CrossMark

Building a Materials Data Infrastructure

TERRY T. WONG[1,2]

1.—Aerojet Rocketdyne, Los Angeles, CA, USA. 2.—e-mail: terry.wong@rocket.com

It is not uncommon to read a glowing review of the latest sports car in a car magazine but it would be unusual to find in that magazine an article praising the layout of the freeway infrastructure in Los Angeles. Tours are often given for people to admire a building that Frank Lloyd Wright designed, but it is probably not likely that anyone would be interested in a tour about electrical infrastructure that gives power to that building. Foundations and infrastructures are rarely noticed let alone praised. In many cases, infrastructures are designed to be unnoticed; for example, no one wants to be aware of the wastewater and sewage infrastructure for the city they live in. While unnoticed and rarely praised, no one would doubt the necessary role that infrastructures play.

The Integrated Computational Materials Engineering (ICME) Committee of The Minerals, Metals & Materials Society (TMS) has as its mission to promote the emerging area of ICME. Material models and computational tools are key aspects of ICME. However, these models and tools depend on a variety of computational and experimental data as both inputs and for validation. The materials data infrastructure is a foundational aspect of ICME and the Materials Genome Imitative (MGI)[1] which allows for data to be stored and shared. The following papers are grouped into two aspects of this data infrastructure; the first group of papers deal with informatics infrastructure needed to share data and the second group covers the development of a "language" needed to share a specific type of materials data.

A materials informatics infrastructure is needed to facilitate the sharing of data. Kyle Michel et al. show how the platform that Citrine Informatics has developed facilitates the storage of structure and unstructured data and the ease by which that data may be retrieved. The collaboration platform developed at the University of Michigan called "The Materials Common" is described by Brian Puchala et al. An important feature of Materials Common is the ability for researchers to electronically duplicate their workflow and capture their data as a natural part of the work they are doing. The data captured can then be shared with others. One of the hurdles in sharing data is the difficulties associated with publishing both the data and metadata that gives meaning and context to the data. The Materials Data Facility (MDF) described by Ben Blaiszik et al. shows how MDF offers researchers the ability to store both the data and metadata. As the use of MDF grows, the authors suggest that other researchers will build upon what was previously stored. The last paper in this first group is written by Alden Dima and his colleagues at the National Institute of Standards and Technology (NIST). Dima et al. describe the work that NIST has been doing to meet two major needs in the informatics infrastructure: the need for materials researchers to store data and metadata-based community-developed standards and templates; and the need for researchers to be able to find data in a decentralized way.

While the first group of papers deal with the infrastructure needed to store and share data in general, the second group of papers revolve around the need to have a common language or semantics to share particular types of material data. Dominik Steinberger et al. note that there is a difficulty in analyzing and comparing dislocation microstructure in a concise manner. They offer a universal microstructure language that will help facilitate this gap. In addition to the need to have a common language to share dislocation microstructure, Elke Schaberger-Zimmermann et al. show how 3D microstructure data may be shared via HDF5 using hypereutectic AlSi alloy as an example case. Lastly, Maciol and Regulski understand both the difficulty in capturing various states and data associated with multiscale models. In their paper they describe a process by which semantic descriptions may be developed for multiscale models using thermos-mechanical treatment of metal alloys as an example.

The following papers being published under the topic of Building a Materials Data Infrastructure provide excellent details and research on the subject. To download any of the papers, follow the url http://link.springer.com/journal/11837/68/8/page/1 to the table of contents page for the August 2016 issue (vol. 68, no. 8).

- "The Citrination Platform: A Comprehensive Materials Data Infrastructure" by Jordan O'Mara, Bryce Meredig, and Kyle Michel.
- "The Materials Commons: A Collaboration Platform and Information Repository for the Global Materials Community" by Brian Puchala, Glenn Tarcea, Emmanuelle. A. Marquis, Margaret Hedstrom, H.V. Jagadish, and John E. Allison.
- "The Materials Data Facility: Data Services to Advance Materials Science Research" by B. Blaiszik, K. Chard, J. Pruyne, R. Ananthakrishnan, S. Tuecke, and I. Foster.
- "An Informatics Infrastructure for the Materials Genome Initiative" by Alden Dima, Sunil Bhaskarla, Chandler Becker, Mary Brady, Carelyn Campbell, Philippe Dessauw, Robert Hanisch, Ursula Kattner, Kenneth Kroenlein, Marcus Newrock, Adele Peskin, Raymond Plante, Sheng-Yen Li, Pierre-François Rigodiat, Guillaume Sousa Amaral, Zachary Trautt, Xavier Schmitt, James Warren, and Sharief Youssef.
- "A Universal Approach Towards Computational Characterization of Dislocation Microstructure" by Dominik Steinberger and Stefan Sandfeld.
- "Hypereutectic AlSi Alloy: Gathering of 3D Microstructure Data" by E. Schaberger-Zimmermann, M. Mathes, and G. Zimmermann.
- "Development of Semantic Description for Multiscale Models of Thermomechanical Treatment of Metal Alloys" by Piotr Maciol and Krzysztof Regulski.

REFERENCE

1. National Science and Technology Council, Materials Genome Initiative for Global Competitiveness (Executive Office of the President, Washington, DC, 2011), http://www.whitehouse.gov/sites/default/files/microsites/ostp/materials_genome_initiative-final.pdf.