

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

James A Warren

Director, Materials Genome Program

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*

# THE MATERIALS GENOME INITIATIVE: A NATIONAL PRIORITY

“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.”

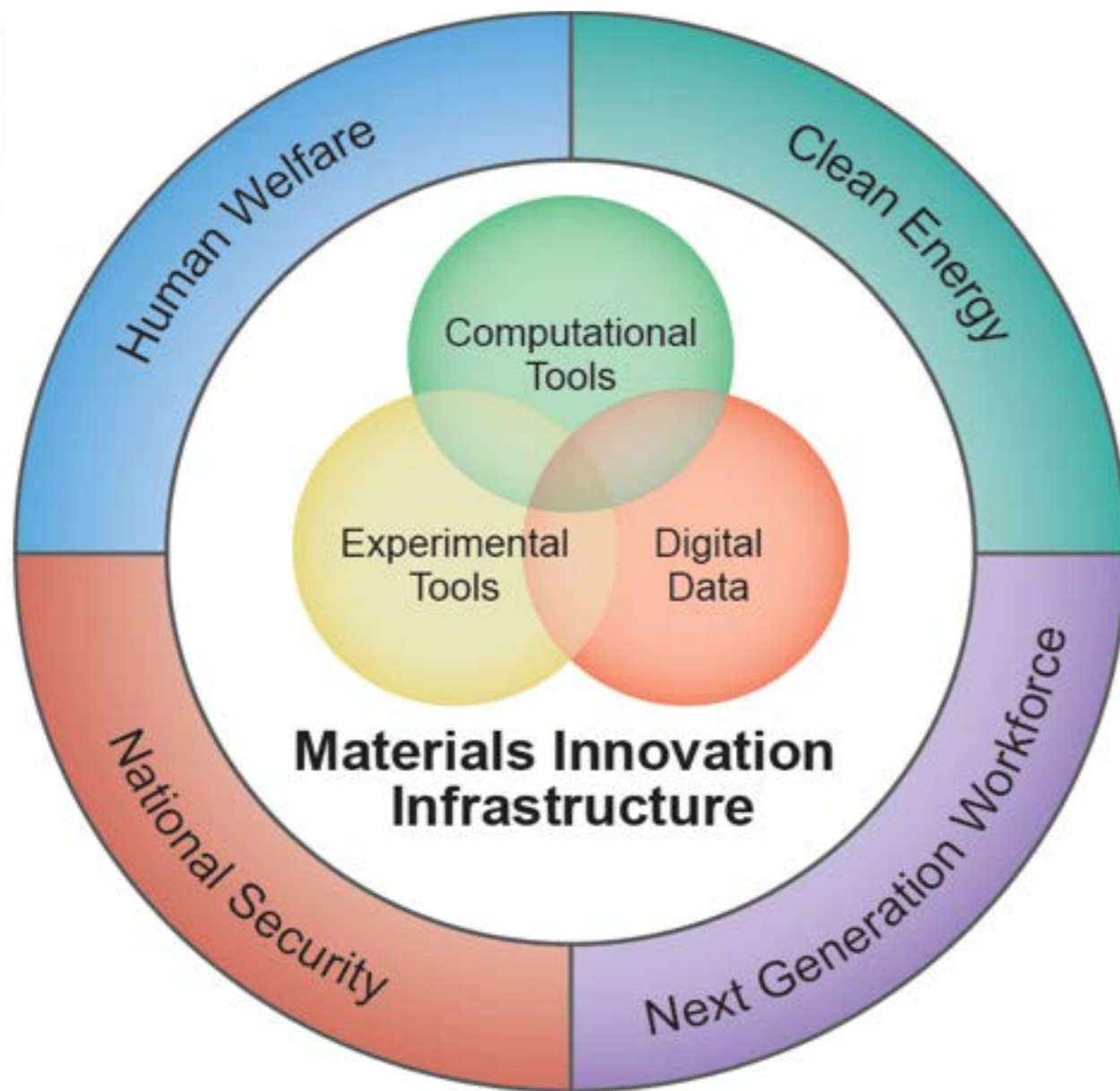
-President Obama (6/11)



There are two groups of people  
that don't like the name  
Materials Genome Initiative

# 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*



# A Multi-Agency Effort

# THE MGI SUBCOMMITTEE (SMGI)

## who we are

- MGI Subcommittee, Committee on Technology, NSTC
- First meeting April, 2012
- Membership includes the Federal agencies: NIST, DOE, DOD, NSF, NASA, NIH, US Geological Survey, National Nuclear Security Administration, DARPA, and Office of Management and Budget
- Co-chairs: OSTP (Whitman), AFRL (Ward), 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

SIGN UP FOR OUR STAKEHOLDER EMAIL LIST!

[www.mgi.gov](http://www.mgi.gov)

# 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

[https://www.whitehouse.gov/sites/default/files/microsites/ostp/NSTC/mgi\\_strategic\\_plan\\_-\\_dec\\_2014.pdf](https://www.whitehouse.gov/sites/default/files/microsites/ostp/NSTC/mgi_strategic_plan_-_dec_2014.pdf)

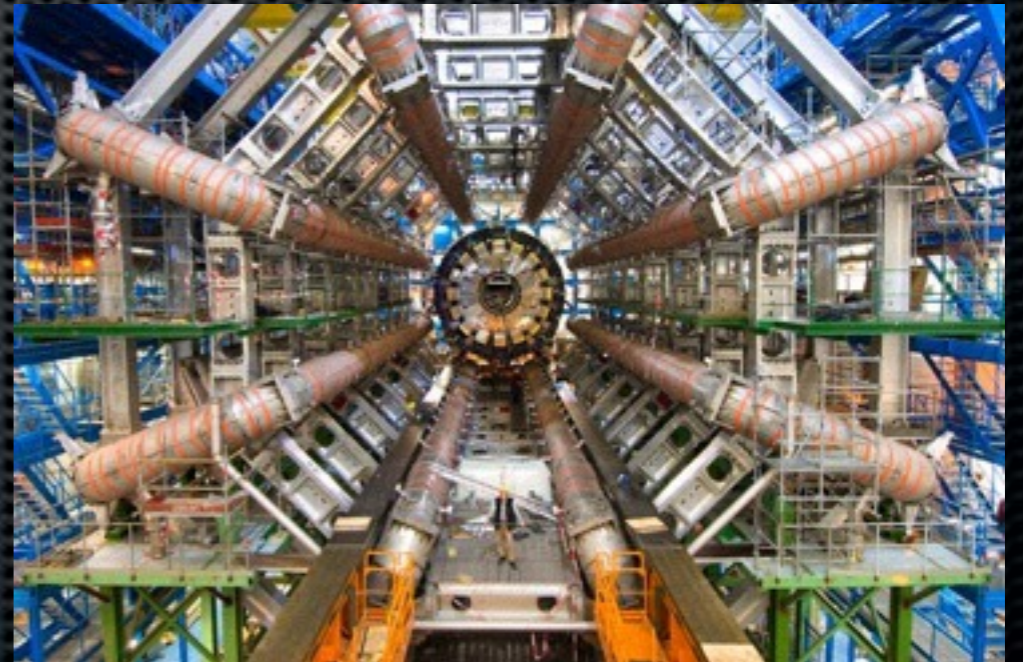
# The Relationship between Models, Measurement, Data, and Publication

A simple example to illustrate the key questions,  
gaps, obstacles to realizing materials by design



# Better Models = Less Data

- ✦ Take the LHC as an example
- ✦ Data produced at 1PB/sec !
- ✦ Reduced data saved: 300 MB/s
- ✦ That's a darn good model
- ✦ Other end : Biology?
- ✦ Materials: In the middle



- ✦ But you need to know the model to make sense of the data!

# What should we be doing?

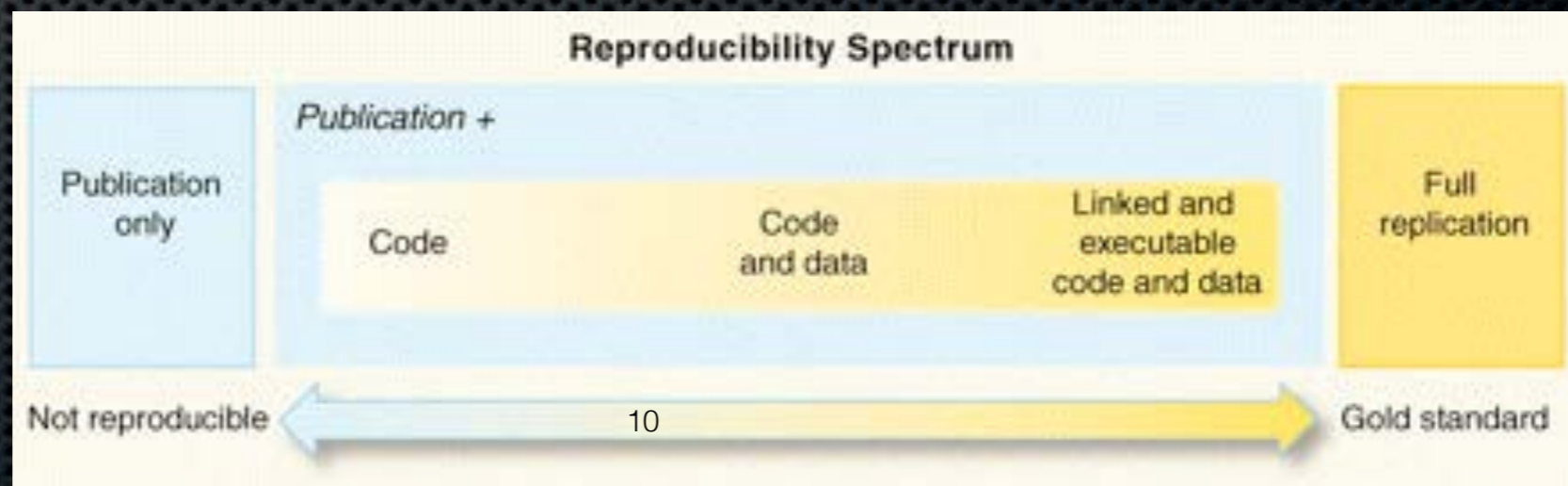
**Science is characterized by the iteration of experiment and models, yielding higher fidelity with lower uncertainty.**

The measurement or computation of a quantity (data) is generally *meaningless* without the associated quantifying model that defines both the data and its uncertainties.

**Thus** dissemination of data is ideally **the dissemination of the following information:**

- 1. Measured quantities,**
- 2. Associated quantifying models, and**
- 3. Raw data, including the protocols (and equipment) by which it was obtained**

## PUBLISHING



see R. Peng *Science* 2  
December 2011:  
1226-1227

# Consequences of the traditional approach

- High barrier to adoption of methods and results
- Extra expense due to duplication of effort
- Lost opportunities to enable discovery & further science

So Now What?

# Formulating the NIST Role in MGI

Some background and further thoughts

# NIST Products and Services

## Measurement Research

⑩ ~ 2,200 publications per year

## Standard Reference Data

⑩ ~ 100 different types

⑩ ~ 6,000 units sold per year

⑩ ~ 226 million data downloads per year



© Robert Rathe



## 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

# Data mandates in the Federal Government



OSTP “Public Access”  
Memo Feb 22, 2013

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 Research

Executive Order  
May 9, 2013

For Immediate Release

May 09, 2013

## Executive Order -- Making Open and Machine Readable the New Default for Government Information

EXECUTIVE ORDER

MAKING OPEN AND MACHINE READABLE THE NEW DEFAULT  
FOR GOVERNMENT INFORMATION

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 easy 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 navigation systems, weather newscasts and warning systems, location-based applications, precision farming tools, and much more, improving Americans' lives in countless ways and leading to economic growth and job creation. In recent years, thousands of Government data resources across fields such as health and medicine, education, energy, public safety, global development, and finance have been posted in machine-readable form for free public use on Data.gov. Entrepreneurs and innovators have continued to develop a vast range of useful new products and businesses using these public information resources, creating good jobs in the process.

