varying bodies are recommended for open ecosystems. These standards are crucial to supporting interoperable, scalable architectures to be adopted by stakeholders in industry and governmental institutions.

IV. CONCLUSION AND OUTLOOK

The drafting of a data scheme for the Digital Reference Material Document (DRMD) at BAM is progressing in alignment with the more mature Digital Calibration Certificate (DCC). The XML schema evaluates existing approaches to data structures and definitions, integrating generic work such as the Digital System of Units (D-SI). While data modeling is ongoing, the next challenges include ensuring data integrity, verifying the identity and credentials of the issuing body, and enabling sovereign sharing in open ecosystems. These challenges are similar to those faced by other digital conformity documents.

The current data structures will be part of a novel modular D-X schema [10]. To facilitate quick industrial adoption, we continue our collaboration within the IDTA Working Group "Digital Quality Documents" to co-create modular submodel templates for AAS. This collaboration is essential for developing a robust and adaptable framework.

Our efforts include testbed and prototyping activities with stakeholders to support market diffusion. Integration into digital QI process chains and digital ecosystems of the industry and international value chains is a priority. Additionally, international harmonization and cooperation with other national institutes are pivotal.

Previous work has shown promising results, with market maturity on the horizon and a large share of reference materials covered. The deployment of thousands of reference materials using the DRMD will lead to more efficient processes in laboratories and industry applications. Together with other digital QI tools, such as the digital calibration certificate and request, and embedded within the digital ecosystem (including AAS and data spaces), the DRMD holds significant potential. This alignment and further development will be pursued through close collaboration within the QI-Digital initiative and with international partners. The ongoing efforts aim to create a comprehensive and harmonized digital framework that supports the efficient and secure exchange of digital reference material documents, ultimately benefiting the broader industry and scientific communities.

V. ACKNOWLEDGEMENTS

QI Digital is an initiative of the central actors in German quality infrastructure (QI): BAM, DAkkS, DIN, DKE and PTB. The Federal Ministry of Economic Affairs and Energy (BMWE) recognizes the pivotal role of QI Digital in safeguarding the success of innovative technologies, products and processes.

REFERENCES

[1] Greg Bernstein et al. "Verifiable Credentials Overview". May 2025. URL: https://www.w3.org/TR/vc-overview/.

[2] Birgit Boss et al. "Verwaltungsschale". In: *Handbuch Industrie 4.0*. Springer Berlin Heidelberg, 2024, pp. 291–317. DOI: 10.1007/978-3-662-58528-3_139.

[3] Bureau International des Poids et Mesures. *Le Système international d'unités/The International System of Units [Brochure]*. 9th ed. (V3.01, updated August 2024). 2019. URL: https://www.bipm.org/en/publications/si-brochure.

[4] D. Camara and I. Dominguez Mendoza. "The Digital NIST: Update on the Digital Transformation of Reference Materials". en. In: *4th International DCC-Conference 2024-02-27 to 2024-02-29 Proceedings* (2024). DOI: 10.7795/810.20240612.

[5] William Dinis Camara et al. "The Digital NIST: Challenges And Opportunities in the Digital Transformation of NIST's Reference Materials". en. In: IMEKO TC6 International Conference on Metrology and Digital Transformation, Berlin, DE, Sept. 2022. URL: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=935320.

[6] Edward Curry, Simon Scerri, and Tuomo Tuikka. "Data Spaces: Design, Deployment, and Future Directions". In: *Data Spaces*. Springer International Publishing, 2022, pp. 1–17. DOI: 10.1007/978-3-030-98636-0_1.

[7] DAkkS. "The digital accreditation symbol". 2025. URL: https://www.dakks.de/en/the-digital-accreditation-symbol.html.

[8] Thomas Engel. "Digital Quality Documents and Certificates". In: *Handbook of Nondestructive Evaluation 4.0*. Springer Nature Switzerland, 2025, pp. 1–29. DOI: 10.1007/978-3-030-48200-8_86-1.

[9] Thomas Engel. "From a digital calibration certificate to a digital quality infrastructure". In: *EPJ Web of Conferences* 323 (2025). Ed. by J. Lopez, p. 01003. DOI: 10.1051/epjconf/202532301003.

[10] Thomas Engel and Alexandru Popescu. "GEMIMEG – Digitalization of calibration processes in metrology". In: *Measurement: Sensors* 38 (May 2025), p. 101474. DOI: 10.1016/j.measen.2024.101474.

[11] European Commission. "European Blockchain Services Infrastructure (EBSI)". (accessed July 15, 2025). 2025. URL: https://ec.europa.eu/digital-building-blocks/sites/display/EBSI/Home.

[12] Federal Ministry for Economic Affairs and Energy (BMWE). *Plattform Industrie 4.0*. (accessed July 15, 2025). 2025. URL: https://www.plattform-i40.de/IP/Navigation/EN/Home/home.html.

[13] Siegfried Gustav Hackel et al. "The Digital Calibration Certificate". en. In: *PTB-Mitteilungen. Volume 127 (2017)* Issue 4 (2017), Page 75 to 81. ISSN 0030–834X. DOI: 10.7795/310.20170403.

[14] Industrial Digital Twin Association (IDTA). *How to create a submodel template - Specification v1.1*. June 2025. URL: https://industrialdigitaltwin.org/wp-content/uploads/2025/06/IDTA_How-to-write-a-SMT-v1.1.pdf.

[15] International Digital Twin Association (IDTA). "AAS Submodel Templates". (accessed July 15, 2025). 2025. URL: https://industrialdigitaltwin.org/content-hub/teilmodelle.

[16] *International vocabulary of metrology – Basic and general concepts and associated terms (VIM)*. Joint Committee for Guides in Metrology (JCGM) member organizations. BIPM, IEC,IFCC, ILAC, ISO, IUPAC, IUPAP and OIML, 2008. DOI: 10.59161/JCGM200-2012.

[17] Bundesanstalt für Materialforschung und -prüfung (BAM). *Code d'Indexation des Matériaux de Référence (COMAR)*. (accessed July 15, 2025). 2025. URL: https://www.comar.org.

[18] Michael Melzer et al. "The Digital RM Certificate Derived from the DCC". In: *CCQM Online Workshop on Digital and FAIR Chemical and Biological Reference Data and Certificates:Challenges and Opportunities*. Sept. 2024. URL: https://opus4.kobv.de/opus4-bam/frontdoor/index/index/docId/61888.

[19] National Research Council Canada (NRC). "Canadian Reference Materials and Methods". (accessed July 15, 2025). 2025. URL: https://nrc-digital-repository.canada.ca/eng/home/collection/canadian-reference-materials-methods/.

[20] Boris Otto, Michael ten Hompel, and Stefan Wrobel, eds. *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage*. Springer International Publishing, 2022. DOI: 10.1007/978-3-030-93975-5.

[21] "QI-Digital initative". (accessed July 15, 2025). 2025. URL: https://www.qi-digital.de.

[22] Tatyana Sheveleva, Gisa Foyer, and Dorothea Knopf. "Conceptual harmonisation between D-CoC and DCC data structures representing result documents in conformity assessment". In: *Measurement: Sensors* 38 (May 2025), p. 101472. DOI: 10.1016/j.measen.2024.101472.

[23] Paul Voigt and Axel von dem Bussche. *The EU General Data Protection Regulation (GDPR): A Practical Guide*. Springer Nature Switzerland, 2024. DOI: 10.1007/978-3-031-62328-8.

