

Digital Metrological Expert – design of a software for automated key comparison data analysis in a digital world

Daniel Hutzschenreuter¹, Clifford Brown¹, Wafa El Jaoua¹, Moritz Gafert¹, David Urban²

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.

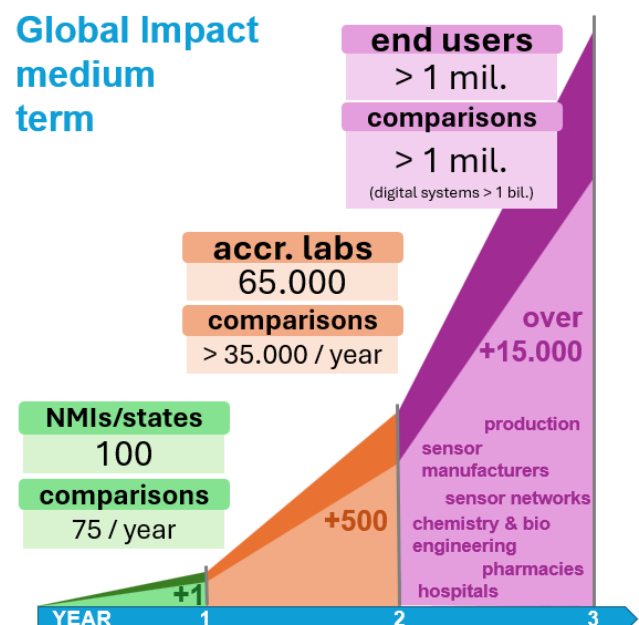


Fig. 1. Estimation of the medium-term impact of comparison data analysis software.

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.

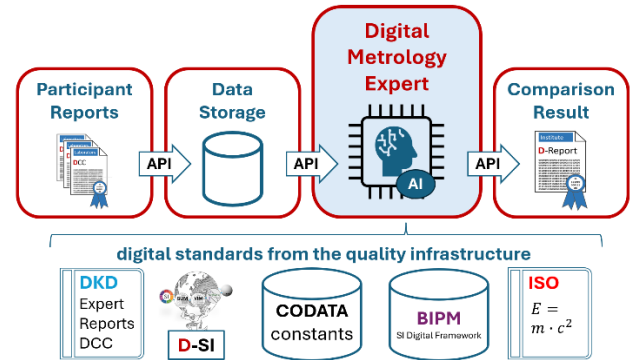


Fig. 2. Basic set-up of the DME in a digital workflow.

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 Electrotechnical 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].

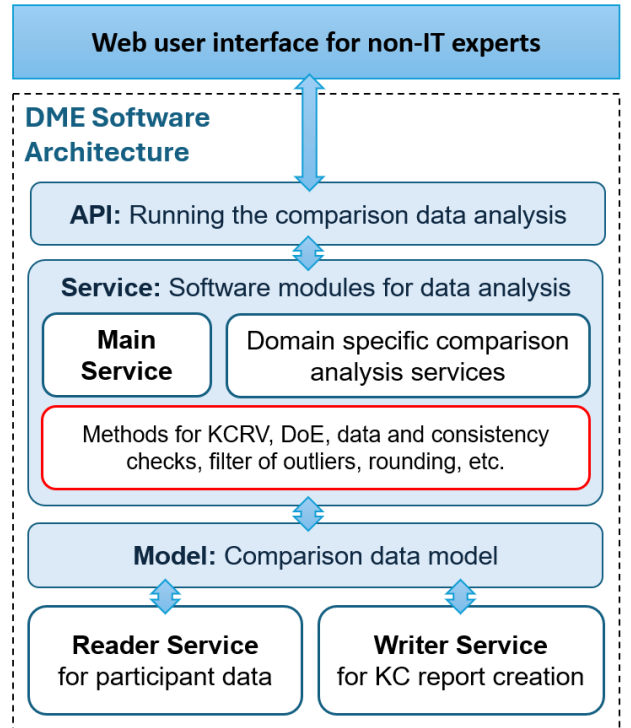


Fig. 3. Outline of the major components of the universal DME software architecture for comparison data analysis web-service.

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.

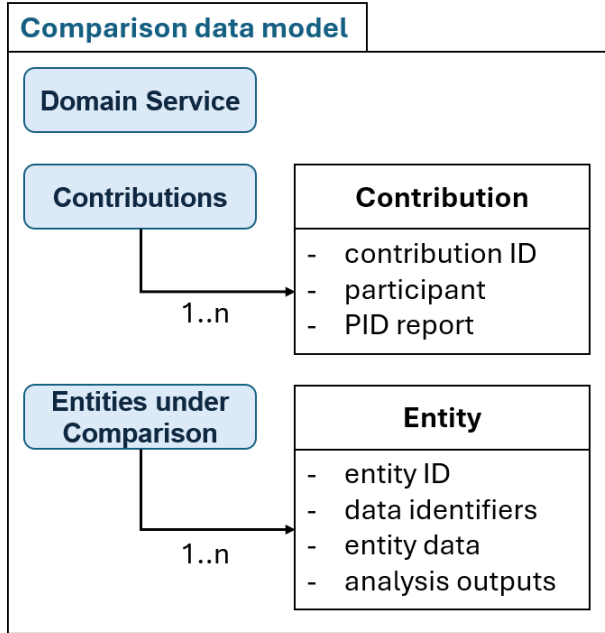


Fig. 4. Outline of the data model within the DME.

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 exchanged. 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. ACKNOWLEDGEMENTS

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/>