OMB “Open Data”  
Memo May 9, 2013



EXECUTIVE OFFICE OF THE PRESIDENT  
OFFICE OF MANAGEMENT AND BUDGET  
WASHINGTON, D.C. 20503

May 9, 2013

THE DIRECTOR

M-13-13

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: Susan M.

DATA Act, P.L. 113-101  
May 9, 2014

S. 904

## One Hundred Thirteenth Congress of the United States of America

AT THE SECOND SESSION

Began and held at the City of Washington on Friday,  
the third day of January, two thousand and fourteen

### An Act

To amend the Federal Funding Accountability and Transparency Act of 2006 to increase accountability and transparency in Federal spending, and for other purposes.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

#### SECTION 1. SHORT TITLE.

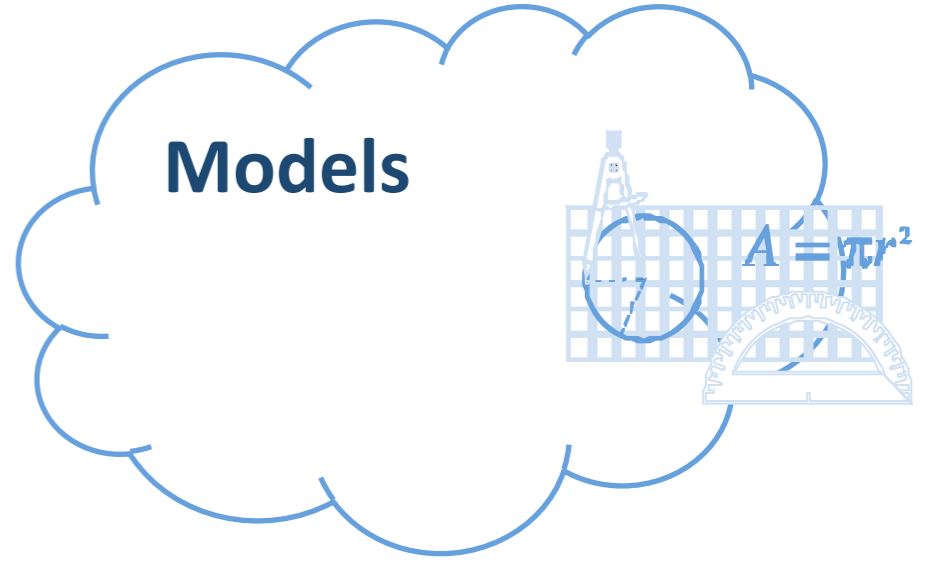
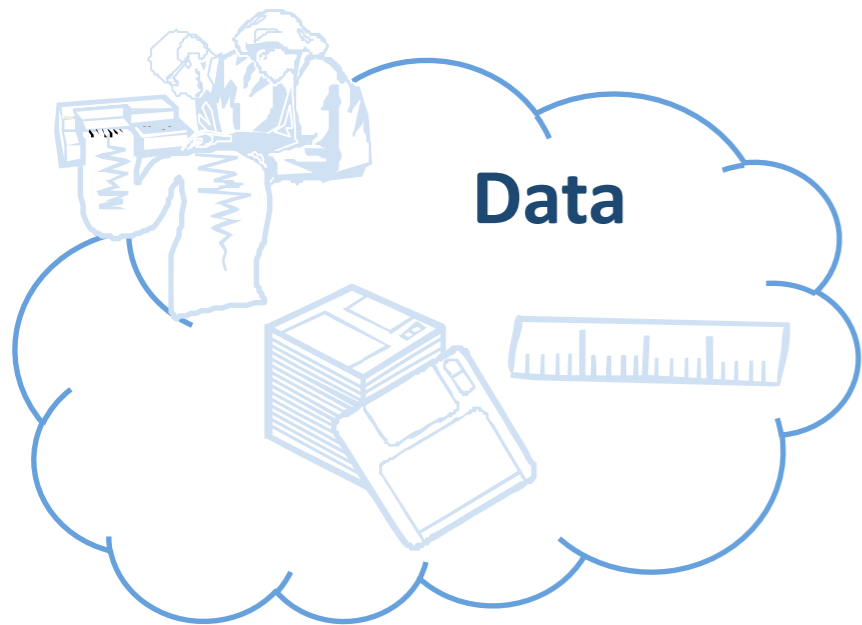
This Act may be cited as the “Digital Accountability and Transparency Act of 2014” or the “DATA Act”.

#### SEC. 2. PURPOSES.

The purposes of this Act are to—

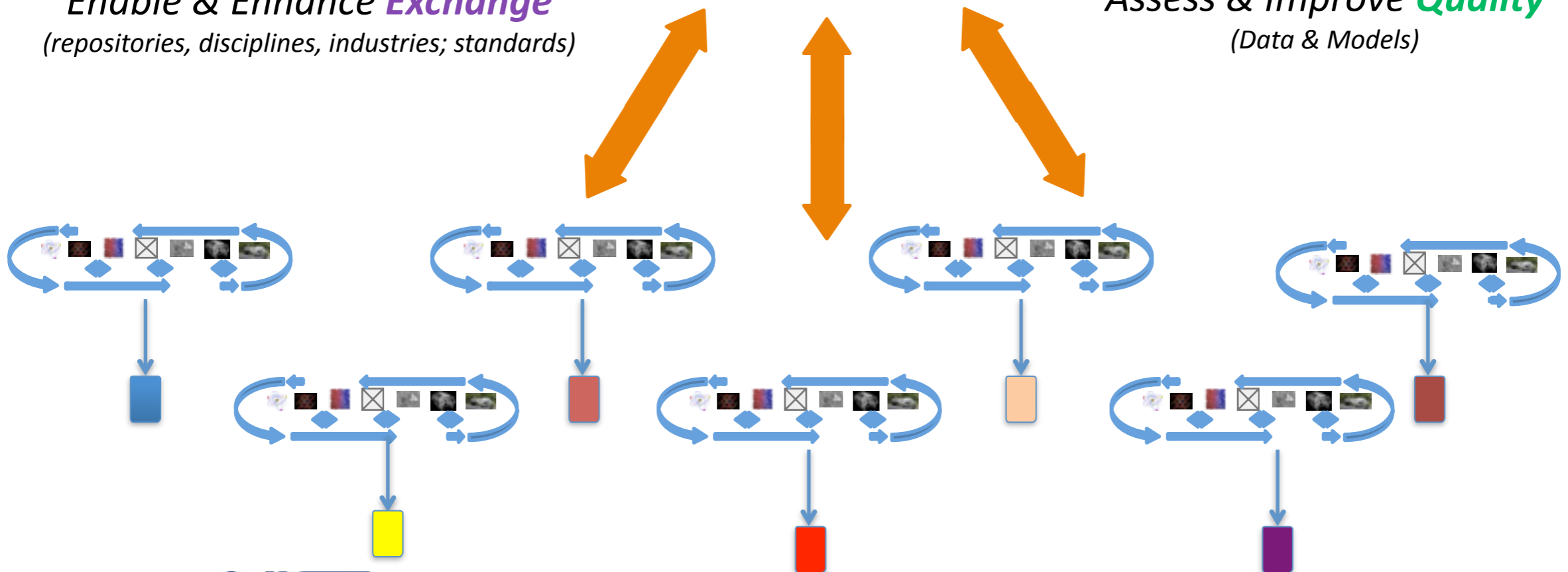
(1) expand the Federal Funding Accountability and Transparency Act of 2006 (31 U.S.C. 6101 note) by disclosing direct Federal agency expenditures and linking Federal contract, loan, and grant spending information to programs of Federal agencies to enable taxpayers and policy makers to track Federal spending more effectively;

(2) establish Government-wide data standards for financial data and provide consistent, reliable, and searchable Govern-



**NIST**  
 Enable & Enhance **Exchange**  
 (repositories, disciplines, industries; standards)

**NIST**  
 Assess & Improve **Quality**  
 (Data & Models)



**NIST**

New **Methods and Metrologies**  
 (data driven analysis and models)

**Materials w/ Targeted Properties**

# SCOPE: Goals of Initiative

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*

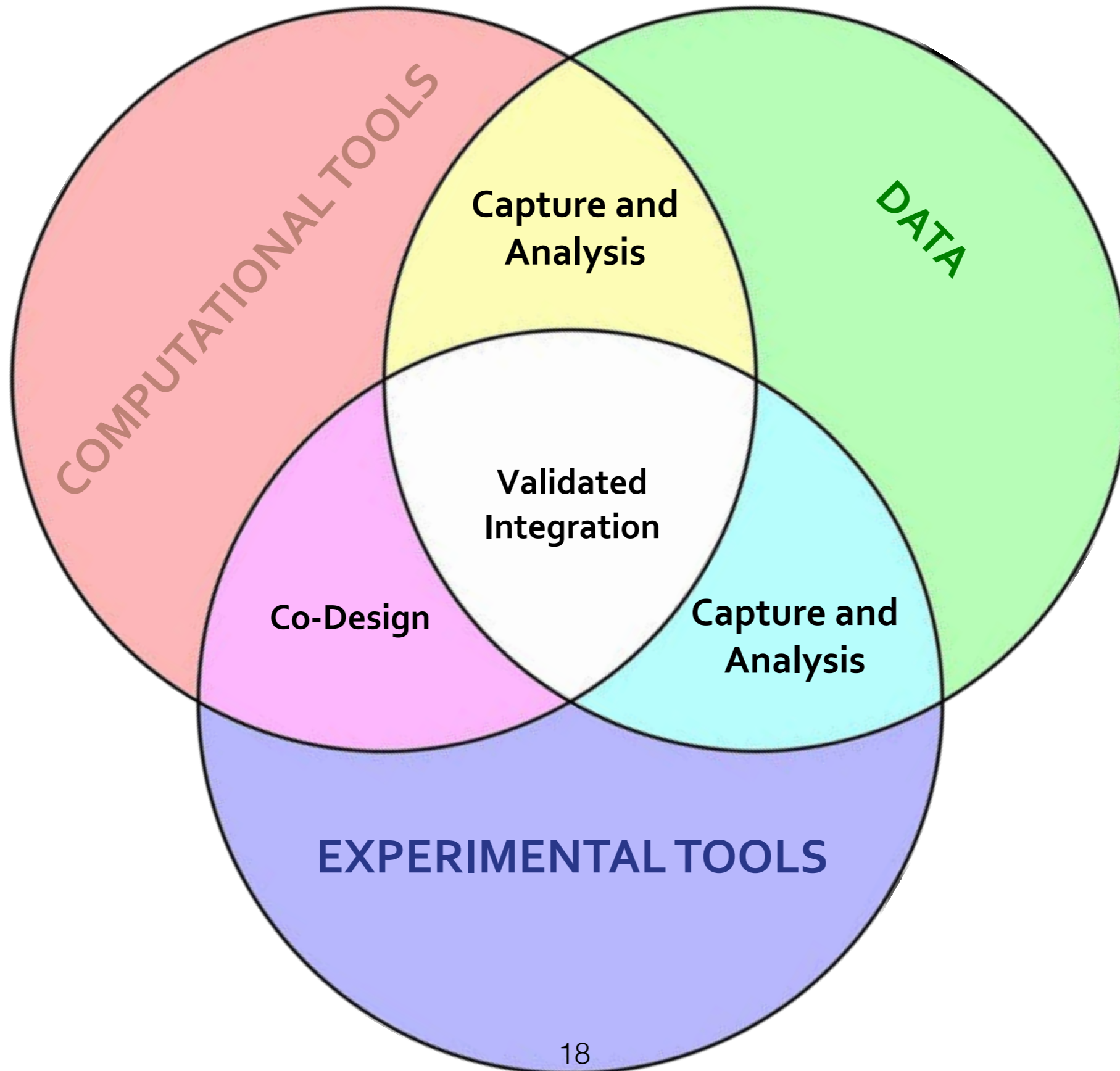
**Goal 3:** NIST establishes *new methods, metrologies and capabilities necessary for accelerated materials development.*



