

Digital Transformation for the United States Air Force Metrology and Calibration Program

Jeremy Latsko¹, Michael Brockway², Salvatore Capra³, Evan Elliott⁴

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.

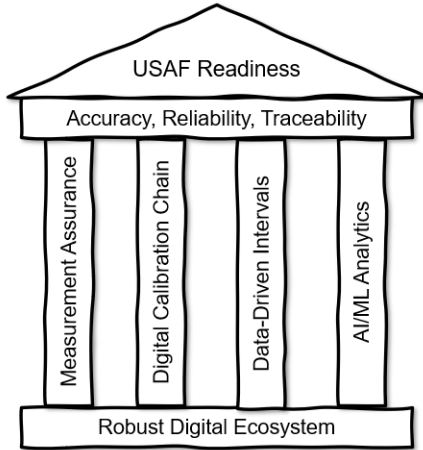


Fig. 1. AFMETCAL's DT framework, built upon a digital foundation and four Lines of Effort. These directly enable USAF readiness through improved quality of A-R-T.

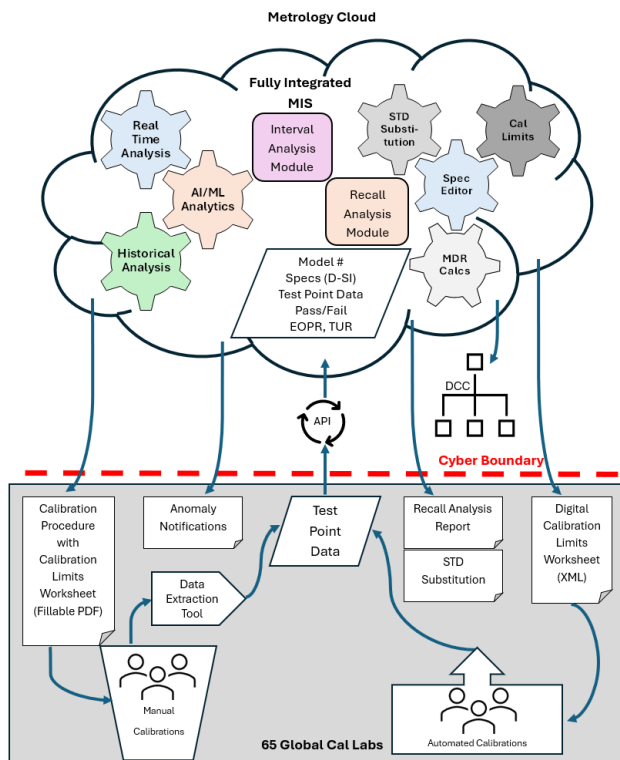


Fig 2. Technical overview of Digital Transformation.

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 minimize 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.

- [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.