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Enhancing digital twin reliability using FAIR principles and data quality assessment

Miguel Burg Demay[1], Luiz Eduardo de Farias[2], Gustavo Donatelli[3], Andre Luiz Meira de Oliveira[4]

[1] *Certi/Centro de Metrologia e Instrumentação, 88040-535, Florianópolis/Brazil, mbd@certi.org.br*
[2] *Certi/Centro de Metrologia e Instrumentação, 88040-535, Florianópolis/Brazil, lzr@certi.org.br*
[3] *Certi/Centro de Metrologia e Instrumentação, 88040-535, Florianópolis/Brazil, gd@certi.org.br*
[4] *Certi/Centro de Metrologia e Instrumentação, 88040-535, Florianópolis/Brazil, aeo@certi.org.br*

***Abstract* – The use of digital tools such as digital twins is spreading throughout the O&G sector. For integrity management, digital models of relevant asset degradation phenomena have been used to estimate and predict its health, aiming to improve maintenance planning and to provide information about the risk of failure and its evolution over time. The input data of models for O&G integrity monitoring are commonly found in different databases and present very different characteristics, such as sampling rate, temporal stability and variability, influencing the data quality in different ways. This work addresses the use of FAIR principles and data quality assessment for O&G asset integrity management. A brief review of established concepts is discussed, and a practical case study is presented, which illustrates the very important role that data quality assessment and the use of FAIR principles play in digital models' reliability.**

I. INTRODUCTION

The integrity management of subsea assets has traditionally depended on inspections performed by tools such as pigs and ROVs. While these methods provide accurate measurements at specific locations, the entire process, from planning to execution, is both costly and time-consuming, resulting in less frequent monitoring than would be ideal. Furthermore, assets design limitations can prevent inspection tools from accessing the most critical areas, which need the most attention.

Digital solutions provide a powerful alternative to overcome these inherent limitations. They involve creating digital models that simulate the key degradation processes impacting subsea assets. When integrated into a digital twin, these models enable continuous integrity monitoring. By analyzing real-time production data and fluid properties, the models can evaluate an asset's health and its future degradation at a significantly higher frequency and lower cost than traditional methods, though with greater uncertainty [1][2].

The reliability of an asset degradation model's estimates or predictions relies heavily on the quality of the input data. In the oil and gas industry, the data used for these models come from various sources and have distinct characteristics, such as sampling rate and temporal variability. In addition, gaps often appear on the data due to sensor malfunctions and variables with different sampling rates within the same dataset.

This paper explores how the application of FAIR principles and the evaluation of data quality can be beneficial in this context. Section II introduces the concept of a digital twin for oil and gas asset integrity management. Sections III and IV discuss the FAIR principles for measurement data and data quality assessment. Sections V and VI illustrate how the FAIR principles and data quality assessment can be applied for a digital twin for O&G integrity management. Finally, Section VII presents the main conclusions of this work.

II. DIGITAL TWIN FOR SUBSEA ASSET INTEGRITY MONITORING

A digital twin (DT) can be described as a set of digital models representing a physical asset, designed to simulate and predict its current and future states and behaviors, while enabling control through feedback to the physical asset. Data and digital models form the foundation of this concept. [3][4]

As shown in Figure 1, data and information serve as the link between the physical and digital worlds. In the physical domain, materials, machines, operators, and the environment are key elements in the production process.

Otherwise, in the digital domain, various models representing different aspects of the production process support accurate predictions and informed decision-making. These models not only describe asset behaviors, but also utilize historical and real-time data, as well as human experience and knowledge, to predict the future conditions of assets [5]

To ensure good performance, synchronized behavior is expected from both physical and digital twins. Through simulations, the digital model optimizes operations and improves the performance of the physical process by offering feedback. This ongoing enhancement requires the co-evolution of both the physical asset and its digital twin [6][7][8].



Fig. 1. Physical and digital interaction in Digital Twin.

Developing a digital twin management system involves integrating and evaluating the quality of data used by the degradation models associated with the asset's function. The digital twin helps assess the asset's health by monitoring key parameters for each degradation model, allowing for the identification of potential failure modes and supporting risk-based decision-making [2].

In the context of subsea assets, erosion is one of the most important degradation phenomena. It impacts the internal surfaces of pipelines, gradually reducing wall thickness and shortening their useful life. Several digital models have been developed to estimate the loss of pipeline wall thickness due to erosion, with one of the most widely used being the model presented in DNV-O501[9].

The DNV-O501 model relies on over 30 input parameters, sourced from multiple databases which use varying data formats and sampling rates. These data differ in characteristics such as temporal stability and variability, each influencing the model's output in distinct ways.

In this complex scenario, applying FAIR principles and assessing data quality could enhance the reliability of digital model outputs, as well as improve the accuracy of digital twin estimates, predictions, and recommendations.

III. FAIR PRINCIPLES

The FAIR principles are a set of guidelines intended to support the development of data management systems for data, metadata, and infrastructure. These principles aim to enhance data management and prepare data for sharing and reuse across stakeholders from different domains, disciplines, or organizations by ensuring that data is Findable, Accessible, Interoperable, and Reusable [9][10][11].

A findable dataset means that data (and metadata) should be easily discoverable by humans and machines. Machine-readable metadata are essential for the automatic discovery of datasets and services. For that, data should be described with comprehensive metadata, and both should be assigned to a globally unique and persistent identifier.

An accessible dataset is one where users know precisely how to access the data, including any necessary authentication and authorization. For that, data (and metadata) should be retrievable by means of an open, free and standardized communication protocol.

Interoperability refers to the requirement that data must be compatible and integrated with other data. Additionally, the data must be able to interact with applications or workflows for analysis, storage and processing. For that, data and metadata should use a formal, accessible, shared and widely applicable language for knowledge representation.

Data reuse aims to maximize the value of data by enabling its reuse. To accomplish it, data and metadata should be well-described to ensure they can be reliably replicated and/or integrated across different contexts.

By following these four principles, datasets become accessible to a broader range of data users, allowing them to easily find, access, understand, and automatically process the data.

The interest in aligning with the FAIR principles has been increasing since they ease data integration, which is essential for linking and combining data across various fields. Adhering to the FAIR enhances the value of data, by making it more discoverable through unique identifiers and more easily integrable through standardized and shared knowledge representations [11][12].

IV. DATA QUALITY

Data is essential in the industry for monitoring, controlling and optimizing production processes, as well as for conducting analyses and simulations that support decision-making. To ensure the results meet the necessary levels of reliability, the data and models must meet distinct quality criteria.

Data quality has been gaining increasing relevance in the oil and gas industry, being addressed in at least two recommendations, DNV-RP-0497 and DNV-RP-A204.

Data quality can be understood as the degree to which the data is suitable for the analysis that is intended to be performed. Therefore, the quality of data cannot be assessed without considering the data's intended application.

When it comes to digital twins, which rely on simulations and data analysis, the quality of input data is crucial for obtaining useful and reliable results. To build trust with the user and contribute to the company's business, a digital twin must operate with valid and accurate data, along with representative models. High data quality must be ensured to enable data reuse and the analysis of historical data.

To measure data quality, different metrics can be used, often grouped into dimensions of data quality. These dimensions can also be found in ISO 8000-8, which describes data and information quality concepts and their importance for quality management processes. Among the various existing metrics, interoperability and accessibility stand out, along with [13][14][15]:

- Accuracy: a measure of how much data represent the real-world values
- Completeness: the extent to which data is available.
- Consistency: related to the uniformity of data across different databases and systems.

- Timeliness: a measure of how much data is up-to-date and available

V. FAIR PRINCIPLES AND DATA QUALITY FOR O&G DIGITAL TWINS

Digital twins for integrity management in the oil and gas industry usually have several digital models for estimating asset degradation, which consume data of different types and characteristics, scattered across various databases.

Production data, like pressure and temperature, are acquired by sensors, with a sampling rate of the order of tens of hertz. Fluid characteristic data, such as density and viscosity, are obtained through laboratory tests and are collected monthly, bimonthly, or semi-annually. Asset specification data, such as its dimensions and materials, remain constant over time.

To ensure that data adhere to the FAIR principles, one approach is to build an unified database, from which all digital models can draw the data they need. This database shall mean the only digital truth of the physical twin.

Nonetheless, data read from different databases, formatted, and stored in a single database may present some gaps due to the different sampling frequencies and invalid values. Therefore, issues related to the validity of the data in this dataset must be addressed, to ensure that the models always have valid data to operate.