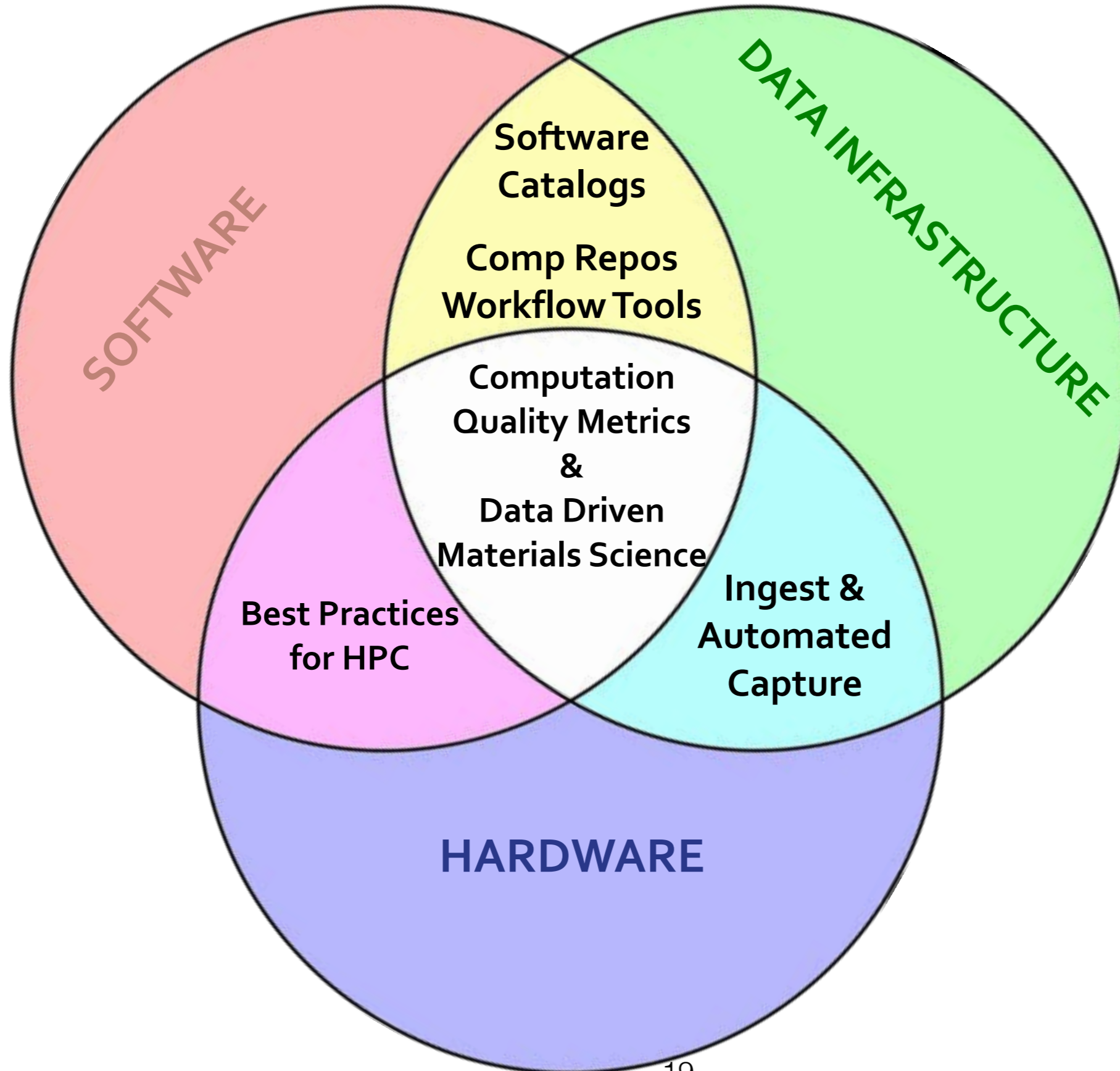
# The NIST Portfolio and Wider Questions about Data

[MGI.NIST.GOV](http://MGI.NIST.GOV)

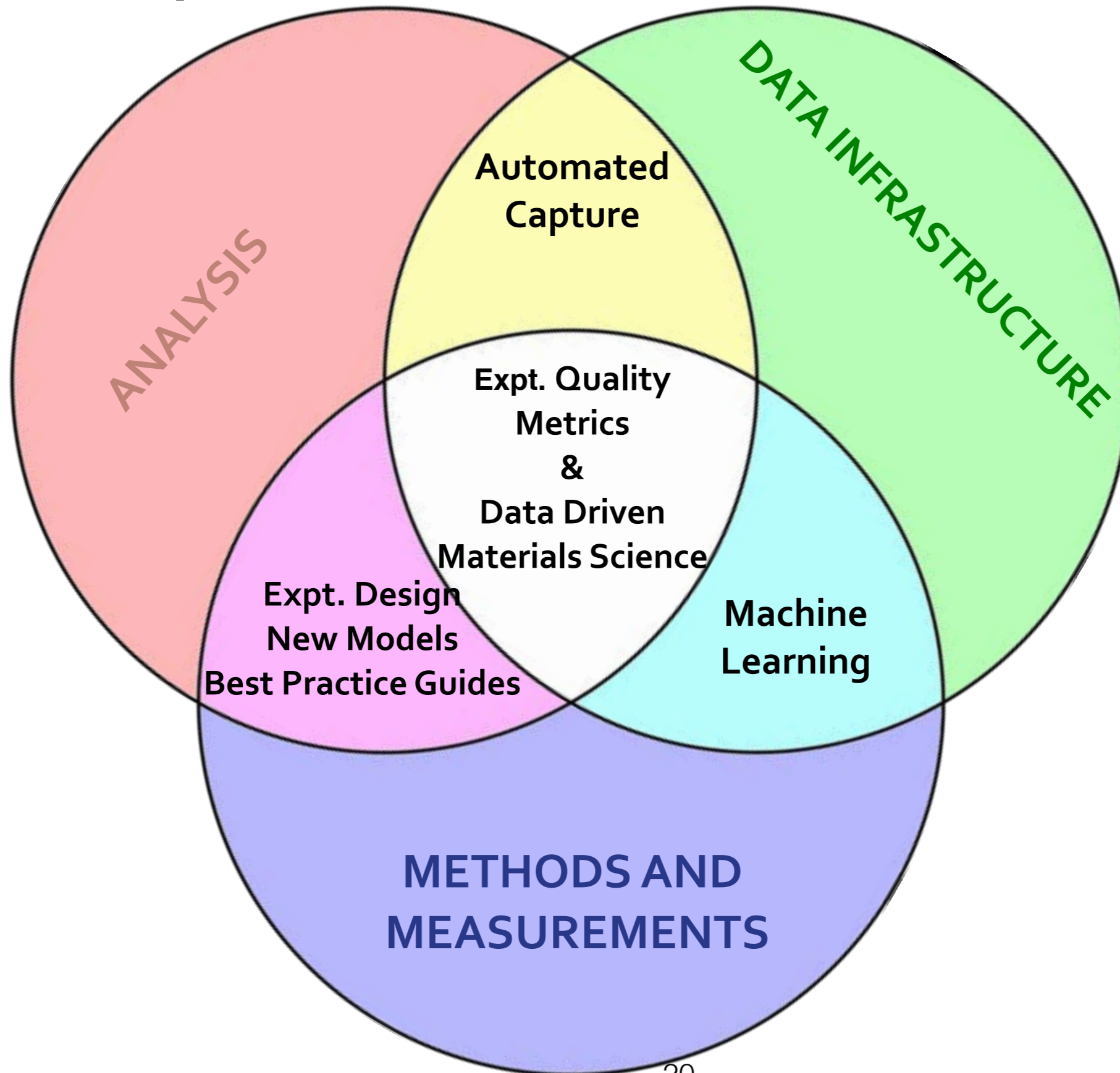
# MGI



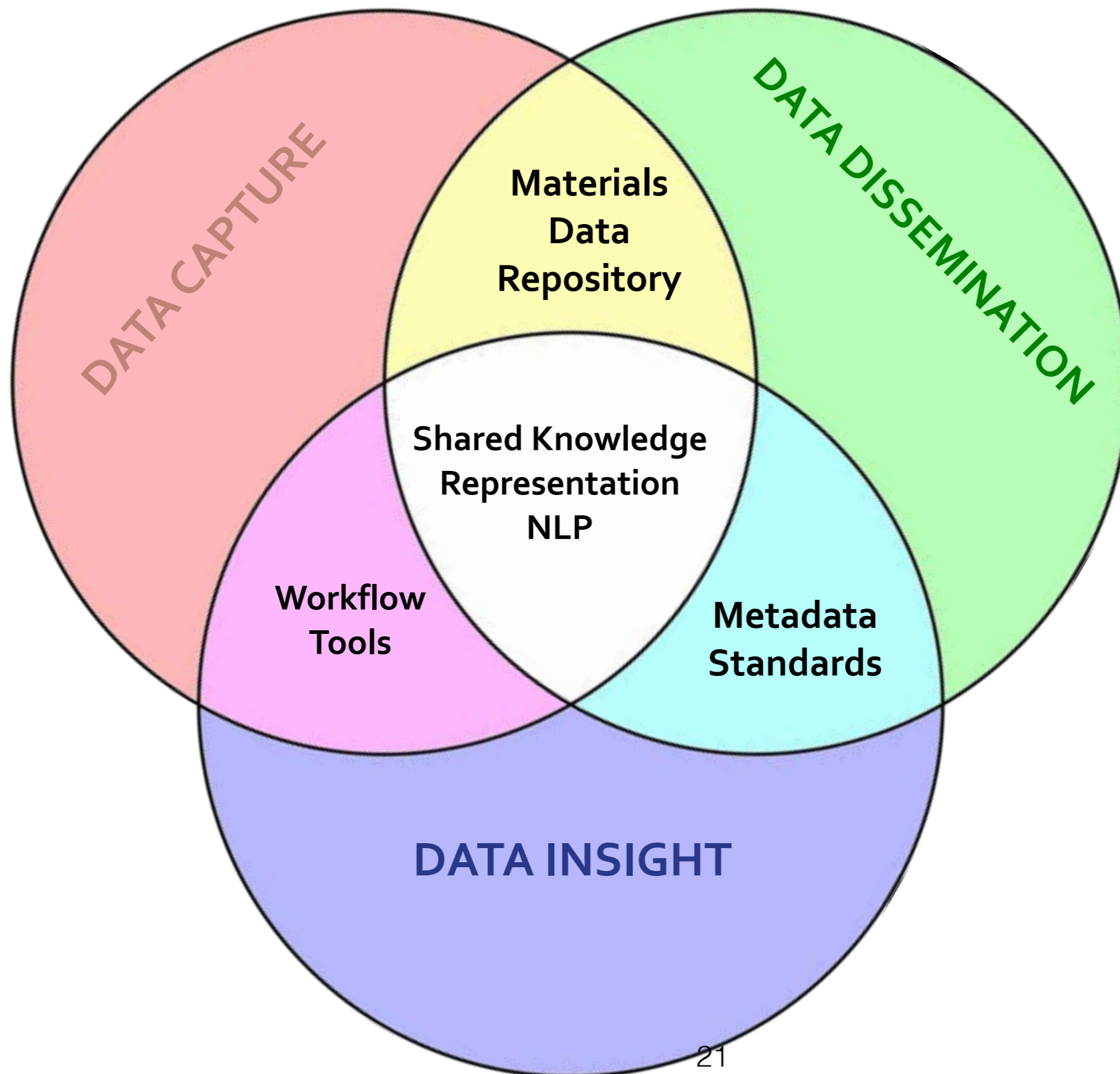
# Computational Tools



# Experimental Tools



# Data



# 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 4<sup>th</sup> 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

# What's NIST, the USG and International Community doing about "Data"

- This is really a helicopter view of the broad issues
- You can hopefully not understand much of this part of talk and it shouldn't matter
- Many issues
  - getting the data "in"
  - getting the data "out"
  - Evaluating the data for quality
  - Using the data for maximum effect
- Who are we working with, who else is a "player"?

