National Materials Data Initiatives

Chuck Ward

Materials & Manufacturing Directorate

Integrity ★ Service ★ Excellence

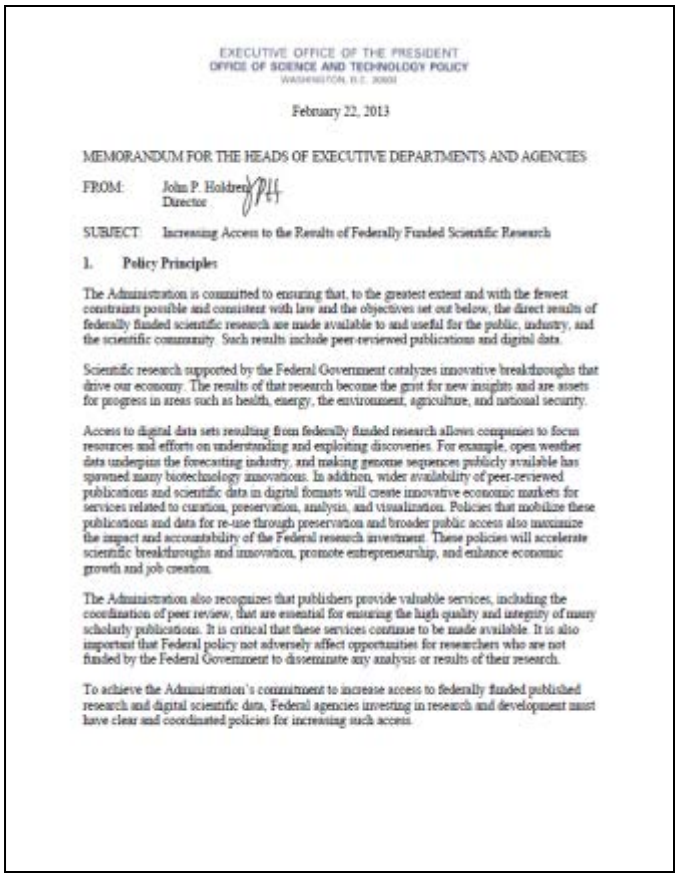
Overview

- 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. 20502

February 22, 2013

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: John P. Holdren
Director

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

1. 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 grist 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.

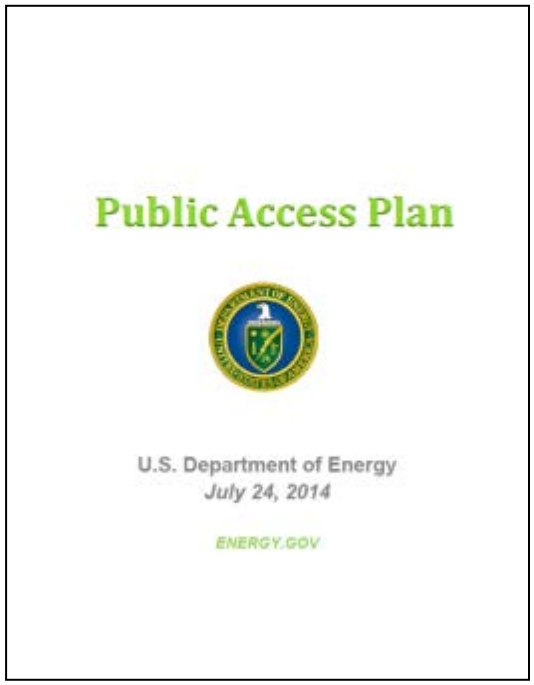
To achieve the Administration's commitment to increase access to federally funded published research and digital scientific data, Federal agencies investing in research and development must have clear and coordinated policies for increasing such access.

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

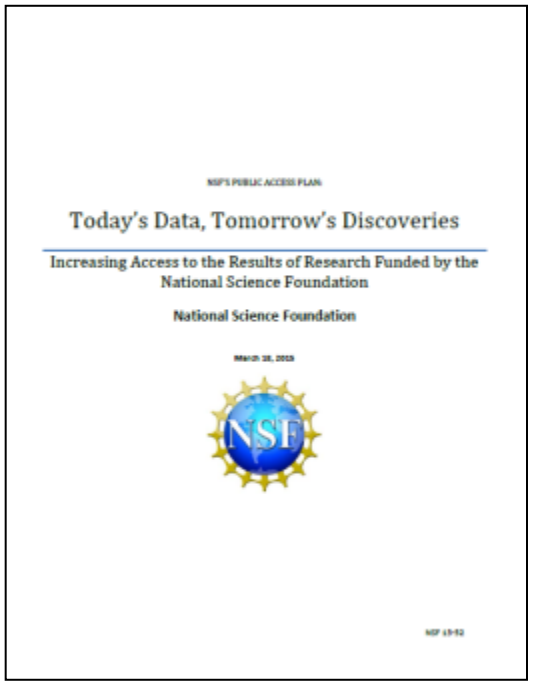
- Peer-reviewed scholarly publications
- Digitally formatted scientific data should be stored and publicly accessible to search, retrieve, and analyze

