Measurement Information Infrastructure (MII) Representing ISO/IEC 17025 Scope of Accreditations in XML Data

Speaker/Author:
David Zajac
Software Engineer, Cal Lab Solutions, Inc.
777 Pettus Rd, Madison, AL 35757
(256) 479-8651
dzajac@callabsolutions.com

Abstract

This paper discusses the progress made on an effort to develop a standardized data infrastructure for metrology. The first major step now under active development by a multi-party initiative is a standardized data infrastructure for ISO/IEC 17025 Scopes of Accreditation. To guide its efforts, the initiative has performed extensive data mining of existing SOA documents. The initiative has iteratively refined a SOA data schema and has developed a number of supporting products including: a standard measurement-quantities-and-unit-of-measure database, open-source unit-of-measure management and conversion tools, an open-source high precision formula interpreter, and an open-source SOA data access object library. User friendly data editors and online SOA search tools are now under active development. As a result of these efforts, the metrology community now has open-source foundational elements in place for dealing with much more than SOA specific data. Because of these developments, tasks such as verifying that final uncertainties on a calibration test report comply with a SOA, can leverage pre-existing software libraries. Tasks such as locating a calibration service provider by means of robust online search tools are on the horizon. A vigorous, self-enforcing standard for terms, taxonomies, and knowledge bases utilized throughout metrology will be developed as a fundamental component of the MII.

Introduction

This paper follows upon the presentation and paper the author made at last year's NCSLI, “Creating a Standardized Schema for Representing ISO/IEC 17025 Scope of Accreditations in XML Data.”[1] The effort to develop a standard for SOA data has now been merged into a greater effort to develop an all encompassing standard framework for the digital representation of measurement data: the MII.[2] The schema presented in the previous paper was sufficient to meet the author’s original objectives, but as a sub-part of the MII, is insufficient to support the MII’s additional, highly worthwhile uses for machine processable SOA data. One of these additional uses, in particular: “enabling the identification of calibration service providers by means of robust online search tools,” drove the need to redesign the schema. A new, more fully encompassing schema has been designed, and around this new schema, a full suite of supporting data sources and open-source software is being actively developed. Much progress has been made and is available online. Following open source practices, much of the developed source code is available under standard open source licensing, and is free to use at no cost.[3] Much has been done, yet much still needs to be done.

Requirements

An Expanded Mission - Becoming a Component of MII

The accomplishment of a digital representation of SOA data is a logical first step for an initiative to implement an entire MII. SoAs exist in the thousands. While this is a large data set, it is much smaller set than the hundreds of thousands of instrument specs and millions of certificates. SOA documents are also publically available and in addition, a schema for SOA data had been developed. However, shortly after activity commenced on this objective, two difficulties soon became apparent.

1. The gamut of highly valuable uses for digitally represented SOA data expanded and quickly many of these new uses became new and firm requirements. Several of these new requirements can not be supported with the original schema. In particular: “enabling the identification of calibration service providers by means of robust online search tools” must be supported, and the original schema, which is limited to data elements that could be commonly found in existing SOA documents, can not support this requirement. It became clear that the gamut of potential uses of SOA data in a standardized digital format exceeds the capabilities of the original schema and an effort to create a new SOA data schema was required. This new schema has an absolute requirement that the data end user need not be a laboratory “insider.” The new users could include any party anywhere in the world and the new schema must be able to provide context for the uncertainty statements that are missing from the original schema and from existing SOA documents.

.

2. A SOA database could never be populated by simply digitizing and data mining the existing SOA documents. No solution, using manpower or any technology, can be envisioned that is capable of adequately solving this problem. No set of pattern recognition rules, translation rules or, even the most powerful artificial intelligence engine that could be envisioned could solve this problem. The most intelligent human, even given a lifetime of experience and expertise in metrology, but lacking an “insider’s” knowledge of the measuring entity’s internal processes, techniques, equipment, and the devices being tested or calibrated, will encounter content within existing SOA documents that are simply incomprehensible. To study the variations in SOA data and to aid in the development of taxonomies, every available online SOA has been manually digitized and data mined as a step in this initiative. This has proven that SOA documents can be economically and manually digitized and data mined and that this can produce a copious

quantity of usable and valuable data. However, it has also proven that without access to their original authors, SOA documents remain a poor data source for MII utilization.