This digital twin's unified database should include metadata, which contain guiding information for data conditioning, such as:

- Type of data: original, conditioned, or synthetic
- Measurement range of each variable
- Physical limits of each variable
- Data quality indicators

The conditioning process involves removing invalid data and filling in gaps, ensuring alignment with the FAIR principles. To realize it, gaps can be filled, for example, with the most recent valid data. The same approach can be applied to data that are not realistic, such as negative pressures or excessively high temperatures.

Thus, conditioning represents a decrease in data quality, as it involves inserting a value based on the best available knowledge, which may not reflect reality. Therefore, each conditioning operation reduces data quality, but it is better than using invalid values.

Additionally, when data is replaced by the most recent valid value, larger gaps are filled with the same value. As a result, the data quality diminishes over time or with the number of sequential replacements, since the time between the replaced data and the most recent valid value increases, thereby reducing its representativeness.

In this sense, it is very important to understand the deep relationship between data quality and uncertainty, and model sensitivity.

The output of a digital model contains uncertainties that arise from the model input data and the modelling process. The former, called aleatoric uncertainty, is related to measurement uncertainties, process and fluid variations, which compose the natural variability of input data and should not impact its data quality. However, outliers and unknown variations should be considered when assessing data quality.

The uncertainty related to the modelling process is named epistemic uncertainty. It is associated with the completeness of the knowledge represented by the model. In other words, it is related to the representativeness of the model in comparison to the real behavior of a degradation phenomenon. [16][17][18][19]22].

This kind of uncertainty is associated with input data quality through the sensitivity of each model parameter. Since each variable impacts differently the model outputs, the quality of input data influences the quality of output data according to its sensitivity [18][19].

VI. DATA QUALITY INDICATOR EVALUATION

Based on these premises, a quality indicator value can be defined for each data in the dataset. For each input data of the DNV-O501 erosion model considered in this paper, a quality indicator can be assigned, which is a measure of reliability of the data used as input in erosion models for subsea equipment.

Data quality indicator (DQI) can be set ranging from 0 to 1, where 0 represents a complete lack of quality and 1 represents the highest quality. To evaluate data quality indicators, four parameters have been considered:

- Data validity: a measure of the validity of each input data, which is decreased by the presence of null or invalid values. For each day, the data quality is calculated as the average of the daily valid records. If all the records are valid, the quality of the input data is 1; otherwise, the quality is reduced proportionally to the number of invalid records.
- Physically possible values: it checks if the data falls within a range of physically possible values for each variable. For example, the temperature of an underwater equipment cannot be lower than 0 °or higher than 100 °C. If the data is within the range, the quality is 1; otherwise, it is 0.
- Coefficient of variation: it evaluates the variation of the data over time. If it is greater than the historical median or equal to zero, the quality is reduced. This helps identify fluctuating or inconsistent data.
- Timeliness: it is also possible to include a quality degradation rate, which reduces the quality of the data over time if it is not corrected or updated. For example, if a sensor fails and is not repaired, the quality of the data recorded by that sensor will gradually decrease until it reaches a minimum value. This approach ensures that the data quality reflects not only its consistency at the time of collection but also its reliability over time.

Based on these parameters, each model's input data can have its quality quantitatively assessed. However, for management purposes, a qualitative approach can be interesting. For that, the classification system presented in table 1 can be used.

Table 1. Qualitative Data Quality Indicators

Quality class	Condition
High quality data	DQI ≥ 0.85
Intermediate quality data	0.5 ≤ DQI < 0.85
Low quality data	DQI <0 .5

Based on these definitions, a dataset of an O&G asset was assessed for degradation caused by erosion. For a 10-year time window, the DNV-O501 erosion model data inputs were recorded, prepared and stored in a database, as well as its DQI as metadata.

Figure 2 illustrates a DQI of a X-tree pressure signal over time. Daily values were obtained and represented in the graph, which shows how the quality of the pressure signal can vary over the years.

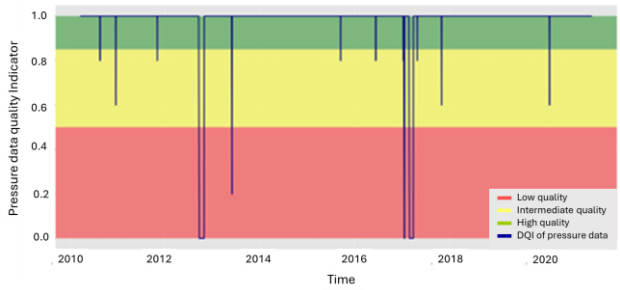


Fig. 2. Data Quality of pressure data

For each input variable, its DQI is evaluated through a weighted average of the daily quality assessments, where the weight of each rating (high, intermediate or low) corresponds to its frequency during the analyzed period.

Then, for each input data, a DQI could be evaluated and stored. Among all input variables of the DNV erosion model, nine relevant DQIs are presented in table 2.

Table 2. Data Quality of DNV-O501 model input data

Input data	DQI	Quality class
Oil rate	0.72	Intermediate quality data
Pressure	0.97	High quality data
Temperature	0.61	Intermediate quality data
GOR	0.72	Intermediate quality data
BSW	0.72	Intermediate quality data
Oil density	0.50	Intermediate quality data
Oil viscosity	0.50	Intermediate quality data
Gas viscosity	0.39	Low quality data
Sand content	0.50	Intermediate quality data

According to table 2, pressure has a high data quality; gas viscosity presents low data quality, and all the others have an intermediated data quality.

Based on this procedure, the quality of each model input data can be assessed. While this is important for asset's integrity management, the quality of the estimated erosion, the model's output, based on these input data is very relevant for improving the reliability of decision-making based on erosion estimates.

During simulation, all input data are combined throughout the model and influence its output in different ways. Therefore, the DQI of each input data affects the DQI of the model's output according to the associated sensitivity.

Thus, it is demanded to know the sensitivity of each model's input data to evaluate the DQI of its output. It is performed by sensitivity analysis, which assesses the impact of each variable on the outcome of the erosion model. Variables with high sensitivity have a significant effect on the model's estimates, while variables with low sensitivity have a lesser impact.

Sensitivity can be estimated using advanced methodologies such as Shapley or GSA, which quantify the contribution of each variable to the model's uncertainty regarding model's nonlinearities and interactions. This analysis is essential for prioritizing which variables require closer assessment and monitoring, particularly when complemented by qualitative evaluation. [19][20][21].

Using Shapley, the sensitivity parameters (SP) presented in table 3 are obtained. As shown, the oil rate and the pressure are the most relevant input data, followed by the temperature, the sand content, the gas-oil ratio and the BSW. Other variables are not relevant.

Table 3. Sensitivity parameters

Input data	SP
Oil rate	28.97
Pressure	28.01
Temperature	13.04
GOR	4.24
BSW	0.97
Oil density	0.00
Oil viscosity	0.00
Gas viscosity	0.00
Sand content	6.25

By combining the sensitivity parameter with the DQI of each model's input data according to equation 1, it is possible to evaluate a DQI of the model's output.

$$DQI_M = \frac{\sum_{i=1}^{N} DQI_i \cdot SP_i}{\sum_{i=1}^{N} SP_i} \quad (1)$$

where:

- DQI_M: data quality indicator of model's output
- DQI_i: data quality indicator for each model's input data "i"
- N: number of model's input data
- SP_i: sensitivity parameter of the input parameter "i".

Thus, evaluating a DQI of each model's input data and a SP for model parameter, equation 1 allows the evaluation of data quality of the model output, as shown in table 4.

Table 4. Data Quality of DNV-O501 model output

Erosion estimate:	3.28 mm
Data quality indicator	0.77
Data quality class:	Intermediate quality data

The advantages of evaluating the model´s output DQI are associated with providing information about the reliability of the model's estimate. It is a measure of how trustworthy the model output is, as well as the decision-making based on it.

Moreover, it can be used to monitor model reliability over time, and to provide insights about how to improve model estimates assertiveness by means of acting on bad quality data, or its source.

VII. Conclusion

This work explores the application of FAIR principles and data quality assessment within the context of digital twins. By utilizing an unique database where all data is conditioned and ready for use by each model, it becomes possible to assess the quality of data inputs for each model.

