Results of the 2012 TMS/ASM Intersociety Scoping Session on Materials Data Management

Organized by The Minerals, Metals, & Materials Society (TMS) and ASM International
In conjunction with the MS&T'12 Conference

February 11, 2013
Based on the workshop conducted on October 10, 2012

INTRODUCTION

The Materials Genome Initiative (MGI) seeks to provide the United States with the infrastructure, workforce, and innovations needed to discover, develop, manufacture, and deploy advanced materials significantly faster, and at a fraction of the cost. Key enablers of the MGI Materials Innovation Infrastructure (MII) are the development of advanced computational and experimental tools, and the management of digital data across such platforms and amongst a variety of organizations. Thus, the future MII necessitates representation from government, academia, and industry. Professional societies can play a leading role in driving the materials science and engineering (MSE) community throughout these organizations toward the management and sharing of digital data and tools, within a collaborative atmosphere.

In this regard, in conjunction with the MS&T'12 Conference in 2012, ASM International and The Minerals, Metals, & Materials Society (TMS) jointly convened a session composed of more than 25 members from an array of professional societies that represent various categories of MSE-topics and a wide range of data types and issues. This session was in large part stimulated by an intersociety MGI-strategic scoping session held at the MS&T'11 conference (October 2011)[1] in which a number of the final report recommendations emphasized the importance of management and sharing of digital data across various organizations as required in the establishment of a future MII. The workshop was also planned in coordination with the "Materials Genome Initiative Workshop – Building the Materials Innovation Infrastructure (MII): Data and Standards"[2] that was led by the National Institute of Standards and Technology (NIST) and held on May 14-15, 2012 in Washington, DC.

A list of the attendees of the MS&T12 data management scoping session is included in Appendix A. Participants assembled into small groups for four hours to address the goals of the workshop. These goal were to:

- Identify the **primary categories of data** associated with MSE with respect to structural materials
- Assess the **current states of** the sharing of primary data and tools in terms of
    - **Familiarity** within the MSE community regarding available data

---

[1] http://materialsinnovation.tms.org/docs/pdfs/2011MGIScopingSession.pdf
[2] http://nvlpubs.nist.gov/nistpubs/ir/2012/NIST.IR.7898.pdf

- **Accessibility** of available data to the MSE community
- **Maturity and uptake** of the data

DISCLAIMER

This report was prepared through the collaborative efforts of The Minerals, Metals, & Materials Society (TMS) and ASM International. The views and opinions of the individuals participating in the workshop do not necessarily state or reflect those of the professional societies with which they are affiliated.

PRIMARY CATEGORIES OF DATA

In a facilitated brainstorming session, participants were asked to present the primary categories of data related to materials science and engineering. Members of professional societies within or related to the MSE community typically bring a diverse assembly of perspectives to the table regarding which kinds of data are of greatest importance to them, and how such data can best be classified. This list of data categories was vetted and condensed to reduce the overlap of similar data classes. Additionally, the workshop participants debated the proper level of specificity to reach a balance without letting their designations become too broad or narrow in scope. Table 1-1 below gives the final list of primary data categories. A description of each of these data categories is provided in Appendix B.

Primary Categories of Materials Data	
*Acoustical Data	Materials Emission Data (e.g., photons, electrons)
*Aesthetic Data	Materials Meta-data
Atomic Potential Data	Microstructure Data
Composition Data	Mobility Data (e.g., diffusion)
Constituent Material Data	*Nuclear property data
Cost Data	Optical Data
Dynamic Mechanical Property Data (e.g., viscosity)	Oxidation/Corrosion Data
Electrical Data	Phases Data
End-Use Mechanical Property Data	Processing Data
Environmental/Health/Safety Data (e.g., legislation)	*Qualitative Data
Failure Data (wear/fracture/corrosion)	*Statistical Data
Fundamental Mechanical Property Data	Surface Morphology Data (e.g., roughness)
Kinetic Data	Thermodynamic Data
*Magnetic Data	Thermophysical Data
Manufacturing Performance Data (e.g., materials fidelity)	*Validation Data

Table 1-1: The primary categories of materials data selected by the workshop participants.

**Categories marked with an asterisk were not further rated for metrics (see below) because discussion time was limited and the workshop participants decided to focus in the time allotted on the categories without asterisks.*

TYPES/VARIETIES OF DATA

Workshop participants next considered these primary categories in terms of how the data can be used or characterized. Data in their initial form can be considered raw, or can be further derived and redefined. It is possible for data to be defined in terms of the techniques used to gather them (e.g., X-ray data) or by the properties characterized by the data (e.g., microstructure properties). They can further be defined by whether the values are fundamental in nature or derived from specific applications. Table 1-2 lists some of the ways that primary data categories can be expressed.

