

# MatDB—Promoting the Conservation and Exchange of Engineering Materials Research Data

T Austin <sup>(1)</sup>, J Rombouts <sup>(2)</sup>, H Over <sup>(3)</sup>

<sup>(1)</sup> *SOASYS Limited*

78 York Street, W1H 1DP, London, UK

Email: tim.austin@soasys.eu

<sup>(2)</sup> *Delft University of Technology*

TU Delft Library, Prometheusplein 1, 2628 ZC Delft, The Netherlands

Email: j.p.rombouts@tudelft.nl

<sup>(3)</sup> *European Commission Joint Research Centre Institute for Energy and Transport*

Postbus 2, 1755 ZG, Petten, The Netherlands

Email: hans-helmut.over@jrc.nl

## ABSTRACT

In the engineering sciences, the European Commission Institute for Energy and Transport hosts an Online Data and Information Network (ODIN) in support of energy and transport research. The facility consists of a collection of online databases organized into four main categories: documents, engineering, nuclear, and product information. In the engineering category, MatDB Online is a database for engineering materials test data that has benefited from many hundreds of person years invested in the development of a robust data model, comprehensive test support, and an intuitive user interface. Recently, it has been enabled to deliver innovative added value services on top of its basic data management features, which together with community building efforts, are designed to promote the adoption of effective data management practices in the engineering materials research sector. The services include support for ICT Standards delivered by a recently completed CEN Workshop, RESTful Web Services in support of systems integration and data pipelining, and DOIs for data. This paper describes the individual services, the technical approach to integrating with OEM systems and software, and the potential impact of the added value services on the engineering materials research community.

Keywords: engineering, alloy, test, data, standards, datacite, doi

## INTRODUCTION

As recent studies indicate, there are various obstacles to promoting the effective capture, conservation and exchange of data in the engineering disciplines, not least of which is that any suggestion of sharing data may compromise long-established but fragile links with the industrial sector [1][2]. However, while there are certainly challenges, these are unlikely to prove insurmountable. Looking to the life sciences for example, it could well be argued that patient confidentiality and competition in the pharmaceutical sector pose greater barriers to developing practices for conserving and sharing data, and yet the life sciences are at the vanguard of the emerging disciplines of web and data science. Further, as indicated in recent survey of engineering materials researchers [3], while ease of use and security are higher on the requirements list of researchers, there is certainly an interest to share and reuse data. Given this interest, and considering the very significant resources invested in generating engineering materials tests data, it is reasonable to assume that the availability of well-designed data management systems will play an important role in the realization of effective data management practices in the engineering sector.

## **Engineering Materials Test Data**

Engineering materials are an integral part of many manufacturing processes, from the fabrication of nanoscale computer processors to the production of plasma containment vessels in fusion reactors. Huge resources are invested in the development, production, and qualification of materials, and yet the data generated by these activities are rarely conserved and reused. Instead, the tests to qualify materials are often repeated, so that not only is the original investment squandered but also the cycle of wasted investment continues. The conservation of data is important for many reasons, including validation, traceability, reuse, and re-evaluation.

## **MatDB Online**

At <https://odin.jrc.ec.europa.eu> the European Commission Institute for Energy and Transport hosts an Online Data and Information Network (ODIN) in support of energy and transport research. The facility consists of a collection of online databases organized into four main categories: documents, engineering, nuclear, and product information. In the engineering category, MatDB Online is a sophisticated database for engineering materials data that has benefited from many hundreds of person years invested in the development of a robust data model, comprehensive test support, and an intuitive user interface. It contains over 40.000 test results and provides a web-interface for data content, data entry, data retrieval, and analysis routines. The database covers mechanical and thermo-physical properties data of engineering alloys generated in accordance with international material standards and recommendations.

## **ENGINEERING MATERIALS RESEARCH DATA MANAGEMENT**

In any discipline, delivering effective data management solutions is as much about engaging with the end-user and delivering services that add real value to the research process as it is about technology. While there is considerable merit in the efforts of the digital curation community to promote research data management, critical review indicates that the research community has yet to be convinced of the merits of embedding data conservation and exchange into mainstream research. One problem identified by a recent review of information use and exchange in the life sciences is that the technologies and tools developed by data management practitioners often fail to meet the requirements of the broader research community whom the technologies are intended to serve [4]. Coupled with continued skepticism of the merits of conserving and sharing data [5], embedding effective data management practices into mainstream research is a problem that remains unresolved. Although the delivery of well constructed and robust repositories for data is a necessary component of a sustainable data management solution, it is simply the foundation, and added value services are required to ensure that researchers benefit in a tangible way from conserving and sharing the fruits of their labour. In recognition of this issue, MatDB Online has been enabled to deliver innovative added-value services on top of its basic data management features, which together with community building efforts, are designed to promote the adoption of effective data management practices in the engineering materials research sector. The services include support for ICT Standards for materials test data delivered by CEN WS/ELSSI-EMD [6], RESTful Web Services in support of systems integration and data pipelining, and DOIs for data [7][8].