Google Materials

Search



# Data Discovery, What do we Need?

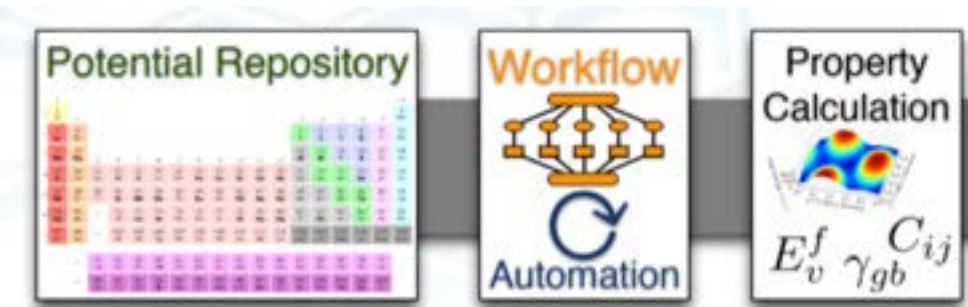
1. Some sort of interface/api that the user interacts with to deposit their information such that (2) is enabled. This entails
  - a. The establishment of **repositories** to store the data
  - b. The reposed data should then be **marked up with sufficient metadata** to inform someone else how the data was created, including attribution or provenance information relevant to citation
  - c. Assignment of a **persistent digital identifier** (like the DOI for journals) so the data can be cited and discovered by others
  - d. Tools must be developed to
    - i. enable the ingest of data from computation or experiment for deposition in a repository
    - ii. simplify metadata collection, e.g. by
      1. Automatically assigning some basic metadata
      2. Implementing automatic extraction of metadata to the extent possible
    - iii. Assign persistent digital identifiers
2. Some sort of interface/api that the user interacts with to find needed information. This requires (at least)
  - a. the **registration of the availability of the data into some sort of “registry”** to enable discovery without prior knowledge of the existence of the repository/specific data features .
  - b. Various types of **policy** enforcement
  - c. Tools to evaluate the **“quality”** of the data
    - i. Enabling determinations of the data’s uncertainty, validity, sensitivity, as well as other metadata qualities
    - ii. Enabling assessments of the relevance of the data to the question at hand, such as [quick-look plotting or imaging capabilities]
  - d. Tools for data manipulation
  - e. **Terminologies and ontologies** to enhance search

# Repositories

# Property Calculations on the Interatomic Potential Repository

Zachary Trautt, Materials Measurement Science Division  
Chandler Becker, Materials Science and Engineering Division

<http://www.ctcms.nist.gov/potentials/>



NIST | NIST Time | NIST Home | About NIST | Contact Us | A-Z Site Index | Search

Material Measurement Laboratory

About MML | Publications | Topic/Subject Areas | Products/Services | News/Multimedia | Events | Programs/Projects | Facilities

NIST Home > MML > CTCMS > Potentials > Properties

## Interatomic Potentials Repository Project

Properties | Methods | References | FAQ | Resources | People | Contact

### Property Calculations

#### Overview

These pages contain property calculation methods and available results for a number of interatomic potentials. The calculations are fully documented and available to users.



The focus of this work is automation and reproducibility of atomistic simulations. Researchers are welcome to use scripts, methods, and data with appropriate disclaimers and attributions.

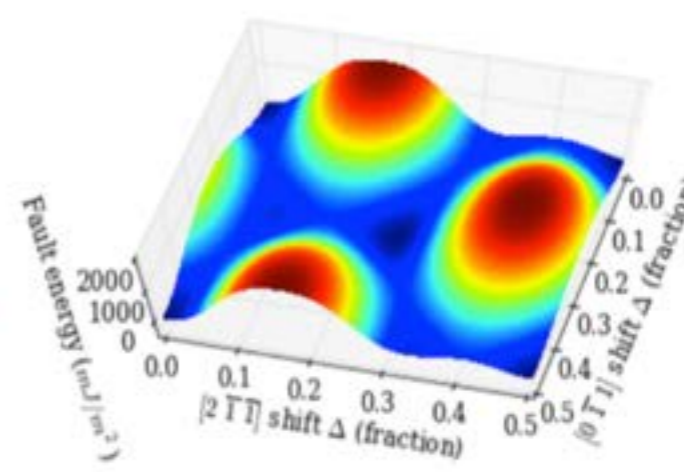
#### Calculation Methods Summary

Descriptions of the property calculation methods used on this site are available on the [Calculation Methods Summary](#) page. Included are any applicable derivations, discussions, and sample scripts.

#### Available Results (by Property)

Select a link below to examine the results for a given property calculation method:

-  **Generalized planar fault energy at 0K** – Calculation of generalized planar fault energy through energy minimization at 0 K.
-  **Bulk mechanical property multi-test at 0K** – This test computes the lattice constants, cohesive energy, bulk modulus, and elastic constants at a temperature of 0K.



The 3D surface plot shows fault energy (mJ/m<sup>2</sup>) on the vertical axis (0 to 2000) versus shift Δ (fraction) on the horizontal axes (0.0 to 0.5). The plot displays several peaks and valleys, indicating energy minima and maxima for different shift values.

# WebFF-A Smart Force-Field Repository of Soft Materials

<http://www.nist.gov/mml/msed/polymers/webff.cfm>

Frederick R. Phelan Jr. , *Material Science & Engineering Division, Polymers & Complex Fluids Group*

*NIST , Gaithersburg, MD*


Huai Sun , *Aeon Technology Inc. and School of Chemistry and Chemical Engineering, Shanghai Jiao*

*Tong University, Shanghai*

## WebFF

### Smart force-field repository

- Assignment Mode
  - Integrated database with software assignment engine, creates workflow
    - Input: Single molecule or molecular system
    - Output: LAMMPS/Gromacs data files
  - Multiple input and output formats
  - Fills missing parameters
- Insertion Mode
  - Extensible, multiple table format allows insertion of new FF data and encourages and enables data sharing
  - Creates central hub for expanding FF coverage (missing parameter problem)



**Assignment Interface**

Select Force Field Type :  AMBER  CFF

Filter available force fields by:

- Chemical group (alkyl, carbonyl, etc.)
- Molecular names
- Molecular formulas

Input: alkyl, amino, alcohol Search

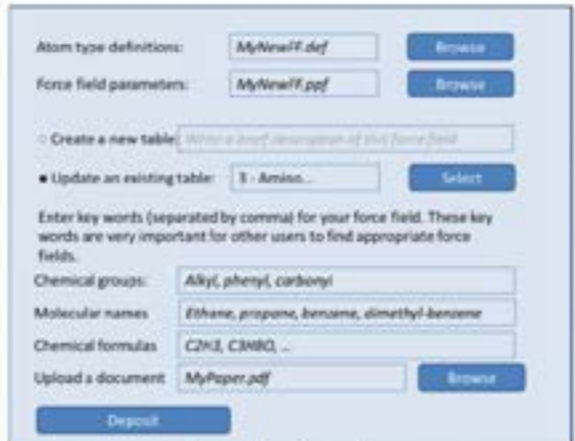
Available force fields:

ID	Description	Developer	Rating	Doc	Order
1	Basic over...	Tom	★★★★	link	Up Down
3	Amino...	Smith	★★★★	link	Up Down
8	Graphene	Dave	★★★	link	Up Down

Load molecular model : MyMolecules.mol2 Browse

Select Simulation Engine :  LAMMPS  GROMACS

Get Parameters Deposit or Modify My Force Field



**Insertion Interface**

Atom type definitions: MyNewFF.def Browse

Force field parameters: MyNewFF.ppf Browse

Create a new table: Write a brief description of this force field

Update an existing table: E - Amino... Select

Enter key words (separated by comma) for your force field. These key words are very important for other users to find appropriate force fields.

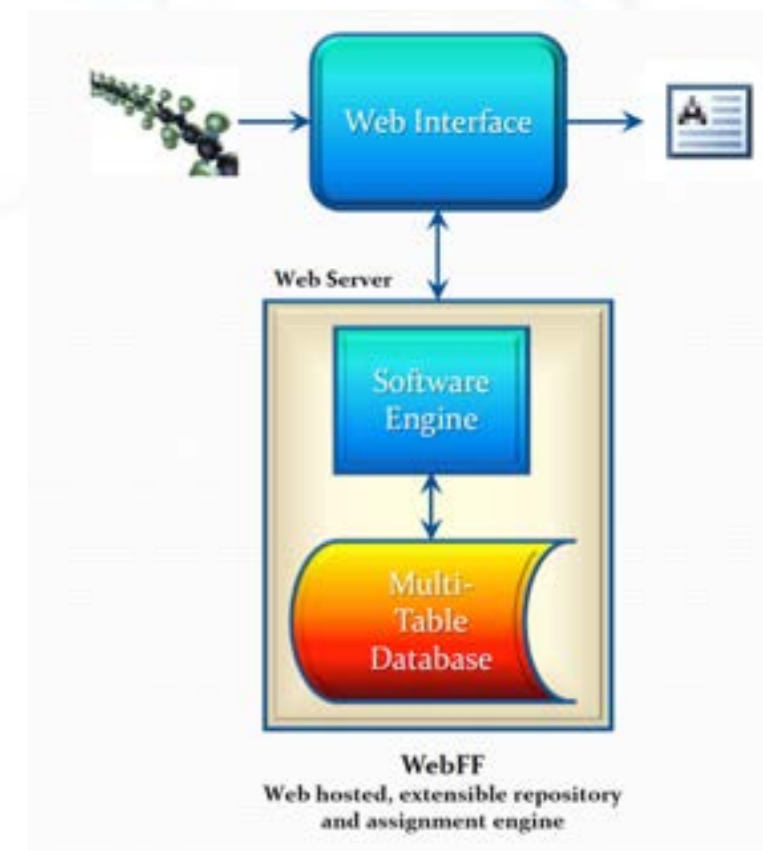
Chemical groups: Alkyl, phenyl, carbonyl

Molecular names: Ethane, propane, benzene, dimethyl-benzene

Chemical formulas: C2H2, C2H6, ...

