IMEKO TC6 International Conference on Metrology and Digital Transformation
September 19 – September 21, 2022, Berlin, Germany

SUPPORTING PROVENANCE OF DIGITAL CALIBRATION CERTIFICATES WITH TEMPORAL DATABASES

Vashti Galpin[a,*], *Ian Smith*[b], *Jean-Laurent Hippolyte*[b]

[a] School of Informatics, University of Edinburgh, UK, Vashti.Galpin@ed.ac.uk
[b] National Physical Laboratory, UK, Ian.Smith@npl.co.uk, Jean-Laurent.Hippolyte@npl.co.uk
* Corresponding author. E-mail address: Vashti.Galpin@ed.ac.uk

Abstract – Trust in current and historical calibration data is crucial. The recently proposed XML schema for digital calibration certificates (DCCs) provides machine-readability and a common exchange format to enhance this trust. We present a prototype web application developed in the programming language Links for storing and displaying a DCC using a relational database. In particular, we leverage the temporal database features that Links provides to capture different versions of a certificate and inspect differences between versions. The prototype is the starting point for developing software to support DCCs and the data with which they are populated and has underlined that DCCs are the tip of the iceberg in automating the management of digital calibration data, activity that includes data provenance and tracking of modifications.

Keywords: digital calibration certificates, provenance, temporal databases, XML, prototype

1. INTRODUCTION

Information relating to the calibration of an instrument or artefact is captured in documents referred to as calibration certificates. These documents are typically provided either as physical paper-based documents or in electronic form, for example, in archivable Portable Document Format (PDF-A). A downside of both approaches is that the information they contain is not machine-readable, that is it is not available in a form that can be read and processed by computer. Using the information therefore requires, to some degree, human involvement and as such is prone to error, for example, arising from the transcription of numerical information from paper to computer.

Recent initiatives have considered the replacement of paper-based and electronic calibration certificates by fully-machine readable calibration certificates, referred to as "digital calibration certificates", often abbreviated to "DCCs". The European Metrology Programme for Innovation and Research (EMPIR) [1] funded the "SmartCom" Joint Research Project (2018-2021) [2] which developed a framework for DCCs. The framework allows the key calibration certificate components of administrative data and measurement results to be stored. The recipient of a DCC will have, or be able to develop, software to ingest and use the information stored therein.

Ensuring the integrity and longevity of current and historical calibration data is essential for calibration laboratories to maintain long-term trusted relationships with their customers. It is also a requirement of ISO/IEC 17025 accreditation [3], along with the preservation of information about the methods, processes, software, equipment, and staff involved in the calibration task. This accompanying information, that is essential to trace and potentially repeat the conditions in which the calibration is performed, is more generally known as "provenance information" [4]. Laboratories must be able to track and compare the changes between successive calibrations of the same artefact made to data (for example, newly observed measurement results) or to provenance (for example, a different employee now signs off the calibration results).

Since DCCs are containers for the trusted exchange of calibration data, their potential to increase the productivity of laboratories relies on automated, integrated, and time-based curation of this data and associated provenance.

The structure of this extended abstract is as follows: we provide background on DCCs before we go on to describe our prototype. We discuss the choices made in the development of the prototype and illustrate its usage with screenshots. We conclude with ideas for further work.

2. DIGITAL CALIBRATION CERTIFICATES

"A *digital calibration certificate (DCC)* serves for the electronic storage, the authenticated, encrypted and signed transmission and the uniform interpretation of calibration results." [5].

The structure of a DCC reflects the information that is required by ISO/IEC 17025 [3] for reporting calibration results. A DCC comprises two compulsory sections. The first compulsory section contains the administrative information including that which is generally presented on the first page of a paper-based certificate. The second compulsory section contains measurement results. The measurement results section itself relies upon a framework, referred to as the "Digital SI", developed specifically for the storage of measurement data. The framework ensures that quantity values, units of measurement and uncertainty information can all be represented. A DCC may also include two optional sections. The first optional section contains information presented purely for humans, for example, calibration-specific data sheets, and other auxiliary, machine-interpretable data, for example, relational tables. The second optional section contains an encoded version of a human-readable version of the certificate.

