The Materials Genome Initiative, Data, Open Science, and NIST

James A Warren
Technical Program Director for Materials Genomics
Material Measurement Laboratory
National Institute of Standards and Technology
Executive Secretary, NSTC Subcommittee on MGI

Science advances one funeral at a time -Max Planck
The Perfect is the Enemy of the Good -Voltaire
NIST’s Mission

To promote U.S. innovation and industrial competitiveness by advancing measurement science, standards, and technology in ways that enhance economic security and improve our quality of life.
NIST at a Glance

Major Assets
- ~ 2800 employees ~(50/50 technical/admin)
- ~ 2600 associates and facilities users
- ~ 1600 field staff in partner organizations (Manufacturing Extension Partnership)

Major Programs
- NIST Laboratories
- Baldrige Performance Excellence Program
- Hollings Manufacturing Extension Partnership
NIST Products and Services

Measurement Research
- ~ 2,200 publications per year

Standard Reference Data
- ~ 100 different types
- ~ 6,000 units sold per year
- ~ 226 million data downloads per year

Standard Reference Materials
- ~ 1,300 products available
- ~ 30,000 units sold per year

Calibration Tests
- ~ 18,000 tests per year

Laboratory Accreditation
- ~ 800 accreditations of testing and calibration laboratories
The Materials Genome Initiative, Data, Open Science, and NIST

James A Warren
Technical Program Director for Materials Genomics
Material Measurement Laboratory
National Institute of Standards and Technology
Executive Secretary, NSTC Subcommittee on MGI

Science advances one funeral at a time - Max Planck
The Perfect is the Enemy of the Good - Voltaire
LAG FROM DISCOVERY TO APPLICATION OF NEW MATERIALS...


- Teflon
- Velcro
- Titanium production
- Polycarbonate
- GaAs
- Diamond-like thin films
- Amorphous soft magnets
- Lithium-ion batteries
- Core-shell electro-catalysts for fuel cells
- Catalysts for olefin metathesis

We can do better!

“To help businesses discover, develop, and deploy new materials twice as fast, we’re launching what we call the Materials Genome Initiative. The invention of silicon circuits and lithium ion batteries made computers and iPods and iPads possible, but it took years to get those technologies from the drawing board to the market place. We can do it faster.”

-President Obama (6/11)
There are two groups of people that don’t like the name Materials Genome Initiative
Develop a Materials Innovation Infrastructure

Achieve National goals in energy, security, and human welfare with advanced materials

Equip the next generation materials workforce
THE MATERIALS GENOME INITIATIVE

to decrease time-to-market by 50% while <$$

Develop a Materials Innovation Infrastructure

Achieve National goals in energy, security, and human welfare with advanced materials

Equip the next generation materials workforce

Materials Genome Initiative for Global Competitiveness
DESIGNING MATERIALS TODAY
DESIGNING NEW MATERIALS IN THE FUTURE

Materials Innovation Infrastructure

- Computational Tools
- Experimental Tools
- Digital Data

- Human Welfare
- Clean Energy
- National Security
- Next Generation Workforce

Discovery | Development | Property Optimization | Systems Design and Integration | Certification | Manufacturing | Deployment
Formulating the NIST Role in MGI
Today’s Approach to Computational Materials Design
SCOPE: Goals of Initiative at NIST

To foster widespread adoption of the MGI Paradigm both across and within the multitude of materials development ecosystems

Goal 1: NIST establishes *essential materials data and model exchange protocols*

Goal 2: NIST establishes the *means to ensure the quality* of materials data and models

Goal 3: NIST establishes *new methods, metrologies and capabilities* necessary for accelerated materials development.
Enable & Enhance Exchange

• Develop and deploy repositories
• Develop and disseminate materials informatics infrastructure
  ─ Enable data discovery through tools and standards
  ─ Capture data from scientific workflows and archival sources
  ─ Engage with stakeholders to determine needs and disseminate best practices
• Integrate across length and time scale
• Build and Test infrastructure through Pilots
Assess & Improve Quality

• Validate Experiments and Models
• Verify Model accuracy
• Quantify Uncertainty
• Quantify Sensitivity
• Define Next Generation of Experiments and Models
New **Methods and Metrologies**

- Develop Data Driven Materials Science
- Integrate with Modeling Expertise
- Enable Crowdsourced/Open Science
- Achieve targets in Materials by Design/ICME
Enable & Enhance Exchange (repositories, disciplines, industries; standards)

Assess & Improve Quality (Data & Models)

New Methods and Metrologies (data driven analysis and models)

Materials w/ Targeted Properties
Data Sharing is Important Beyond MGI & NIST

OSTP “Public Access” Memo
Feb 22, 2013

Executive Order
May 9, 2013

1. Policy Principles

The Administration is committed to ensuring that, to the extent possible and consistent with law and the public interest, the results of federally funded scientific research are made available to the scientific community. Such results include peer-reviewed publications and scientific data in digital formats.