Additionally, with the help of sensitivity parameters evaluated through methods like Shapley, the quality of the model's outputs can also be assessed. The ability to evaluate data quality indicators for both inputs and outputs enhances the amount of reliable information, improving the accuracy and trustworthiness of the model's estimates and predictions.

Ultimately, this leads to greater reliability, higher assertiveness and reduced uncertainty in decision-making, making maintenance planning more reliable and effective.

References

[1] G. Pauli, M. B. Demay, A. M. da Mata, S. S. Rodrigues, J. M. Xavier, J. de O. Braga, G. D. Donatelli, E. Margotti. "On the Impact of Temporal Resolution on Nonlinear Model Accuracy for Predicting Wear Due to Solid Particle Erosion in Digital Twins of Oil and Gas Equipment." Paper presented at the Offshore Technology Conference Brasil 2023 (OTC Brazil 2023). https://doi.org/10.4043/32797-MS

[2] S. S. Rodrigues, J. de O. Braga, J. M. Xavier, G. Pauli, A. M. da Mata, E. Margotti, M. B. Demay, G. D. Donatelli. "Digital Twin of Subsea Assets as a Tool for Integrity Management and Risk-Based Inspection: Challenges and Perspectives." Paper presented at the Offshore Technology Conference Brasil 2023 (OTC Brazil 2023). https://doi.org/10.4043/32757-MS

[3] J. D. Hochhalter, W. P. Leser, J. A. Newman, E. H. Glaessgen, V. K. Gupta, V. Yamakov, S. R. Cornell, S.A. Willard, G. Heber. "Coupling Damage-Sensing Particles to the Digital Twin Concept." NASA/TM–2014-218257. 2014.

[4] F. Tao, Q. Qi, L. Wang, A.Y.C. Nee. "Digital Twins and Cyber–Physical Systems toward Smart Manufacturing and Industry 4.0: Correlation and Comparison", Engineering, v. 5, 2019, p-p 653-661, https://doi.org/10.1016/j.eng.2019.01.014.

[5] Q. Qi., F. Tao. "Digital Twin and Big Data towards Smart Manufacturing and Industry 4.0: 360 Degree Comparison". IEEE Access 6 (2018): 3585–93. https://doi.org/10.1109/access.2018.2793265.

[6] E. Glaessgen, D. Stargel, "The Digital Twin Paradigm for Future NASA and Air U.S. Force Vehicles." 2012. https://doi.org/10.2514/6.2012-1818.

[7] Tao, F. 2018. "Digital twin-driven product design, manufacturing and service with big data". Int. J Adv Manuf. Technol. 94: 3563–3576. https://doi.org/10.1007/s00170-017-0233-1.

[8] DNV. "DNVGL-RP-O501: Managing sand production and erosion". 2015.

[9] J. Top, S. Janssen, H. Boogaard, R. Knapen, G. Simsek-Senel, "Cultivating FAIR principles for agri-food data". Comput. Electron. Agric. 196. 2022. https://doi.org/ 10.1016/j.compag.2022.106909.

[10] N. N. K. Krisnawijaya, B. Tekinerdogan, C. Catal, R.van der Tol, Y. Herdiyeni, "Implementing FAIR principles in data management systems: A multi-case study in precision farming", Computers and Electronics in Agriculture, v. 230, 2025, https://doi.org/10.1016/j.compag.2024.109855.

[11] Gofair. "FAIR Principles". https://www.go-fair.org/fair-principles/. 2025.

[12] A. N. de la Hidalga, J. Goodall, C. Anyika, B. Matthews, C. Richard A. Catlow. "Designing a data infrastructure for catalysis science aligned to FAIR data principles". Catalysis Communications, v. 162, 2022, https://doi.org/10.1016/j.catcom.2021.106384.

[13] DNV. "DNV-RP-0497: Assurance of data quality management". 2023.

[14] DNV. "DNV-A204: Assurance of digital twins". 2023.

[15] ISO. "ISO 8000-8: Information and data quality: Concepts and measuring". 2015.

[16] M. B. Demay, G. D. Donatelli, A. L. M. de Oliveira, S. S. Rodrigues, "On the influence of uncertainty in risk analysis based on digital models", Measurement: Sensors, https://doi.org/10.1016/j.measen.2024.101510.

[17] D. Curto, F. Acebes, J.M. Gonzales-Varona, D. Poza, "Impact of aleatoric, stochastic and epistemic uncertainties on project cost contingency reserves". 2022. Int. J. Production Economics 253.

[18] JCGM, "JCGM 100: evaluation of measurement data — guide to the expression of uncertainty in measurement (ISO-GUM)", 2008.

[19] M. B. Demay, G. D. Donatelli, A. L. M. de Oliveira, P. H. Z. Machado, S. S. Rodrigues. "On the meaning of sensitivity analysis". Measurement: Sensors, 2024, https://doi.org/10.1016/j.measen.2024.101511.

[20] A. Saltelli, S. Tarantola, F. Campolongo, M. Ratto. "Sensitivity analysis in practice: a guide to assessing scientific models", v. 1.. New York: Wiley. 2004.

[21] H. M. Wainwright, S. Finsterle, Y. Jung, Q. Zhou, J. T. Birkholzer, "Making sense of global sensitivity analyses", Computers & Geosciences, v.65, 84-94, 2014, https://doi.org/10.1016/j.cageo.2013.06.00

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

An Overview of Metrology Knowledge Storage: Taxonomies, Ontologies and Constrained Vocabularies

Clifford Brown[1], Julia Neumann[2]

[1] *PTB, Berlin, clifford.brown@ptb.de*
[2] *PTB Berlin, julia.neumann@ptb.de*

***Abstract* – Within metrology new and innovative digital ways of working are being implemented that can offer great benefits to the future science, technology and administration of the area. How metrological knowledge is stored and hence processed will play an important role in maximizing the benefits of digitalisation. In this work a high level overview of current and future ways of storing scientific knowledge, the advantages and disadvantages are discussed. Finally, a potential optimal route to scientific knowledge storage for metrology is proposed.**

I. INTRODUCTION

The administrative processes and procedures that support Metrological Science and Technology are now evolving into a digital ways of working [1]. A core example of this is provided by the development of Digital Calibration Certificates as a means to replace paper based ('analogue') certificates (SmartCom [2]). Although the evolution of digital methods will not immediately provide new science for metrology, the future need to collect and collate more meta data related to the science of measurement processes and procedures will provide the possibility to interconnect scientific information in ways that have not been possible before. Using AI and ML methods to analyse this information can lead to new scientific discoveries for metrology.

It is important to state that there are significant immediate benefits of digital ways of working in metrology including: improved speed and quality of communication and transcription of data; the ability to store additional meta-data beyond what is currently stored and the machine readability for metrology communications.

An important aspect of collecting information (of any kind) is how to store it in a digital way that retains all the benefits and qualities of the information for now and into the future. Unfortunately, data, and for metrology scientific data, manages to stay 'alive' for significant lengths of time [3]. Even digital data struggles to stay active indefinitely. Without the relevant meta-data that provides a clear description of the data being stored alongside the core scientific data, data tends to become unusable even if it is findable. The FAIR principles [4] provide a number of important and clearly defined ways to keep data relevant, extending the life of data, retaining its value into the future..

II. TRADITIONAL VERSUS MODERN METHODS FOR STORING SCIENTIFIC INFORMATION

In general, a Computerised Information System (CIS) of any kind (be it a traditional localised system e.g. Excel, to a new AI system generating radically new information) is fundamentally made up of two parts: data used (either held locally or obtained from some form of distributed storage system); and the process used to manipulate that data in some way to generate some form of new, additional information which is typically displayed to the User of stored somewhere for future use.