The Digital SI and DCC frameworks specify, for example, how numerical values and date and time information should be presented, and uses the BIPM SI brochure [6] and the siunitx package for LaTeX [7] as the basis for the provision of units of measurement.

The Digital SI and the DCC may be implemented in a language chosen by the user, for example, Extensible Markup Language (XML), JavaScript Object Notation (JSON). The SmartCom project has developed and made available an XML schema for the Digital SI framework [8] while the Physikalisch-Technische Bundesanstalt (PTB), the National Metrology Institute of Germany, has made available an XML schema for the DCC [9].

We are aware that the frameworks (and related schemas) may be subject to future updates, for example, if the BIPM SI brochure is updated, or if the requirements for the contents of a calibration certificate change.

3. DCC PROTOTYPE

Temporal databases capture *when a piece of data is valid.* In the case of a relational database management system (RDBMS), we are interested in the time period for which a row in a database table is valid, and this can be achieved by adding additional fields to capture this time period [10,11]. We are interested in provenance of DCCs and hence we want to capture when data changes in the database, namely *transaction-time* information. Temporal databases can also record *valid-time* information which is information about when something is true in the real world. Our prototype is developed in Links which is a strongly-typed statically-typed functional programming language that supports the development of web applications with database back-ends in a single language [12,13]. It has recently been extended with temporal features that have been demonstrated in the context of curated scientific databases [14].

3.1. The Links programming language

Links is a *cross-tier* language: it removes the need for the developer to write JavaScript or a particular database query language (in this case, PostgreSQL [15]) for the different tiers of a web application providing data from a database. In particular, Links assists in the writing of correct software by providing *language-integrated query* which allows the programmer to write high-level queries that are transformed to correct SQL queries with known performance for execution on the database.

Furthermore type-checking variables and functions before program execution reduces the likelihood of run-time errors.

We work with DCCs formatted according to the XML schema [9] made available by PTB. XML is tree-structured, and each element in an XML document is either a named tag with zero or more elements as children or a text node which is a leaf (hence has no children). An XML schema describes the tags that can be used in a document and how they can be nested [16].

For this paper, a synthetic DCC, containing dummy administrative data and measurement results, is used.

3.2. Implementation

The temporal features of Links are currently supported by an RDBMS, hence it is necessary to map from the XML schema to this database. There were two choices here: either to use the DCC XML schema to create the appropriate tables in the RDBMS or to treat the data in a schema-agnostic manner and capture it as an XML document. We choose the latter for reasons we describe below. The Dynamic Dewey (DDE) labelling scheme supports the labelling of XML documents and has good query and insertion performance [17].

The prototype provides an interface for viewing a single DCC. Straightforward extensions include the ability to edit the DCC and to validate that it adheres to the DCC XML schema. Further obvious extensions are the ability to work with many DCCs, to compare different DCCs, and to support the signing of DCCs. In the longer terms, we will require systems that support the whole calibration workflow including automated collection and processing of machine-readable calibration data.

3.3. Interface

Using the interface, it is possible to view the current state of the document, the state at a previously specified date and time, and to compare two versions at two specified dates and times. It is possible to hide or show specific subtrees of the document, and when doing the comparison, to see the subtrees where changes have been made.

The Change Analysis option allows the user to see a history of all changes over the life of a DCC. Figure 1 shows how changes are highlighted as well as a cumulative total of insertions and changes for each tag in the document. There are various options to view subtrees for a specific tab or to control the display of the whole document. Figure 2 shows how details of changes are displayed.