Agency Public Access Plans

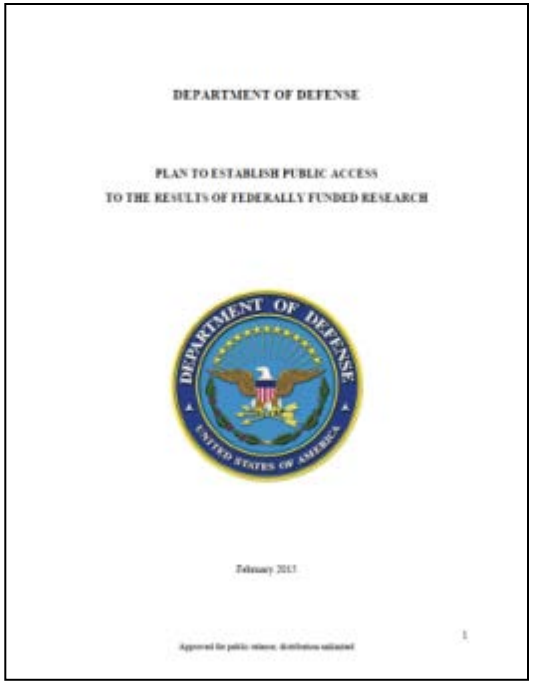
Digitally Formatted Scientific Data



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

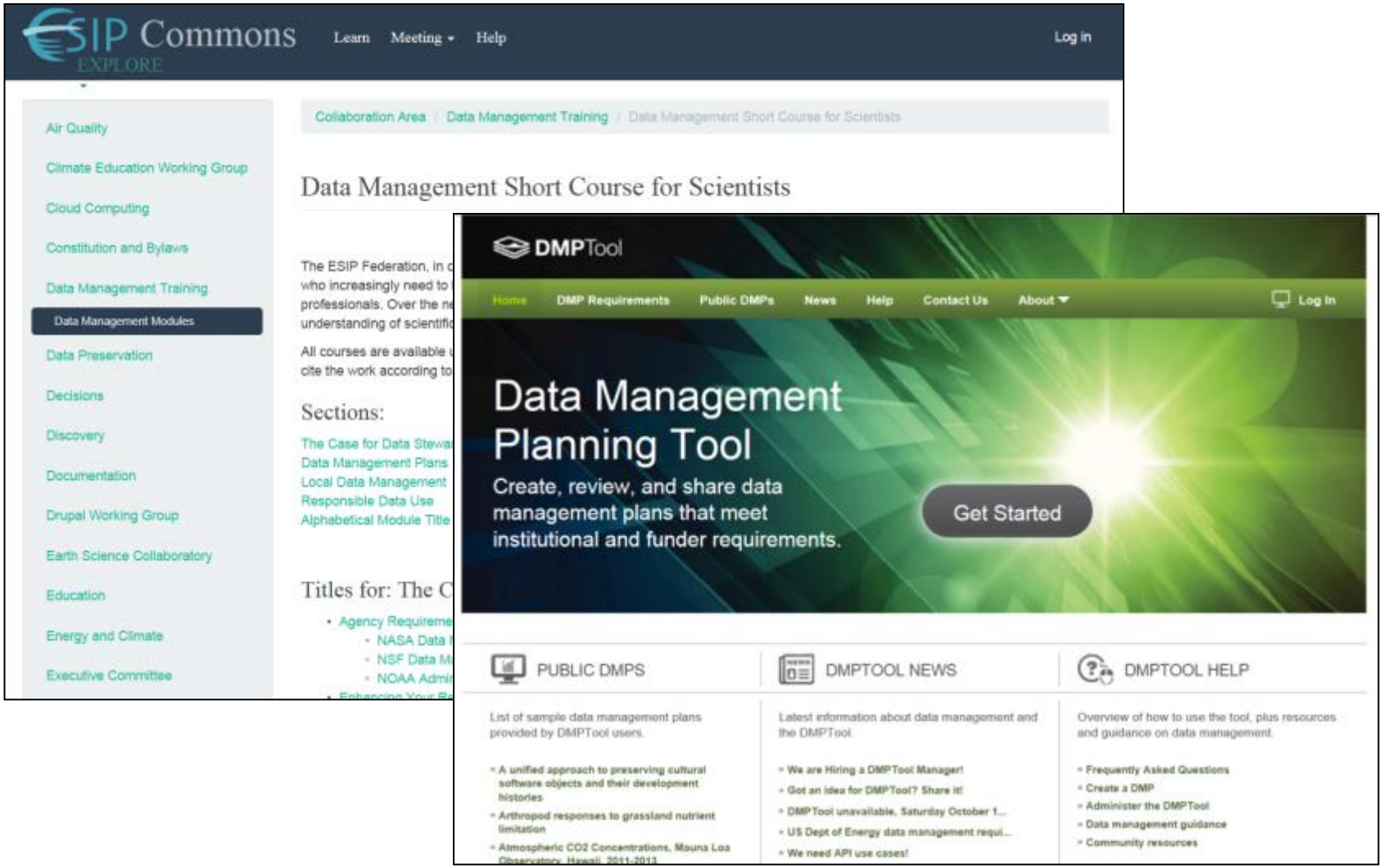


- Should be stored and publicly accessible
- DMP with proposal
- Deposit data per publication policy or proposal DMP



