The International Journal of Digital Curation
Volume 8, Issue 1 | 2013

Prenormative Research into Standard Messaging Formats for Engineering Materials Data

Tim Austin,
European Commission Joint Research Centre Institute for Energy and Transport

Chris Bullough,
Alstom Power

Dimitri Gagliardi,
Manchester Institute of Innovation Research,
University of Manchester

David Leal,
CAESAR Systems

Malcolm Loveday,
Beta Technology

Abstract

To qualify materials for specific applications comprehensive testing is necessary, and consequently the engineering materials community has developed an extensive collection of documentary testing standards to define test conditions, specimen configurations, and post processing and reporting procedures. Unfortunately, in the absence of corresponding data formats, test results are rarely conserved and their value diminishes as the material pedigree, test conditions, and results become disassociated. In an effort to address this issue, prenormative research has demonstrated the viability of deriving data formats from documentary testing standards and thus the possibility to realize a standards-based data infrastructure for the engineering materials community.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by UKOLN at the University of Bath and is a publication of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

 doi:10.2218/ijdc.v8i1.245

Introduction

As evidenced by a recent report from the Publishing Research Consortium, of the various categories of information available to the scientific community – including clinical guidelines, doctoral theses/dissertations, datasets/data models/algorithms, software programs, conference proceedings, and research article – the ratio of ease of access to importance of information is always lowest for the data category, irrespective of the type of organization, geographic location or discipline (PRC, 2010).

The engineering sciences lag behind the life sciences, the natural sciences, and the humanities when it comes to research data management, as indicated by recent Digital Curation Centre (DCC) and ERIM[1] reports (Ball and Neilson, 2010; Howard et al., 2010a, 2010b) and by the fact that the Engineering and Physical Sciences Research Council (EPSRC) is the only remaining UK Research Council that does not require researchers to submit data management or sharing plans in grant applications (DCC, 2012). However, with the engineering sector generating large volumes of data and its graduates typically being highly computer literate, there is considerable potential for improved data management practices to contribute to innovation and knowledge discovery. This is perhaps a motivating factor in the recent decision by the EPSRC to introduce and begin to implement a data policy (EPSRC, 2011).

While initiatives to promote improved data preservation in the engineering sciences are to be welcomed, it is important to recognize that an underlying information and communications technology (ICT) infrastructure is required to enable the preservation, transfer and reuse of research data in the various sub-disciplines of the engineering sciences. One such sub-discipline, and one which is relevant to all domains of engineering, is that of engineering materials. To qualify materials for specific applications, thorough testing is necessary. Consequently the engineering materials community has developed a comprehensive collection of documentary testing standards to define test conditions, specimen configurations, and post processing and reporting procedures. However, in the absence of corresponding data formats, test results are rarely conserved and their value diminishes as the material pedigree, test conditions and results become disassociated. In an effort to address this issue, a European Committee for Standardization (CEN) Workshop was started in 2009 with the objective of undertaking prenormative research into ICT standards for engineering materials data.

Standards-Compliant Data Formats

The CEN Workshop on the Economics and Logistics of Standards-compliant Schemas and ontologies for Interoperability of Engineering Materials Data (CEN/WS ELSSI-EMD) began in 2009 with the objective of establishing the viability of data formats derived from documentary testing standards for engineering materials. It combined three main strands intended to establish the viability of such standards-compliant data formats from technical, economic and standardization perspectives, as follows:

[1] The Engineering Research Information Management (ERIM) Research Project: http://www.bath.ac.uk/idmrc/erim

- **Technical:** Investigate the viability of standards-compliant schemas and ontologies for systems interoperability and the long-term preservation of test data in the engineering materials sector.
- **Economic:** Establish the business case for standards-compliant schemas and ontologies in the engineering materials sector.
- **Standardization:** Investigate the implications of standards-compliant schemas and ontologies for the documentary standards to which they correspond.

Of these three strands of activity, the technical work was concerned solely with the methodology for deriving electronic representations of engineering materials data and the delivery of accompanying reference implementations, while the economic and standardization strands were concerned with adoption by the engineering materials community.

The three strands of activity took place simultaneously over a 15 month period, and the findings of CEN/WS ELSSI-EMD were reported in CEN Workshop Agreement 16200:2010 (CWA, 2010b).