Scientific research supported by the Federal Government drives our economy. The results of that research benefit progress in areas such as health, energy, the environment, and national security.

Access to digital data sets resulting from federally funded research and development is essential to understanding and exploiting the potential of the emerging science. It underpins the forecasting industry, and it is essential to the development of new products and services related to curation, preservation, analysis, and interpretation. These data sets provide the platform for the re-use and re-use of the data and the results of the research. The ability of researchers to access and use the data is important to their ability to advance science.

By the authority vested in me as President by the Constitution and the laws of the United States of America, it is hereby ordered as follows:

Section 1. General Principles. Openness in government strengthens our democracy, promotes the delivery of efficient and effective services to the public, and contributes to economic growth. As one vital benefit of open government, making information resources easier to find, accessible, and usable can fuel entrepreneurship, innovation, and scientific discovery that improves Americans’ lives and contributes significantly to job creation.

Decades ago, the U.S. Government made both weather data and the Global Positioning System freely available. Since that time, American entrepreneurs and innovators have utilized these resources to create new products and services. These products and services have been sold to businesses and governments in all corners of the world.

A strong science and technology infrastructure provides the foundation for innovation and growth. Federal investments in research and development have driven America’s technological progress, creating new industries and creating jobs across the country.

This Executive Order directs Executive agencies to ensure that Federal research and development (R&D) investments in science and technology result in the production of products and services that can be made available to the public for use in commercial applications.

This policy is consistent with the goals of the OSTP Public Access Memo (Feb 22, 2013) and the OMB Open Data Memo (May 9, 2013).
Office of Data and Informatics
Material Measurement Laboratory

Robert J. Hanisch, ODI Director
About ODI

• Overall goal: establish the technical infrastructure and foster a culture of first-class data management for MML, eventually for all of NIST
• Near term initiatives
  • Update Standard Reference Data collection
    • Web-based user interfaces
    • Application Programming Interfaces (APIs)
  • Implement OMB/OSTP open access data policy
    • Laboratory-wide Data Management Plans
    • NIST-wide Enterprise Data Inventory, data.gov
  • Build solution inventory for data management systems
    • Storage
    • Metadata
    • Electronic Lab Notebooks
  • Informatics / data analytics consulting
• ODI works closely with Materials Genome Initiative
end
Some Initial Partnerships and Modalities
NIST Data Efforts

Collaborations

- ASM International: Structural Data Demonstration Project
- DOE/EERE Kinetics of Cast Mg Alloys
- Journals collaboration
  - IMMI
  - Others under discussion

Data Dissemination

- DSpace: WebFT Repository
- ThermoML Archive
- Interatomic Potential Repository

Data Processing

- Uncertainty Analysis
- Data Analytics
- Bench marking activities (DFT)

Data Capture

- TRC: Guided Data Capture
- Materials Data Curator

Ontologies Schemas (XML based)

NLP

Workflow Tools

Semantic Media Wiki

Meta data standards

Bench marking activities (DFT)
Another way of telling the story

• Industry needs good data and models
  – How do they get them? (exchange)
  – Can they trust them? (quality)
  – How can they use them best and who has the talent? (new methods)
NIST Supports New Open Collaboration Modalities (Move to LinkedIn, Join us!)

The MGI Digital Data Community
Developing the Materials Innovation Infrastructure to Support the Materials Genome Initiative

What's New
ANNOUNCEMENTS

Welcome to the MGI Digital Data Community
By: David Howe, Mar 12, 2013 12:32 PM

The MGI (Materials Genome Initiative) Digital Data Community is now open for community-building, discussions, and more. This is a great forum for advancing the technical conversation surrounding the digital data topics and challenges facing the materials science and engineering field and its allied disciplines. Make

Our Community
LATEST DISCUSSIONS

RE: What is Digital Data?
By: Tony Fast, Mar 8, 2013 10:50 AM
Posted in: MGI Digital Data Community