Upload a document: MyPaper.pdf Browse

Deposit



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

Login

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**
    - [In-Situ Si Composites \(Si-Cr-Al\)](#)
  - **Polymer Nanocomposites**
    - [Data for Polymer Nanocomposites](#)
  - **Precipitation Strengthened Alloys**
    - [Co-base Alloys](#)
    - [Shape Memory Alloys](#)
- **Computational File Repository**
  - [Atomistic Simulations](#)
  - [CALPHAD Assessments](#)
  - [First Principles Phase Stability \(FPPS\) Files](#)
  - [Other Computational Methods](#)
- **Experimental Data Repository**
  - [Diffusion Data](#)
  - [Molar Volume/Thermal Expansion Data](#)
  - [Other Experimental Data](#)
  - [Phase Equilibria and Thermodynamic Data](#)
  - **Mechanical Properties**
    - [Elasticity Data](#)

### Search NIST Repositories

[Advanced Search](#)

### Browse

[All of NIST Repositories](#)  
[Communities & Collections](#)  
[Subjects](#)  
[Titles](#)  
[Authors](#)

### My Account

[Login](#)

Customized DSpace repository for materials  
➤ Enables sharing of a variety of data types, including text, images, and video

Data Citation:

Campbell, Carelyn; Zhao, J-C; Henry, M. F.

Examination of Ni-base superalloy diffusion couples containing multiphase regions (2014-04-02)

<http://hdl.handle.net/11256/22>

→ Digital Identifier

Affiliation: National Institute of Standards and Technology, Metallurgy Division, Gaithersburg, MD 20899-8555, USA  
General Electric Company, GE Global Research, 1 Research Circle, Niskayuna, NY 12309, USA  
Contact Email: carelyn.campbell@nist.gov

Primary Publication Citation:

Materials Science and Engineering A 407 (2005) 135-148  
<http://dx.doi.org/10.1016/j.msea.2005.07.015>

→ Related Work

Related Publications by Author:

Campbell CE, Boettinger WJ, Kattner UR (2002) Development of a diffusion mobility database for Ni-base superalloys. Acta Mater 50:775-792 DOI: [http://dx.doi.org/10.1016/S1359-6454\(01\)00383-4](http://dx.doi.org/10.1016/S1359-6454(01)00383-4)

Campbell CE, Zhao JC, Henry MF (2004) Comparison of experimental and simulated multicomponent Ni-base superalloy diffusion couples. J Phase Equil Dif 25 (1):8-15. DOI: <http://dx.doi.org/10.1361/10549710417968>

Abstract:

Four Ni-base superalloy diffusion couples with multiphase regions were studied. The diffusion couples contained single-phase (gamma), two phase (gamma + MC carbide) and three-phase (gamma + gamma prime + MC carbide) regions. Measured average composition profiles were in good agreement with the diffusion simulation predictions. The measured and predicted phase fraction profiles showed similar trends; however, there were some discrepancies in the predicted position of the gamma + gamma prime + MC/ gamma + MC boundary. Phase fraction profiles and optical metallography were used to determine the type and direction of the moving phase region boundaries.

Search

Search input field with buttons: Search, This, Advanced

Browse

- All of h
- Co
- Su
- Ti
- Aut
- This C
- Su
- Ti
- Aut

My Ac

Login

Funding Agency & Award No.:

This work was supported by the Defense Advanced Research Project Agency (DARPA) under the accelerated Insertion of Materials (AIM) Program (Grant number F33615-00-C-5215) with Dr. L. Christodoulou as the project manager and Dr. Rollie Dutton as the project monitor. The authors would like to express their appreciation to N. Saunders for the use of his thermodynamic database for Ni alloys and to Louis A. Pluso, Rubenna C. Goss, Mitch Hammond and Karen Denike for their experimental support.

→ Data files

Files in this item



Name: R95-R88-expsimul.txt  
Size: 25.32Kb  
Format: Text file

View/Open

Description:  
Experimental and simulate composition profiles for the R88/R95 diffusion couple at 1150 C for 1000 h



Name: r95r88-1000h-labe ...  
Size: 58.76Mb  
Format: TIFF image

View/Open

Description:  
Micrograph of R95/R88 diffusion couple after 1000 h at 1150 C

The following license files are associated with this item:

- Creative Commons

This item appears in the following Collection(s)

- Diffusion Data

Offer licenses with attribution 3.0



Except where otherwise noted, Universal



Related items

Showing items related by title, author, creator and subject.

Further Studies on the Nickel-Aluminum System. I. The  $\beta$ -Ni<sub>2</sub>Al<sub>3</sub> Phase Fields

Taylor, A; Doyle, N.J. (1972-01-31)

New lattice parameter and density results have been obtained for alloys in the  $\beta$ -NiAl and  $\beta$ -Ni<sub>2</sub>Al<sub>3</sub> phase fields of the nickel-aluminum system. The lattice parameter of the  $\beta$ -NiAl phase (CsCl-type) falls linearly from ...

Elemental vacancy diffusion for fcc and hcp structures

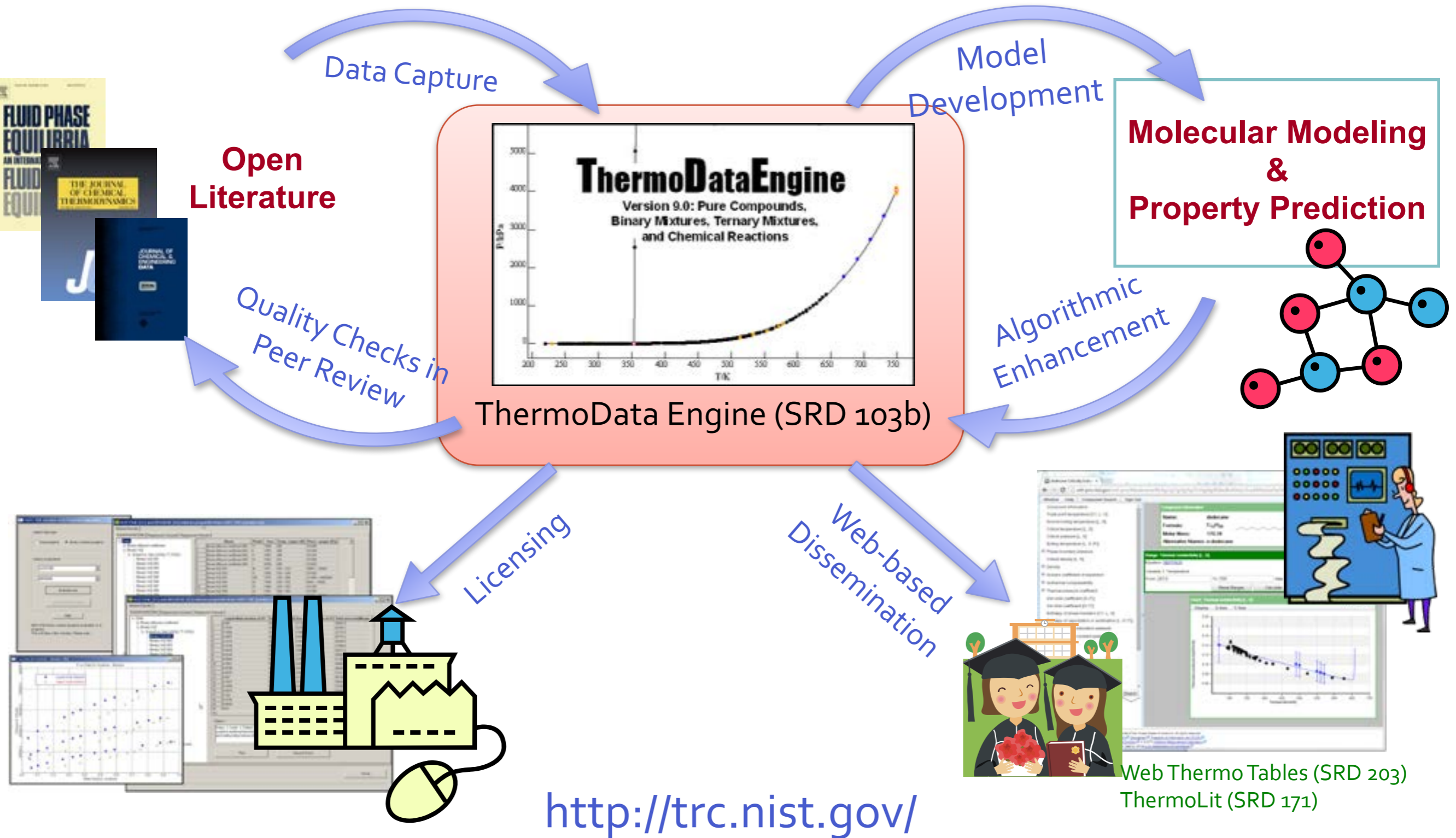
Angsten, Thomas; Mayeshiba, Tam; Wu, Henry; Morgan, Dane (2014-08-08)

This work demonstrates how databases of diffusion-related properties can be developed from high-throughput ab initio calculations. The formation and migration energies for vacancies of all adequately stable pure elements ...

# Metadata Curation

Markup

# THERMODYNAMICS RESEARCH CENTER



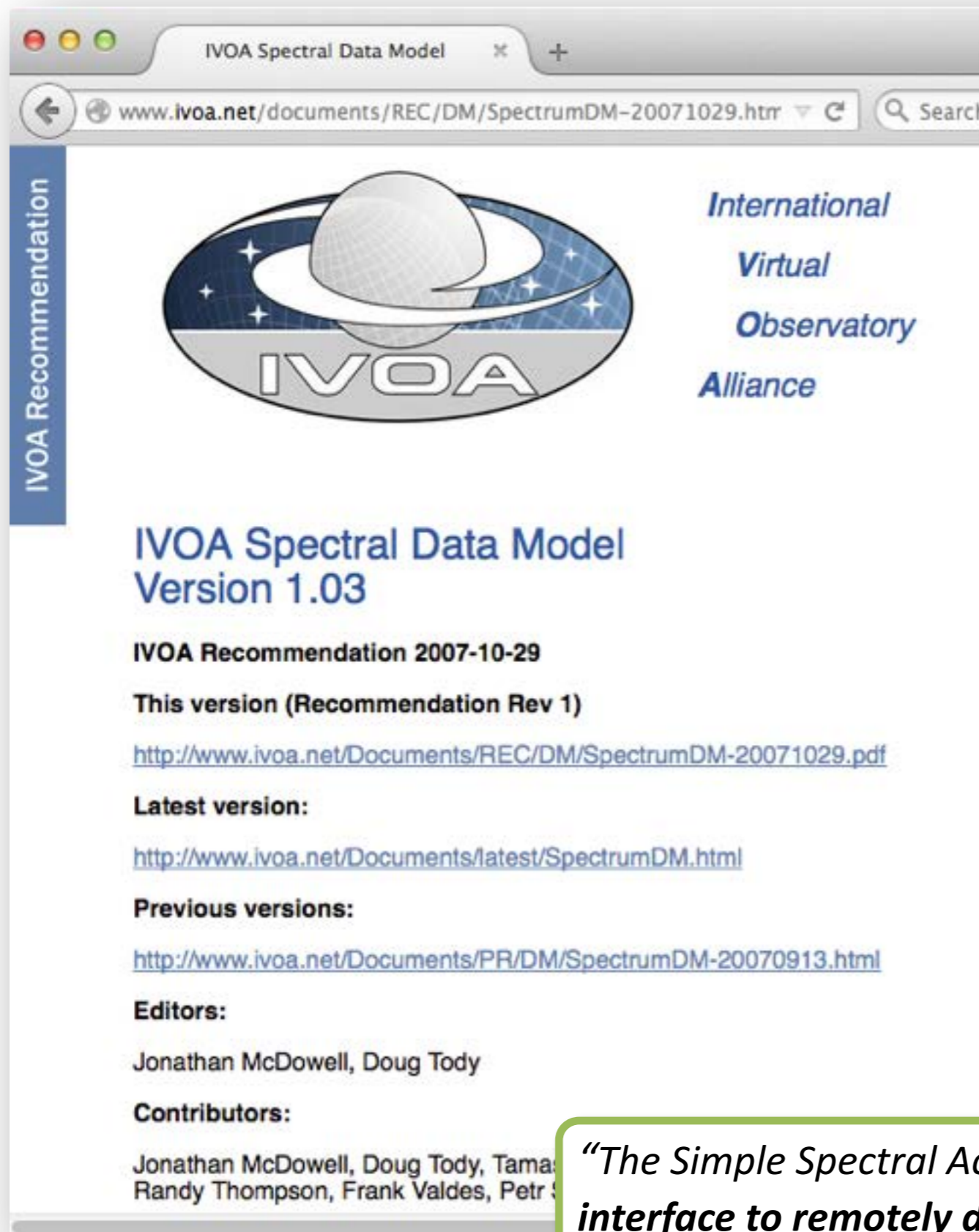
- Expanding to metallic systems
- Initial focus on phase equilibria data and thermochemical property data .