The Information Context Deficiency

The primary deficiency in the original schema and existing SOA documents is lack of rigorously-organized, in-situ context for the uncertainty statements. If the user of the original schema or of a typical existing SOA document is familiar with the measurement laboratory's internal processes, techniques, equipment, and the devices being tested or calibrated, the original schema could be generally sufficient and the typical existing SOA document can be, for the most part, comprehended. If the user (in particular a software developer) has full contextual knowledge of the uncertainty data entries, the original schema allowed the automated location and retrieval of all the required data needed to perform final test report uncertainty calculations. However, an assumption of requisite contextual knowledge by the SOA data end user is completely inappropriate if the end user is a party having no pre-established relationship with the measurement laboratory.

It is most common for the SOA document to reside publicly online. This implies that there is at least some intent for the SOA document to be useful for users that are not internal to the measuring entity. In fact, as the MII initiative members brainstormed the possible uses of SOA data it became clear that SOA data, once converted into digitally processable form did indeed have high value uses to entities external to the laboratory. Yet, far too often, it is simply not possible for an external party to comprehend important content contained within these public documents. For MII, merely representing the existing SOA data in digitally processable form would not be not sufficient. Whether intentional or not, the far too typical requirement to have undocumented contextual knowledge of the laboratory's internal processes, techniques, equipment, and the devices being tested or calibrated, is simply not valid. Within the MII, the potential audience of all data, including that of an SOA, exists in the broadest sense possible. For the MII initiative, the existing state of SOA data is not acceptable. To fully realize the MII vision, the existing SOA data source has to be fundamentally re-engineered to enable the inclusion of currently missing contextual information.

Solving the Missing Context Problem

Defining The Technique

The original schema has a data element for a technique. The problem is that this data element is not explicitly defined to hold anything more than an informational text entry. As such, the original schema treats the technique data element as ancillary information, not as data intended to be queried. In the new schema, the technique element is foundational to comprehension of all uncertainty statements. As such, in the new schema all uncertainty statements must be linked to a technique.

The technique now:

- Defines the technique result(s) type by measurement quantity, and all measurement quantities must exist in a standard reference measurement-quantity-unit-of-measurement database.
- Defines the minimum and maximum expected values for each process result.
- Defines the name and type of every parameter utilized to either make a selection or calculate a value. If the underlying value is numeric, it defines the measurement quantity type for the parameter. If the underlying value type is a list of values represented by names, it defines that list.
- Defines every piece of required equipment both by equipment type and by role served in the test setup.
- Defines every formula utilized within the technique with sufficient detail to allow automated assignment of values into each parameter and subsequent calculation.

The new schema additionally encourages the documentation of techniques by having data elements reserved for HTML based documentation.

Organizing Techniques - Formal Metrology Taxonomies

The original schema contains a purely informational data element for process types. In the new schema, the ProcessType element is given a formal role: It conveys formal standard metrology taxonomy, structure, and nomenclature to techniques. Every technique must be linked to a ProcessType element and every ProcessType must exist in a standard metrology taxonomy. The process type serves as a content-rich grammatical adjective. The ProcessType may be thought of as an unimplemented generic technique of a specific type.

The data structures for ProcessType are well defined and exist primarily to create a means of categorizing and templatizing techniques. For example, if the technique measures a safe to handle alternating voltage, the technique's ProcessType might be named:

"Measure.Voltage.Sinusoidal,LowFrequency,LowVoltage"

Many techniques exist that will make this "kind" of test or measurement. The ProcessType is the "kind" of test or measurement. By requiring every uncertainty specification in a SOA to be tied to a technique, and in turn, requiring every technique to be tied to a process type, and finally by having standard taxonomies for all of these nomenclature elements, the clarity of the contents of SOA data will become viable for comprehension by the expanded set of users the MII envisions.