Types/Varieties of Data
Raw Data
Metadata (descriptors about the raw data)
Derived Data
Method (used to obtain the data)
Model Verification and Validation Data
Design Values
Statistical Data

Table 1-2: The variety of ways that data can be designated and/or used.

STATES OF SHARING DATA

Participants then organized into smaller groups to assess the current states of sharing data, and tools to develop such data, within the final list of primary data categories. To accurately represent these states of sharing, the data categories were rated for familiarity, accessibility, and maturity of the data:

Familiarity: *How familiar is the MSE community with the available data (and tools to develop such data) within the primary data categories?* Familiarity is often a function of how regularly the data are used or accepted by the MSE community. There are cases where the data may already be accessible and mature, yet the larger community may still not be aware of the data or tools, or comfortable with their use.

Accessibility: *Is it difficult to obtain or access the necessary data and tools of interest?* Accessibility can be defined by limitations such as proprietary classifications, export controls, data-transfer limitations (hardware or software), or other constraints. Data may have some familiarity and reputability to those intimately using the data, but the greater MSE community cannot easily procure those data.

Community Uptake/Maturity: *Are the data and tools mature or ready for commercial implementation?* Data requires provenance to strengthen its reliability with regard to its "readiness" to be implemented into commercial applications. Maturity is a consequence of extensive research, experimentation, documentation, and validation that prove the accurate characterization of a material.

Using the final list of primary data categories (Table 1-1), the groups agreed on suitable ratings to assess the state of sharing for the primary data categories. While many topics were deemed "primary," not all of them were chosen during the rating portion of the workshop, primarily due to time restrictions. This approach allowed the groups to independently select the data categories of most relevance to them, and then work in parallel.

Figure 1-1 presents a portrayal of the current state of sharing data within these primary categories. Each primary category of data can also encompass a significant number of subtopics. Some examples of segmenting into subtopics include materials classifications (e.g., ceramics, polymers, metals), or industry-specific applications (e.g., mining, aerospace). Therefore, some counterparts for specific applications are also presented in Figure 1-1, below the primary data category. The primary categories that are parenthetically denoted as “All” thus represent a collective rating of all specific and nonspecific subtopics. Some additional comments on how the results in Table 1-1 and Fig. 1-1 were obtained, and potential limitations on interpreting these results, are presented in Appendix C.

The results presented in Table 1-1 and Fig. 1-1 not only provide guidance as to the various data categories that are important to the session participants representing a variety of sub-disciplines and professional societies, but also provide insight into the relative ability and/or need to enable data access and sharing within a future MII. For example, Fig. 1-1 suggests that corrosion data may be ripe for data sharing across organizations and communities, whereas materials-related cost data or ceramics and polymers mechanical property data may present more significant challenges that first need to be addressed. Also, Fig. 1-1 suggests that there are many types of data with which scientists and engineers may be quite familiar, but that familiarity often does not directly correlate with the level of maturity or accessibility of such data. The reverse trend - low familiarity and high accessibility or maturity – is not observed.

The observations in this report comprise a building block or first step from which to develop an open structure for sharing materials-related data across organizations. Before undertaking a contribution to such an effort, the interested party should identify the category of data targeted and the level of familiarity, accessibility, and maturity (or community uptake) of shared data in that category. Data categories that possess high metrics in these areas may be ripe for data sharing and interaction, while data within categories that scored low in these areas may require significant efforts to raise these metrics before these data categories become candidates for interactivity across a MGI MII. Alternatively, data categories that rate low with respect to community familiarity, accessibility, and maturity may offer important opportunities for new contributions to the field, if the need for data sharing within the MSE community is sufficiently high. In either case, building the digital data component of the materials innovation infrastructure will take a concerted, collaborative effort throughout the sub-communities and professional societies of the participants of this workshop, and beyond.