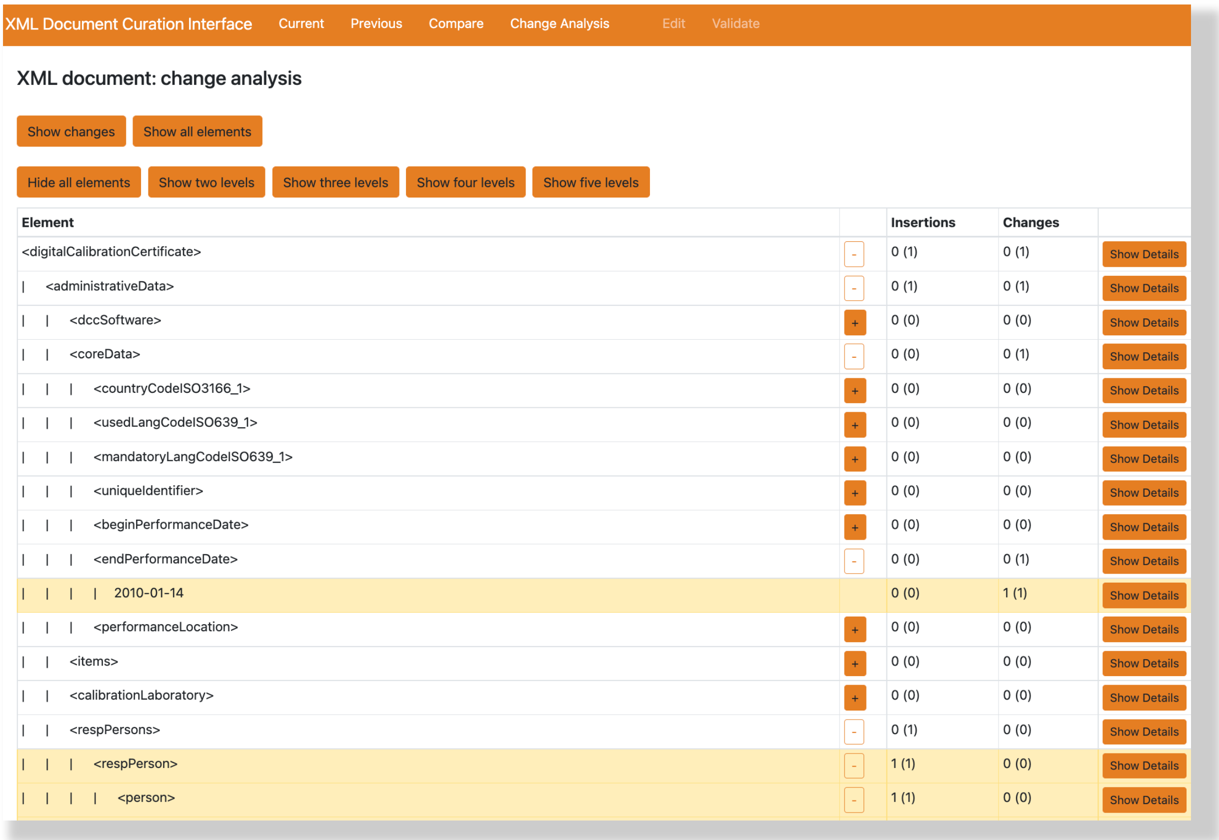

Fig. 1. A screenshot of the Change Analysis interface showing the highlighting of changes.