In creating a new, quality assured computerised system there is a tendency to assume that the input data is a already given, defined, quality assured and fixed, and that the major work to be performed in generation of the new CIS revolves around the software development of the process to manipulate the data. There is no question that the processing part is very important and does require significant effort and testing to make sure it functions correctly. However, the problem tends to be, that the data aspect is put to one side until after the process part has been created. Only relatively small quantities of data tends to be used in testing a CIS. But, the quality and hence the capability of the dataset as a whole, is a significant factor in the creation of a successful and quality CIS [Ref]. This type of issue tends to be more relevant when a traditional, functional decomposition of the core problem is considered.

Another way to develop CIS systems is to apply the object-oriented (OO) paradigm to the core problem [5], and is known as Object-Oriented Analysis and Design

(OOAD). The Unified Software Development Process (USDP [6]) is an example of an OOAD software engineering practice that encompasses the Agile style of CIS development. In this technique the data as well as the processing part is considered together in developing the system. The core problem is broken down into a hierarchical set of interacting objects, where each object has its own data and processing capabilities . The OOAD provides a data model that can automatically be converted into a relational database structure, or taxonomy (Ref wiki taxonomy [7]). A taxonomy for a structure that is well defined follow a number of self referential rules [8]. One of the most important rules states that any piece of information should only be stored in one location, and this ensures that the data held is self consistent and a good taxonomical design makes it more robust against accidental corruption.

III. TAXONOMIES

As mentioned above the OOAD methodology performed correctly provides an object oriented data model that can automatically be converted into a relational database structure, or taxonomy. A relational database structure or schema consists of a number of two dimensional tables linked together using linking columns of identical information between the tables. The data information held in a row of a table typically describes the data attributes of a single object instantiation of a class from an OO model. An individual table with a relational schema or structure describes the full details of a particular core concept within the functionality or purpose of the system. For example, in metrology, a measurement would be an important concept. The relevant data would probably include: the magnitude or size of the measurement as a decimal value, and the unit associated with the measurement either as a string or maybe a link to another table containing a list of units. (Figure ?)

Functional breakdown methodologies can also lead to taxonomies describing the underlying data but the process of determining the structure requires an additional process to be performed and the output will need to be made consistent with the functional decomposition. Frequently the needs of the underlying data and the functional decomposition contain contradictions requiring a round of modifications to be made that tend to lead to results based on undesirable compromise solutions.

Taxonomies are ubiquitous structures for storing scientific information and are the most common method used for storing information. For example, when collecting data from a fixed in time experiment that, by definition, is limited in concept, not requiring additional, new types of information to be stored. Referential databases can be queried using Structured Query Language (SQL). SQL is a logic-based language using the names of the main concepts in a system whose information is being stored in a referential database.

IV.

V. TAXONOMY ISSUES

Apart from the process dependent issues of determining a good taxonomical structure that defines the scientific information for a system in a meaningful and effective way, inevitably any taxonomy will be limited regardless of how it is generated. Even well-defined taxonomies are invariably limited in time. Science is an ever-evolving subject with new ideas and associated data being uncovered continuously. Occasionally data models can be extended in a simple fashion to include new types of data for a system, but invariably a major revision of a taxonomy is required to include new types of information. Also, the code behind the taxonomy will need to be updated to work with the new data taxonomy. This tends to be a time and resource expensive procedure and is only undertaken when there is scientific justification and associated funds to do so. In addition, most likely, there will also be the need to provide some form of backwards compatibility with the old data structure.

VI. ONTOLOGIES

Ontologies are very different to a taxonomies. It stores information in an unstructured way (Ref my paper last year). In theory, you can store any information in a single ontology; even information that is totally disconnected. An ontology is infinitely flexible in terms of the information it can store. This has advantages and disadvantages.

Ontologies work using the mathematics of First Order Logic (Ref. Horrocks [9]) which allows for simple statements such as:

“Measurement 1” “has Magnitude” “0.579”

“Measurement 1” “has Unit” “kilogram

to be stored in an instance of an ontology.

This triad, or ‘triplet’ of information, “subject”, “predicate”, “object” where “subject” and “object” are effectively nouns describing “things” in system of interest, and the “predicate” is a “verb” or action that connects the two objects. Using this reasoning, it is completely possible to store all the information that exists in an example taxonomy in an equivalent ontology which is literally and simply a list of triples. However, where ontologies exceed the capability of taxonomies is in the fact that it is self-describing and hence completely extendable. There is no limit on how to extend the ontology. So, if new

information needs to be stored, the pattern for the extension is stored in the ontology, as a triple, and then the new type of information is also added, as triples.

This ability to store new information with minimal effort would appear to be the perfect solution for storing ever-changing scientific information. However, there are problems associated with the ontological solution.

VII. ONTOLOGY ISSUES – CONCEPT DRIFT

The determination of a taxonomy to describe the data associated with an information system leads to a group of linked table structures. These table structures typically represent core concepts with the problem space. They are necessary and sufficient to describe the problem under consideration; they represent implementation of the important nouns in a requirements document for the system.

Ontologies do not, in principle, need to use concepts to operate. As mentioned above information in the form of triplets can be added to an ontology simply by adding them to the full list of triplets. However, the resulting ontology will be disconnected, random facts and will not be able to generate any interesting results other than the basic information it stores. Structured storage like taxonomies, which use core concepts, join information together within the linked table structures, and can, using SQL, generate interesting information beyond the basic facts input. So, the main benefit of Ontologies, i.e. their capability to store any type of information, is also the source of its main weakness.

The middle ground is to impose the use of concepts within an ontology. By using the same methods of discovering the concepts for a taxonomy from the problem requirements, the concept (or class) meta information can be stored within the ontology. The concepts, or classes, of information can then be used, (almost like a pastry cutter to create multiple biscuits for cooking), to create actual instances of pertinent objects. In the example shown earlier, the concept of Measurement could be created with two attributes "Magnitude" and "Unit", and Measurement1 would be ('is a') an instantiation of the Measurement class.

An ontology can therefore provide the ability to evolve as new information becomes available and can embody concepts to group connected information together into higher level meaningful entities, or objects. Probably the most appropriate storage system for scientific knowledge. However, this potentially ideal storage system is compromised by the reality and complexity of how scientific knowledge evolves. Scientific knowledge is filled with contradictions. Even some of the most successful theories e.g. relativity and quantum mechanics in particle physics do not complement each other cohesively.

Even something as well defined as the systems of units for measurements suffers from knowledge uncertainty. The QuDT ontology system of units [10] developed originally by NASA out of a satellite failure due to the use of inconsistent units, arguably the most developed Unit Ontology has developed in complexity to the point that its use requires significant research into the unit concepts it contains. This research tends to uncover multiple versions of complex compound units which might be considered to be the same concept but contain subtle differences in their definitions. In this way concepts become blurred due to multiple interpretations resulting in confusion and users sometimes deciding (using the 'Open World Assumption' Ref. [11]) to create even further versions of a concept to ensure it is what they need and understand. This, concept drift tends to bloat ontologies almost to the point of becoming unusable. However, this situation reflects the normal reality of the evolution of scientific knowledge, this variation of ideas is necessary, but can lead to uncontrolled complexity.

VIII. HYBRID METHOD: TAXONOMY, ONTOLOGY AND CONSTRAINED VOCABULARY

Dealing with the complexity of concepts that are used within a problem space as described above can lead to ontologies becoming bloated to the point where their use becomes difficult. The 'Open World Assumption', (OWA, Ref.[11]) where in principle 'anyone can say anything about anything', is an idealistic concept that was developed in order to allow ontologies to evolve and so contain ever changing new information. How to manage this complexity whilst still allowing for the flexibility that ontologies provide is the challenge. Constrained Vocabularies (Ref [12]) provide one way to do this.

In a constrained vocabulary the core concepts of a problem space which are well-defined but fixed are referred to as a 'Closed World Assumption' (CWA Ref. [11]). In this approach the information required to understand and process a problem space is assumed to be well known, self-consistent and complete. There is no need to have further knowledge concepts added to the system and in this respect is more like a taxonomy. However, this approach still leaves open the opportunity to store and treat this information semantically in an ontology: allowing the flexibility to interconnect knowledge in new ways. The closed nature of constrained vocabularies controls and manages the complexity for problem spaces, but does not allow new concepts to be automatically added to the system.