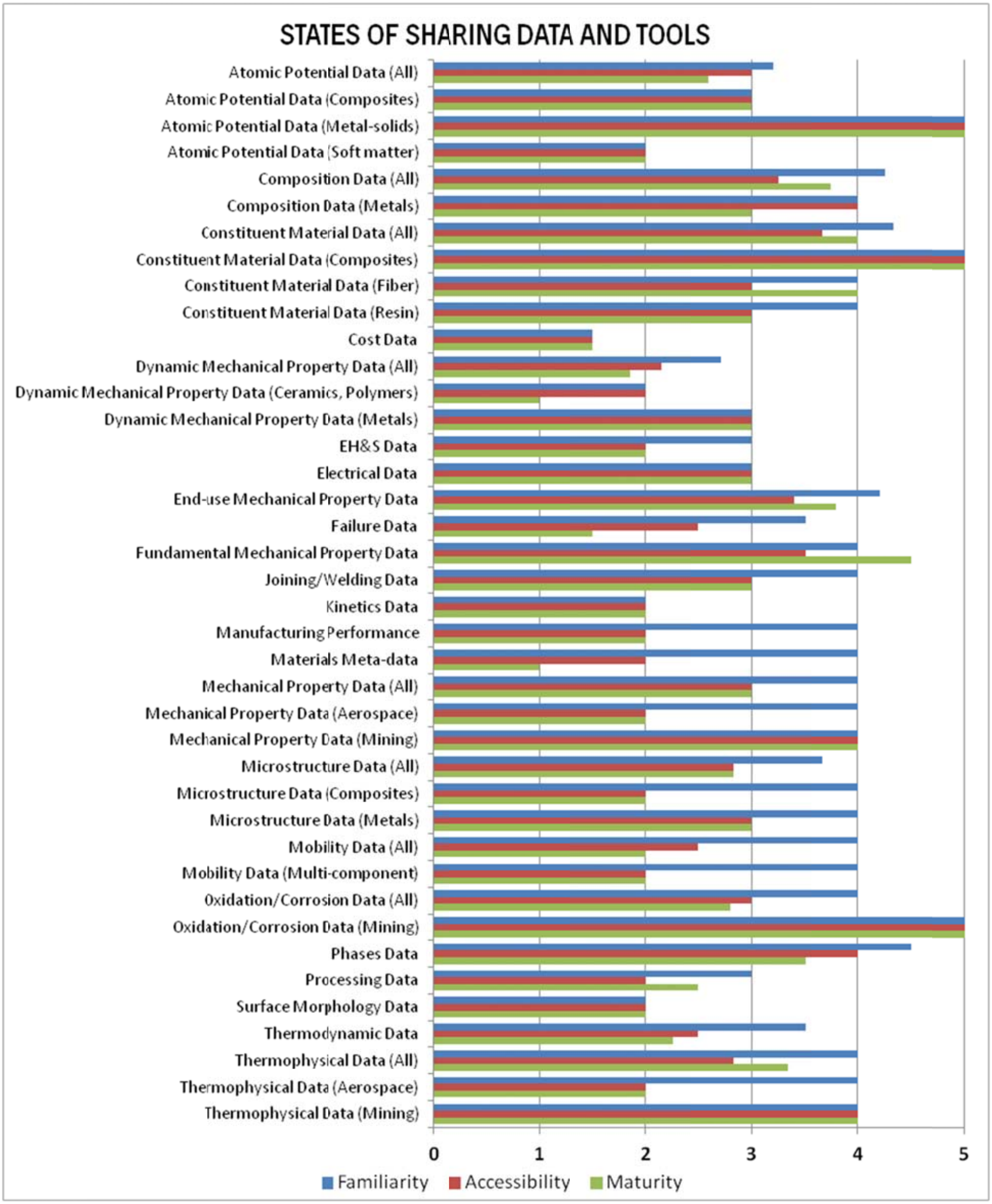

Figure 1-1: Average ratings of familiarity, accessibility, and maturity of each data category. (0 = lowest; 5 = highest)

NEXT STEPS

Participants from the various professional societies expressed a commitment to further communicate with their respective societies and sub-communities to discuss the workshop results as they relate to the MGI and the need to establish a MII to share data and data generation tools. Furthermore, the participants voiced interest in brainstorming and examining methods for working with their respective societies to incentivize the sharing of data and tools by the society members. Such approaches will be particularly important as the materials community moves toward sharing more complex data, including three- or four-dimensional data pertaining to both existing and new materials and processing approaches. The most commonly mentioned course of action by participants was to immediately present the workshop results to the appropriate management personnel within their organizations, and identify which of the primary data categories their professional societies have both interest and activity in, with regard to using and generating digital data in these areas.

ASM International and The Minerals, Metals & Materials Society intend to continue working with the MSE community to support the development of the MGI MII. In particular, TMS will use the results of this report to support the digital data (and data generating tools) elements of the MII through the TMS Cyberinfrastructure Portal[3], and a current collaboration with the NIST to develop an on-line MGI Digital Data Community. ASM intends on using these workshop results to bolster their efforts in establishing a Computational Materials Data Network (CMDN)[4]. The CMDN is a center for data collection and distribution in the MSE community, and is directly related to the interests of the MGI.

This report will be provided to the workshop attendees and steering committees, and will also be disseminated to the broader materials science and engineering community.

