



# National Materials Data Initiatives

**Chuck Ward**

**Materials & Manufacturing Directorate**

***Integrity ★ Service ★ Excellence***



# Overview



- Policy drivers
- Discipline-based approaches
- Repository efforts
- Collaborative environments
- Supporting technology/efforts

***We don't have a big data problem ...***

***... we have a big problem with data.***



# OSTP Direction on Public Access



EXECUTIVE OFFICE OF THE PRESIDENT  
OFFICE OF SCIENCE AND TECHNOLOGY POLICY  
WASHINGTON, D.C. 20503

February 22, 2013

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: John P. Holdren, *JPH*  
Director

SUBJECT: Increasing Access to the Results of Federally Funded Scientific Research

## I. Policy Principles

The Administration is committed to ensuring that, to the greatest extent and with the fewest constraints possible and consistent with law and the objectives set out below, the direct results of federally funded scientific research are made available to and useful for the public, industry, and the scientific community. Such results include peer-reviewed publications and digital data.

Scientific research supported by the Federal Government catalyzes innovative breakthroughs that drive our economy. The results of that research become the grant for new insights and are assets for progress in areas such as health, energy, the environment, agriculture, and national security.

Access to digital data sets resulting from federally funded research allows companies to focus resources and efforts on understanding and exploiting discoveries. For example, open weather data underpins the forecasting industry, and making genome sequences publicly available has spawned many biotechnology innovations. In addition, wider availability of peer-reviewed publications and scientific data in digital formats will create innovative economic markets for services related to curation, preservation, analysis, and visualization. Policies that mobilize these publications and data for re-use through preservation and broader public access also maximize the impact and accountability of the Federal research investment. These policies will accelerate scientific breakthroughs and innovation, promote entrepreneurship, and enhance economic growth and job creation.

The Administration also recognizes that publishers provide valuable services, including the coordination of peer review, that are essential for ensuring the high quality and integrity of many scholarly publications. It is critical that these services continue to be made available. It is also important that Federal policy not adversely affect opportunities for researchers who are not funded by the Federal Government to disseminate any analysis or results of their research.

To achieve the Administration's commitment to increase access to federally funded published research and digital scientific data, Federal agencies investing in research and development must have clear and coordinated policies for increasing such access.

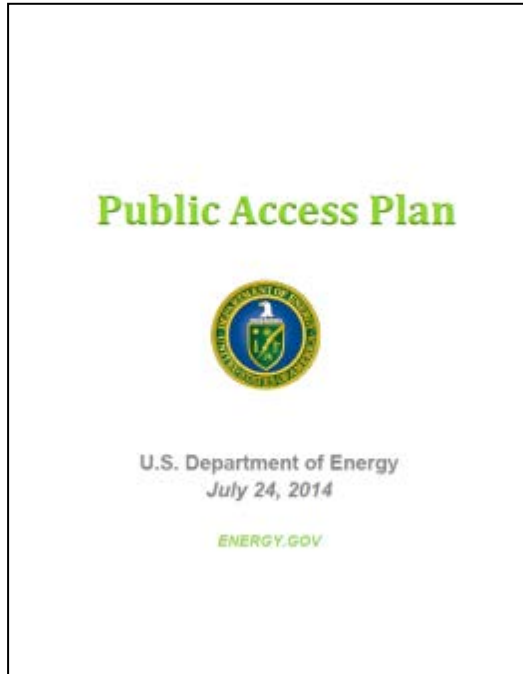
Make publically available the products of research supported wholly or in part by Federal funding:

- Peer-reviewed scholarly publications
- Digitally formatted scientific data should be stored and publicly accessible to search, retrieve, and analyze

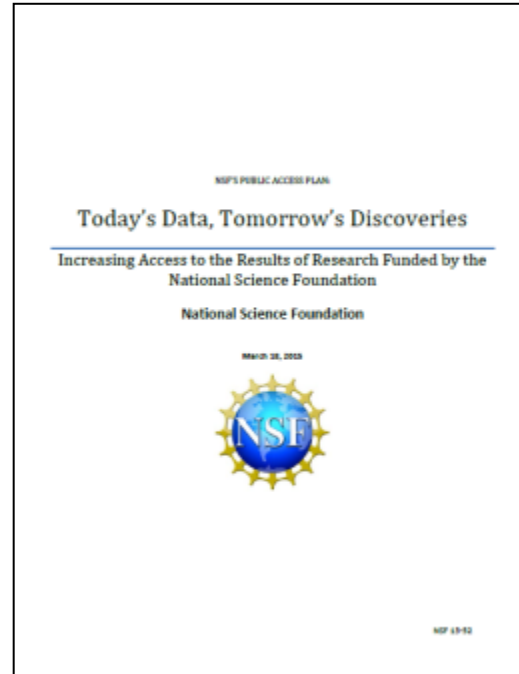


# Agency Public Access Plans

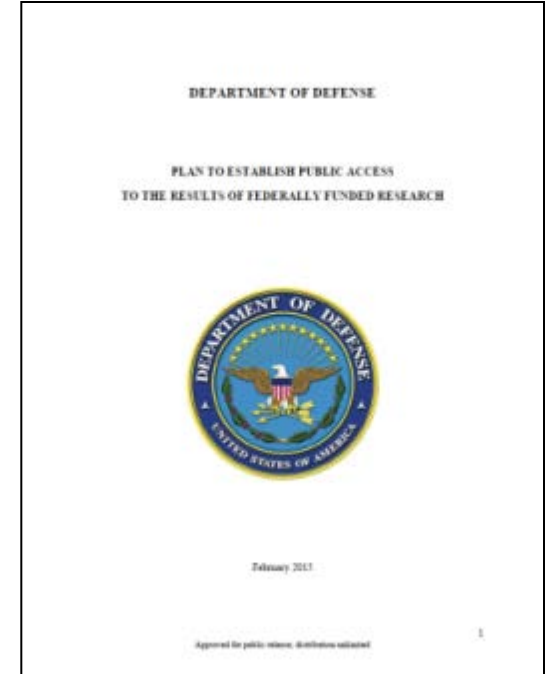
## *Digitally Formatted Scientific Data*



- Should be stored and publicly accessible
- Data management plans (DMP) with proposal
- Deposit data per proposal DMP



- Should be stored and publicly accessible
- DMP with proposal
- Deposit data per publication policy or proposal DMP



- Should be stored and publicly accessible
- DMP with proposal
- Store data in public repositories, centralized data catalog/locator at DTIC



# General DMP Help is Available



ESIP Commons EXPLORE Learn Meeting Help Log in

Collaboration Area / Data Management Training / Data Management Short Course for Scientists

## Data Management Short Course for Scientists

The ESIP Federation, in collaboration with other organizations, provides courses for scientists who increasingly need to manage their data. Over the next few years, we will offer courses that will help you understand the importance of data management and how to cite the work according to best practices.

All courses are available on a self-paced basis. You can complete the work according to your own schedule.

Sections:

- The Case for Data Stewards
- Data Management Plans
- Local Data Management
- Responsible Data Use
- Alphabetical Module Title

Titles for: The Case for Data Stewards

- Agency Requirements
  - NASA Data Management
  - NSF Data Management
  - NOAA Administrative
- Enhancing Your Research

DMPTool

Home DMP Requirements Public DMPs News Help Contact Us About Log in

## Data Management Planning Tool

Create, review, and share data management plans that meet institutional and funder requirements.