## **Standards-compliant Data Formats**

The underlying MatDB database schema defines five entities (source, materials, specimen, condition, and test) and for each test type the database structure reflects international test Standards [9][10]. This compliance with Standards extends to support for the Standards-compliant data formats delivered by CEN/WS ELSSI-EMD, thereby enabling interoperability of heterogeneous systems.

## **Data Pipelining**

Delays storing test data result in their value diminishing as the sample pedigree, test conditions and results become disassociated. The opportunity to transfer data directly from the test facility offers an effective means to ensure data are captured and conserved. A recently implemented prototype (available

at <https://odin.jrc.ec.europa.eu/dataentry/XMLConsole.html>) demonstrates how the use of Web Services has enabled MatDB Online to integrate with testing facilities without the need to modify OEM software.

## Data Citation

Motivations for researchers to share their data include establishing a means to acknowledge their contribution to the research process in a tangible way. Making data citeable in much the same way as publications offers just such an opportunity. The DataCite initiative makes this possible, and MatDB Online supports assigning DOIs to individual data sets. As the following example demonstrates, this allows researchers in the engineering materials sector to cite their data in much the same way as a publication:

1. Data for a Strain-Controlled LCF test performed at 650°C and 0.053 Hz on Udimet, JRC Petten Institute for Energy and Transport, <http://dx.doi.org/10.5290/1000000010021>.

The DOI in the example is resolved (by the DOI lookup service) to a page hosted at ODIN that displays a summary of the data set and links to the full data set.

## CONCLUSION

In the engineering sector, considerable resources are invested in the generating engineering materials test data. MatDB Online is an internationally recognized data centre designed to manage experimentally measured engineering alloys data. It is, however, recognized that added value services are required if effective data management practices are to become embedded in mainstream research. Thus, beyond the basic data entry and retrieval features, the robust data model, and extensive test type support that MatDB offers, its added-value services include a RESTful test facility to database data pipelining service, standards-based systems interoperability, and citable data (using DOIs).

## REFERENCES

- [1] Curation of research data in the disciplines of Engineering (2010). Retrieved 12<sup>th</sup> Septmeber, 2011 from [http://www.dcc.ac.uk/sites/default/files/documents/publications/case-studies/SCARP\\_B4812\\_EngCase\\_v1\\_2.pdf](http://www.dcc.ac.uk/sites/default/files/documents/publications/case-studies/SCARP_B4812_EngCase_v1_2.pdf).
- [2] Howard, T., Darlington, M., Ball, A., Culley, S., McMahan, C., 2010. Understanding and Characterizing Engineering Research Data for its Better Management. Project Report. Bath, UK: University of Bath, (ERIM Project Document erim2rep100420mjd10).
- [3] Materials Data Centre - Report on User Requirements. Retrieved 12<sup>th</sup> September, 2011 from [http://www.materialsdatacentre.com/uploads/7/3/0/0/730051/jisc\\_mdc\\_user\\_requirements\\_report.pdf](http://www.materialsdatacentre.com/uploads/7/3/0/0/730051/jisc_mdc_user_requirements_report.pdf).
- [4] Research Information Network and the British Library, Patterns of information use and exchange: case studies of researchers in the life sciences, November 2009. Retrieved 12<sup>th</sup> September, 2011 from <http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/disciplinary-case-studies-life-sciences>.
- [5] Steven Wiley, Why Don't We Share Data? The Scientist, Volume 23, 2009, p 33-36.
- [6] CWA 16200:2010 A Guide to the Development and Use of Standards-compliant Data Formats for Engineering Materials Test Data. Retrieved 12<sup>th</sup> September, 2011 from [ftp://ftp.cen.eu/CEN/Sectors/List/ICT/CWAs/CWA16200\\_2010\\_ELSSI.pdf](ftp://ftp.cen.eu/CEN/Sectors/List/ICT/CWAs/CWA16200_2010_ELSSI.pdf).
- [7] DataCite. Retrieved 12<sup>th</sup> September, 2011 from <http://www.tib-hannover.de/fileadmin/datacite/index.html>.
- [8] J Starr and A Gastl. isCitedBy: A Metadata Scheme for DataCite. Retrieved 12<sup>th</sup> September, 2011 from <http://www.dlib.org/dlib/january11/starr/01starr.html>.
- [9] M Nagy, HH Over and E Wolfart. XML related data exchange from the test machine to a web-enabled MAT-DB. Data Science Journal, Volume 4, 8 December 2005.
- [10] T Ojala and HH Over. Approaches in using MatML as a common language for materials data exahcnge. Data Science Journal, Volume 7, November 2008.