An intermediate state between the CWA and OWA is the Partial Closed World Assumption (PCWA). In this case, the problem space is divided into two states: for situations where knowledge is considered well understood and complete, the CWA is used; for other cases the OWA is applied allowing new knowledge concepts to be added. In this way the PCWA approach provides a way to control complexity but still allow for concept innovation in areas of developing science and technology. For metrology, the PCWA could therefore provide a sensible but still flexible way to store knowledge. The international SI system of units in its D-SI format [13], is a fundamental cornerstone of metrology, and could be treated using a CWA.

IX. CONCLUSIONS

A number of strategies for storing scientific knowledge including: taxonomies, ontologies and constrained vocabularies have been discussed. Taxonomies provide a well-controlled, closed world assumption (CWA) environment in which to store knowledge from scientific and technical areas that are well understood, but form an expensive option when the concepts need to be modified and or enhanced. Ontologies provide a very flexible route to storing the ever changing knowledge behind scientific discovery, but can suffer from concept drift and uncontrolled complexity growth due to the associated Open World Assumption (OWA). The Partial Closed World Assumption (PCWA) provides a hybrid option: closing and controlling areas (CWA) of science evolution that are considered well understood, but allowing other areas to evolve under the OWA. The PCWA approach could provide the optimal route for storage of scientific and technical knowledge.

REFERENCES

[1] Digital Ways of Working, https://www.bipm.org/en/-/2025-03-12-metrology-in-the-digital-age.

[2] SmartCom, https://www.ptb.de/empir2018/smartcom/project/,

[3] Historical Data, …

[4] FAIR Principles, https://www.go-fair.org/fair-principles/

[5] Object-Oriented Paradigm, https://en.wikipedia.org/wiki/Objectoriented_programming

[6] Unified Software Development Process, https://en.wikipedia.org/wiki/Unified_process

[7] Taxonomy, https://en.wikipedia.org/wiki/Taxonomy

[8] "An Introduction to Database Design" C. J. Date ISBN 978-0321197849

[9] Horrocks, First Order Logic, https://www.cin.ufpe.br/~in1006/HorrocksDLIntro.pdf

[10] QuDT, https://qudt.org/

[11] Closed World Assumption, https://en.wikipedia.org/wiki/Closed-world_assumption

[12] Constrained vocabularies, …

[13] D-SI Format, https://si-digital-framework.org/

2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Lightweight Passive Monitoring for Soft Anomaly Classification in Wired Networks on Resource Constrained Microcontrollers

Prabin Dhakal[1], Francesco Picariello[2], Basanta Joshi[3], Nanda Bikram Adhikari[4*]

[1,3,4]*Institute of Engineering, Pulchowk Campus, Lalitpur, Nepal*
[2]*Univeristy of Sannio, Benevento, Italy*
[4*]*adhikari@ioe.edu.np*

***Abstract* – This paper presents a method for detecting soft anomalies in Ethernet cables using time domain signal analysis and machine learning. A dataset consisting of oscilloscope captured signals from a CAT5e cable using passive measurement technique under four controlled scenarios was used. To reduce data dimensionality while preserving statistical characteristics, a histogram-based feature extraction process was applied prior to classification. Classification was performed using Decision Tree, Random Forest, and Support Vector Machine (SVM) algorithms under various downsampling rates. Models performances were good at generalizing and predicting the classes using those features. The results demonstrate that histogram features are effective in distinguishing between different anomaly types, even at significantly reduced sampling rates showing potential for real time implementation on low powered microcontroller platforms.**

I. INTRODUCTION

Cyber Physical System (CPS) are an integral part of industry 4.0 and 5.0. CPSs are intelligent systems that closely integrate cyber and physical processes, and exchange data and information in real time [1]. While these systems significantly improve the efficiency, reliability, and functionality of the infrastructure, they also introduce new security challenges that must be addressed [2]. CPSs operate at the intersection of the digital and physical worlds. Thus, vulnerabilities in the cyber domain can have a direct impact on physical operations [3].

Effectively monitoring physical layer of network helps detect issues such as signal degradation, physical damage, and unauthorized access with tapping [2]. The targets of physical layer monitoring for cable communication security are the following: 1) detection; 2) localization; and 3) classification of the anomalies. The type of anomalies can be classified as follows: 1) hard and 2) soft [4]. Hard anomalies are severe and typically cause complete failure of the cable. These anomalies are characterized by physical damage or significant degradation that interrupts the electrical or signal continuity (e.g., open and short circuits) [4]. Soft anomalies are less severe and usually cause degradation in performance rather than complete failure. These anomalies can result in intermittent issues or degraded signal quality (e.g., insulation damage, tapping, and connector issues) [4].

Reflectometry is a technique for diagnosing cable conditions, which exploits the signal characteristics of the reflections of a transmitted signal in the wireline channel [5]. It is beneficial for identifying anomalies, discontinuities, and physical degradation within cables and connectors [6]. The traditional reflectometry techniques are active, i.e., based on injecting a test signal into the cable under test and analyzing the reflected waveforms in different domains [5]. Based on the type of incident signal and the considered domain of analysis, reflectometry can be divided into time-domain reflectometry (TDR), frequency-domain reflectometry (FDR), and time-frequency-domain reflectometry (TFDR) [6], [7].

However, these active methods require interrupting the operability of the communication system. To overcome this limit, passive methods have been proposed by considering link quality indicators, such as bit error rate, state transition, eye pattern, and signal-to-noise ratio (SNR) [6]. These methods can detect partial or full open circuits or short circuits within the cable. Nevertheless, they cannot provide good information about soft anomalies (e.g., insulator degradation) [8]. Kallel et al. [8] proposed a machine learning (ML)-based and cost-effective system for a distributed monitoring solution, which can detect anomalies and their types using a histogram-based technique. However, the cable anomalies were only open/short circuits faults. The research by Balestrieri et al. [5] proposed a new method for detection and diagnosis of three types of soft anomalies using different types of features obtained from samples acquired at several sampling rates. This research is the extension of the work by [5] for lightweight diagnosis mechanism.

The preprocessing using frequency-based measure is always demanding for computation. The research aims to solve this problem by using simpler feature extraction mechanism using time domain analysis that can be further

implemented in a low powered micro controller device.

In this paper, Section II covers the theoretical background of impact of soft anomalies in acquired signal from cable. Section III reviews related works on current and past state of art methods on active and passive anomaly detection. Section IV details the proposed methodology, including dataset, feature extraction, and machine learning classifiers. Section V presents results, comparing classifier performance under various scenario and assessing deployment feasibility on microcontrollers. Section VI concludes the research's findings and, Section VII presents current work limitation and suggests directions for future improvement.

II. RELATED THEORY

Different faults have different impacts in signal in time domain. High-impedance tapping introduces minimal load on the transmission cable, causing small signal reflections due to impedance mismatches. Such faults often go undetected by conventional protection systems [9]. Their behavior is described by the telegrapher's equations:

$$\frac{\partial V(x,t)}{\partial x} = -RI(x,t) - L\frac{\partial I(x,t)}{\partial t} \tag{1}$$

$$\frac{\partial I(x,t)}{\partial x} = -GV(x,t) - C\frac{\partial V(x,t)}{\partial t} \tag{2}$$

These equations model voltage and current along the line, accounting for resistance (R), inductance (L), capacitance (C), and conductance (G) [10].

Water ingress increases the local dielectric constant, raising capacitance and reducing the characteristic impedance Z_0:

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \tag{3}$$

Air exposure can cause oxidation or insulation degradation, leading to minor parameter shifts. Though subtle, such effects may accumulate and degrade signal quality. Higher-order statistical features and time-varying models are often employed to detect these anomalies [11].

III. RELATED WORKS

The passive measurement is the process of obtaining the measurement statistics of cables conditions without injecting test signals into the channel.

In [2], a passive, noninvasive monitoring system is proposed for Fieldbus networks to detect unauthorized physical access by identifying changes in network impedance caused by intrusion devices (i.e., impedance-matching tapping). Wang et al. [2] achieved a classification accuracy of approximately 99% in detecting device intrusion across three positions along the network.