Get Started

PUBLIC DMPs	DMPTOOL NEWS	DMPTOOL HELP
<p>List of sample data management plans provided by DMPTool users.</p> <ul style="list-style-type: none"> <li>A unified approach to preserving cultural software objects and their development histories</li> <li>Arthropod responses to grassland nutrient limitation</li> <li>Atmospheric CO2 Concentrations, Mauna Loa Observatory, Hawaii, 2011-2013</li> </ul>	<p>Latest information about data management and the DMPTool.</p> <ul style="list-style-type: none"> <li>We are Hiring a DMPTool Manager!</li> <li>Got an idea for DMPTool? Share it!</li> <li>DMPTool unavailable, Saturday October 1...</li> <li>US Dept of Energy data management requi...</li> <li>We need API use cases!</li> </ul>	<p>Overview of how to use the tool, plus resources and guidance on data management.</p> <ul style="list-style-type: none"> <li>Frequently Asked Questions</li> <li>Create a DMP</li> <li>Administer the DMPTool</li> <li>Data management guidance</li> <li>Community resources</li> </ul>



# “Typical” Agency DMP Elements



*DMPs are key to defining the who, what, where, how, and why associated with the management of research data*

**Expected data**

**Preservation criteria**

**File formats**

**Data & metadata standards**

**Data storage**

**Dissemination**

**“Long-term” access**

**Ethics/privacy/proprietary/IP/security**

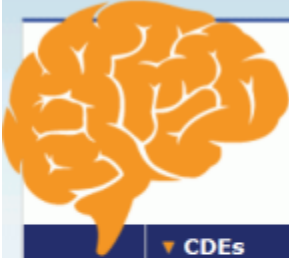
*Agency DMP requirements are necessary but insufficient to alone ensure the reusability of published research data*



# Agency Based Support



PROJECT OVERVIEW CDE SEARCH CRF SEARCH FORM BUILDER CONTACT



## NINDS Common Data Elements

Harmonizing Information. Streamlining Research.


▼ CDEs    ▼ Tools    ▼ Learn

### Streamline Your Neuroscience Clinical Research

using content standards that enable clinical investigators to systematically collect, analyze, and share data across the research community.


The NINDS strongly encourages researchers who receive funding from the Institute to ensure their data collection is compatible with these common data elements (CDEs). [Learn more about the CDE Project.](#)

CDEs Now Available	CDEs Under Review	CDEs in Development
General (CDEs that cross diseases) ^		
Amyotrophic Lateral Sclerosis		
Epilepsy		
Friedreich's Ataxia		
Headache		
Huntington's Disease		
Mitochondrial Disease <i>NEW!</i>		
Multiple Sclerosis		
Neuromuscular Diseases		
Congenital Muscular Dystrophy <i>NEW!</i> < >		




#### Launch Your Own Studies Faster

- ▶ Case report form modules
- ▶ Standardized data element definitions
- ▶ Instrument recommendations



#### Incorporate CDEs Into Systems

- ▶ Search for current CDEs
- ▶ Download CDE metadata
- ▶ Download Case Report Forms



#### Learn About the CDE Project

- ▶ Project overview and background
- ▶ Meetings and Presentations
- ▶ Collaboration with developers around the world

Privacy Statement | NeuroQOL | NIH Toolbox | PROMIS



# Community Based Support



## Geospatial Data Preservation

### Featured Resource

[Geospatial Data Transfer: Archival Data Transfer, Validation and Dataset Functional Verification \[Web\]](#)

Quick Links For

- Data Managers
- System Developers
- Researchers

### What's New

- [Suggest resources!](#)
- [Data Management Short Course for Scientists](#)
- [NDSA Levels of Digital Preservation @ USGS: An interview with John Faundeen](#)
- [Geospatial Data Stewardship: Key Online Resources](#)

### Featured Practice

## Format Descriptions for Geospatial Data

Explore descriptions of formats used for geospatial data and how to assess them for potential use.

Find resources via free text search on titles, authors, and

EDUCATION & TRAINING	TOOLS & SOFTWARE
BROWSE BY TOPIC	BROWSE BY TYPE

### Featured Website

## Geospatial Multistate A and Preservation Partn (GeoMAPP)

Explore the results of the GeoMAPP project and learn recommended practices for geospatial data steward



## Reporting Protein Identification Data

THE NEXT GENERATION OF GUIDELINES

Ralph A. Bradshaw, Alma L. Burlingame, Steven Carr, and Ruedi Aebersold

In June 2004, in response to a growing concern among both editors and reviewers, we published a set of instructions to authors, developed in house by a committee chaired by Steve Carr, which spelled out criteria for reporting protein identifications determined by mass spectrometry. The overall response from the community at large to this effort was quite favorable, and we were encouraged to expand on this document by soliciting broader input from a larger group of stakeholders with the view of fashioning a universal set of guidelines that might be useful, in general, for publishers of journals reporting the type of data. Accordingly, with support from the American Society for Biochemistry and Molecular Biology (ASBMB) and the organizational skills of Mike Baldwin, including important input from Steve Carr, Ruedi Aebersold, and Al Burlingame, a two-day workshop was held in mid-May 2008 in the Maison de la Chimie in Paris, France, that was attended by 30 of the ~60 invited scientists, engineers, bioinformaticians, and authors/publishers representing academia, research/government institutes, and the private sector (see below). This group reflected an excellent cross-section of individuals involved in the development of mass spectrometry and its proteomic applications. The attendees spent a day presenting/discussing the issues,<sup>1</sup> and then, using the Carr guidelines as a point of departure, divided into four sub-groups, with each working through a section. The four representatives, in some cases, new) sections were reworked into a single document and were then subjected to further post-meeting editing. In mid-July the draft document was released and broadly disseminated for public comment, suggestions, and criticism. In October, these were collated and returned to the original Paris group for further editing. The final document was assembled in late January/early February 2008. In mid-February, the final guidelines were read and discussed at an open forum at the Association of Biomolecular Resource Facilities (ABRF) meeting in Long Beach, CA. They have been subsequently sent to the editors of journals interested in this type of data.

The guidelines are now posted on the MCP website (<http://www.mcponline.org>) and have been officially adopted by the journal as the instructions to authors for papers in this area. They eventually will be shifted to that part of the journal, but we felt initially that it was appropriate to make them as visible as possible. At the same time that the Paris guidelines were being developed, MCP incorporated a mechanism for evaluating papers for their compliance with the guidelines initially the Carr guidelines, but now those developed in Paris that inserts an independent evaluation in the review process. A simplified outline of the guidelines in the form of a checklist is used and will be returned to authors by the Associate Editor that documents any areas of deficiency. The checklist will also be sent to the reviewers, along with the submitted materials to facilitate the review and to make compliance with the guidelines more consistent. Additional comments also may be added to help clarify the problems detected. This assessment is not considered to be a review, and unless the deficiencies are deemed to be so great by the Associate Editor, after receiving the checker's report, that in the Associate Editor's opinion the manuscript cannot be appropriately reviewed until the deficiencies are corrected, they will be returned along with the usual reviewer critique, and the compliance issues can be dealt with at the time any other corrections/changes are made.