Data Formats

While the engineering materials community relies on an extensive collection of documentary testing standards, the work to develop data formats focussed on just one standard: the ISO 6892 Part 1 Standard for Ambient Temperature Tensile Testing (ISO, 2009). While the fact that the tensile test is one of the more common test types was an important factor in the selection of ISO 6892 Part 1, another important factor is that its annexes already incorporate rudimentary data formats (intended specifically for the qualification of algorithms for calculating derived values).

As anticipated, the ISO 6892 Part 1 standard provided a specification for defining computer readable representations of a tensile test. Although the schemas and ontology were developed largely independently, the results argue in favour of a common development process, as follows:

1. An ontology is created and documented using natural language text;
2. A unified modelling language (UML) model is created to depict the ontology;
3. An XML schema is created with the semantics of the elements defined by reference to the ontology, the UML model, or both.

In this process, the ontology is a development of the technical dictionary based on the terms and definitions defined in clause three of ISO 6892 Part 1, and the terms and symbols defined in clause four. The ontology adds relationships between the classes, which correspond to the terms in the technical dictionary, and rules. The relationships and rules are established through examination of the individual clauses that describe the test procedure, including test piece, conditions and properties, as described in clauses 6, 10, and 11 through 21, respectively.

 doi:10.2218/ijdc.v8i1.245

Based on ISO 6892 Part 1, CEN/WS ELSSI EMD delivered an ontology and a schema, both of which provide a computer readable interpretation of the documentary standard. Both provide a way to record tensile test data, either as OWL statements or as an XML document instance. For both the ontology and the schema, the UML diagrams allow the interpretation of the documentary standard to be reviewed by domain experts. Whilst the clarity and scope of the underlying standard contributed to the ease of the development process, the dialogue with engineering materials experts proved indispensable to clarify ambiguity (either of the meaning of a term or of a procedure).

All the processes involved in the computerization of the documentary standard are manual at present, including the creation of the technical dictionary, the ontology, the XML schema and the UML diagrams. Nonetheless, the links between the different modelling artefacts and their basis on the technical dictionary ensures that there is an explicit link back to the documentary standard. This approach has been documented in the ISO TC184/SC4 future standardization architecture proposal (ISO, 2010).

Business Analysis

The business analysis focussed on the impact of a transition from paper-based to electronic test certificates. It examined the test certificate lifecycle from material production to delivery to the original equipment manufacturer (OEM). A survey was undertaken to establish the value chain and gauge reactions to the proposed introduction of electronic test certificates. While the findings were positive, with broad agreement that the general concept is desirable, a risk was identified for software houses and organizations with well-developed test certificate generation and delivery processes.

Standardization Practices

The scope of the standardization practices efforts extended to assessing the implications for the documentary standards from which data formats are derived; establishing a mechanism for publishing the standards-compliant data formats; and to promoting their adoption by the ISO work group responsible for ISO 6892 Part 1, namely TC 164/SC 1/WG 4.

An assessment of the conventional path to standardization indicates that the adoption of standards-compliant data formats by technical committees responsible for the corresponding documentary standards has implications for all stages of the workflow. Since prenormative activities will include the development of an accompanying data format, these will need to be approved together with the natural language versions. This approval process will likely require a rigorous testing framework to validate the data formats. Further, for the lifetime of the documentary standard, there will be a need to synchronize the data formats with changes to the implementation technologies. Since these tasks require specialist services, their execution is best delegated to a competent authority. For ISO standards, there is a procedure already in place that allows tasks of a specialist nature to be delegated to a maintenance agency.

The preferred mechanism for publishing the data formats is as permanently available HTTP URIs hosted by CEN. Subsequent to CEN assessment on the use of URIs in the context of referencing digital objects that are part of a standards, digital object identifiers (DOIs) were selected as the most appropriate referencing system to be implemented in the context of referencing to digital objects referred to in CEN standards. A pilot phase is in progress whereby DOIs are assigned to freely available standards (Hoel, 2012).

The response to CEN/WS ELSSI-EMD from TC 164/SC 1/WG 4 has been positive but cautious. While there has been a commitment to reference the data formats from the ISO 6892 Part 1 informative annex on computer-controlled testing, there are development and maintenance implications that preclude the transfer of ownership of the CEN/WS ELSSI-EMD data formats at the present time. To ease the route to adoption, CWA 16200:2010 (CWA, 2010b) includes an Annex A template for incorporation into the documentary testing standard.