- 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



“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

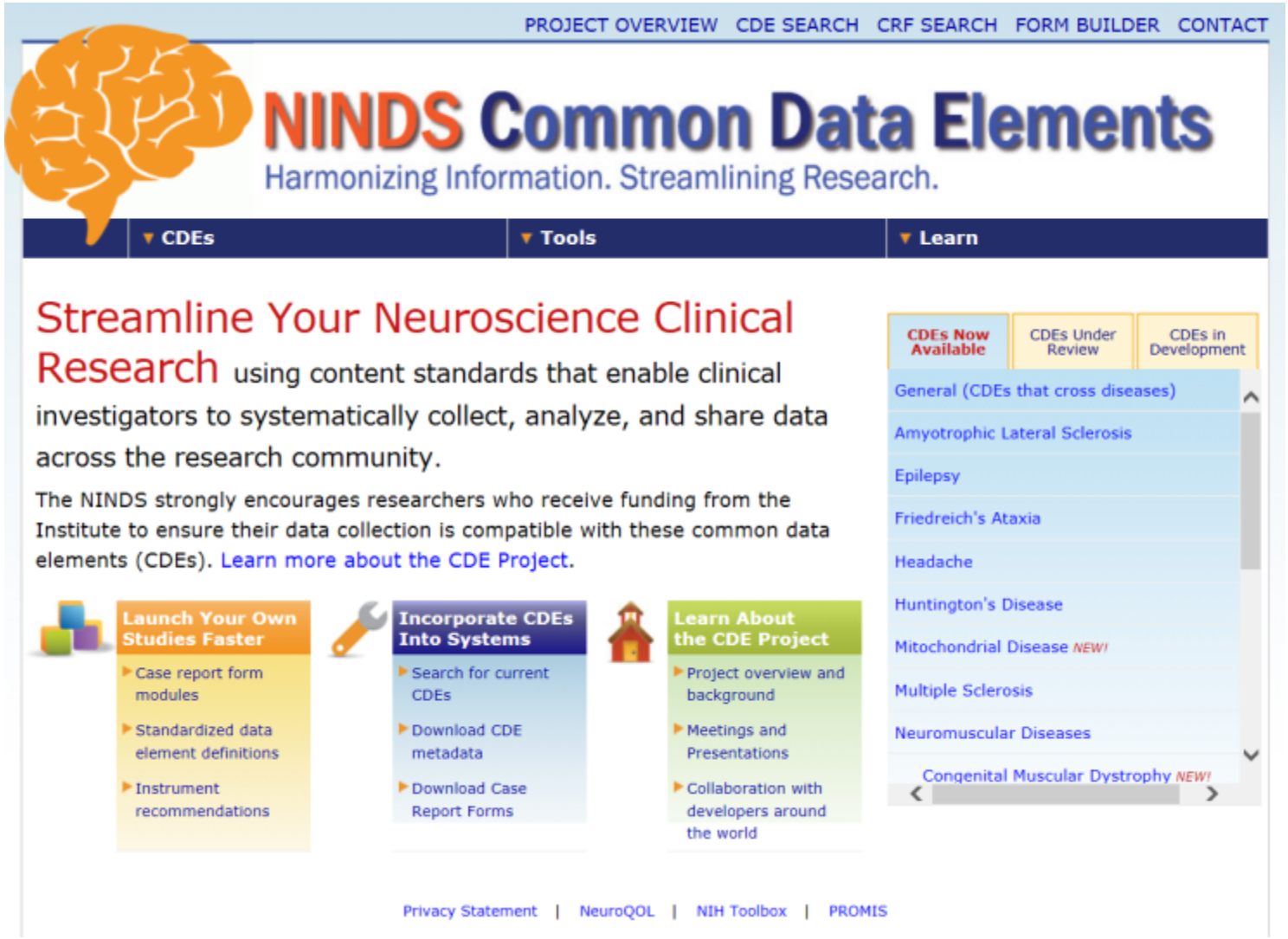
Dissemination

“Long-term” access

Ethics/privacy/proprietary/IP/security

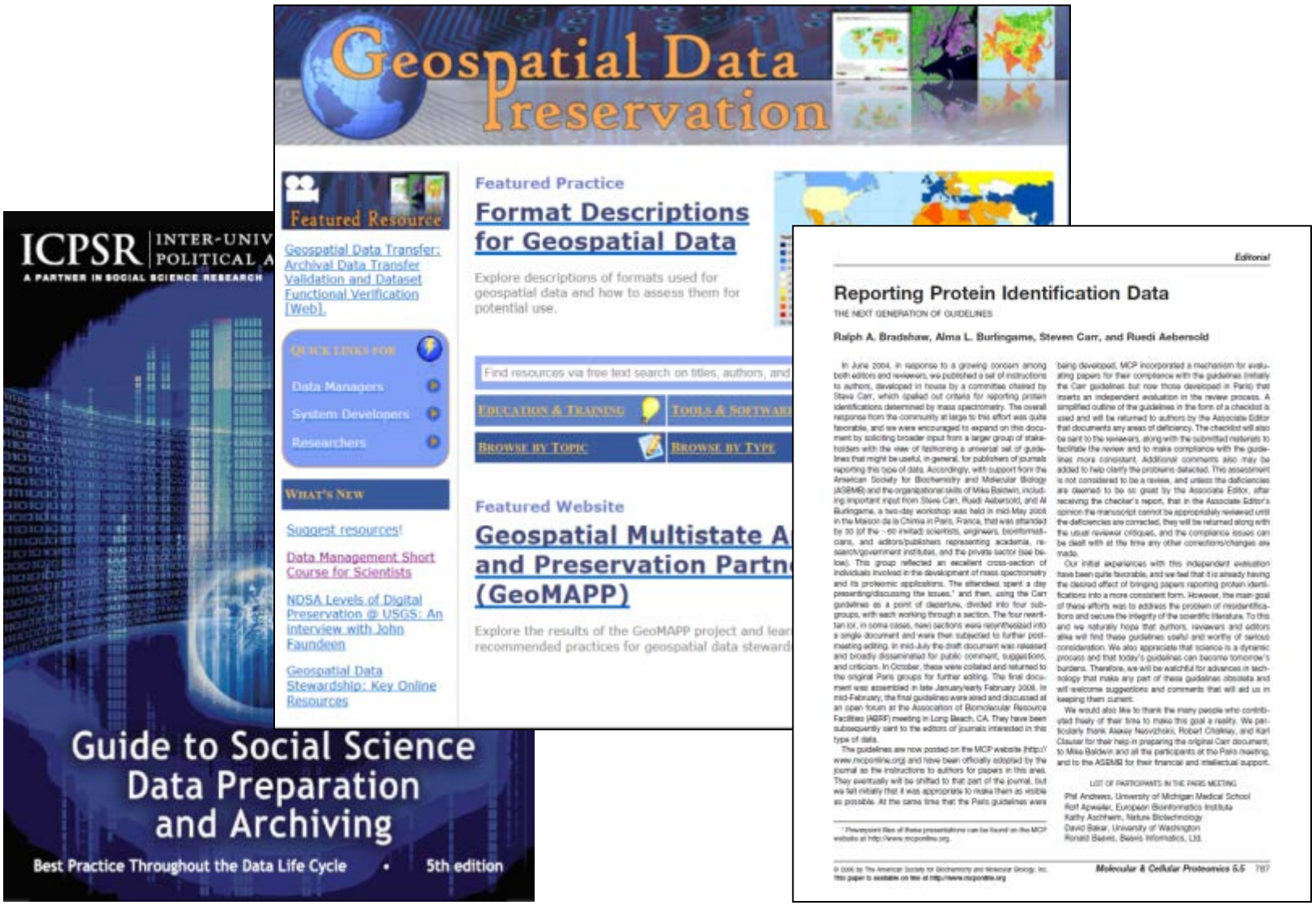
Agency DMP requirements are necessary but insufficient to alone ensure the reusability of published research data