Our initial experiences with the independent evaluation have been quite favorable, and we feel that it is already having the desired effect of bringing papers reporting protein identifications into a more consistent form. However, the main goal of these efforts was to address the problem of inconsistencies and secure the integrity of the scientific literature. To this end we naturally hope that authors, reviewers and editors alike will find these guidelines useful and worthy of serious consideration. We also appreciate that science is a dynamic process and that today's guidelines can become tomorrow's burdens. Therefore, we will be watchful for advances in technology that make any part of these guidelines obsolete and will welcome suggestions and comments that will aid us in keeping them current.

We would also like to thank the many people who contributed freely of their time to make this goal a reality. We particularly thank Aleksey Novitskiy, Robert Chalkley, and Karl Clubber for their help in preparing the original Carr document; to Mike Baldwin and all the participants at the Paris meeting; and to the ASBMB for their financial and intellectual support.

Editorial

## Reporting Protein Identification Data

THE NEXT GENERATION OF GUIDELINES

Ralph A. Bradshaw, Alma L. Burlingame, Steven Carr, and Ruedi Aebersold

In June 2004, in response to a growing concern among both editors and reviewers, we published a set of instructions to authors, developed in house by a committee chaired by Steve Carr, which spelled out criteria for reporting protein identifications determined by mass spectrometry. The overall response from the community at large to this effort was quite favorable, and we were encouraged to expand on this document by soliciting broader input from a larger group of stakeholders with the view of fashioning a universal set of guidelines that might be useful, in general, for publishers of journals reporting the type of data. Accordingly, with support from the American Society for Biochemistry and Molecular Biology (ASBMB) and the organizational skills of Mike Baldwin, including important input from Steve Carr, Ruedi Aebersold, and Al Burlingame, a two-day workshop was held in mid-May 2008 in the Maison de la Chimie in Paris, France, that was attended by 30 of the ~60 invited scientists, engineers, bioinformaticians, and authors/publishers representing academia, research/government institutes, and the private sector (see below). This group reflected an excellent cross-section of individuals involved in the development of mass spectrometry and its proteomic applications. The attendees spent a day presenting/discussing the issues,<sup>1</sup> and then, using the Carr guidelines as a point of departure, divided into four sub-groups, with each working through a section. The four representatives, in some cases, new) sections were reworked into a single document and were then subjected to further post-meeting editing. In mid-July the draft document was released and broadly disseminated for public comment, suggestions, and criticism. In October, these were collated and returned to the original Paris group for further editing. The final document was assembled in late January/early February 2008. In mid-February, the final guidelines were read and discussed at an open forum at the Association of Biomolecular Resource Facilities (ABRF) meeting in Long Beach, CA. They have been subsequently sent to the editors of journals interested in this type of data.

The guidelines are now posted on the MCP website (<http://www.mcponline.org>) and have been officially adopted by the journal as the instructions to authors for papers in this area. They eventually will be shifted to that part of the journal, but we felt initially that it was appropriate to make them as visible as possible. At the same time that the Paris guidelines were being developed, MCP incorporated a mechanism for evaluating papers for their compliance with the guidelines initially the Carr guidelines, but now those developed in Paris that inserts an independent evaluation in the review process. A simplified outline of the guidelines in the form of a checklist is used and will be returned to authors by the Associate Editor that documents any areas of deficiency. The checklist will also be sent to the reviewers, along with the submitted materials to facilitate the review and to make compliance with the guidelines more consistent. Additional comments also may be added to help clarify the problems detected. This assessment is not considered to be a review, and unless the deficiencies are deemed to be so great by the Associate Editor, after receiving the checker's report, that in the Associate Editor's opinion the manuscript cannot be appropriately reviewed until the deficiencies are corrected, they will be returned along with the usual reviewer critique, and the compliance issues can be dealt with at the time any other corrections/changes are made.

Our initial experiences with the independent evaluation have been quite favorable, and we feel that it is already having the desired effect of bringing papers reporting protein identifications into a more consistent form. However, the main goal of these efforts was to address the problem of inconsistencies and secure the integrity of the scientific literature. To this end we naturally hope that authors, reviewers and editors alike will find these guidelines useful and worthy of serious consideration. We also appreciate that science is a dynamic process and that today's guidelines can become tomorrow's burdens. Therefore, we will be watchful for advances in technology that make any part of these guidelines obsolete and will welcome suggestions and comments that will aid us in keeping them current.

We would also like to thank the many people who contributed freely of their time to make this goal a reality. We particularly thank Aleksey Novitskiy, Robert Chalkley, and Karl Clubber for their help in preparing the original Carr document; to Mike Baldwin and all the participants at the Paris meeting; and to the ASBMB for their financial and intellectual support.

LIST OF PARTICIPANTS IN THE PARIS MEETING  
Phil Andrews, University of Michigan Medical School  
Rolf Apweiler, European Bioinformatics Institute  
Kathy Aschinger, Maxima Biotechnology  
David Baker, University of Washington  
Ronald Bales, Babio Informatics, LLC

<sup>1</sup> The present list of these presenters can be found on the MCP website at <http://www.mcponline.org>.