# Vision of Materials Data Curation System



# A Successful Example



IVOA Recommendation

**International  
Virtual  
Observatory  
Alliance**

**IVOA Spectral Data Model  
Version 1.03**

**IVOA Recommendation 2007-10-29**

**This version (Recommendation Rev 1)**

<http://www.ivoa.net/Documents/REC/DM/SpectrumDM-20071029.pdf>

**Latest version:**

<http://www.ivoa.net/Documents/latest/SpectrumDM.html>

**Previous versions:**

<http://www.ivoa.net/Documents/PR/DM/SpectrumDM-20070913.html>

**Editors:**

Jonathan McDowell, Doug Tody

**Contributors:**

Jonathan McDowell, Doug Tody, Tamas  
Randy Thompson, Frank Valdes, Petr

*“We present a **data model describing the structure of spectrophotometric datasets** with spectral and temporal coordinates and **associated metadata**. This data model may be used to represent spectra, time series data, segments of SED (Spectral Energy Distributions) and other spectral or temporal associations.”*

*“In this document we present a proposed abstraction for spectral data and serializations in VOTABLE, FITS, and XML, for use as a **standard method of spectral data interchange**.”*

*“This is a Proposed Recommendation, developed with the intention to **support the Simple Spectral Access Protocol**.”*

*“The Simple Spectral Access (SSA) Protocol (SSAP) defines a uniform **interface to remotely discover and access one-dimensional spectra**.”<sup>34</sup>*

# Astronomy vs. Material Measurement

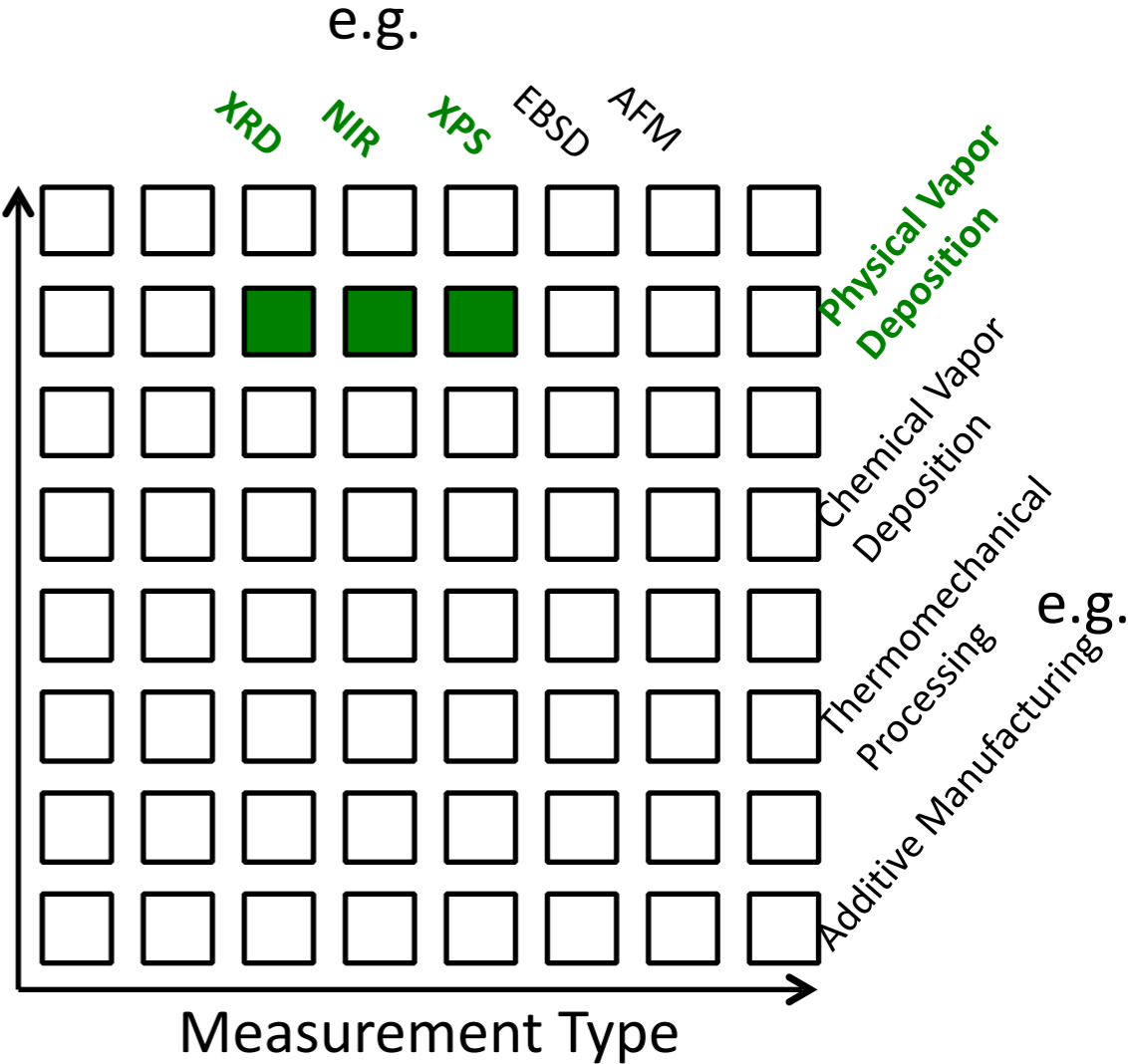
Position in the sky



Measurement Type

Singular data models

Material synthesis and processing history



Modular data models

# Materials Data Curation System

Welcome, cecamp. Thanks for logging in.

Logout | My Profile | Help

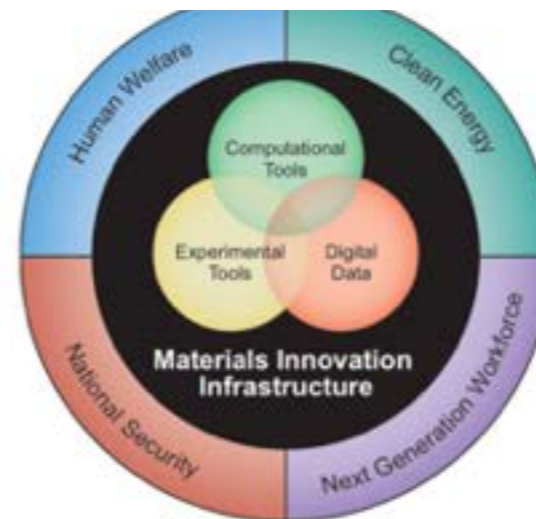
Home Data Curation Data Exploration Composer

## Materials Data Curator

This system allows for the curation of Material Data in a repository using predefined templates and a prototype ontology.

This is being developed at the National Institute of Standards and Technology and is made available to solicit comments from the Material Science community. Please do not enter any proprietary data into this system.

<https://github.com/usnistgov/MDCS>



## Available Options

[All Options »](#)



### [Curate your Materials Data](#)

Click here to select a form template and then fill out the corresponding form.



### [Explore the repository](#)

Click here to search for Material Data in the repository using flexible queries.



### [Compose a template](#)

Click here to compose your own template.

## Most Recent Templates

[Browse All »](#)

Diffusion-demo | demo.diffusion.xsd

- Written in python Django
- Backed by MongoDB
- SPARQL Query interface
- XML-based Schema
- Ability to store templates
- Schema management tools
- REST API interface
- Features in progress: Schema composer and Links to Dspace repository

# Data Curation: Tracer Diffusivity Test Schema

## Material Genome Initiative

XML Form Editor

Home Register Experiment Data Exploration

Enter Data View XML

### Data Entry

In this step, you have to fill in the form. During the process, you can view the XML. Once you have fill every field, you can view the XML.

Navigation: [Home] [Back] [1] [2] [3] [4] [5] [6] [7] [Next] [End]

**Experiment**

- ExperimentType
  - Choose:
- Id:
- Citation
  - Choose:
  - Citation:
  - Doi:

Navigation: [Home] [Back] [1] [2] [3] [4] [5] [6] [7] [Next] [End]

© 2012 NIST - MGI - XML Form Editor 0.3f | Privacy Policy | Terms

~data

Select Template

Enter Data

View Data

Select Element

1	H
2	Li
3	Na
4	K
5	Rb
6	Cs
7	Fr

Lanthanides

Actinides

Electron Configuration

Navigation: [Home] [Back] [1] [2] [3] [4] [5] [6] [7] [Next] [End]

© 2012 NIST - MGI

Materials Data Curation System

Welcome, cecamp. Thanks for logging in.

Logout My Profile Help

Home Data Curation Data Exploration

Select Template Enter Data View Data

### Data Curation

- Select Template
- Enter Data
- View Data

### XML View

This is a preview of the curated data in XML format. As part of this demo, please save your data to the repository so that we can better evaluate our system. You can also click the download button for a copy on your local machine.

Download XML Save to repository

```
<experiment>
  <specieGroup>
    <symbolOrNumber>
      <wyckoffSequence>
        <sequence>
          <Composition>
            <quantityUnit>mass fraction
            <constituents>
              <element>Al
              <quantity>1
              <purity>0.9999
              <error>
            <materialForm>
              <polycrystalline>
                <averageGrainSize>10
                <length>millimeter
            <diffusingSpecies>
              <element>Cu

```

Save to Repository

Save As:

Save Cancel

# Development of Composer

Materials Data Curation System  
Welcome, cecamp. Thanks for logging in.

Logout | My Profile | Help

Home | Data Curation | Data Exploration | **Composer**

Start Template | **Compose Template**

### Composer

1 Start Template

2 Compose Template

### Legend

name

Type

(minOccurs, maxOccurs)

## Compose Template

Please click on an element of the tree to start composing the template. A menu will appear and you will be able to interact with that element.

Download | Save as Template | Save as Type

```
xsd:schema
├── xsd:element : root Root (1,1)
│   └── xsd:complexType Root
│       └── xsd:sequence
```

*User develops needed template from smaller schema*

## Material

- Metals
  - Al Alloys
  - Fe Alloys
  - Ni Alloys
  - ....
- Polymers

## Processing

- Heat Treatment
- Casting
- Thermomechanical processing
- Sintering
- Rapid Cooling
- ...