MII's Relation to the International Vocabulary of Metrology[4] (VIM)

MII must be built upon a specific standard metrology vocabulary. What, therefore, is the relationship between MII and the VIM? The VIM's reason for existence is both similar yet different from that of the MII. Both may exist side by side. Both exist to foster communication. The MII can and likely will leverage the VIM. However, while the VIM is limited by its nature to being a recommended best-practice, a standard taxonomy within MII will not be optional; compliance with its standards will be a fundamental requirement of its machine to machine problem domain.

Unlike the VIM, in MII all terms will be either formally defined in a database schema schema or as entries in a reference database.. In the MII, terms will include names for complex data structures and sub-components that will all exist within an overarching superstructure. The pinnacle for MII's test and measurement data will be the ProcessType. All specifically metrological data exchanged in MII must be linked to a ProcessType in order to provide the context crucial to its comprehension. While the concept of a process type is not new to metrology, the concept of a formal standardized reference dataset of process types to which all else is related, is new to metrology.

The vocabulary maintained in the VIM is a start, but it is far from sufficient. First, the VIM is not organized as a dataset and the VIM does not define data structures. The MII targets a software, machine-to-machine communication problem domain. While at best, the VIM recommends standards terms, the MII will have a much more powerful means of enforcing both adoption and compliance. The MII is an information technology based infrastructure for computer exchange of metrology information. All data sources utilized in MII must be computer accessible and processable datasets. As a result, any non-compliance will directly result in non-functionality.

Schema Changes

Restructuring For Reusability

Bearing in mind that this effort is currently in active development and remains subject to change, the following data element definitions are merely intended to convey a sense of the exactness with which data must be represented within the MII.

In the redesigned SOA schema, data elements that are likely to be seen again in follow-on MII projects are split out from the SOA schema. These reusable data elements now reside in the schema "MetrologyTaxonomyCatalog.xsd." The reusable data elements currently defined in the

“MetrologyTaxonomyCatalog.xsd” are: DeviceType, DeviceTypes, Value_type, RangeLimit_type, Result, Function_type, Function, Parameter, ProcessType, and Technique.

- **DeviceType:** The value for a DeviceType is the name of a specific device or family of devices. As a standard metrology taxonomy develops, these values should correlate with a standard online reference.

- **DeviceTypes**: A DeviceTypes element simply holds a collection of DeviceType elements.

- **Value_Type:** A Value_Type is a reusable definition for any data element holding a measurement value. Any data element derived from a Value_type will hold a combination of a numeric value, an optional unit-of-measure, and optional presentation formatting.

- **RangeLimit_type:** A RangeLimit_Type is a reusable definition for any data element holding a value that can either be the start or end of a measurement value range. Any data element derived from a RangeLimit_Type holds the combination of a Value_Type and a test. A test is a boolean operator that is equivalent to “is equal to,” “is less than,” “is greater than,” “is greater than or equal to,” or “is less than or equal to.”

- **Result:** The Result element defines the type of any result produced by a Function, ProcessType or Technique. The type of a result is defined by specifying its measurement quantity type.

- **Function_type and Function:** A Function_type is a reusable definition for any data element holding a mathematical symbolic expression in which all symbols are fully defined. A Function is a data element that is directly derived from a Function_type.

- **Parameter:** A Parameter holds a named variable type definition. The variable definition can be either for a variable that can hold a measurement value or named item. If the variable definition is for a measurement value, the Parameter element holds the variable’s name and measurement quantity type. On the other hand, if the variable definition is for a variable that can hold a named item, the Parameter element holds the variable’s name as an enumerated list of valid item names.

- **ProcessType:** A ProcessType data element holds a name, Result(s), Parameter(s), Function(s), and Documentation. To be useful in MII, Process types must exist in a standard universally accessible reference - a metrology taxonomy organized as a public online dataset. Process types serve as grammatical adjectives which specify a specific yet still high level domain for test and measurement techniques.