**ICPSR** INTER-UNIV POLITICAL A  
A PARTNER IN SOCIAL SCIENCE RESEARCH

## Guide to Social Science Data Preparation and Archiving

Best Practice Throughout the Data Life Cycle • 5th edition

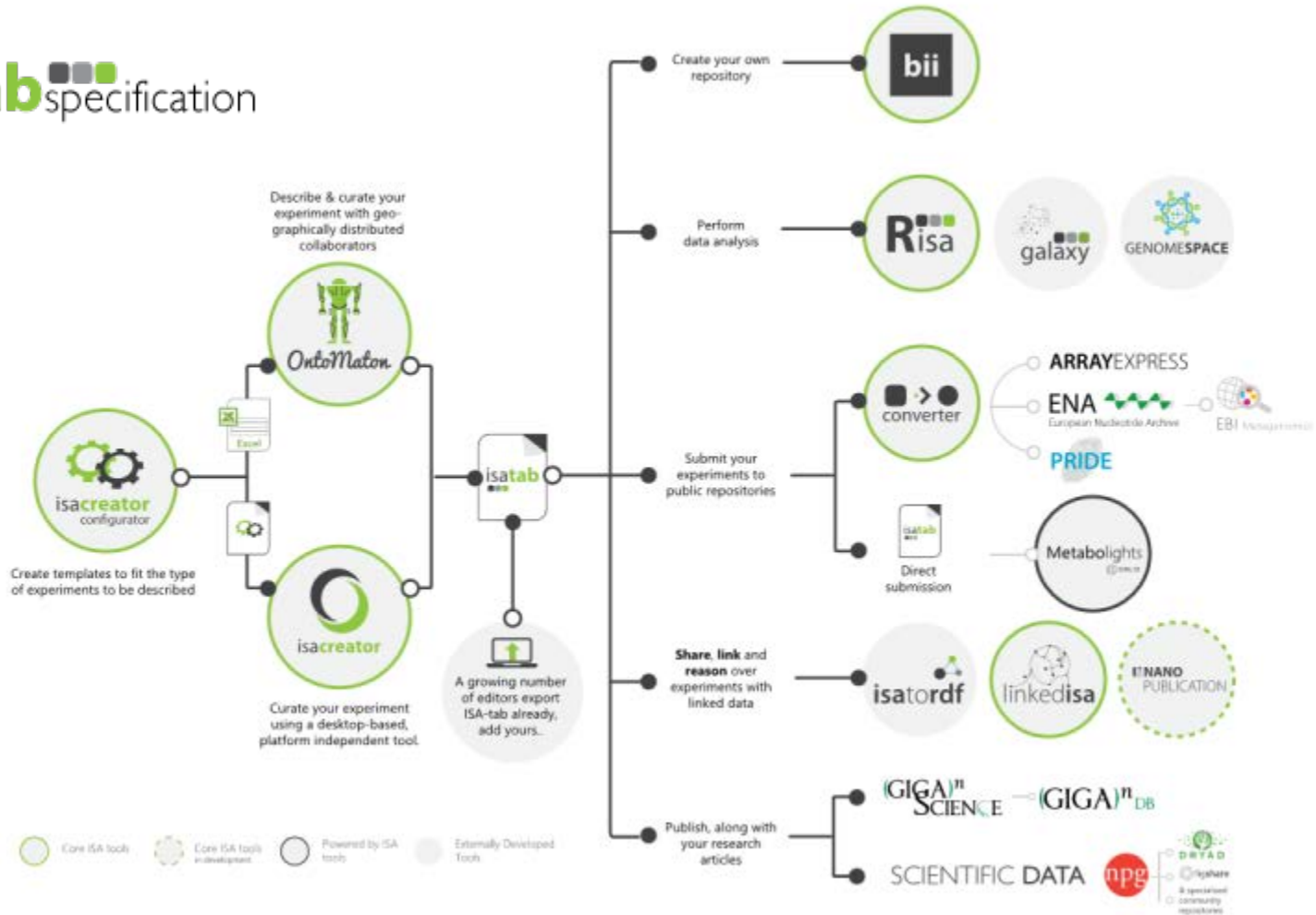




# Community Based Support



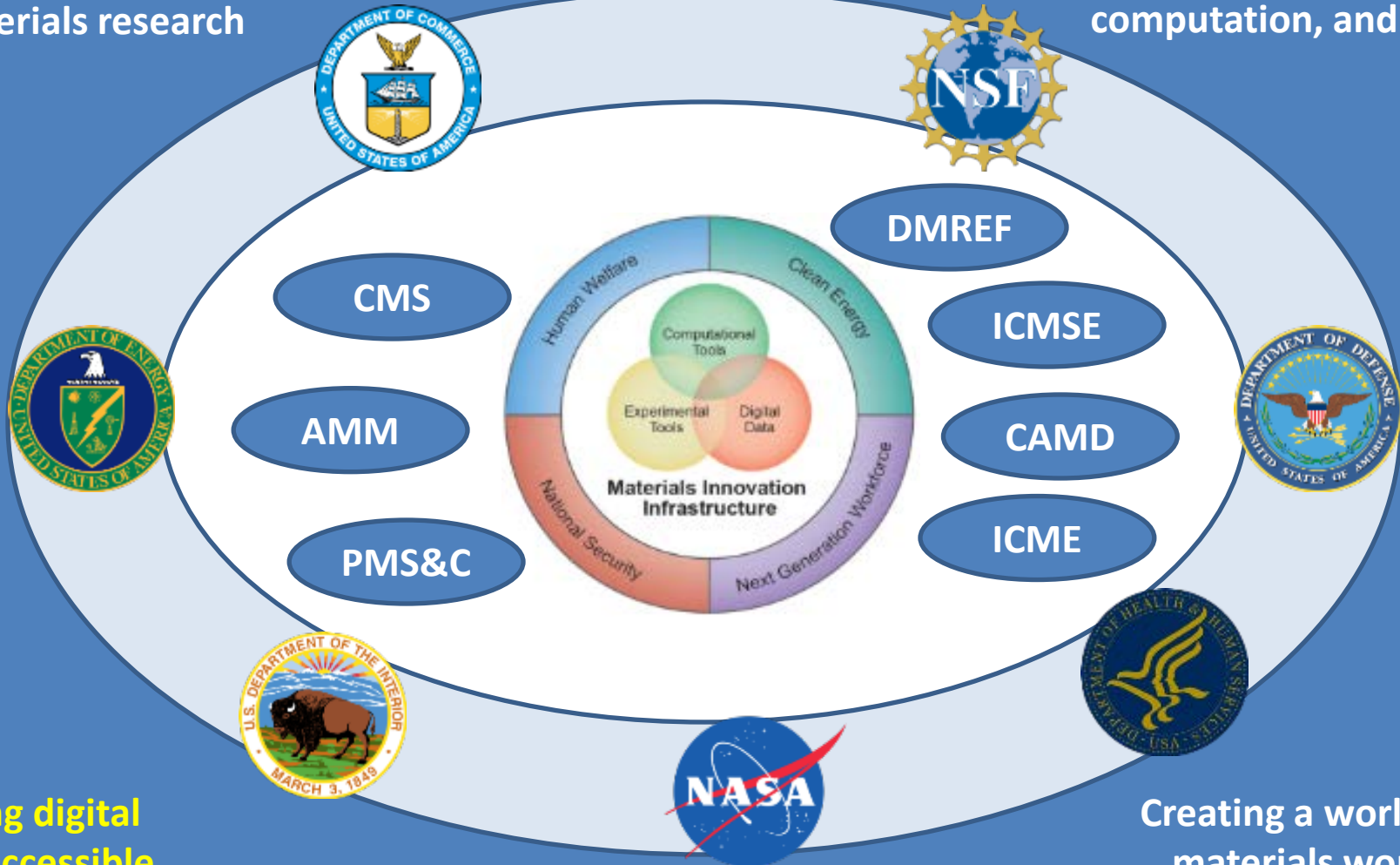
isatab specification



# Materials Genome Initiative

Leading a culture shift  
in materials research

Integrating experiment,  
computation, and theory



Making digital  
data accessible

Creating a world-class  
materials workforce

*Enabling the discovery, development, manufacturing, and deployment of advanced materials at least twice as fast as possible today, at a fraction of the cost.*



# Requirements for a Materials Data Infrastructure



- Repositories for materials data
  - Accessible, federated, affordable
- Standards for data exchange
  - Formats, data and metadata standards, vocabularies/ontologies
- Data quality metrics
  - Pedigree, provenance, verification, validation, uncertainty, sensitivity
- Citation and attribution protocols
  - Persistent identification
- Intellectual property and liability determinations
  - Export control education, Licensing clarity

*Findable, Accessible, Interoperable, Reusable*



# MGI Data Workshop, July 2014

## *Priorities*



### Community

- Develop and deploy standards for data and federated/collaborative environments
- Communicate value and need for digital materials data
- Define critical data to be compiled
- Engage Community
- Explore and leverage data solutions developed outside materials community
- Train students in these emerging areas
- Establish community norms for publishing materials data

### Government

- Lead data strategy/approach development
- Incentivize data deposition
- Facilitate data deposition (policy and technology)
- Support data sharing and deposition (provide resources)
- Support data creation
- Enhance data management competency
- Provide quality datasets for model development/validation



# Agency Data Gateways



## DOE Data Explorer

Discover science, technology, engineering research and data collections from the US Department of Energy

<http://www.osti.gov/dataexplorer/>

- Very rudimentary
- Limited data available



<http://maptis.nasa.gov/>

- Restricted access
- Engineering-oriented
- Extensive collection



# NIST Materials Data Repository



The screenshot shows the NIST Materials Data Repository website. At the top, there is a blue banner with the NIST logo and the text "Material Measurement Laboratory" and "materialsdata.nist.gov". Below the banner, there is a navigation bar with "NIST Repositories" and "Community List". The main content area is titled "NIST Repositories" and contains a paragraph about the National Institute of Standards and Technology's efforts in data exchange. Below this, there is a list of repository categories and sub-categories. On the right side, there is a search box with a "Go" button and a "Browse" section with links to "All of NIST Repositories", "Communities & Collections", "Subjects", "Titles", and "Authors". There is also a "My Account" section with a "Login" link.