Agency Based Support

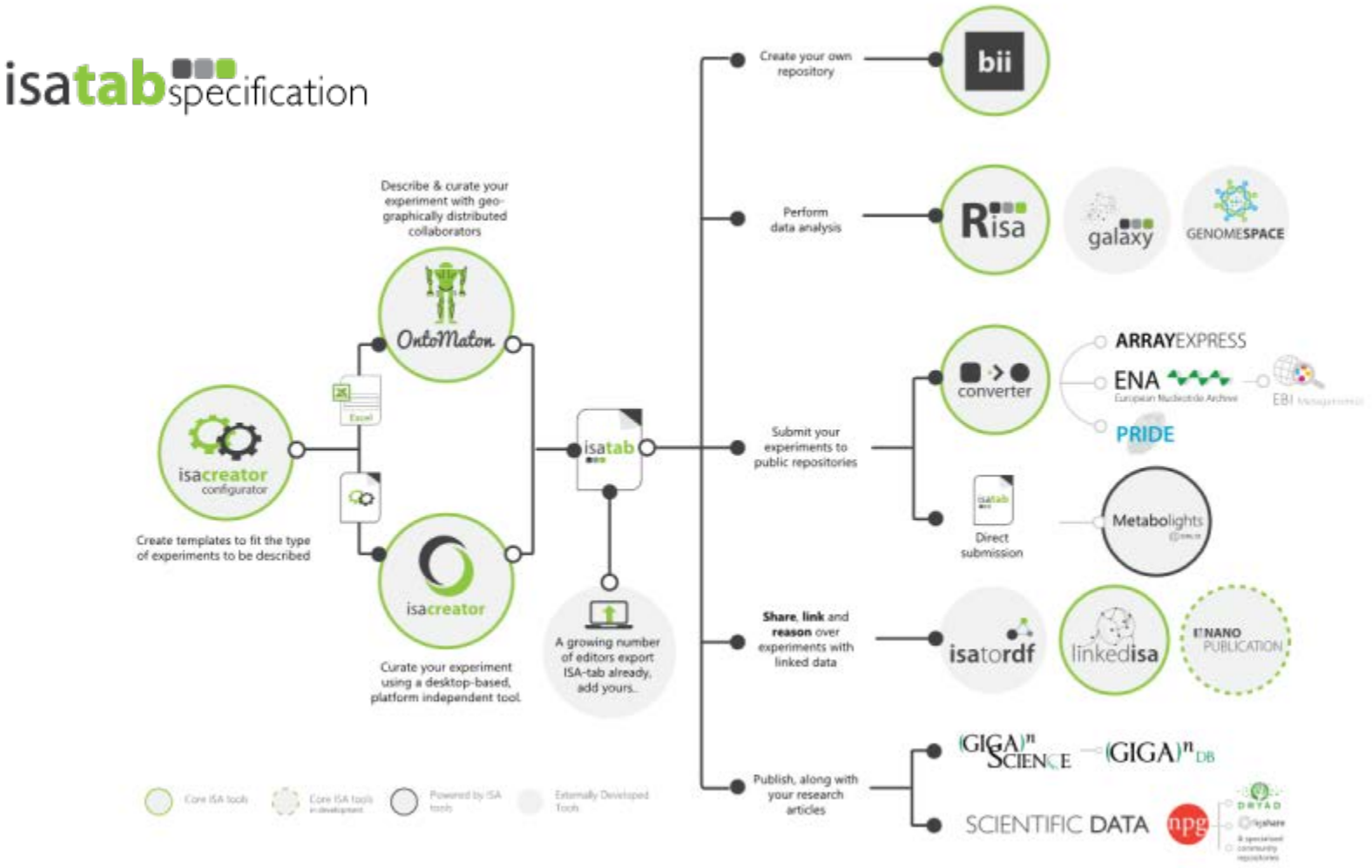


http://www.commondataelements.ninds.nih.gov/

Community Based Support

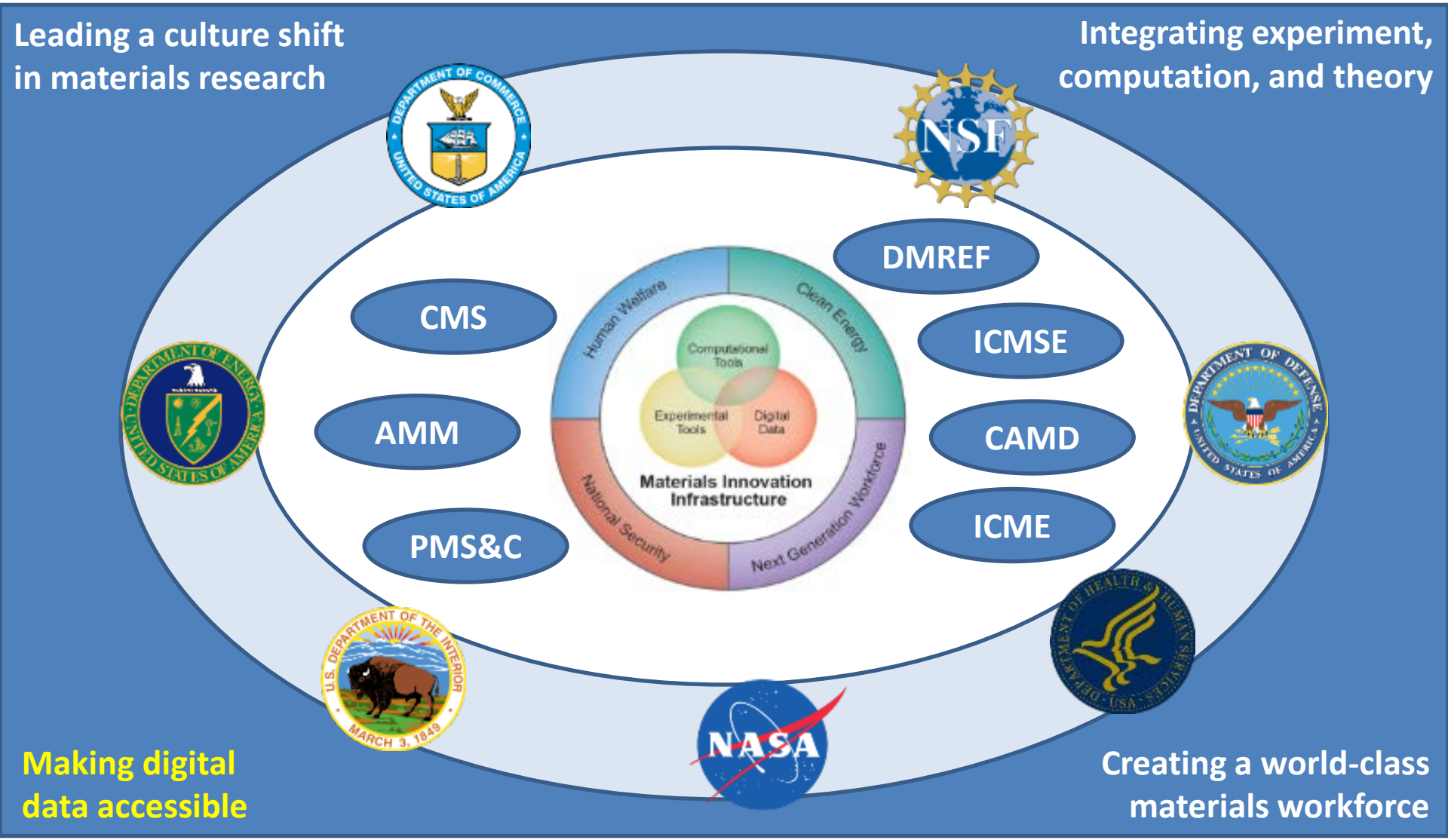


Community Based Support



http://isatab.sourceforge.net/tools.html

Materials Genome Initiative



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

***F**indable, **A**ccessible, **I**nteroperable, **R**eusable*

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

http://dx.doi.org/10.6028/NIST.IR.8038

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

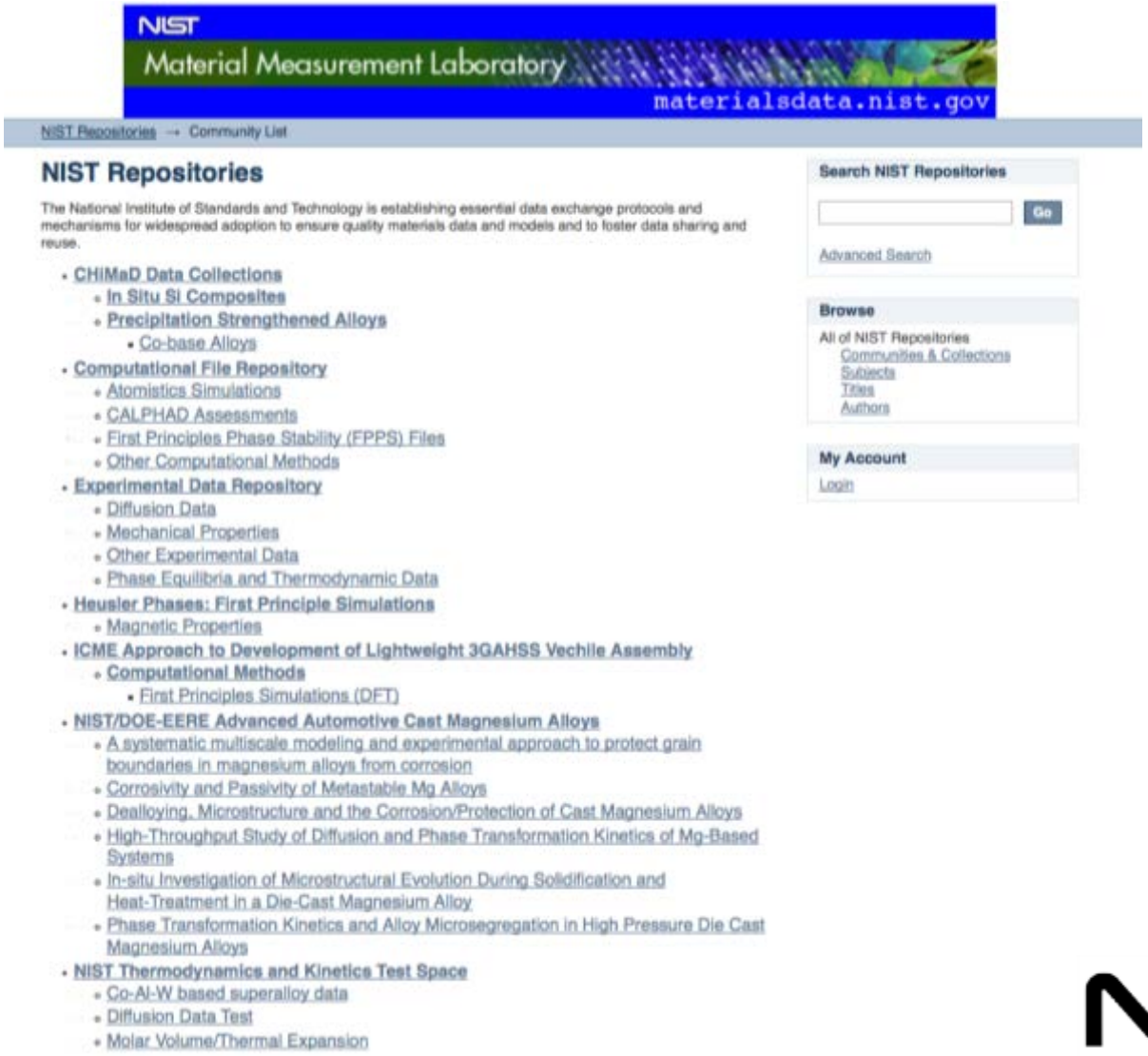
MAPTIS

MATERIALS AND PROCESSES TECHNICAL INFORMATION SYSTEM

http://maptis.nasa.gov/

- Restricted access
- Engineering-oriented
- Extensive collection

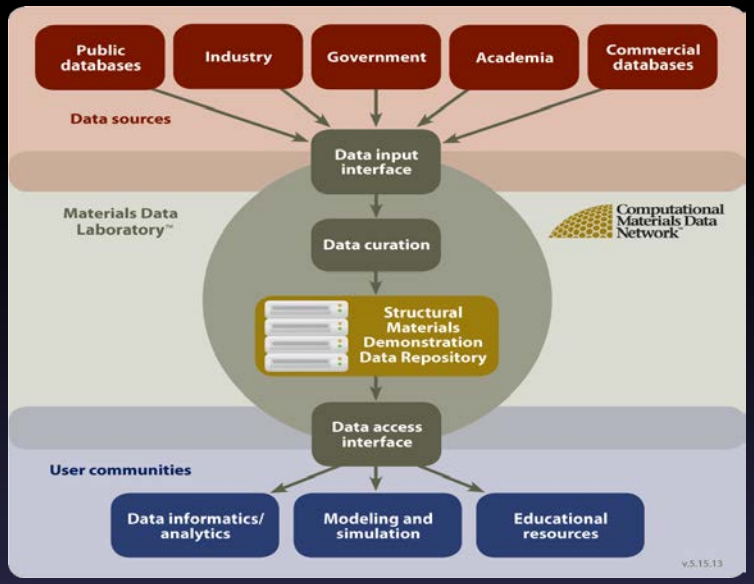
NIST Materials Data Repository



NIST

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

Structural Data Demonstration Project: Al6061



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

U.S. DEPARTMENT OF ENERGY | Energy Efficiency & Renewable Energy

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_2 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

NIST National Institute of Standards and Technology U.S. Department of Commerce

- 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.