## Sample Preparation

- Polishing
- Etching
- Electro-polishing
- .....

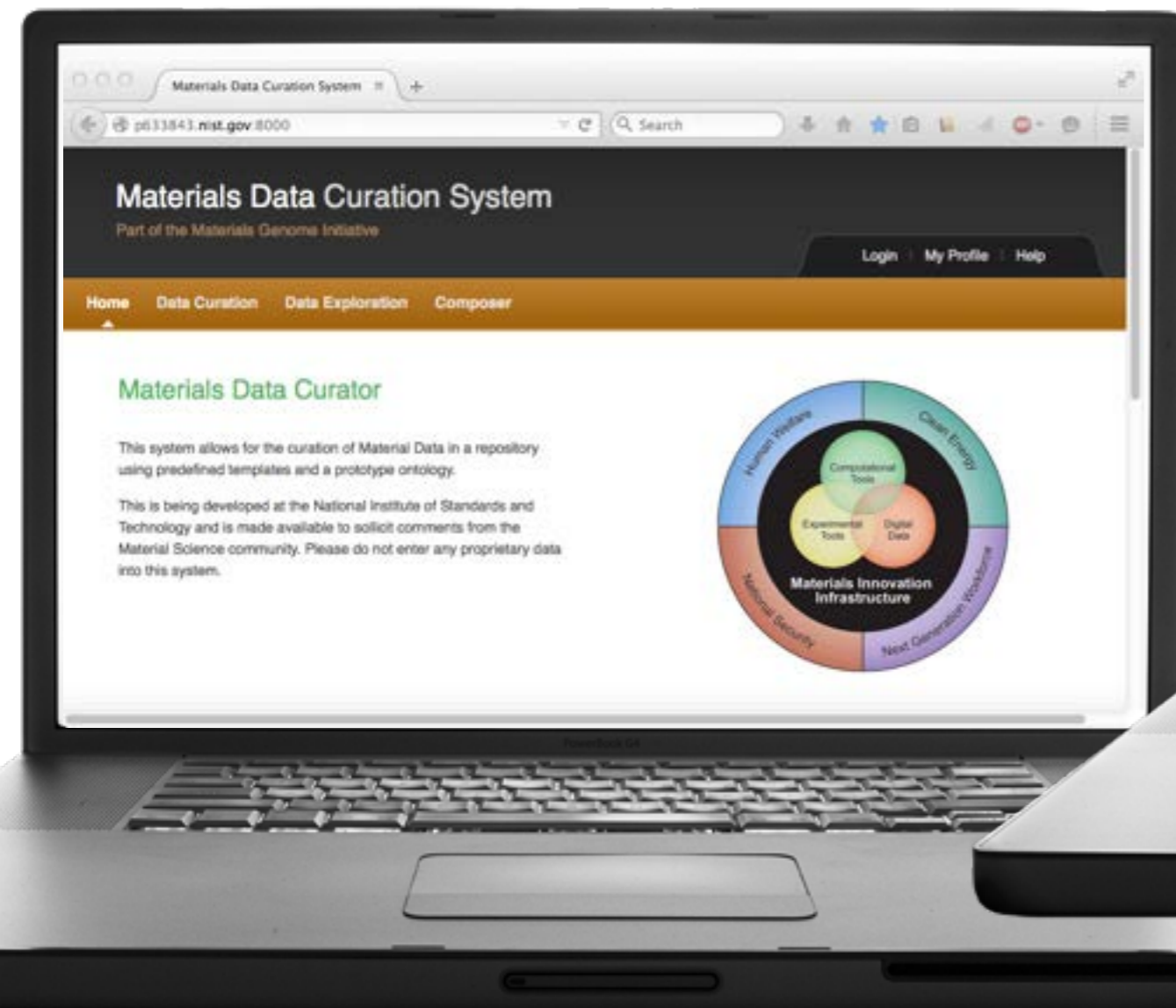
## Measurement Method

- EPMA/EDS
- EMSD
- DSC/DTA
- Dilatometer

# Application Programming Interface *in a nutshell*

**WWW** interface to the Curator:

**API** interface to the Curator:



**Hand data entry**

**Automated data entry**





# Registries

# The case for a Materials Resource Registry

Register. Discover. Access.  
Research Data Alliance

National Institute of Standards and Technology (NIST)

National Data Service (NDS)

Center for Hierarchical Materials Design (CHiMaD)

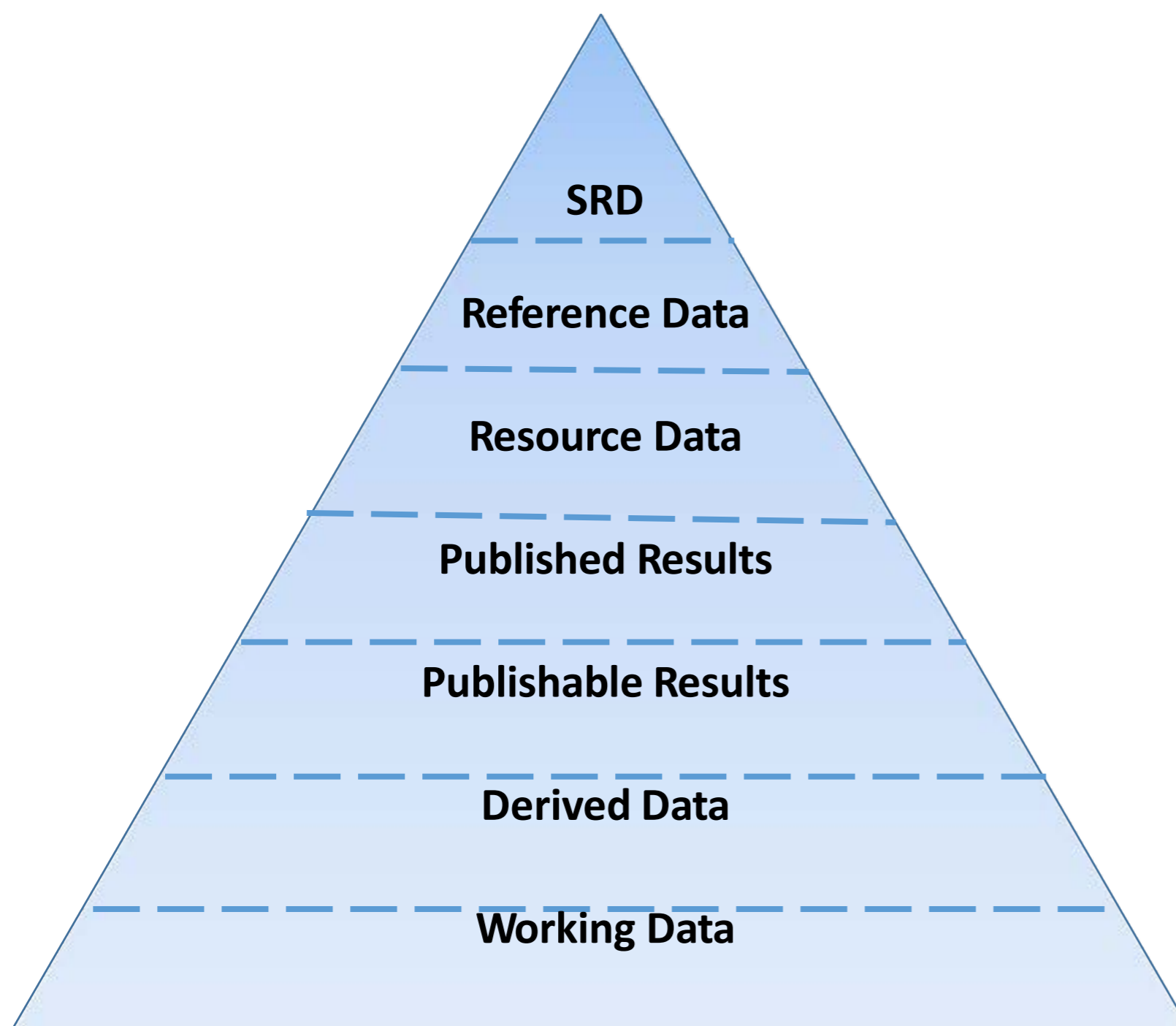
Research Data Alliance (RDA)

**Collaborators**

# Data Policy

A NIST Approach (in progress)

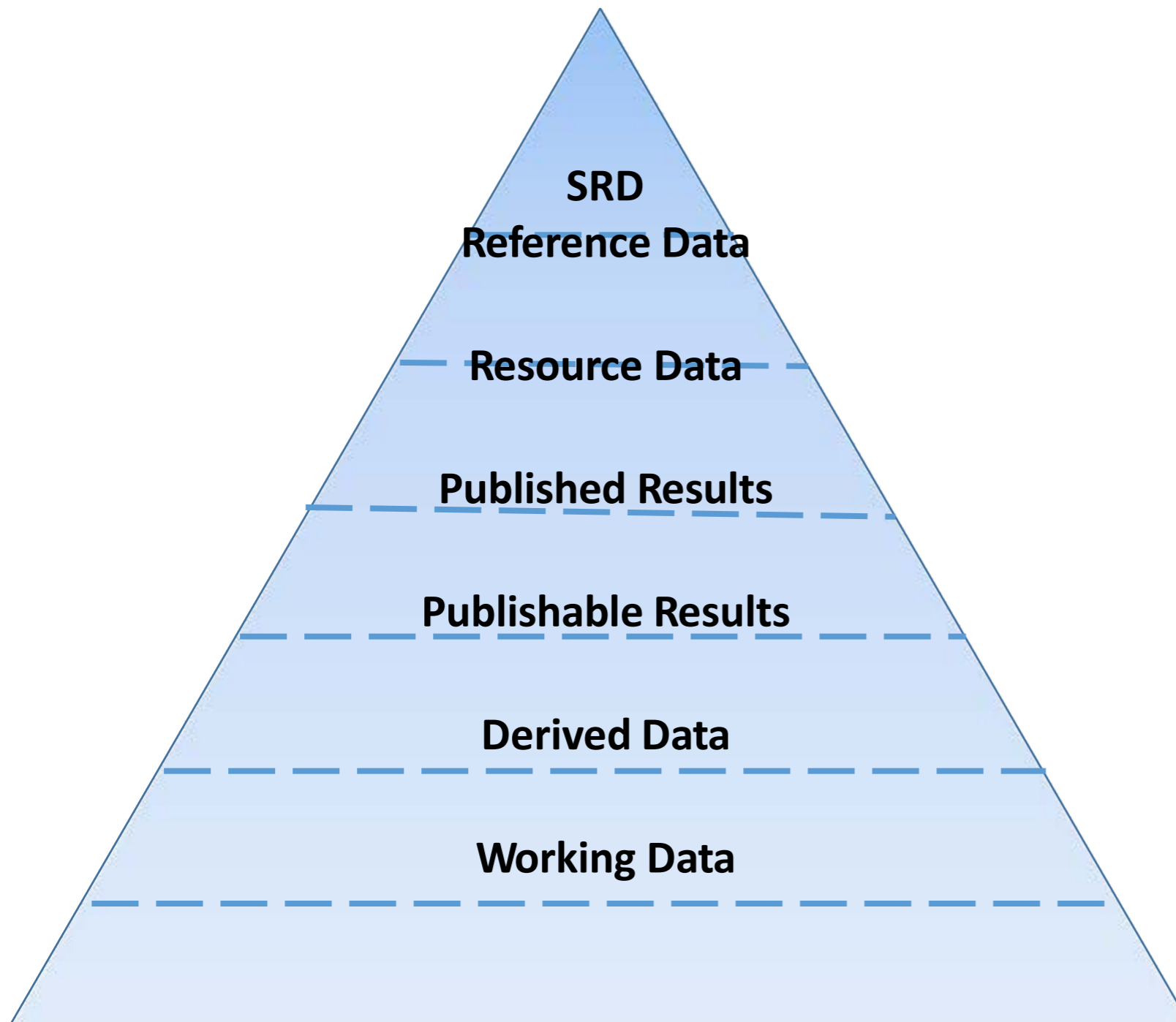
# Data Taxonomy: STRAWMAN-DRAFT



## Working Data

The digital equivalent of entering data in a laboratory notebook. Working data may be raw observational data that is acquired directly from an instrument or a measurement system, or digital values acquired or generated during experiments or simulations. In some cases the researcher responsible for generating the working data may determine that this data has immediate value and is worth preserving, or the researcher expects the data will have value after it has been manipulated or further evaluated, and the data has the potential to develop into a publication or will be used to draw conclusions. In other cases working data may be recognized as not appropriate for broader use in its present form. It may have value to the data producers and their collaborators, but it should be recognized that the data could be easily misinterpreted by people not closely involved in its production because some metadata and important facts about its status or acquisition are not readily available beyond the immediate research team (i.e. adequate metadata for re-purposing is not attached to the data itself, expending resources to codify needed metadata is not justified, etc