**NIST**  
Material Measurement Laboratory  
materialsdata.nist.gov

NIST Repositories → Community List

## 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.

- **CHIMaD Data Collections**
  - [In Situ Si Composites](#)
  - [Precipitation Strengthened Alloys](#)
    - [Co-base Alloys](#)
- **Computational File Repository**
  - [Atomistics Simulations](#)
  - [CALPHAD Assessments](#)
  - [First Principles Phase Stability \(FPPS\) Files](#)
  - [Other Computational Methods](#)
- **Experimental Data Repository**
  - [Diffusion Data](#)
  - [Mechanical Properties](#)
  - [Other Experimental Data](#)
  - [Phase Equilibria and Thermodynamic Data](#)
- **Heusler Phases: First Principle Simulations**
  - [Magnetic Properties](#)
- **ICME Approach to Development of Lightweight 3GAHSS Vehicle Assembly**
  - [Computational Methods](#)
    - [First Principles Simulations \(DFT\)](#)
- **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**
  - [Co-Al-W based superalloy data](#)
  - [Diffusion Data Test](#)
  - [Molar Volume/Thermal Expansion](#)

**Search NIST Repositories**

[Advanced Search](#)

**Browse**

All of NIST Repositories

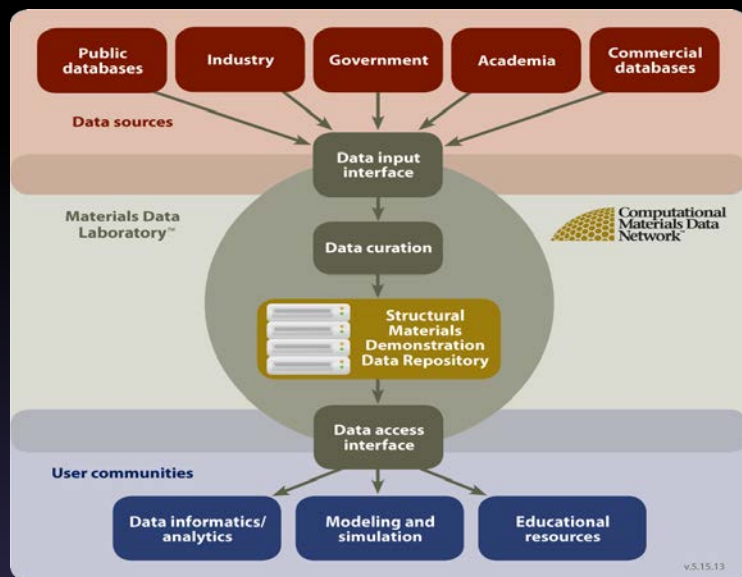
- [Communities & Collections](#)
- [Subjects](#)
- [Titles](#)
- [Authors](#)

**My Account**

[Login](#)

**NIST**

<https://materialsdata.nist.gov/dspace/xmlui/>



March 2014: Phase 1 release.  
June 2014: Phase 2 release.  
Dec 2014: Project Completion

**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

# DOE Vehicle Technologies – Supporting the MII

*Objectives: Generate thermodynamic, kinetic, and corrosion data for automotive Mg die casting alloys to fill significant gaps in the reported properties and to enable design of high performance alloys. Partner with NIST to structure data and deliver via NIST Dspace repository.*



**PI: J. Allison**

Coupled modeling and experiment to determine liquid- and solid-state kinetics in die castings

DOE Funding: \$600k



**PI: K. Sieradzki**

Synthetic microstructures and atomistic modeling to explore microstructure effects on bulk corrosion

DOE Funding: \$500k



**PI: J.C. Zhao, A. Luo**

High-throughput measurement of binary, ternary, and quaternary Mg alloy kinetics in liquid and solid

DOE Funding: \$600k



**PI: M. Horstemeyer**

Model development and experimental validation for coupled H<sub>2</sub> evolution and corrosion damage model

DOE Funding: \$500k



**PI: A. Rohatgi**

Dynamic-TEM measurement of Mg liquid- and solid-state kinetics with ~500ns resolution

DOE Funding: \$500k



**PI: G. Song**

Systematic experimental test of passivation behavior for wide range of Mg-X solid solutions

DOE Funding: \$600k



- All PIs will work with NIST to determine best format, content, and meta-data
- All PIs will upload project data to a NIST repository where it will be publicly available, searchable, useable, etc.
- All data will be assigned a persistent identifier for citation and connection with publications





# National Data Service

## National Center for Supercomputing Applications



The National  
DATA SERVICE

Home About **Projects** News Get Involved

Home Projects **The Materials Data Facility**

## The Materials Data Facility

In June 2014, the National Data Services Consortium announced its first pilot project, The **Materials Data Facility** (MDF). It served as a response the White House's **Materials Genome Initiative (MGI)** to accelerate the process for creating new materials. Being able to share data readily through the materials development chain will be critical to achieving this acceleration. The consortium saw this as an opportunity to not only prototype and demonstrate the data publishing capabilities that needed across all research disciplines, but also deliver real value to working scientists.

The MDF will provide a repository where scientists can preserve and share materials research data, produced by both simulations and experiments. The capabilities needed by MGI scientists mirrors closely the broader NDS vision: sharing data privately before publication, creating data collections, publishing, linking with the literature, and connecting with other data resources and databases in the world.

## What you will be able do with the MDF

The Facility will provide capabilities that will be useful throughout the research and publishing process:

### Private storage of data and sharing prior to publication

- Researchers will be able to log into the MDF to securely access personal storage space where they can upload and download their data files.
- Researchers will be able to organize their data products and files into logical collections, and upload files into those collections.
- Researchers will be able to share data files and collections with their collaborators, either by simply sharing a URL to the data or by managing group permissions on the data collections.

### Preparing data for publication

### Enabling data re-use



User home



Publish screen

<http://www.nationaldataservice.org>



# Project-based repositories



**MRL**  
MATERIALS RESEARCH LABORATORY  
the innovation engine for new materials  
UC Santa Barbara

## Energy Materials Datamining

For details, please refer to the original publication: [DOI: 10.1021/cm400893e](https://doi.org/10.1021/cm400893e).

Links : [thermoelectric](#) : [HHI for all elements](#) : [crystal abundance](#) : [about](#) : [methodology](#) : [contributors](#)

Plot MRL dataming data    Plot my own data  (right click, choose Save As, [example.xls](#))

Note: for plotting your own data, a less specific [data visualization tool](#) is also available that only requires x, y and marker values.