User home

What you will be able do with the MDF

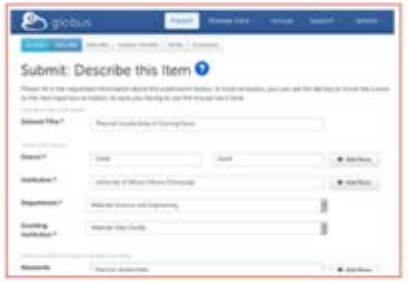
The Facility will provide capabilities that will be useful throughout the research and publishing process:

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

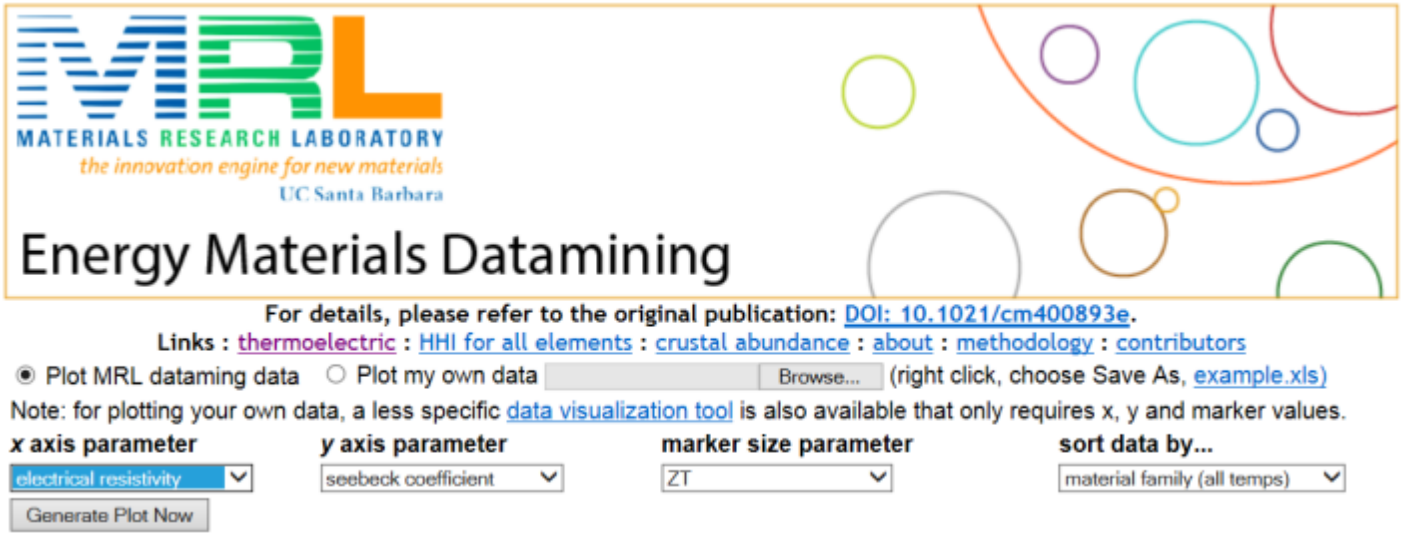
Enabling data re-use



Publish screen

http://www.nationaldataservice.org

Project-based repositories



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

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

The Materials Project

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

AFLOW
Automatic - FLOW for Materials Discovery

Collaborative Environments

Univ. Michigan – Materials Commons
- Store materials data (& provenance)
- Collaborate and share data
- Search and use data, REST API

PRISMS

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

DARPA

COMPOSITES DESIGN & MANUFACTURING HUB

hubzero PLATFORM FOR SCIENTIFIC COLLABORATION

nanoHUB

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

matIN

NIST Data Efforts

Collaborations

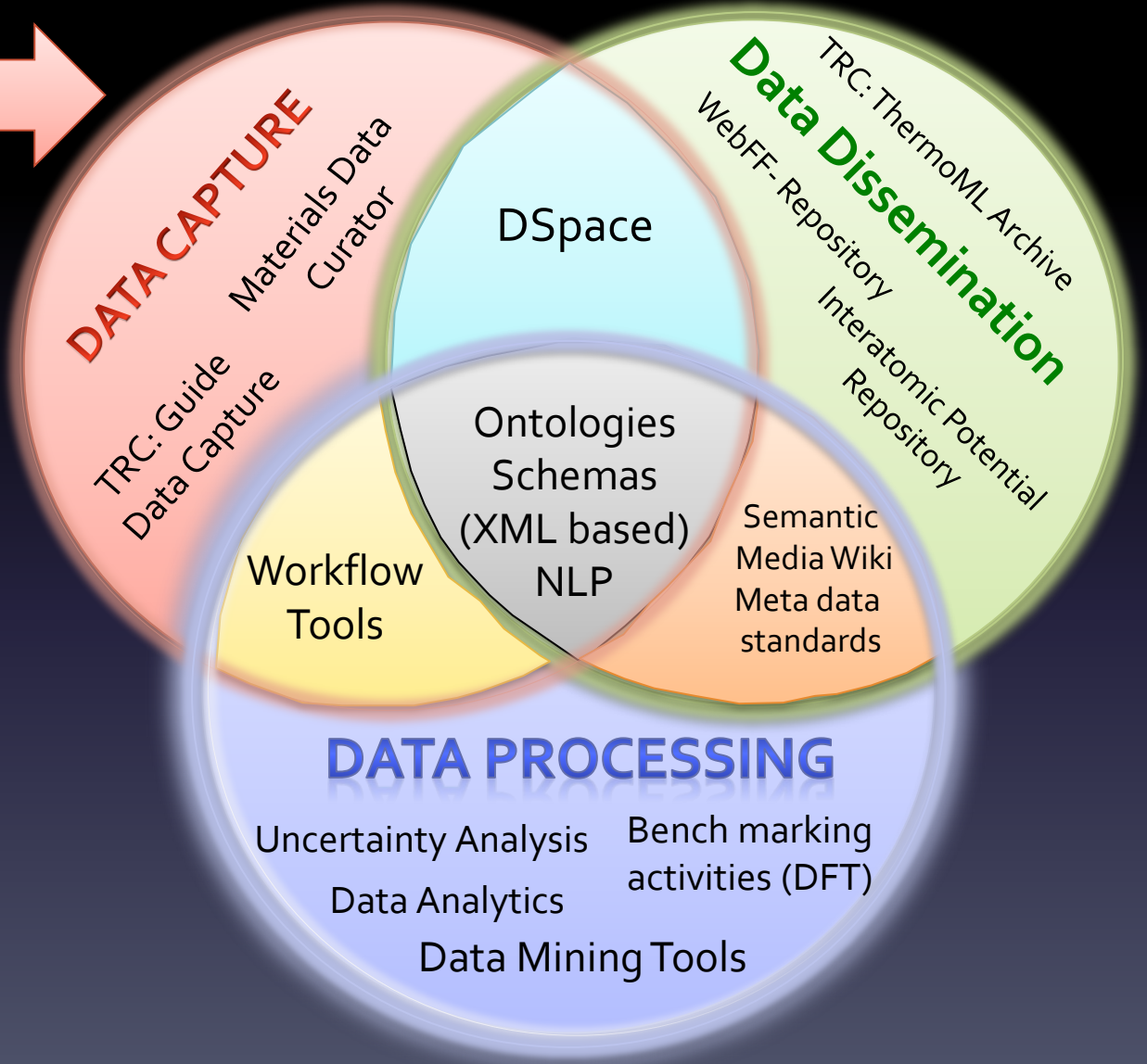
ASM International: Structural Data Demonstration Project