[3] http://materialsinnovation.tms.org/cyberPortal.aspx
[4] www.cmdnetwork.org

MATERIALS SCIENCE AND ENGINEERING PROFESSIONAL SOCIETY AFFILIATIONS OF PARTICIPANTS

- ACerS – The American Ceramic Society
- ACS – American Chemical Society
- AFS – American Foundry Society
- APS – American Physical Society
- ASC – American Society of Composites
- ASCE – American Society of Civil Engineers
- ASM International – "The Materials Information Society," formerly known as the American Society for Metals
- ASME – American Society of Mechanical Engineers
- ASNT – American Society for Nondestructive Testing
- ASTM International – (No acronym), formerly known as the "American Society for Testing and Materials"
- AWS – American Welding Society
- IOM^3 – The Institute of Materials, Minerals and Mining
- MRS – Materials Research Society
- NACE – "The Corrosion Society," formerly known as the National Association of Corrosion Engineers
- SAE International – (No acronym), formerly known as the Society of Automotive Engineers
- SAMPE – Society for the Advancement of Material and Process Engineering
- SEM – Society for Experimental Mechanics
- SME – Society of Mining, Metallurgy and Exploration
- STLE – Society of Tribologists and Lubrication Engineers
- TMS – The Minerals, Metals & Materials Society

Appendix A:

ATTENDANCE

Name	Email
Laura Bartolo	lbartolo@kent.edu
Ross Brindle	rbrindle@nexightgroup.com
Eric Brown	en_brown@lanl.gov
Kristine Chin	krisc@aiche.org
Kathryn Dannemann	kathryn.dannemann@swri.org
Eileen De Guire	edeguire@ceramics.org
Dennis M. Dimiduk	dennis.dimiduk@wpafb.af.mil
Mary Doyle	mdoyle@sae.org
Stephen Freiman	steve.freiman@comcast.net
David Furrer	david.furrer@pw.utc.com
Somnath Ghosh	sghosh20@jhu.edu
John Grubb	john.grubb@atimetals.com
Norma Jane Hair	nhair@astm.org
Scott Henry	scott.henry@asminternational.org
Elizabeth Holm	eaholm@andrew.cmu.edu
Richard Huff	huffrk@cat.com
Jared Kosters	jkosters@nexightgroup.com
Ted Lynch	ted@strategicmi.com
Will Marsden	will.marsden@grantadesign.com
Matt Miusa	mpur4@cornell.edu
Tom Prucha	tprucha@afsinc.org
Robert Quarshie	robert.quarshie@materialsktn.net
Vilupanur Ravi	vravi@csupomona.edu
John Rumble	jumbleusa@earthlink.net
Benjamin Scheiner	bscheiner@sae.org
Greg Schoeppner	greg.schoeppner@wpafb.af.mil
Scott Shuey	scott.shuey@newmont.com
Dan Smith	dsmith@astm.org
George Spanos	gspanos@tms.org
Curtis Taylor	curtis.taylor@ufl.edu
Doug Ward	doug.ward@ge.com
Ashley White	ashley.ann.white@gmail.com
Terry Wong	terry.wong@pwr.utc.com

Appendix B:

DESCRIPTIONS OF PRIMARY DATA CATEGORIES
(Note: these are descriptions and not all encompassing definitions)

- **Acoustical Data:** Data describing the generation and/or transmission of sound in different types of materials, as well as related materials properties.
- **Aesthetic Data:** Quantitative or qualitative descriptions of the degree to which various products appeal to customers or users that have bearing on engineering decisions.
- **Atomic Potential Data:** Interatomic potential data comprise information regarding energies of position for different types of materials and provide a basis for ab initio materials modeling.
- **Composition Data:** This type of data comprises measurements of the elemental or molecular content of a material.
- **Constituent Material Data:** Data that describes the properties of constituent materials when they are used in a multi-material system (i.e., a composite).
- **Cost Data:** The data used in the cost-analysis aspect of decision-making. Includes the costs associated with different engineering scenarios from project inception to culmination.
- **Dynamic Mechanical Property Data (e.g., viscosity):** Mechanical property data associated with the *rate* of mechanical processes in a material; particularly high rates of deformation.
- **Electrical Data:** Values such as the resistivity and conductivity of various kinds of materials under specified conditions.
- **End-Use Mechanical Property Data:** The mechanical properties of a material such as ductility and strength, after all processing has been completed and the final product is in use.
- **Environmental/Health/Safety Data (e.g., legislation):** Data such as environmental regulations, information on health effects of different materials, and instructions for safe handling of materials. These types of data provide important boundary conditions for materials science and engineering activities at all stages.
- **Failure Data (wear/fracture/corrosion):** Data describing the properties of various types of materials relating to types of failure such as wear, fracture, fatigues, and corrosion.
- **Fundamental Mechanical Property Data:** Related to fundamental mechanical processes such as deformation. Such data includes elastic coefficients, stress tensors, etc.
- **Kinetic Data**: Data describing the movement of materials or of particles within a material at various scales and under different conditions.
- **Magnetic Data:** Data describing the magnetic properties of materials of various classes and how these properties depend upon material structure and composition.
- **Manufacturing Performance Data (e.g., materials fidelity):** Data describing the consistency in the production of products and the quantification of deviation from specifications.
- **Materials Emission Data (e.g., photons, electrons):** Data surrounding the quantity and types of particles (such as photons and electrons) emitted by materials under specified conditions.
- **Materials Meta-data:** Data used to document the pedigree and/or interpret materials data. This includes data that does not always make it into a graph or table such as material processing