**x axis parameter**   **y axis parameter**   **marker size parameter**   **sort data by...**

<http://www.mrl.ucsb.edu:8080/datamine/thermoelectrics.jsp>

 **Hydrogen Storage Materials Database**

[Request Access to Add Data](#)   [Contact Us](#)

<http://hydrogenmaterialssearch.govtools.us/>

*Issues: data discoverability, machine processing, data curation/longevity*



# Computationally-derived Data Repositories



## MIT & LBNL

- 58,123 compounds
- 41952 band structures
- 1,243 elastic tensors
- Thermodynamic props
- Crystal structure
- Phase diagrams
- REST API



## Duke

- 51,367 compounds
- 308,975 calculations
- Thermodynamic props
- Magnetic props
- Crystal structure
- Phase diagrams
- REST API





# Collaborative Environments



## Univ. Michigan – Materials Commons

- Store materials data (& provenance)
- Collaborate and share data
- Search and use data, REST API



## Univ. Illinois

- **T2C2 Curator**, real-time acquisition and curation of data from instruments
- **T2C2 Coordinator**, filter, identify correlate data



**Timely and Trustworthy  
Curating and  
Coordinating Data  
Framework (T2C2)**

## Purdue University

- Platform for simulations
- Project space
- Expanding to data management



## Georgia Tech

- e-collaboration space
- Models, data, sharing
- GitHub, Plot.ly, figshare, Authorea



# NIST Data Efforts

## Collaborations

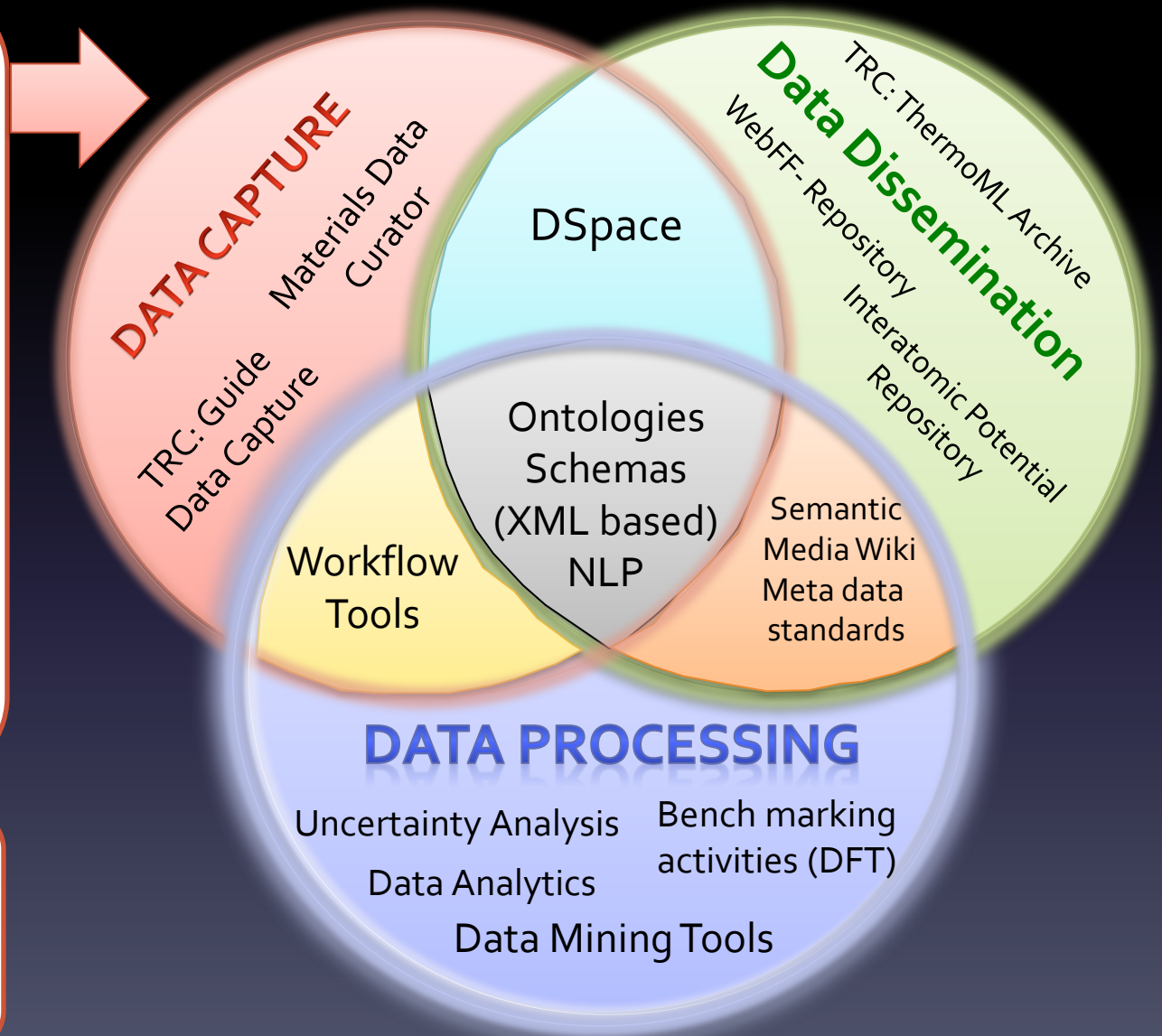
ASM International:  
Structural Data  
Demonstration Project