DOE/EERE Kinetics of Cast Mg Alloys

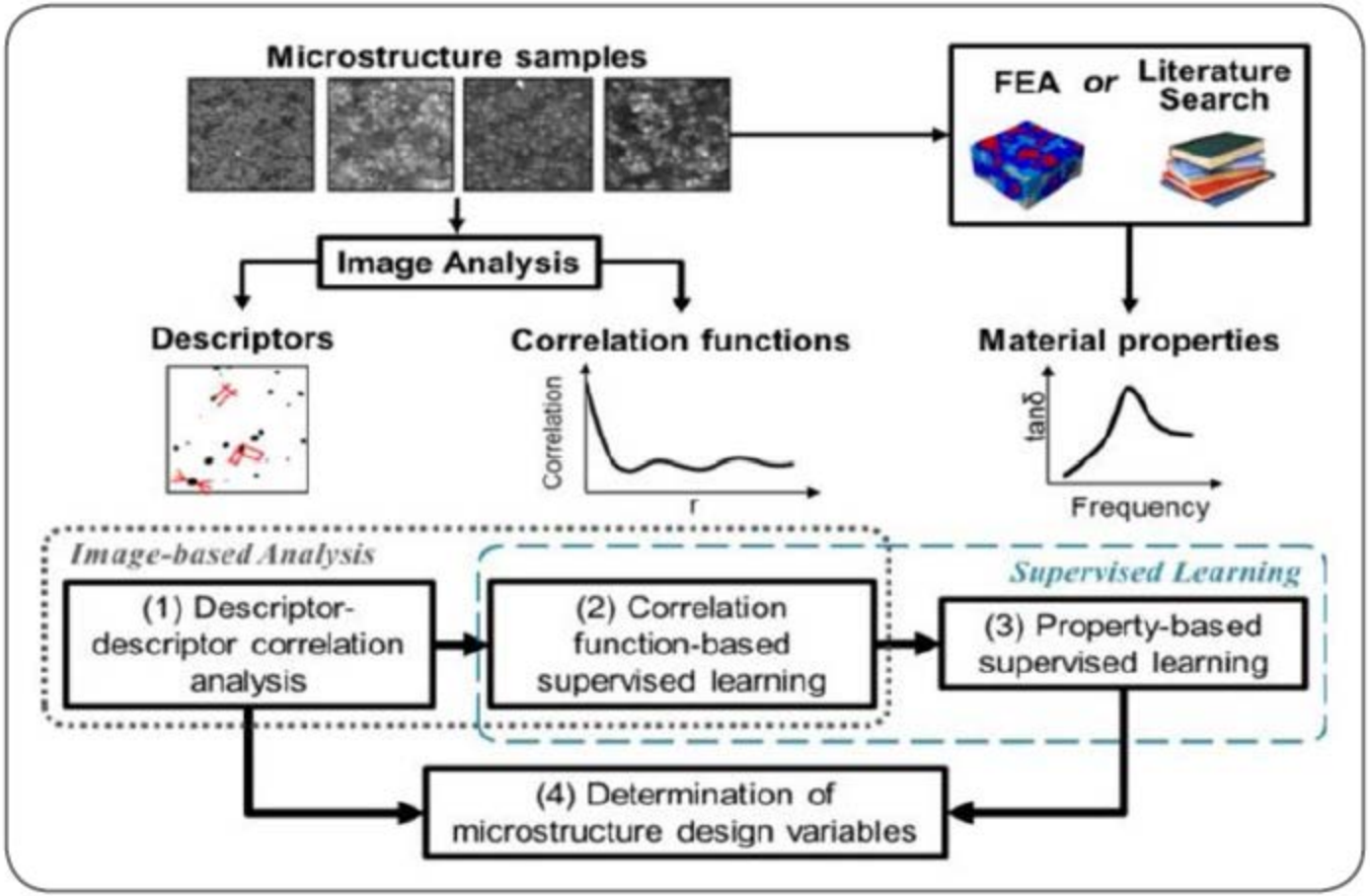
Journals collaboration

- IMMI
- *Others under discussion*

Coordinated by new Office of Data and Informatics



ChiMaD Data Mining

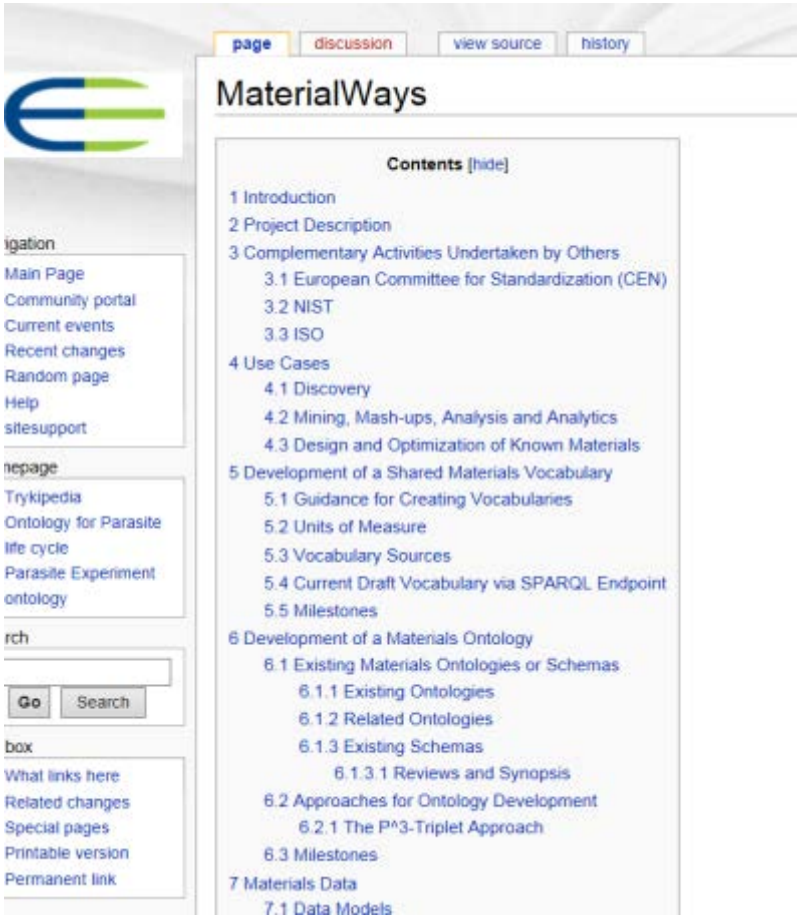


CHiMaD

NIST

AFRL Data Efforts

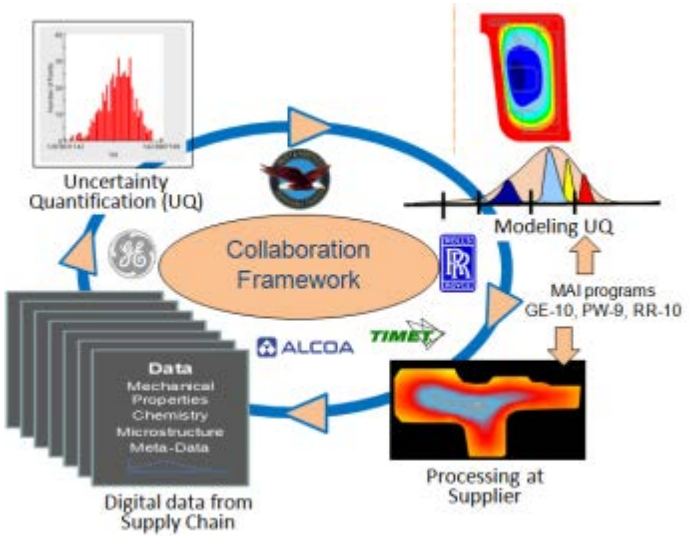
Wright State University



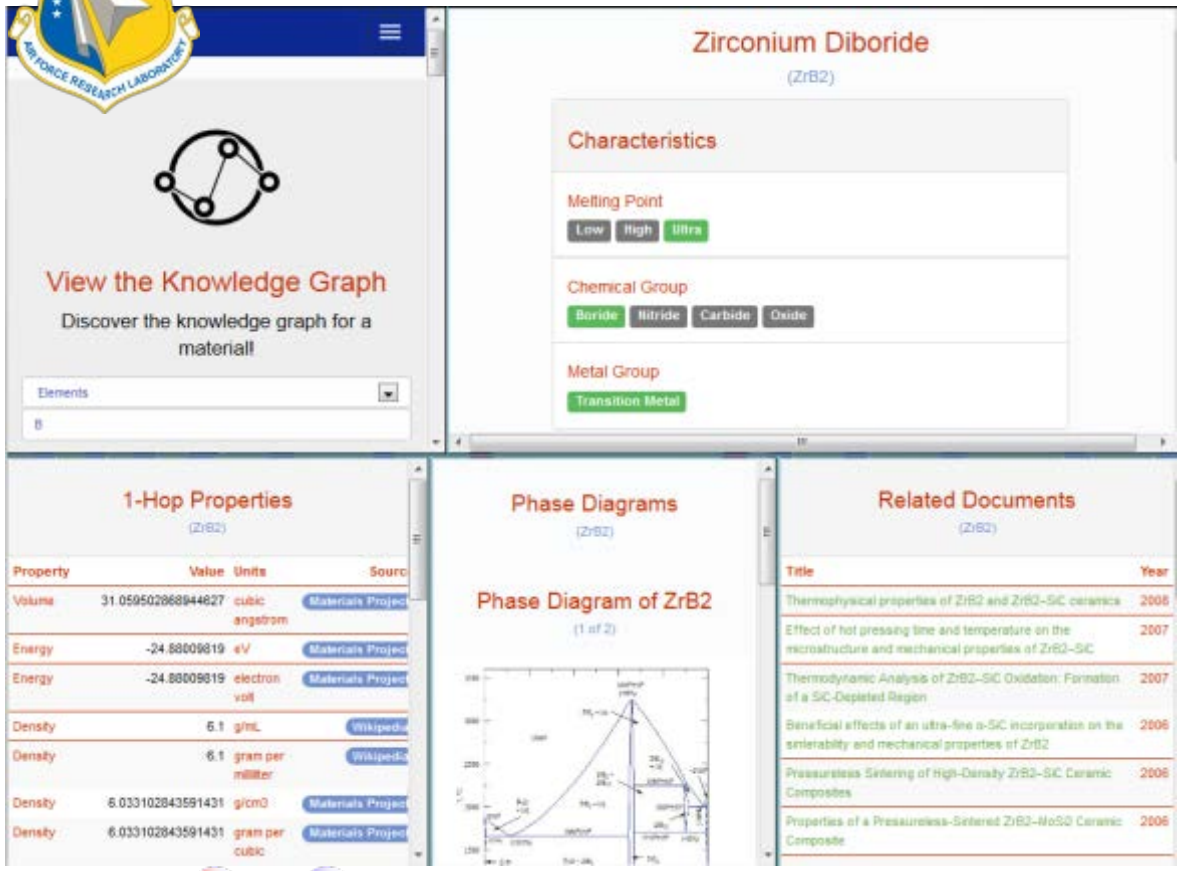
http://wiki.knoesis.org/index.php/MaterialWays

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

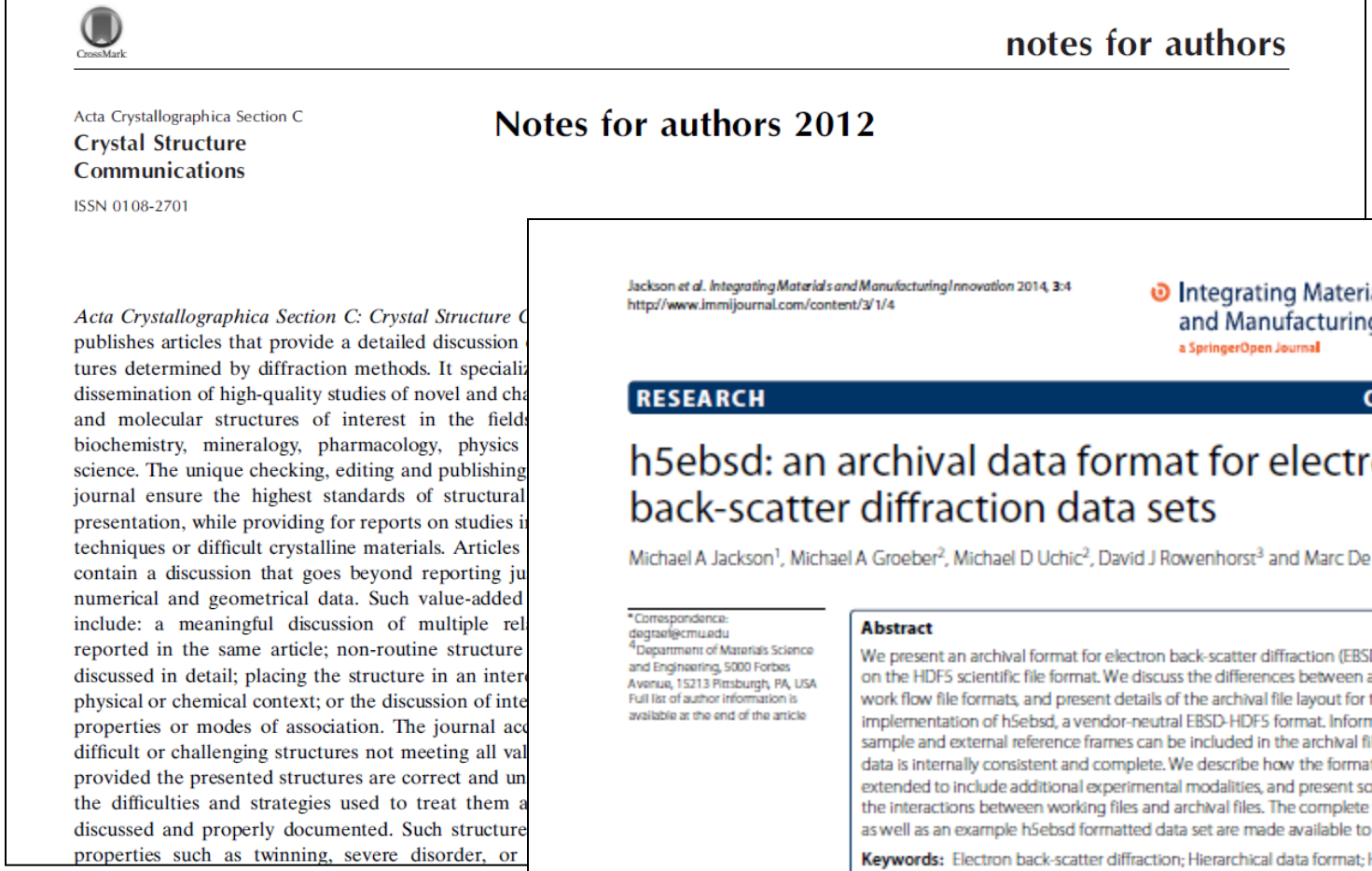


- **Develop Robust Mid-level Materials Ontology Ready for Crowd-sourcing and Initial Experimental Research Use**
- **Explore Sophisticated Low-level Ontologies**
- **Significantly Expand Size of, and Integration Across Data Stores**
- **Propose and Demonstrate Computational Approaches for Establishing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

Matonto

iNovex Information Systems

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 C
publishes articles that provide a detailed discussion
tures determined by diffraction methods. It specializ
dissemination of high-quality studies of novel and cha
and molecular structures of interest in the fields
biochemistry, mineralogy, pharmacology, physics
science. The unique checking, editing and publishing
journal ensure the highest standards of structural
presentation, while providing for reports on studies i
techniques or difficult crystalline materials. Articles
contain a discussion that goes beyond reporting ju
numerical and geometrical data. Such value-added
include: a meaningful discussion of multiple rel
reported in the same article; non-routine structure
discussed in detail; placing the structure in an inter
physical or chemical context; or the discussion of inte
properties or modes of association. The journal acc
difficult or challenging structures not meeting all val
provided the presented structures are correct and un
the difficulties and strategies used to treat them a
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.immijournal.com/content/3/1/4

Integrating Materials and Manufacturing Innovation
a SpringerOpen Journal

RESEARCH Open Access

h5ebsd: an archival data format for electron back-scatter diffraction data sets

Michael A Jackson[1], Michael A Groeber[2], Michael D Uchic[2], David J Rowenhorst[3] and Marc De Graef[4*]

*Correspondence:
degraef@cmu.edu
[4]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 work flow file formats, and present details of the archival file layout for the implementation of h5ebsd, 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 h5ebsd 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:

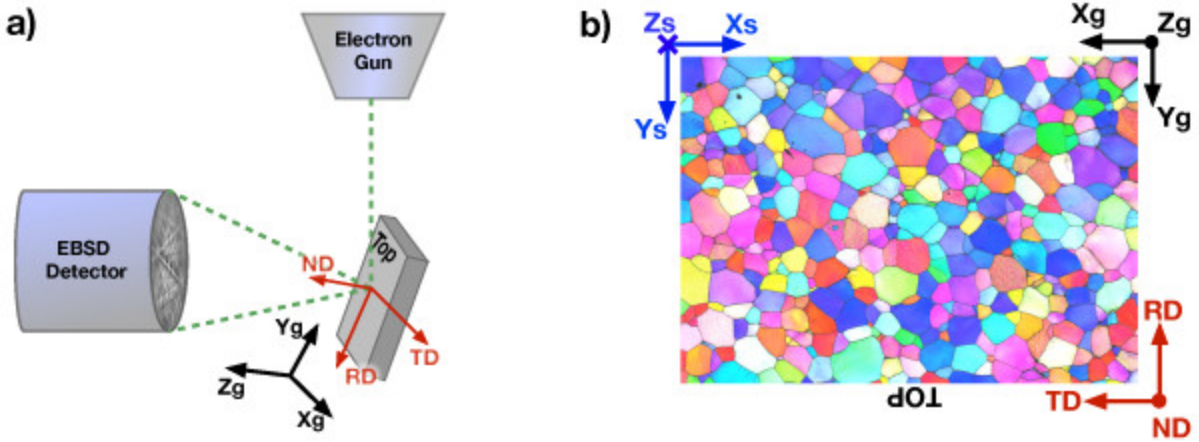
– Data generated from scanning electron microscopy (SEM) have become much more complex and content rich, but there is no standard format used to facilitate sharing and machine interpretation

Objectives:

– Better preservation and re-use of SEM data across the services

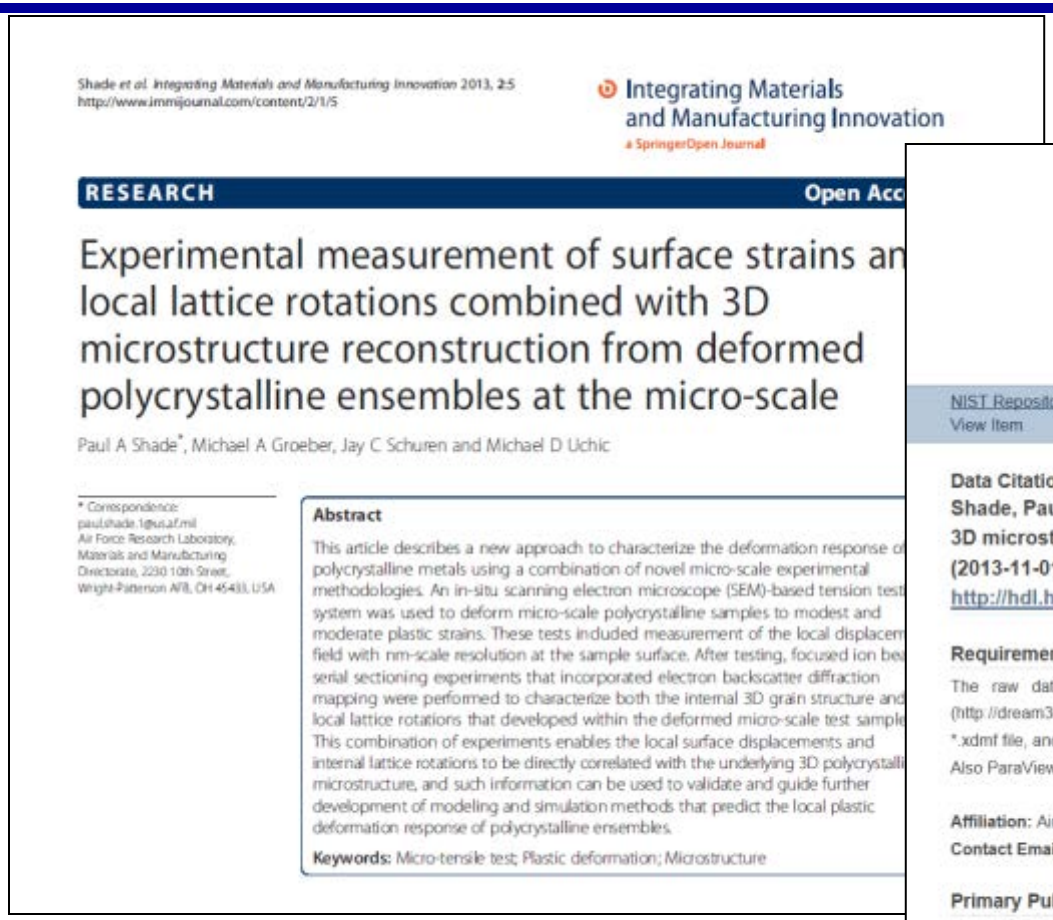
– Improve automatic metadata capture

– 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.immijournal.com/content/2/1/5

Integrating Materials and Manufacturing Innovation
a SpringerOpen Journal

RESEARCH Open Acc

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

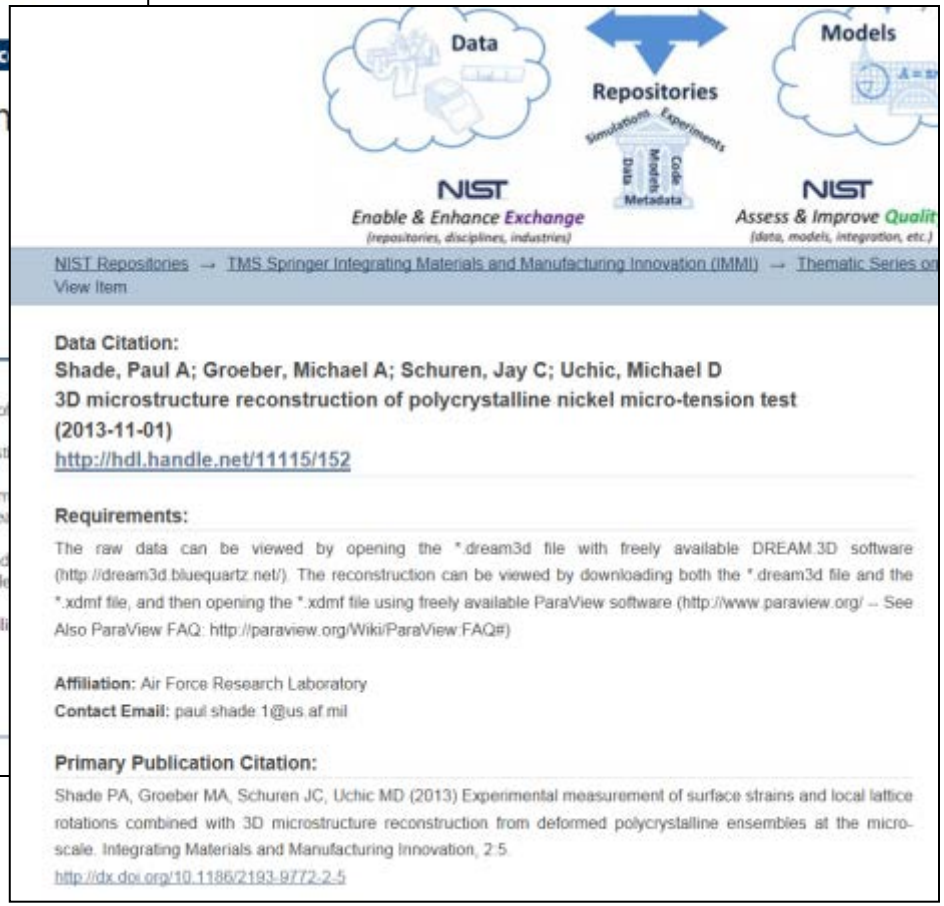
Paul A Shade*, Michael A Groeber, Jay C Schuren and Michael D Uchic

* Correspondence:
paul.shade.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 o 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 displacem field with nm-scale resolution at the sample surface. After testing, focused ion be 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 polycrystalli 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



http://www.immijournal.com/authors/instructions/datadescriptorarticle

Many Players in the Field

MDMC.net

cen

ASTM INTERNATIONAL

ISO

RDA RESEARCH DATA ALLIANCE Research Data Sharing without barriers

ICSU 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


AIR FORCE RESEARCH LABORATORY