The research [8] used a histogram to analyze the received signal of fast Ethernet cables using STM32-based microcontroller. Two anomalies, open and short circuit were accurately identified with Classification accuracy of 100% [8].

The research [5] used the passive measurement for diagnosis of the cable anomalies using frequency and time domain based features. The soft anomalies were tapping, water exposed and air exposed, which were introduced on CAT5e cables [5]. They created many features and trained the Decision Tree classifier with different sampling rates ranging from 62.5 MHz to 6.25 GHz. The classification accuracy at 312.5 MHz was found to be 99.05% [5].

Many literatures have worked on different active measurement methods to detect the anomalies. However, the anomalies detection using passive method are limited. In the research [8] only the detection of hard anomalies is carried out using histograms of signals. The research [5] conducted detection and classification of soft anomalies using frequency and time domain features which are much complex and computationally expensive mechanism that can be beneficial for powerful and bench setup but is not suitable for handheld or low powered devices.

IV. METHODOLOGY

A. Dataset

The dataset used in this study is based on CAT5e 4PR cables and includes four classes: normal, air exposed, water exposed, and tapped cables. Anomalies were introduced by altering insulation lengths and introducing tapping branches. Data acquisition was performed using an oscilloscope, capturing 10,000 time-domain records per class over a 5.6 μs window. Each record has 35,000 samples in a single time window. The dataset comprises 9913 records for normal cables and approximately 9600-9700 records for each type of anomaly, as summarized in Table 1. Full details of the dataset and collection methodology are available in [12].

Table 1. Dataset Summary

Anomaly Type	**Nos. of Records**	**Anomaly Lengths**	**Count**
Normal	9913	N/A	1000
Air-Exposed	9632	5 to 50 cm (5 cm steps)	1000
Water-Exposed	9684	5 to 50 cm (5 cm steps)	1000
Tapped	9684	3 to 15 m (3 m steps)	2000

Fig. 1 and Fig. 2 are histograms of acquired signal for a CAT5e cable, at 6.25 GHz and 312.5 MHz respectively. The x-axis represents voltage bins (time-domain features)

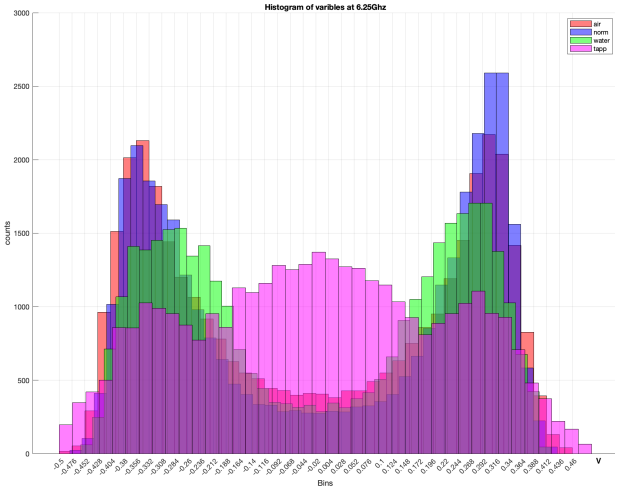


Fig. 1. Histogram Counts of signals at 6.25 GHz

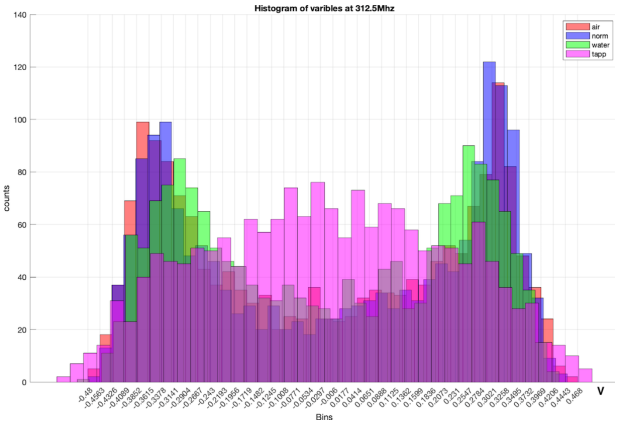


Fig. 2. Histogram Counts of signals at 312.5 MHz

and the y-axis displays the count of occurrences. Four cable conditions types Water exposed, Tapped (Ttap), Air exposed (Tair), and Normal were taken for detection of anomalies in cable. The Water exposed cable curve shows distinct peaks around voltage bins -0.33 V and 0.3 V, indicating significant deviations from the expected -0.38 V and 0.34 V seen in the normal cable signal curve. These shifts might be due to abnormal behavior or unexpected fluctuations in the cable's performance due to water exposure which might be due to change in impedance (Z_0). In contrast, the Tapped (Ttap) cable condition maintains a consistently low and stable count across voltage bins, lacking the pronounced peaks observed in the normal cable curve which might be due to several factors like signal reflection caused by impedance mismatch. This pattern suggests lower sensitivity to signal variations but potential vulnerability due interception. The Air-exposed (Tair) signal displays moderate peaks, with some overlap with both the Water-exposed and Normal cable signals, indicating a potential correlation or shared influence among these conditions. The distribution of the Water-exposed, Air-exposed, Tapped, and Normal cable signal curves across a broader range of voltage bins reflects increased variability which enhances the potential for anomaly detection using machine learning based techniques.

B. Feature Extraction

To reduce data dimensionality and retain essential signal characteristics, a decimation process is applied prior to feature extraction. Decimation consists of low-pass filtering followed by downsampling of acuquired signal. Given a discrete-time input signal x[n], the filtered output is computed as:

$$x_f[n] = \sum_{k=-\infty}^{\infty} h[k] \cdot x[n-k] \tag{4}$$

where h[k] is the impulse response of the low-pass filter.

The filtered signal is then downsampled by a factor r:

$$y[m] = x_f[m \cdot r] \tag{5}$$

To avoid aliasing, the filter cutoff frequency is set to

$$(w_c) = \pi / r \tag{6}$$

The decimated signal is subsequently processed using histogram binning, where bin counts serve as features. This approach compresses the dataset while preserving its statistical distribution, enabling efficient and informative feature representation which results in reduced burden to classification process too.

C. Classification

The dataset was partitioned using a standard 80/20 train-test split prior to classification. Three machine learning classifiers Random Forest (RF), Decision Tree (DT), and Support Vector Machine (SVC) were evaluated across various downsampling rates, ranging from 31.25 MHz to 6.25 GHz.

For each of the machine learning algorithms, classification performance was assessed using confusion matrices and other performance metrics. Additionally, a training size comparison was conducted to evaluate the adequacy and representativeness of the dataset.

V. RESULTS AND DISCUSSION

The Table 2 shows the performance of three machine learning classifiers, Random Forest, Decision Tree, and Support Vector Classifier (SVC) evaluated across various downsampling factors. These factors represent the rate at which the original data, sampled at 6.25 GHz, is reduced.

Table 2. Performance comparison (validation accuracy) of Random Forest, Decision Tree, and SVC for different downsampling factors

Factor	Rate(MHz)	RF	DT	SVC
1	6250	99.99	99.69	99.99
2	3125	99.99	99.69	99.99
4	1562.5	99.96	99.5	99.99
8	781.25	99.93	99.34	99.97
10	625	99.97	98.32	99.97
20	312.5	99.33	97.37	98.94
200	31.25	79.03	61.10	79.59

Low Downsampling Factors (1, 2, 4, 8): The classifiers demonstrate consistently high performance with accuracies exceeding 99%. This indicates that the essential features of the signal are preserved at these effective sampling rates.

Moderate Downsampling Factors (10, 20): A slight drop in accuracies is observed, especially for the Decision Tree classifier. This suggests potential loss of discriminative features due to smoothing and reduction of details as the sampling rate decreases, affecting simpler models like decision tree more significantly than other two models.