Discussion

The key outcome of CEN/WS ELSSI-EMD has been to demonstrate that the nature of documentary testing standards for engineering materials is such that they provide a specification from which corresponding data formats can be derived. Considering that there is an extensive collection of documentary testing standards for a range of test types and various sub-classes of engineering material, the outcome of CEN/WS ELSSI-EMD offers the engineering materials community a consistent approach to developing data formats for test data. Since documentary standards for qualification and testing are not limited to engineering materials, the approach could extend to other disciplines. For example, scientists in the life sciences rely on protocols to allow experiments to be recreated in their own laboratory and to ensure consistency of procedures. Although such protocols do not share the same degree of rigour in their maintenance and ratification as those of documentary testing standards for engineering materials, there is clearly the potential to derive corresponding data formats.

The fact that there are so many documentary standards for materials testing may itself prove to be an issue, insofar as maintaining data formats for a large collection of documentary standards may not be sustainable. Although this is a concern, the engineering materials community has already established a framework of processes and workflows for maintaining its extensive collection of testing standards, and concerns about the additional overheads that maintaining accompanying data formats will introduce will likely be resolved through consultation with stakeholders and the relevant ISO technical committees. For example, in circumstances where an ISO technical committee requires specialist services to develop a standard, ISO processes allow for delegation to a maintenance agency. It is also the case that the standardization process tends to consolidate complementary standards. The ISO 6892 Part 1 standard for ambient temperature tensile testing is in fact an example of exactly such consolidation. Bearing this in mind, it is likely to be the case that only a small proportion of the total number of engineering materials standards are in widespread use. Further, it is inevitable that testing standards will evolve to reflect the extent that electronic devices now pervade science and technology, and so it should be expected that standards bodies will be lobbied to extend the scope of testing standards accordingly. In this respect, CEN/WS ELSSI-EMD has demonstrated that by engaging

 doi:10.2218/ijdc.v8i1.245

stakeholders from both the information engineering and the engineering materials communities, it is perfectly feasible to extend documentary standards to meet the demands of the information age, including improved data preservation and traceability, and support for electronic transactions.

When developing technologies for preserving and sharing data, the importance of engaging with stakeholders cannot be underestimated. As other disciplines have discovered to their cost, unless end users are consulted during the development process, solutions can fail to achieve widespread adoption and hence fail to be effective (RIN and the British Library, 2009). In this respect, undertaking the work to demonstrate the viability of standards-compliant data formats in the context of a CEN Workshop proved to be particularly effective in engaging with a broad cross-section of stakeholders. As registered participants in the Workshop, stakeholders representing more than 30 organizations from different sectors of the engineering materials community worked together to contribute to, guide, and eventually approve the work of the CEN/WS ELSSI-EMD Project Team.

The measure of success of any standardization effort is the extent of adoption by the community it is intended to serve and ultimately the commitment of that community to continue to maintain the standard. By this measure, prior efforts to develop standards for representing engineering materials have met with limited success (OASIS, 2006; Swindells, 2009). As the cautious interest shown by ISO TC 164/SC 1 suggests, the early indications for the work of CEN/WS ELSSI-EMD are more promising, most likely because of the efforts made to frame the work in the standardization domain.

A commitment by standards organizations to maintain and develop ICT standards for engineering material data is just one step that needs to be taken towards their broader adoption by the engineering materials community. Additionally, the challenge of transitioning from the paper-based world of test certificates and reports to one based on electronic data will involve addressing the potential disruption that the new technologies will pose for established business practices. Although the findings of the CEN/WS ELSSI-EMD business analysis suggest the industrial sector recognizes the added value of a transition to electronic reporting, the concerns of larger organizations with well developed test certificate generation and delivery processes need to be addressed if widespread adoption is to be realized. A potential solution is to encourage the participation of such organizations in the standardization process, so that the in-house solutions and best practices developed by industrial organizations are incorporated into the ICT standards. Although such contributions to the standardization process may seem counterintuitive, insofar as being potentially damaging to commercial interests, there are precedents in the process industry based on the procedures described in CWA 16180:2010 (ORCHID) on standardizing information across the plant engineering supply chain (CWA, 2010a). The likely motivation for contributing internal standards is that the commercial advantage they offer is not sustainable in isolation, meaning that in the short term there are commercial benefits consequent to improved internal procedures but over the longer term these commercial advantages cannot outweigh the costs of maintaining the internal standards. In this circumstance there is benefit in rolling them out to other actors in the supply chain, so that the efficiency of the whole supply chain improved.