- **Technique:** A Technique element holds information about techniques. The information held includes: a name for the technique, the name of the process type the technique implements, ranges for results, (if required) an extension to the list of parameters contained in its process type, ranges for all numeric parameters, a list of required equipment, (if required) additional functions to those defined in its process type, functions that are specifically annotated for utilization in a SOA, and documentation that should go beyond the generic documentation of its process type.

Updates to other SOA Elements

The level of detail given the preceding section of this paper was intended to impart a taste of the degree of exactness and rigor in which data will be formalized within the MII. However, since as of this date, this effort is not yet at an approved version release, further presentation of data elements at this level of detail is not appropriate for this paper.

More Than Data

In parallel with the development of the MII data structures, software is being developed that is able to work with these data structures. This parallel development drives an iterative refinement of both the formal definitions of the data elements and the ongoing discovery of new potential value as the MII is coming to fruition.

The Data Access Object Library

An open source data access object library currently exists with the following features:

- **Ability to load data from both local files or internet based XML files**
- **Raw data structure independent access to all data values**
- **Ability to add, remove, and/or modify individual data values**
- **Ability to save updated data back to file**
- **Support for common data queries**
- **Built in formula interpretation**
- **Substantial data validation and automatic conflict resolution**
- **Ability to integrate with higher level software**
- **Open source**
- **Dual licensing**

Data Editors

Data editors are being developed to allow manipulation of data without having to open the XML files in a text editor.

- **Unit-of-Measurement Database Editor**

- **SOA Database Editor**

Online SOA Search Tools[5]

- **Geolocation of calibration services providers with search filters for full SOA contents including:**
  - **Quantity**
  - **Range**
  - **CMC**
  - **Device Type**
- **Business identification**
- **Point of Contact identification**

Next Steps - Collaboration and Formalization

The immediate and most pressing need at this point is the creation of advanced, standardized, metrology taxonomies. SOA data can now exist in and be readily accessed from highly organized and rigorously defined data structures. What is now required is the achievement of advanced, standardized, metrology taxonomies so that terms stored within these data structures follow defined naming conventions and successfully represent standardized meaning. It is time to bring library science to metrology!

References

1. D. Zajac, "Creating a Standardized Schema for Representing ISO/IEC 17025 Scope of Accreditations in XML Data," in *Proc. NCSL Int. Workshop & Symposium* (NCSLI), 2016.

2a. M. J. Kuster, "Toward a Measurement Information Infrastructure," *Metrologist*, **6.1-pres.**, 2013-pres.
2b. M. J. Kuster, "Metrology: Standardize and Automate!," *Cal Lab: The Int. J. of Metrology*, **20.2**, pp. 26-34, 2013.
2c. M. J. Kuster, "Toward a Measurement Information Infrastructure: Smart Accreditation Scopes," *Metrologist*, **8.1**, 2015.
2d. C. Walker, M. J. Kuster, M. Schwartz, "A Measurement Information Infrastructure (MII): Inception, Vision and Progress," presented at *Measurement Science Conference*, 2017.
2e. M. Schwartz, "AUTOMATION CORNER: SAO/CMC Editor & Search Tools," *Cal Lab: The Int. J. of Metrology*, **24.1**, p. 48, 2017.

3a. https://github.com/CalLabSolutions/Metrology.NET_Public
3b. http://testsite2.callabsolutions.com/UnitsOfMeasure/UOM_Database.xml
3c. http://schema.metrology.net/Uom_Database.xsd
3d. http://schema.metrology.net/Uncertainty.xsd
3e. http://schema.metrology.net/MetrologyTaxonomyCatalog.xsd

3f. http://schema.metrology.net/SOA_Master_Datafile.xsd

4. JCGM, "International vocabulary of Metrology–Basic and general concepts and associated terms (VIM)," 3rd edition (2008 version with minor corrections), *JCGM 200:2012*.

5. https://search.qualer.com