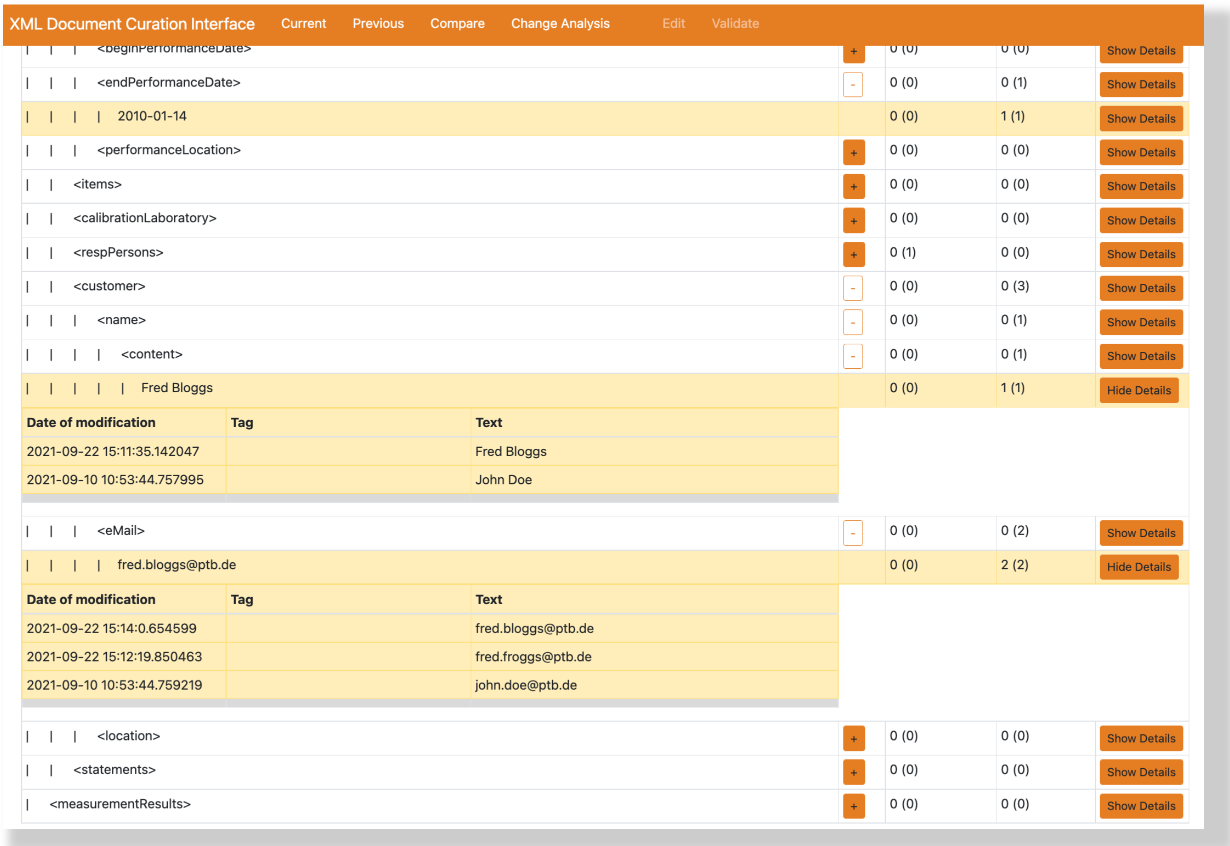

Fig. 2. A screenshot of the Change Analysis interface showing details of two changes.

4. FUTURE WORK AND CONCLUSIONS

The prototype we have developed demonstrates that the combination of DCCs and temporal databases result in streamlined curation of calibration data. The prototype leverages features of the Links language

(a) to facilitate the development of a user-friendly web front-end,
(b) to integrate relational database queries on a database that maps with the DCC XML tree structure, and
(c) to enable automated tracking of changes made to calibration records over time, as well as the explanation and timeline for those changes.

Considering our choice to work with a relational database, Links may in future support XML with temporal features and hence the translation would not be necessary. However, since the DCC is only one component of the envisioned digital calibration workflow, and can be seen as output and data exchange from some form of data storage, there is no need to use an XML storage format. Dealing with temporal data in a relational database is better understood and less experimental so we believe it makes sense to use a RDBMS for a system for digital calibration data. Working in a schema-less fashion with the XML document allows to cope with changes in the DCC XML schema without affecting the storage in the RDBMS.

