Materials Discovery xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Materials Discovery

journal homepage: www.elsevier.com/locate/md

Towards a digital infrastructure for engineering materials data

Q1 Tim Austin

European Commission, Joint Research Centre (JRC), Institute for Energy and Transport (IET), Innovative Technologies for Nuclear Reactor Safety Unit, Westerduinweg 3, 1755 LE Petten, Netherlands

ARTICLE INFO

Article history:
Received 16 December 2015
Accepted 18 December 2015
Available online xxx

Keywords:
Engineering materials
Data formats
Standardization
Data citation
Interoperability

ABSTRACT

The industrial and research sectors make significant investments in developing and producing engineering materials. These materials are manufactured and qualified in accordance with a body of product and testing standards that have evolved over many decades to meet ever more demanding requirements. Yet the very significant volume of data that result from these activities remains largely unavailable. While efforts to establish a digital infrastructure for engineering materials data can be traced back more than three decades, widespread adoption of machine-readable formats to enable the routine transfer of engineering materials data has yet to be realized. Given the reliance on electronic systems in all aspects of engineering materials development, manufacture, and qualification, it is an anomaly that simply preserving and transferring engineering materials data remains an issue. This anomaly is accentuated by the fact that in recent years other business domains have benefited from the integration of web technologies into established business models. To address these shortcomings a digital infrastructure is needed that allows and encourages the seamless transfer of engineering materials data between different systems. It is in this context that renewed efforts to develop standard formats for engineering materials data are taking place in the frame of CEN Workshops. While building on prior activities at ASTM and ISO, this work leverages existing product and testing standards with a view to engaging the engineering materials community. With preliminary technical specifications having already been demonstrated to streamline the data transfer process, attention is turning to the long-standing challenge of promoting a culture of data sharing. Whereas previously the motivations for researchers and industrial organizations to share data were lacking, the initial impacts of the DataCite framework for data citation on the utilization of the European Commission materials database hosted at http://odin.jrc.eu.europa.eu are suggestive of a sea-change in data sharing and reuse. This paper describes the status of the work to develop data formats for engineering materials in the frame of CEN Workshops and reports on the added value of data citation beyond simply ensuring that data creators are properly accredited for their work. It also reports the outcome of work to enable the European Commission materials database to support standards compliant data formats and data citation, whereby the barriers to systems integration have been considerably reduced and, irrespective of the level of confidentiality, organizations in both the industrial and research sectors now routinely enable their data sets for citation. Together with recent innovations in digital publishing, a renewed interest in the development of standards for engineering materials data offers new prospects for discovery, exchange, and reuse of engineering materials data. Taken along with other data centric initiatives, such as physics-based and multi-scale modelling, Open Data, and linked data, it can reasonably be argued that standard data formats and data publishing herald a transition towards a digital infrastructure for engineering materials data.

1. Introduction

Q2 The drivers for conserving and exchanging data are no longer confined to the vested interests of individuals and organizations wanting to preserve their intellectual and financial investments in data creation activities. Instead, government, institutional, funding agency, and publisher policies demand more responsible approaches to data management. Coupled with increased opportunities for innovation made possible by the web and technologies for publishing and transferring data, organizations and communities that fail to adopt business practices that leverage Web-connectivity risk losing advantage. Although technologies for engineering materials data exist, to date they have largely failed to find widespread

E-mail address: timothy.austin@ec.europa.eu

http://dx.doi.org/10.1016/j.md.2015.12.003
2352-9245/

adoption [1]. Similar failings with semantic technologies in the life sciences suggest that the failure to achieve widespread adoption can be traced to insufficient engagement with the broader community [2]. There is thus a strong case for revisiting the development of technologies for engineering materials data with a view to ensuring the wider engineering materials community is properly engaged. The task at hand is thus to establish a digital infrastructure for engineering materials data that will allow organizations to compete effectively in a digital market. This paper aims to demonstrate that in the context of its testing and product standards there exists a robust framework on which to develop data formats and that in combination with data citation, there are very real opportunities to gain added value.

1.1. European policy

In the scope of its nuclear safety standardization interests, the Joint Research Centre of the European Commission supports the development of standard formats for engineering materials data and the use of these data formats for efficient storage and transfer of nuclear materials information [3]. More generally, the European Commission is also committed to ensuring that data resulting from publicly funded research become publicly accessible, usable and re-usable through digital e-infrastructures; that datasets are made easily identifiable and can be linked to other datasets and publications through appropriate mechanisms, and additional information is provided to enable their proper evaluation and use; and that institutions responsible for managing public research funding and academic institutions that are publicly funded assist in implementing national policy by putting in place mechanisms enabling and rewarding the sharing of research data [4,5].