A particularly effective means to rollout the internal standards is to transfer their ownership to standardization organizations.

Although the outcomes reported in CWA 16200:2010 (CWA, 2010b) demonstrate the viability of standards-compliant data formats for engineering materials test data, further work is necessary to deliver a solution that is sufficiently comprehensive. Most notably, due to the failure to ratify the work of the Organization for the Advancement of Structured Information Standards (OASIS) Material Markup Committee (OASIS, 2006) there is an absence of any recognised representation of an engineering material. For example, in the case of a tensile test, while CWA 16200:2010 provides the formats for the test data, there is no recognised means to represent the data corresponding to the material on which the test was performed, such as manufacturing route and heat treatment. Thus, there is a requirement for a standard representation of an engineering material. This and other recommendations of CWA 16200:2010 are the subject of a follow up CEN Workshop on Standards for Electronic Reporting in the Engineering Sector (CEN/WS SERES), the objective of which is to develop and demonstrate standards for representing and reporting engineering materials data (European Committee for Standardization, 2012), as depicted in Figure 1.

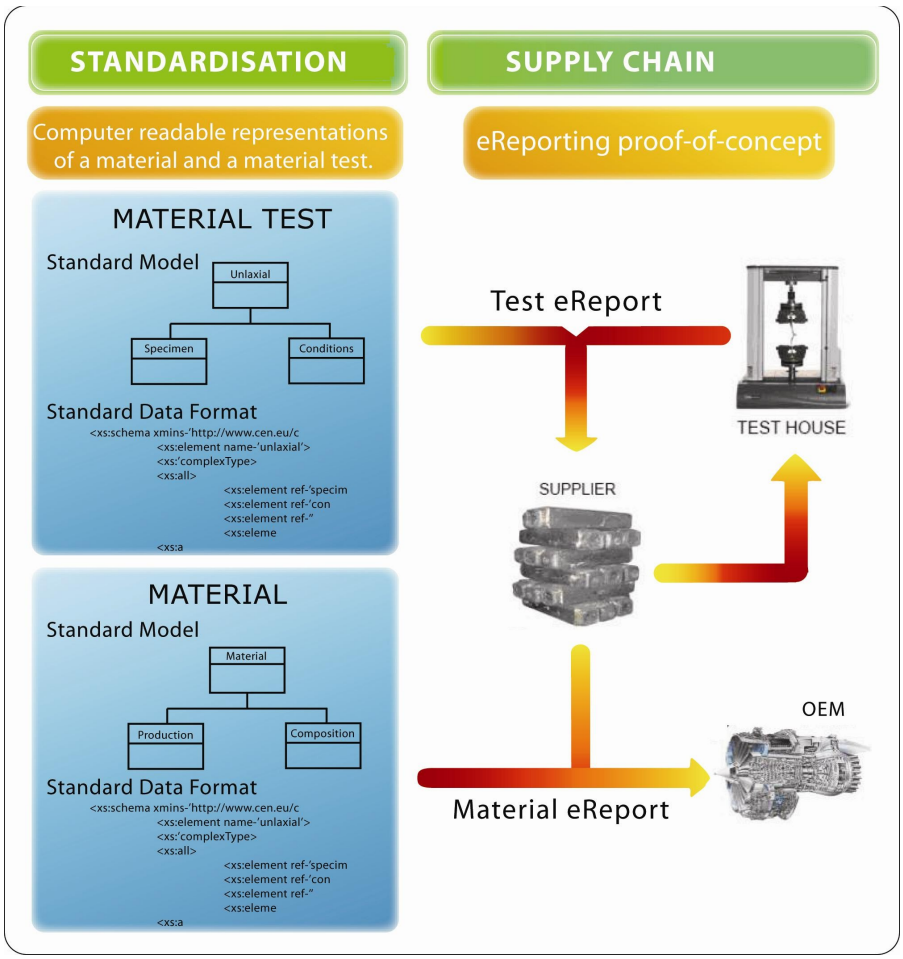

Figure 1. The use of standards for representing and reporting engineering materials data in the supply chain.