I agree completely... More

Atomistics workshop to focus on validating interatomic...
By: Chandler Becker, Mar 4, 2013 10:13 AM
Posted in: Model Validation

Site Navigation and Information
ABOUT THE WEBSITE

The MGI Digital Data Community allows users to build and join communities surrounding specific technical disciplines and topics, especially focused on the creation and sharing of data. These communities provide a forum for discussions; sharing documents, slide shows, and videos; notifying other members of upcoming events, and more. Create a user profile and join the MGI
Questions to Consider
(MRS/TMS Survey)

• Federal Agencies will soon (or may already) require, the archiving of materials science and engineering research digital data generated with Federal funding. How will you respond? What policies would be achieve the greatest impact?

• What actions should stakeholders (e.g., professional societies, universities, research institutions, libraries, scientific publishers, industry) take to archive key digital data generated in materials science and engineering?

• What categories of materials science and engineering digital data would be best maintained by specific stakeholders?

• What role could professional societies play in the development of standards for digital data and metadata for materials science and engineering?

• Are professional societies and peer-reviewed journal editors the most appropriate stakeholder to establish community practices for ensuring digital data discoverability and citation in materials science and engineering?

• What quality metrics are needed (e.g., verification, validation, uncertainty quantification, etc.) to adequately ensure confidence in materials science and engineering data generated from both experiment and computation and how should these metrics be associated with the data set?

• What additional actions can peer-reviewed journals take to ensure materials science and engineering publications well-describe the experimental and computational details associated with generating digital data that are necessary for reproducibility?
Office of Data and Informatics

**SRD**
- continue existing SRD distribution
- Quality Framework
- SRD Modes
- assess external need
- new product ideas
  - SRMDS
  - data streams
  - alternative delivery methods
- Open Data Initiative
- Open Govt Directive
- Data.gov

**Research Data**
- deal w/ data deluge
- provide advice to MML bench staff
- gather best practices
- interpret external rules & regulations
- reduce redundancy
- promote cooperation and coherent action
- manage changes in scholarly publishing
- coordinate with
  - WERB
  - Library
  - JResNIST

**Lead/Liaison**
- partner with ITL
- represent MML
  - NIST committees
  - NSTC & IWGs
  - NIH, NSF, DOE
  - other Fed Govt
  - Research Data Alliance (RDA)
- data standards
- champion proposals
  - budget initiatives
  - IMS
  - inter-agency, RDA

**Data Science**
- The 4th paradigm?
  - will it stand next to
    - theoretical
    - experimental
    - computational
  - Cloud
  - Statistical Learning
  - Big Data
  - Knowledge Discovery
  - very fast growing
  - many new jobs
  - new degrees/depts
Goal: Establish well-pedigreed and curated demonstration datasets for non-proprietary metallic structural materials data over all length scales.

NIST’s role
- Provide data schemas and meta-data formats for diffusion and phase equilibria data.
- Provide sample diffusion and phase equilibria data for the Al-Mg-Si system.
- Use expanded TRC Guided Data Capture program with available binary and ternary phase equilibria literature.
- Expand use and implementation of DSpace Repository.
- Link with developing ontology and semantic web tools.

March 2014: Phase 1 release.
June 2014: Phase 2 release.
Dec 2014: Project Completion.
• High-Throughput Study of Diffusion and Phase Transformation Kinetics of Magnesium-Based Systems for Automotive Cast Magnesium Alloys: J-C Zhao and A. Luo, The Ohio State University

• Phase Transformation Kinetics and Alloy Microsegregation in High Pressure Die Cast Mg Alloys: John Allison, University of Michigan

• In-situ Investigation of Microstructural Evolution During Solidification and Heat- Treatment in a Die-Cast Mg Alloy: Aashish Rohatgi, Pacific Northwest Laboratory

• A systematic multiscale modeling and experimental approach to protect grain boundaries in Mg alloys from corrosion: Mark Horstemeyer at Mississippi State and Santanu Chaudhuri at Washington State

• Corrosivity and Passivity of Metastable Mg Alloys ---An Introductory Study to Future Stainless Mg Alloys: Guang-Ling Song, ORNL

• Dealloying, Microstructure and the Corrosion/Protection of Cast Mg Alloys, Karl Sieradzki, Arizona State U.
New Research Data Alliance IG

RDA/CODATA Materials Data, Infrastructure & Interoperability IG

James A Warren and Laura Bartolo, Co-Chairs
THE NATIONAL DATA SERVICE

The National Data Service is an emerging vision of how scientists and researchers across all disciplines can find, reuse, and publish data. It is an international federation of data providers, data aggregators, community-specific federations, publishers, and cyberinfrastructure providers. It builds on the data archiving and sharing efforts under way within specific communities and links them together with a common set of tools.