1.2. Research data management

In recent years the sciences have witnessed a fundamental change in attitude to research data management. This is evidenced by various phenomena, including the widespread adoption of data management policies; the ever increasing posts for data specialists; and innovations in data publishing, all of which is indicative of data management becoming integral to the mainstream research process.

1.2.1. Data management policies

A 2010 report by the Publishing Research Consortium (PRC) on importance versus ease of access to scientific outputs indicated that irrespective of discipline, geographic location, organization type, and respondent demographics, data sets are regularly identified as having high value but poor accessibility [6]. Such findings played a role in convincing funding agencies to develop policies for research data management [7], so that today there is a global trend towards more responsible research data management practices. For example, all UK Research Councils have adopted data management policies [8]. In combination with support from agencies providing the necessary technical support, such as the Digital Curation Centre (DCC) and the Joint Information Systems Committee (JISC), this circumstance has led to many higher education institutes in the UK now stipulating data management policies [9]. Similarly at the European level, the European Commission is committed to the more effective use of data generated through publicly funded activities [10]. In the domain of European research, this is manifest in the Horizon 2020 Open Data policy [11]. Beyond Europe, the global perspective was set over a decade ago by the OECD, which at the request of Science and Technology Ministers established guidelines to facilitate cost-effective access to digital research data from public funding [12]. More recently, this perspective has been reiterated by the G8 Science Ministers in a statement that makes explicit reference to the fact that open scientific research data should be easily discoverable, accessible, assessable, intelligible, useable, and wherever possible interoperable to specific quality standards [13].

In the context of this widespread recognition of the inherent worth of research data and its potential for reuse, the Research Data Alliance (RDA) was established in March 2013 with the support of the European Commission, the US Administration and Australian Government. The purpose of the RDA is to accelerate international data-driven innovation and discovery by facilitating research data sharing and exchange, use and re-use, standards harmonization, and discoverability. To achieve these objectives, the RDA is focused on the development and adoption of infrastructure, policy, practice, and standards. In the domain of engineering materials, these objectives are being pursued by the recently established RDA/CODATA Materials Data, Infrastructure & Interoperability interest group (MDII IG). While the MDII IG is one of many RDA interest groups tackling discipline-specific issues, there are also a large number of cross-cutting interest groups, several of which are concerned with different aspects of data publishing.

1.2.2. Trends in data publishing

The emergence of data journals and publisher policies on data management are suggestive of a paradigm change in the way publishers perceive data and their place in the publication process. While some publishing houses host their own repositories, a key development in data publishing came with the establishment in 2009 of the DataCite Consortium and its framework for data citation based on digital object identifiers (DOIs). This framework allows any data centre that meets DataCite quality requirements to enable the data sets it hosts for citation. This circumstance provides researchers with a wide choice of potential hosts for their data. As shown in Fig. 1, a query of the form http://search.datacite.org/api?q=*&rows=0&facet=on&facet.range=minted&facet.range.gap=%2B1YEAR&facet.range.start=2009-01-01T00:00:00Z&facet.range.end=2016-01-01T00:00:00Z indicates that of the more than six million DataCite DOIs minted since 2009, over a third have been registered in 2015.

Another example of innovation in data publishing is Data In Brief, an Elsevier journal that provides the opportunity to publish a comprehensive description of data sets that is often not feasible in Section 2 of a traditional scientific publication. As shown in Fig. 2, Data In Brief statistics posted at http://www.sciencedirect.com/science/journal/23523409 indicate an exponential rise in data articles over the six quarters since the journal came into existence, with approximately 300 publications expected in 2016 Q1.

Such growth in data publication and citation is entirely consistent with the results of the PRC study on importance versus ease of access to scientific outputs, the findings of which indicated the importance to researchers of improved access to research data.

1.3. Technologies for engineering materials data

Engineering materials data covers a broad spectrum of data types, including mechanical properties; physical properties; chemical properties; microstructures; phase properties; kinetics; and thermodynamic properties [15]. Recent initiatives such as the Materials Genome Initiative (MGI) and integrated computational materials engineering (ICME) have given new momentum to the development of technologies for engineering materials data. With their significant dependence on ICT, both the MGI and ICME are presently the key drivers for technologies for engineering materials data.

In the scope of the Materials Innovation Initiative (MII) of the US Administration, the MGI aims to reduce the time-to-market of newly developed materials by half. The MGI relies on a better integration of related activities, including experiment, computation,

Please cite this article in press as: T. Austin, Towards a digital infrastructure for engineering materials data, Mater. Discov. (2016), http://dx.doi.org/10.1016/j.md.2015.12.003