Conclusions

Prenormative research into standard messaging formats for engineering materials has demonstrated that documentary standards provide a specification from which corresponding data formats can be derived. While the path to ratification and adoption of such data formats is necessarily lengthy and involved, data formats derived from documentary testing standards have the potential to provide the engineering materials community with an effective and sustainable means of preserving, exchanging and reusing engineering materials test data.

Acknowledgements

The financial support of the EC and EFTA for CEN/WS ELSSI-EMD is gratefully acknowledged, as is the help of Simone Roos at the Institute for Energy and Transport in producing the graphics for this publication.

References

Ball, A. & Neilson, C. (2010). Curation of research data in the disciplines of Engineering. SCARP Case Study 7, Digital Curation Centre. Retrieved from http://www.dcc.ac.uk/sites/default/files/documents/publications/case-studies/SCARP_B4812_EngCase_v1_2.pdf

CWA 16180:2010. (2010). The CEN ORCHID Roadmap standardising information across the plant engineering supply chain: Part 2 Implementation guide. Retrieved from ftp://ftp.cen.eu/CEN/Sectors/List/ICT/CWAs/CWA16180-2_ORCHID implementation.pdf

CWA 16200:2010. (2010). A Guide to the development and use of standards-compliant data formats for engineering materials test data. European Committee for Standardization. Retrieved from ftp://ftp.cen.eu/CEN/Sectors/List/ICT/CWAs/CWA16200_2010_ELSSI.pdf

DCC. (2012). Curation policies and support services of the main UK research funders. Digital Curation Centre. Retrieved from http://www.dcc.ac.uk/webfm_send/778

European Committee for Standardization (2012). *Business plan for the CEN Workshop on Standards for Electronic Reporting in the Engineering Sector (WS SERES)* (CEN/WS SERES Document No. N005). Retrieved from http://www.cen.eu/cen/Sectors/Sectors/ISSS/Activity/Pages/SERES.aspx

EPSRC. (2011). EPSRC policy framework on research data. Engineering and Physical Sciences Research Council. Retrieved from http://www.epsrc.ac.uk/about/standards/researchdata/Pages/policyframework.aspx

Hoel, T. (2012). Unique identifier challenge. Retrieved from http://wiki.teria.no/display/cenwslt/Unique+Identifier+Challenge

Howard, T., Darlington, M., Ball, A., Culley, S. & McMahon, C. (2010a). Understanding and characterizing engineering research data for its better management. ERIM Project Document, University of Bath. Retrieved from http://opus.bath.ac.uk/20896/1/erim2rep100420mjd10.pdf

Howard, T., Darlington, M., Ball, A., Culley, S. & McMahon, C. (2010b). Opportunities for and barriers to engineering research data re-use. ERIM Project Document, University of Bath. Retrieved from http://opus.bath.ac.uk/21166/1/erim3res100805tjh10.pdf

ISO 6892-1:2009. (2009). Metallic materials - Tensile testing Part 1: Method of test at ambient temperature. International Organization for Standardization.

ISO TC 184/SC 4 document N2615. (2010). Industrial data integrated ontologies and models (IDIOM) architecture specification.

OASIS. (2006). Material markup language public review draft 01, 06 June 2006. Organization for the Advancement of Structured Information Standards. Retrieved from http://docs.oasis-open.org/materials/materials-matml-spec-pr-01.pdf

PRC. (2010). Access vs. importance: A global study assessing the importance of and ease of access to professional and academic information phase 1 results. Publishing Research Consortium. Retrieved from http://www.publishingresearch.net/documents/PRCAccessvsImportanceGlobalNov2010_000.pdf

RIN and the British Library. (2009). Patterns of information use and exchange: case studies of researchers in the life sciences. Research Information Network. Retrieved from http://www.publishingresearch.net/documents/RINPatterns_information_use-REPORT_Nov2009.pdf

Swindells, N. (2009). The representation and exchange of materials and other engineering properties. *Data Science Journal, 8,* 190-200. doi:10.2481/dsj.008-007