VISION

It is widely believed that ubiquitous digital information will transform the very nature of research and education. The reasons for this excitement are clear: in essentially every field of science, simulations, experiments, instruments, observations, sensors, and/or surveys are generating exponentially growing data volumes. Information from different sources and fields can be combined to permit new modes of discovery. Data, including critical metadata and associated software models, can capture the precise scientific content of the processes that generated them, permitting analysis, reuse, and reproducibility. By digitizing communication among scientists and citizens, discoverable and shareable data can enable collaboration and support repurposing for new discoveries and cross-disciplinary research enabled by data sharing across communities. Open, shareable data also promise to transform education, society, and economic development.

However, while some communities are making progress in developing discipline-specific data services, the U.S. and international scientific communities lack a unified framework and supporting services for storing, sharing, and publishing data: for locating data; or for verifying data. More specifically, we are lacking standard means of
NIST Repositories

The National Institute of Standards and Technology is establishing essential data exchange protocols and mechanisms for widespread adoption to ensure quality materials data and models and to foster data sharing and reuse.

- Computational File Repository
  - Atomistics Simulations
  - CALPHAD Assessments
  - First Principles Simulations
  - Other Computational Methods

- Experimental Data Repository
  - Diffusion Data
  - Mechanical Properties
  - Other Experimental Data
  - Phase Equilibria and Thermodynamic Data

- NIST/DOE-EERE Advanced Automotive Cast Magnesium Alloys
  - A systematic multiscale modeling and experimental approach to protect grain boundaries in magnesium alloys from corrosion
  - Corrosivity and Passivity of Metastable Mg Alloys
  - Dealloying, Microstructure and the Corrosion/Protection of Cast Magnesium Alloys
  - High-Throughput Study of Diffusion and Phase Transformation Kinetics of Mg-Based Systems
  - In-situ Investigation of Microstructural Evolution During Solidification and Heat-Treatment in a Die-Cast Magnesium Alloy
  - Phase Transformation Kinetics and Alloy Microsegregation in High Pressure Die Cast Magnesium Alloys

- NIST Thermodynamics and Kinetics Test Space

https://materialsdata.nist.gov/
NIST Center of Excellence for Advanced Materials
THE MGI SUBCOMMITTEE (SMGI)

Who we are

- MGI Subcommittee, Committee on Technology, NSTC
- First meeting April, 2012
- Co-chairs: OSTP (Wadia), NIST (Locascio), DOE (Horton)
- Executive Secretary: NIST (Jim Warren)

What we do

- Coordinate across government
- Convene stakeholders to engage in strategy: Grand Challenge Summits
- Development of a National Strategy for MGI
What does a functioning Infrastructure look like?

Figure 1. Schematic of how the design criteria for a given material dictate the needed material properties and thus define the needed experiments, models and data.
Why an MGI Now?

Materials Are Complicated Systems
Modeling is a Challenge

- Advanced materials are complex: multi-component and multi-phase
- Without adequate modeling, informatics and data exchange, the development of next generation materials using empirical approaches is bogged down by their complexity
- The Materials Genome Initiative seeks to advance materials design capabilities to promote faster, cheaper

Alloy cooled from 300 °C

Alloy cooled from 800 °C

- Composition and processing affect properties
- Phases change as a function of processing
- Microstructures consist of mixtures of multiple material phases
- Finer microstructure results in a much stronger alloy
How do we do it?

An infrastructure for Open Science

Materials w/ Targeted Properties
Community-based Curated Repositories

Data

Models

MGI Ecosystems

Materials w/ Targeted Properties
Enable & Enhance Exchange (repositories, disciplines, industries; standards)

Data
Repositories
Models

Experiment
Simulation
Materials w/ Targeted Properties

Quantum Nano Micro Macro
Data
Repositories
Models

Assess & Improve Quality
(Data & Models)

Materials w/ Targeted Properties
MGI NATIONAL STRATEGY: 4 GOALS

- Enable a Paradigm Shift in Culture
- Integrate Experiments, Computation, Theory
- Facilitate Access to Materials Data
- Equip the Next-Generation Materials Workforce

What Should NIST Do?
Halt!