DCCs are the tip of iceberg when considering the whole lifecycle of machine-readable calibration data. Their creation and use depend on software to support storage and provenance of the underlying calibration data. The development of the prototype has raised interesting questions about what such a system should be able to provide, and we will continue with our investigations and development.

ACKNOWLEDGMENTS

Funding: This work was supported by ERC Consolidator Grant Skye [grant number 682315] and by an ISCF Metrology Fellowship grant provided by the UK government's Department for Business, Energy and Industrial Strategy (BEIS).

We thank James Cheney for proposing the iceberg analogy, and Simon Fowler for his assistance in using the temporal extensions to Links.

REFERENCES

[1] EMPIR—EURAMET. https://www.euramet.org/research-innovation/research-empir/(accessed 12 April 2022).
[2] SmartCom—Project. https://www.ptb.de/empir2018/smartcom/home/ (accessed 12 April 2022).
[3] ISO/IEC 17025:2017; General Requirements for the Competence of Testing and Calibration Laboratories. International Organization for Standardization and International Electrotechnical Commission: Geneva, Switzerland, 2017. https://www.iso.org/publication/PUB100424.html
[4] J. Cheney, A. Chapman, J. Davidson and A. Forbes, Data provenance, curation and quality in metrology, in *Advanced Mathematical and Computational Tools in Metrology and Testing XII*. World Scientific, pp. 167–187, 2022. https://doi.org/10.1142/9789811242380_0009
[5] S.G. Hackel, F. Härtig, J. Hornig and T. Wiedenhöfer, The Digital Calibration Certificate, PTB-Mitteilungen 127(4), pp. 75-81, 2017. https://doi.org/10.7795/310.20170403
[6] BIPM. The International System of Units (SI), 9th ed.; Bureau International des Poids et Mesures: Sèvres, France, 2019. https://www.bipm.org/documents/20126/41483022/SI-Brochure-9-EN.pdf (accessed 12 April 2022).
[7] siunitx: A comprehensive (SI) units package, https://ctan.org/pkg/siunitx (accessed 12 April 2022)
[8] Schema for Digital System of Units (D-SI). https://gitlab1.ptb.de/d-ptb/d-si/xsd-d-si (accessed 1 July 2022).
[9] Schema documentation for dcc.xsd. https://www.ptb.de/dcc/v3.1.0/autogenerated-docs/doku_3.1.0_all.html (accessed 12 April 2022).
[10] C. Jensen, R. Snodgrass, Temporal database. In: Encyclopedia of Database Systems. Springer, 2009. https://doi.org/10.1007/978-0-387-39940-9 395
[11] R. Snodgrass, *Developing Time-Oriented Database Applications in SQL*, Morgan Kaufmann, 1999.
[12] E. Cooper, S. Lindley, P. Wadler and J. Yallop, Links: Web programming without tiers, *5th International Symposium on Formal Methods for Components and Objects (FMCO 2006)*, LNCS 4709, 2006. https://doi.org/10.1007/978-3-540-74792-5_12
[13] Links: Linking Theory to Practice for the Web, https://links-lang.org. (accessed 12 April 2022)
[14] V. Galpin and J. Cheney, Curating Covid-19 Data in Links. *8th and 9th International Provenance and Annotation Workshop (IPAW 2020 + IPAW 2021)*, LNCS 12839, 2021. https://doi.org/10.1007/978-3-030-80960-7_19
[15] PostgreSQL: The World's Most Advanced Open Source Relational Database. https://www.postgresql.org/ (accessed 12 April 2022).
[16] XML Technology. https://www.w3.org/standards/xml/ (accessed 12 April 2022)
[17] L. Xu, T.W. Ling, H. Wu and Z. Bao, DDE: From Dewey to a fully dynamic XML labeling scheme, ACM SIGMOD International Conference on Management of Data (SIGMOD 2009), pp. 719-730, 2009. https://doi.org/10.1145/1559845.1559921