Extreme Downsampling (200): At an effective sampling rate of 31.25 MHz (6.25 GHz / 200), the performance of all classifiers drops significantly. The Decision Tree, a relatively simple model, suffers the most, while the SVC shows comparatively better resilience. This highlights the challenge of feature preservation at such low sampling rates. But still the performance is acceptable for SVC and Random forest classifier.

The drop in classification performance could have been due to the reduction in the number of samples in each record. For instance, at the extreme downsampling rate of 31.25 MHz, each record is reduced to approximately 175 samples, in contrast to the original 35,000 samples captured at 6.25 GHz. This significant reduction in temporal resolution could have impacted classification accuracy due to loss of critical signal features. More detailed analysis is required to quantify this effect. A potential mitigation strategy could involve increasing the time window for signal acquisition to preserve sufficient feature representation even at lower sampling rates.

Model Performance: The confusion matrix (see Fig. 3) reflects the Random Forest classifier's ability to distinguish between four classes: Air exposed, Normal, Tapped, and Water exposed cable as class 0-3 respectively at 312.5 MHz sampling rate. The near diagonal structure of the matrix indicates high true positive rates for all classes. The

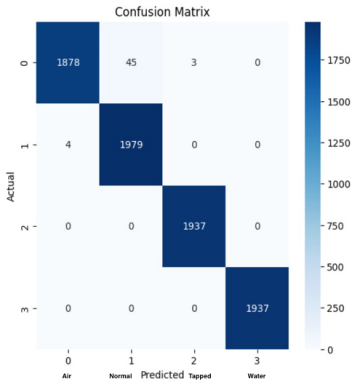


Fig. 3. Confusion Matrix at 312.5 MHz using Random Forest Classifier

Table 3. Classifier overall performance at 312.5 MHz using Random Forest

Class	Precision	Recall	F1-Score	Support
Air(0)	1.00	0.98	0.99	1926
Normal(1)	0.98	1.00	0.99	1983
Tapped(2)	1.00	1.00	1.00	1937
Water(3)	1.00	1.00	1.00	1937

Metric	Score
Accuracy	0.99
Macro Average	0.99
Weighted Average	0.99
Validation Accuracy	0.9933

misclassifications are minimal expect, some Air exposed classes were misclassified as Normal cable, likely due to similarities in signal features as seen in Fig. 1 and 2. Table 3 quantifies the classifier's performance per class using precision, recall, and F1-score for futhuer evaluation of classifier performance at 312.5 Mhz. Classes like Tapped cable and Water exposed cable have perfect scores, suggesting that these anomaly classes produce distinct, consistently identifiable patterns. The slightly lower recall for Air exposed cable may be due to signal patterns that resemble Normal cable behavior under most conditions. Nevertheless, all scores are above 0.98, demonstrating that the model performs balanced and reliable distinction between all the classes. The model globally performs with an overall accuracy of 99.33%, along with high macro and weighted averages. These results confirm that the Random forest classifier generalizes well across all anomaly types.

At a sampling frequency of 312.5 MHz-6.25 GHz, this time-domain based classification achieved a validation accuracy higher than 99% using a Random Forest classifier which is comparable even better than [5] for soft anomaly detection. The support vector based classifier has similar

accuracies at different sampling rates comparing to Random Forest. However, the Decision Tree classifier demonstrated worse performance with extremely reduced sampling size. The metrics of classification also validate the feasibility of deploying this lightweight model on low-powered microcontrollers for real-time, passive anomaly detection in wired networks without intrusive probing in communication channel.

Computational complexity: Compared to the approach in [5], which requires a complex O(N log N) computational process,whereas the calculation of histogram of signal level takes only one execution or iteration through each records of the dataset which reflects O(N) complexity for computational resource. It makes the process computationally inexpensive suggesting usability of the feature extraction process even in low powered microcontroller.

VI. CONCLUSIONS & RECOMMENDATIONS

Overall, Random Forest and Support vector machine based classifiers demonstrate robust performance across all downsampling factors, maintaining high accuracy. The results emphasize the trade-off between computational efficiency and model performance when selecting a model for classification of anomalies using histogram based bins as feature. This shows, the potential for using low-powered devices like micro controllers to detect soft anomalies in network cables can be realized by reducing the sampling rate of the signal. By employing time-domain feature extraction using voltage bins histograms as features, the computational requirements for anomaly detection can be significantly minimized. Demonstrating a notable reduction in computational demand during the feature extraction and classification steps.

This suggests that digital conversion devices like lower-bit Analog-to-Digital Converters (ADCs) with reduced sampling rates could effectively detect soft anomalies using time-domain, passive measurement techniques in wired network cables.

VII. LIMITATIONS AND FUTURE WORKS

The use of histograms derived from signal values has been demonstrated as an effective method for classifying cable anomalies, such as air-exposed, high-impedance tapped, and water-exposed conditions. However, the practical implementation of this approach remains limited when relying on the low MHz sampling rates typically found in microcontroller based ADCs. Future work will investigate the impact of extreme downsampling on classification accuracy and explore the use of extended time windows to preserve critical signal characteristics. Real world deployments in environments such as data centers and complex cabled infrastructures present further promising directions for practical validation.

REFERENCES

[1] Zhenhua Yu, Hongxia Gao, Xuya Cong, Naiqi Wu, and Houbing Herbert Song. A survey on cyber–physical systems security. *IEEE Internet of Things Journal*, 10(24):21670–21686, 2023.

[2] Xiangming Wang, Yang Liu, Kexin Jiao, Pengfei Liu, Xiapu Luo, and Ting Liu. Intrusion device detection in fieldbus networks based on channel-state group fingerprint. *IEEE Transactions on Information Forensics and Security*, 2024.

[3] Kang Yan, Xuan Liu, Yidan Lu, and Fanglu Qin. A cyber-physical power system risk assessment model against cyberattacks. *IEEE Systems Journal*, 17(2):2018–2028, 2022.

[4] Qinghai Shi and Olfa Kanoun. Wire fault diagnosis in the frequency domain by impedance spectroscopy. *IEEE Transactions on Instrumentation and Measurement*, 64(8):2179–2187, 2015.

[5] Eulalia Balestrieri, Pasquale Daponte, Luca De Vito, Francesco Picariello, Sergio Rapuano, and Ioan Tudosa. A passive-measurement method for physical security and cable diagnosis. *IEEE Transactions on Instrumentation and Measurement*, 74:1–12, 2025.

[6] Xing-Yu Zou, Hai-Bao Mu, Hao-Tian Zhang, Lan-Qing Qu, Yi-Fan He, and Guan-Jun Zhang. An efficient cross-terms suppression method in time–frequency domain reflectometry for cable defect localization. *IEEE Transactions on Instrumentation and Measurement*, 71:1–10, 2022.

[7] Pasquale Daponte, Gianluca Mazzilli, Enrico Picariello, Francesco Picariello, and Ioan Tudosa. Online diagnosis of automotive wireline channels: the role of measurements and instrumentation. In *2022 IEEE International Workshop on Metrology for Automotive (MetroAutomotive)*, pages 150–154. IEEE, 2022.

[8] Ahmed Yahia Kallel, Dhia Haddad, Thomas Keutel, and Olfa Kanoun. Real-time monitoring of cables based on network interface controllers for predictive maintenance. *IEEE Transactions on Instrumentation and Measurement*, 71:1–8, 2022.

[9] W. C. Santos, F. V. Lopes, N. S. D. Brito, and B. A. Souza. High-impedance fault identification on distribution networks. *IEEE Transactions on Power Delivery*, 32(1):23–32, 2017.

[10] Clayton R Paul. *Analysis of multiconductor transmission lines*. John Wiley & Sons, 2007.

[11] Yatai Ji, Paolo Giangrande, Weiduo Zhao, Vincenzo Madonna, He Zhang, Jing Li, and Michael Galea. Investigation on combined effect of humidity–temperature on partial discharge through dielectric performance evaluation. *IET Science, Measurement & Technology*, 17(1):37–46, 2023.

[12] Francesco Picariello. Two-wire communication channel with anomalies, 2024.