history, alloy composition, data and time of the generation, as well as experimental or computational platforms used to generate the data.

- **Microstructure Data:** Data describing the internal physical structure of a material at a relatively high resolution.
- **Mobility Data (e.g., diffusion):** Theoretical and/or experimental data of transport (or mobility) phenomena; e.g., diffusion coefficients.
- **Nuclear property data:** Data describing the radioactive behavior (the intensity and type of emissions) of materials under different conditions, and/or the interactions of materials with various types of radiation to which they are exposed.
- **Optical Data:** Data describing the behavior and properties of light as it interacts with various materials.
- **Oxidation/Corrosion Data:** Includes oxidation/reduction potentials of various materials as well as empirical data based on controlled experiments and case studies of objects that experience corrosion.
- **Phase Data:** Data that describes the type, quantity, and properties of the different phases within a material.
- **Processing Data:** Data describing various types of processing steps of a material such as heat-treating, stamping, cold rolling, etc.
- **Qualitative Data:** Data, often descriptive in nature, that has not or cannot be quantified.
- **Statistical Data:** Statistical information for any type of data set that is gathered.
- **Surface Morphology Data (e.g., roughness):** Data describing the shape and structure of materials surfaces and interfaces, such as measurements of curvature, surface roughness, etc.
- **Thermodynamic Data:** Data surrounding the relationships between heat, work, and various types of energy in materials systems.
- **Thermophysical Data:** Differentiated from thermodynamic data in that it relates heat with other physical properties. This includes for example coefficients of thermal expansion and thermal conductivity.
- **Validation Data:** The data needed to establish confidence that a materials model has relevance to a particular physical phenomenon within a specified range. Typically, this will involve comparing experimental data to theoretically obtained values.

Appendix C:

ADDITIONAL FACTORS AND OBSERVATIONS

There are several other factors and observations to consider with regard to how the primary data categories of interest were chosen (Table 1-1), as well has how they were rated (Fig. 1-1):

- Participants observed the level of complexity of data of subtopics within the primary categories of data chosen. “Simple” materials tend to be affordable, attainable, and easier to characterize in terms of data capture, and therefore carry higher metrics for data sharing (Fig. 1-1).
- Pre-competitive materials data tended to receive higher, more accurate ratings as a result of multidisciplinary and cross-sector collaboration. These data sets may be of particular interest to the future MII with regard to their capacity to be shared among the MSE community. However, the ratings were based on a consensus decision by relatively small number of participants offering ratings from their specific discipline and experience base and should be interpreted as such.
- Consideration was given to age of the data. Older data sets tend to exist in paper form, are not digitized, and may have lower resolution compared to data captured with newer techniques. However, older data may have a long pedigree and therefore a greater maturity level. Participants deliberated on exactly how these variables affected the states of sharing the data and tools to ensure a fair assessment. This included the feasibility of transforming certain “paper data” into digital data to provide a proper assessment.
- The ratings of the data and tools of interest were based primarily on existing data with less emphasis on three-dimensional (3D) and time-dependent (4D) data. While participants openly acknowledged the need for advancing 3D/4D data in the interest of the MGI they were more accustomed to evaluating the state of existing data within the time permitted.
- The participants rated a mix of categories with and without subtopics. While each data category included an average rating (denoted as “All”), this collective mix of ratings may not accurately portray the state of sharing data as a true average of all subcategories in that primary data category. This would otherwise require a larger sample size or a stricter rating system.
- Experts within particular subtopics or disciplines may have a richer understanding of the data that they are accustomed to using. The data may be proprietary, utilized often, and also be commercially mature to an exclusive group of experts while the wider MSE community may know little of the data’s existence.