DOE/EERE Kinetics of  
Cast Mg Alloys

Journals collaboration

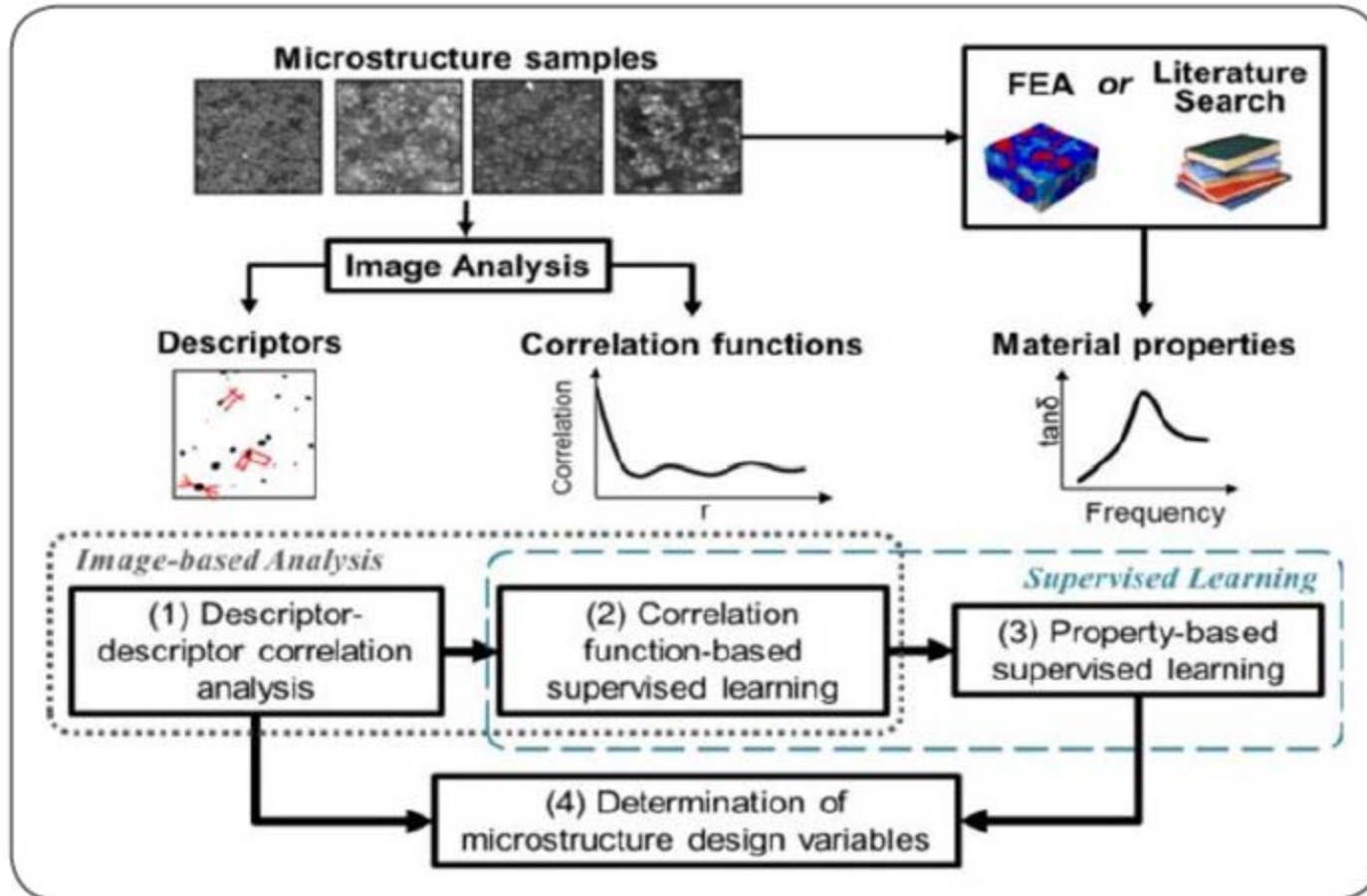
- IMMI
- *Others under discussion*

Coordinated by new  
Office of Data and  
Informatics





# ChiMaD Data Mining





# AFRL Data Efforts

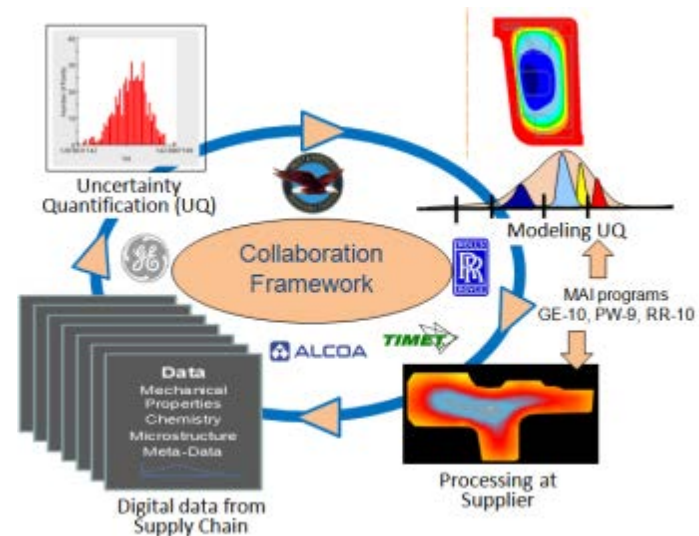


## Wright State University

<http://wiki.knoesis.org/index.php/MaterialWays>

## Metals Affordability Initiative

- Web-based OEM-Supplier collaboration framework
- Data standards and protocols for OEM-Supplier data and model exchange.
- Uncertainty Quantification methods to capture variability in process, microstructure and property data.





# Semantic-based Tools for Materials Data Management



The screenshot displays the Matonto interface for Zirconium Dioxide (ZrO<sub>2</sub>). The main header shows the material name and its chemical formula. Below this, there are several sections:

- Characteristics:** Includes a 'Melting Point' section with buttons for 'Low', 'High', and 'Ultra'. The 'Chemical Group' section has buttons for 'Boride', 'Nitride', 'Carbide', and 'Oxide'. The 'Metal Group' section has a button for 'Transition Metal'.
- 1-Hop Properties:** A table listing various properties of ZrO<sub>2</sub>.
- Phase Diagrams:** A section titled 'Phase Diagram of ZrO<sub>2</sub>' (1 of 2) showing a graph of temperature vs. composition.
- Related Documents:** A list of documents related to ZrO<sub>2</sub>, including titles and years.

Property	Value	Units	Source
Volume	31.0595020688944827	cubic angstrom	Materials Project
Energy	-24.88009819	eV	Materials Project
Energy	-24.88009819	electron volt	Materials Project
Density	6.1	g/mL	Wikipedia
Density	6.1	gram per milliliter	Wikipedia
Density	6.033102843591431	g/cm <sup>3</sup>	Materials Project
Density	6.033102843591431	gram per cubic	Materials Project

Title	Year
Thermophysical properties of ZrO <sub>2</sub> and ZrO <sub>2</sub> -SiC ceramics	2006
Effect of hot pressing time and temperature on the microstructure and mechanical properties of ZrO <sub>2</sub> -SiC	2007
Thermodynamic Analysis of ZrO <sub>2</sub> -SiC Oxidation: Formation of a SiC-Depleted Region	2007
Beneficial effects of an ultra-fine $\alpha$ -SiC incorporation on the sinterability and mechanical properties of ZrO <sub>2</sub>	2006
Pressureless Sintering of high-Density ZrO <sub>2</sub> -SiC Ceramic Composites	2006
Properties of a Pressureless-Sintered ZrO <sub>2</sub> -MoSi <sub>2</sub> Ceramic Composite	2006

- Develop Robust Mid-level Materials Ontology Ready for Crowd-sourcing and Initial Experimental Research Use
- Explore Sophisticated Low-level Ontologies
- Significantly Expand Size of, and Integration Across Data Stores
- Propose and Demonstrate Computational Approaches for Establishing Provenance
- Provide Access Control for Linked Materials data and Computational Functions
- Integrate with Manufacturing or Component Design Domains







# Materials Data Formats



notes for authors

Acta Crystallographica Section C  
**Crystal Structure  
Communications**

ISSN 0108-2701

## Notes for authors 2012

*Acta Crystallographica Section C: Crystal Structure Communications* publishes articles that provide a detailed discussion of structures determined by diffraction methods. It specializes in the dissemination of high-quality studies of novel and challenging structures of interest in the fields of materials, molecular, and biochemical structures. The journal ensures the highest standards of structural presentation, while providing for reports on studies in difficult or challenging materials. Articles contain a discussion that goes beyond reporting numerical and geometrical data. Such value-added content includes: a meaningful discussion of multiple related structures reported in the same article; non-routine structures discussed in detail; placing the structure in an interphysical or chemical context; or the discussion of interphysical properties or modes of association. The journal accepts difficult or challenging structures not meeting all value-added criteria provided the presented structures are correct and unique, the difficulties and strategies used to treat them are discussed and properly documented. Such structure properties such as twinning, severe disorder, or

Jackson et al. *Integrating Materials and Manufacturing Innovation* 2014, 3:4  
<http://www.immjournal.com/content/3/1/4>

 **Integrating Materials  
and Manufacturing Innovation**  
a SpringerOpen Journal

RESEARCH

Open Access

## h5ebds: an archival data format for electron back-scatter diffraction data sets

Michael A Jackson<sup>1</sup>, Michael A Groeber<sup>2</sup>, Michael D Uchic<sup>2</sup>, David J Rowenhorst<sup>3</sup> and Marc De Graef<sup>4\*</sup>

\*Correspondence:  
degrael@cmu.edu  
<sup>4</sup>Department of Materials Science  
and Engineering, 5000 Forbes  
Avenue, 15213 Pittsburgh, PA, USA  
Full list of author information is  
available at the end of the article

### Abstract

We present an archival format for electron back-scatter diffraction (EBSD) data based on the HDF5 scientific file format. We discuss the differences between archival and data workflow file formats, and present details of the archival file layout for the implementation of h5ebds, a vendor-neutral EBSD-HDF5 format. Information on sample and external reference frames can be included in the archival file, so that the data is internally consistent and complete. We describe how the format can be extended to include additional experimental modalities, and present some thoughts on the interactions between working files and archival files. The complete file specification as well as an example h5ebds formatted data set are made available to the reader.