# Data Taxonomy



## Derived Data

Derived data are those that underpin the conclusions provided in a publication or report. Derived data come from working data that have been manipulated, analyzed, processed, or evaluated in some way. The data must have passed some minimal (perhaps ad hoc) evaluation and be considered by the responsible researcher (typically the data producer) to be ready for the next steps in the workflow or project/product development effort.

## **Publishable Results**

All final or summary results that comply with relevant NIST policies (e.g., SI units, uncertainty statements), that have been reviewed internally and approved by an appropriate NIST authority, and that could be published either in a scientific publication or as a standalone data product.

## **Published Results**

Results that are publishable and that are contained in a document that has been reviewed and approved for publication by the necessary NIST organizational authorities, submitted to its intended publisher, and made public. We see Reports of Analysis as fitting in this category but no longer call them out because they seem unique to MML.

- Preservation Levels
  1. No additional requirements
  2. Individual user responsibility
  3. Data backed up using a tested/automated process
- Discoverability Levels
  1. No additional requirements
  2. Persistent Identifier (PID) assigned
  3. Entered in NIST Enterprise Data Inventory<sup>1</sup>
  4. Inventory record flagged for public access

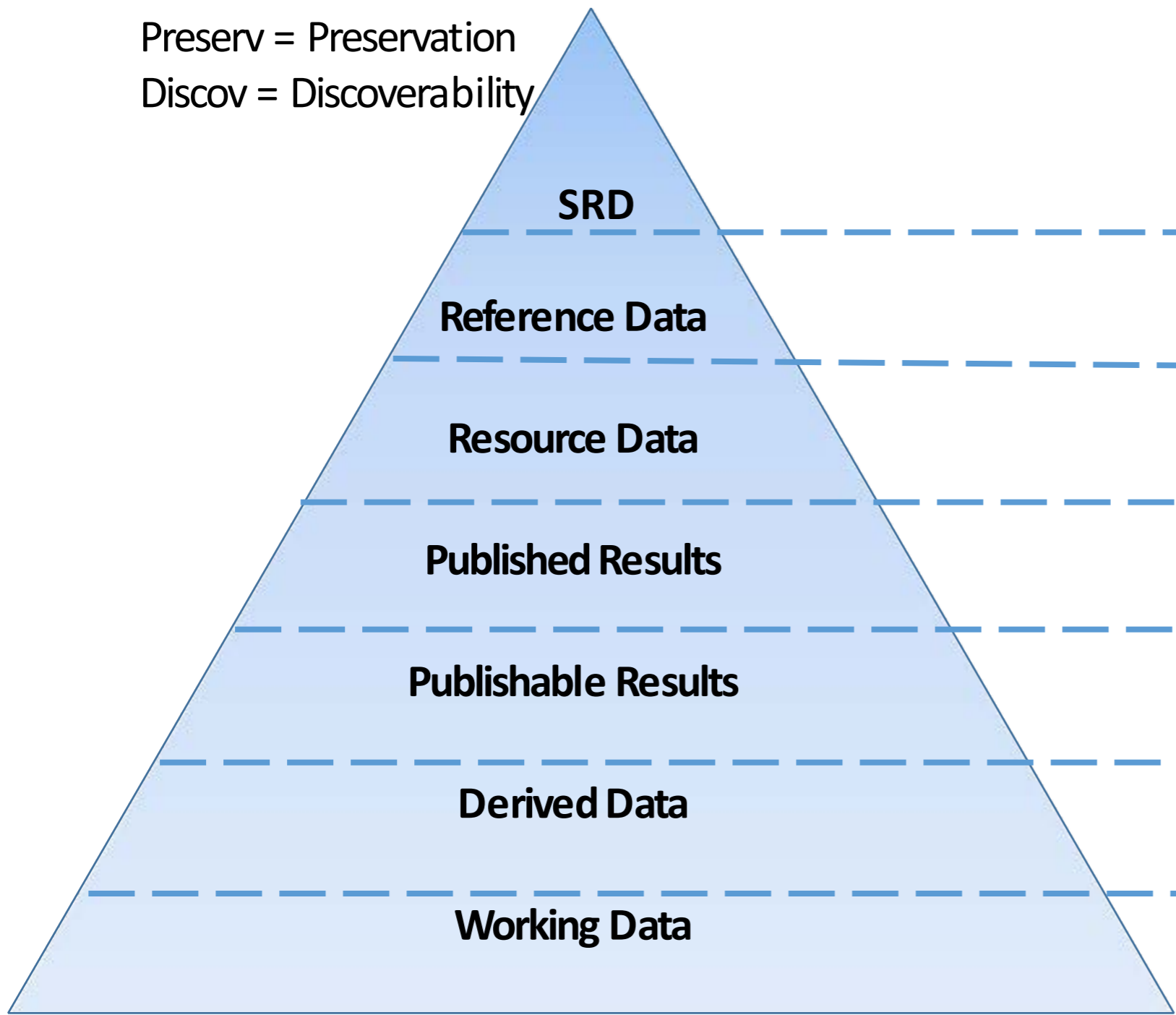
<sup>1</sup>PID assigned + NIST minimum metadata present

# Data policy under discussion

*mapping levels onto categories*

# STRAW MAN-DRAFT

Preserv = Preservation  
Discov = Discoverability



Preserv	Discov
3	4
3	4
3	4
3	4
2	3
1	1
1	1



Collaborations  
Discussions  
Convening



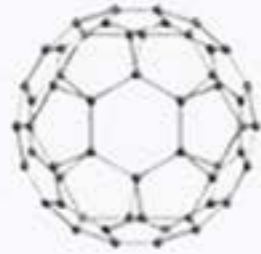
Research Data Sharing  
without barriers

# Research Data Alliance Interest Group

RDA/CODATA Materials Data, Infrastructure &  
Interoperability IG

*James A Warren and Laura Bartolo, Co-Chairs*

Materials Resource Registry, Working Group, *Robert Hanisch, Chair*



World  
Materials Research Institutes  
Forum

世界材料研究所フォーラム

[WMRIF Home](#)

[About the Forum](#)

[Member institutes](#)

[Working groups](#)

[News & events](#)

[Contact](#)

## WMRIF: GETTING TO KNOW US

- [What we want](#)
- [Taking a look back](#)
- [Want to join?](#)

## RECENT POSTS

- [Internal documents](#)
- [6th WMRIF General Assembly – Up-date 3](#)
- [9th International](#)

## WMRIF Home



Exploration of Opportunities



# EMMC



- The aim of the council (for the moment) is to do preparatory networking to establish activities necessary in the materials modelling field.
- The Council could build on existing activities happening in Europe and make use of the network to complement these activities.

# National Data Service

HOME

ABOUT

EVENTS

WIKI

GET INVOLVED



## 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.

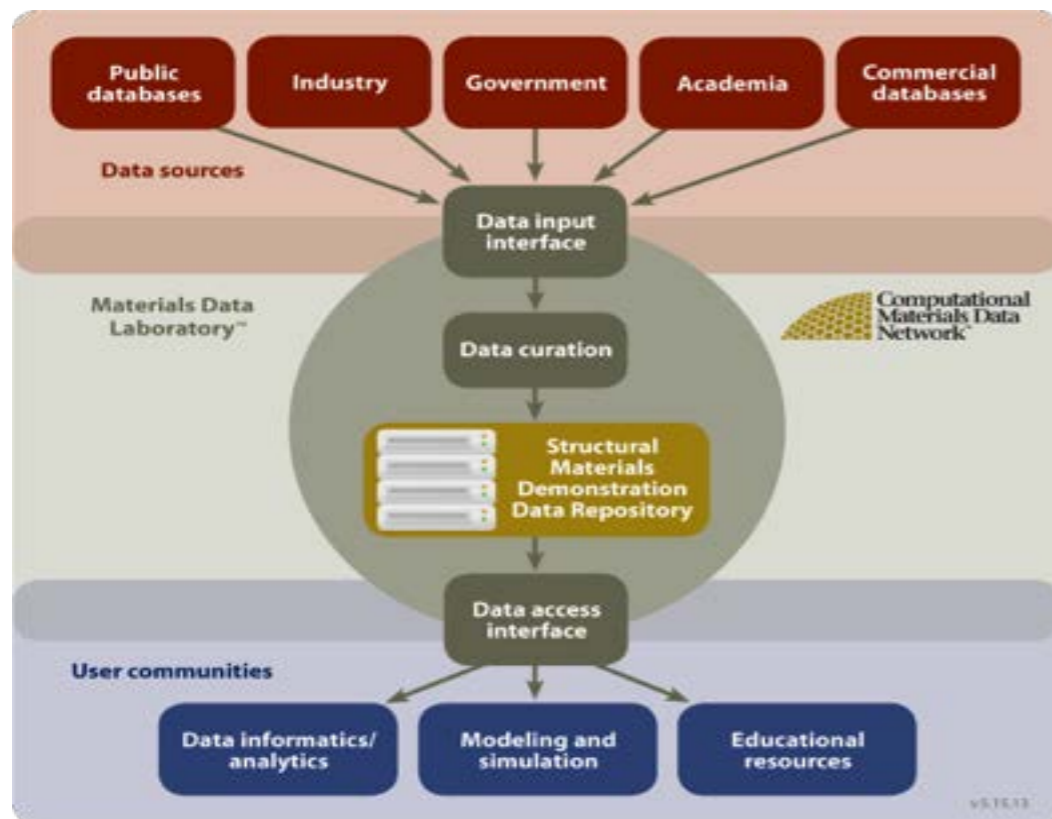
However, while some communities are making progress in developing discipline-specific data services, the U.S. and international scientific communities lack a unified framework and supporting services for storing, sharing, and publishing data; for locating data; or for verifying data. More specifically,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

### NEWS

#### The launch of Materials Data Facility in support of Materials Genome Initiative

06.19.14

The National Data Service Consortium to launch a materials data facility for the advancement of materials science research through open data access and sharing. [Read more](#)



**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



# NIST Center of Excellence for Advanced Materials





First US-Japan  
MGI Workshop,  
June 2015





STOP

MGI.NIST.GOV