**Keywords:** Electron back-scatter diffraction; Hierarchical data format; HDF5



# DoD Materials Project on SEM Data

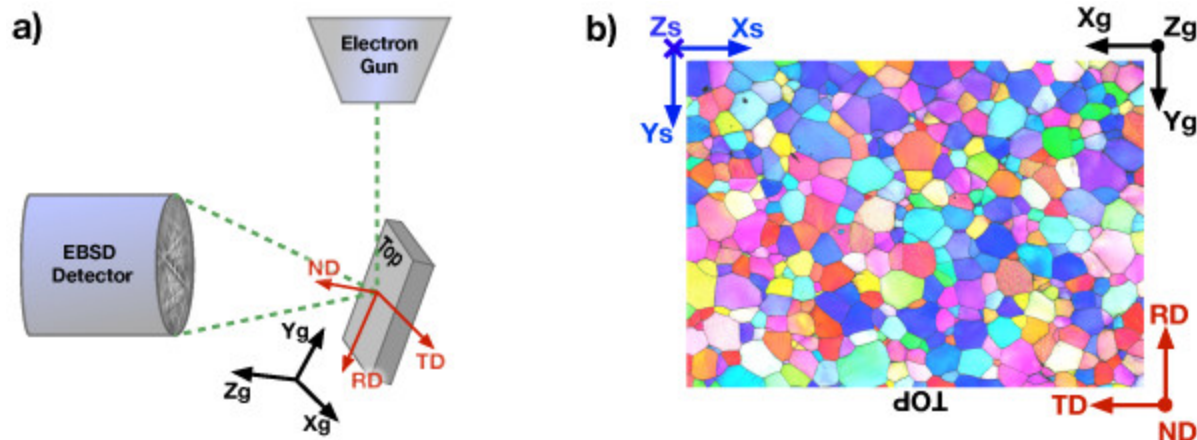


## Problem:

- Data generated from scanning electron microscopy (SEM) have become much more complex and content rich, but there is no standard format used to facilitate sharing and machine interpretation

## Objectives:

- Better preservation and re-use of SEM data across the services
- Improve automatic metadata capture
- Define and improve data flow from research instruments to analysis





# Making Data More Valuable & Accessible

## Supporting Reports of Research with Underlying Data



Shade et al. Integrating Materials and Manufacturing Innovation 2013, 2:5  
<http://www.immjournal.com/content/2/1/5>

 Integrating Materials  
and Manufacturing Innovation  
a SpringerOpen Journal

RESEARCH

Open Access

### Experimental measurement of surface strains and local lattice rotations combined with 3D microstructure reconstruction from deformed polycrystalline ensembles at the micro-scale

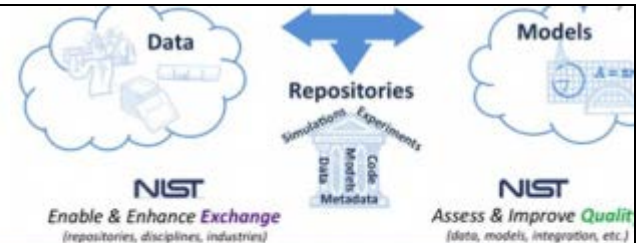
Paul A Shade\*, Michael A Groeber, Jay C Schuren and Michael D Uchic

\* Correspondence:  
paulshade.1@us.af.mil  
Air Force Research Laboratory,  
Materials and Manufacturing  
Directorate, 2230 10th Street,  
Wright-Patterson AFB, OH 45433, USA

#### Abstract

This article describes a new approach to characterize the deformation response of polycrystalline metals using a combination of novel micro-scale experimental methodologies. An in-situ scanning electron microscope (SEM)-based tension test system was used to deform micro-scale polycrystalline samples to modest and moderate plastic strains. These tests included measurement of the local displacement field with nm-scale resolution at the sample surface. After testing, focused ion beam serial sectioning experiments that incorporated electron backscatter diffraction mapping were performed to characterize both the internal 3D grain structure and local lattice rotations that developed within the deformed micro-scale test sample. This combination of experiments enables the local surface displacements and internal lattice rotations to be directly correlated with the underlying 3D polycrystalline microstructure, and such information can be used to validate and guide further development of modeling and simulation methods that predict the local plastic deformation response of polycrystalline ensembles.

**Keywords:** Micro-tensile test; Plastic deformation; Microstructure



NIST Repositories → TMS Springer Integrating Materials and Manufacturing Innovation (IMMI) → Thematic Series on View Item

#### Data Citation:

Shade, Paul A; Groeber, Michael A; Schuren, Jay C; Uchic, Michael D  
3D microstructure reconstruction of polycrystalline nickel micro-tension test  
(2013-11-01)

<http://hdl.handle.net/11115/152>

#### Requirements:

The raw data can be viewed by opening the \*.dream3d file with freely available DREAM 3D software (<http://dream3d.bluequartz.net/>). The reconstruction can be viewed by downloading both the \*.dream3d file and the \*.xdfm file, and then opening the \*.xdfm file using freely available ParaView software (<http://www.paraview.org/> -- See Also ParaView FAQ: <http://paraview.org/Wiki/ParaViewFAQ#>)

**Affiliation:** Air Force Research Laboratory

**Contact Email:** paul.shade.1@us.af.mil

#### Primary Publication Citation:

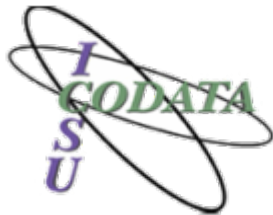
Shade PA, Groeber MA, Schuren JC, Uchic MD (2013) Experimental measurement of surface strains and local lattice rotations combined with 3D microstructure reconstruction from deformed polycrystalline ensembles at the micro-scale. Integrating Materials and Manufacturing Innovation, 2:5.

<http://dx.doi.org/10.1186/2193-9772-2-5>

<http://www.immjournal.com/authors/instructions/datadescriptorarticle>



# Many Players in the Field



**CODATA**

**NIST**

**BOARD ON RESEARCH DATA AND INFORMATION**



# Summary



- A few materials data repositories have been established
- Development and application of commonly accepted standards and tools enabling data discovery, data quality, and reusability is needed
- Educational resources on good data practices are needed for the community
- Community-wide acceptance of the need for materials data stewardship is slow to build

