Science advances one funeral at a time -Max Planck

The Perfect is the Enemy of the Good -Voltaire

JAMES A WARREN
DIRECTOR, MATERIALS GENOME PROGRAM
MATERIAL MEASUREMENT LABORATORY
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
EXECUTIVE SECRETARY, NSTC SUBCOMMITTEE ON MGI

# THE MATERIALS GENOME INITIATIVE, DATA, OPEN SCIENCE, AND NIST

"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)

THE MATERIALS GENOME INITIATIVE: A NATIONAL PRIORITY

# WHY NIST?

# 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

TELL YOUR COLLABORATORS TO SIGN UP FOR OUR STAKEHOLDER EMAIL LIST!

https://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

SOME BACKGROUND AND FURTHER THOUGHTS ON HOW WE DETERMINED OF STRATEGY, AND EXPLAIN OUR EFFORT TO THE WORLD

# FORMULATING THE NIST ROLE IN MGI

- TAKE THE LHC AS AN EXAMPLE
- DATA PRODUCED AT 1PB/SEC !
- REDUCED DATA SAVED: ~ GB/S
- THAT'S A DARN GOOD MODEL
- OTHER END : BIOLOGY?
- MATERIALS: IN THE MIDDLE

# BETTER MODELS = LESS DATA

# What should we be doing?

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

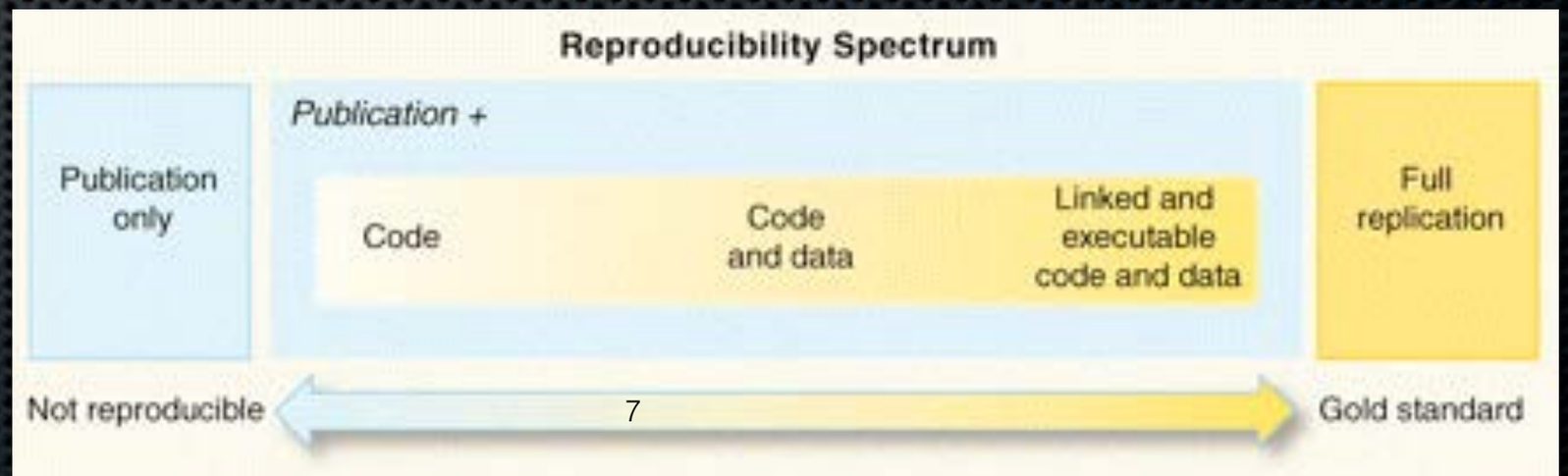
The measurement or computation of a quantity (data) is generally *meaningless* without the associated quantifying model that defines both the data and its uncertainties.

**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's Mission

To promote U.S. innovation and industrial competitiveness by advancing

measurement science, standards, and technology

in ways that enhance economic security and improve our quality of life

©Robert Rathe

# NIST Products and Services

© Robert Rathe

## Measurement Research

- ~ 2,200 publications per year

## Standard Reference Data

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

## Standard Reference Materials

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

## Calibration Tests

- ~ 18,000 tests per year

## Laboratory Accreditation

- ~ 800 accreditations of testing and calibration laboratories

# 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 t… 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

THE DIRECTOR

May 9, 2013

M-13-13

MEMORANDUM FO… ENCIES

FROM: Sylvia M…

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

S. 994

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

AT THE SECOND SESSION

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

**An Act**

To expand 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.**

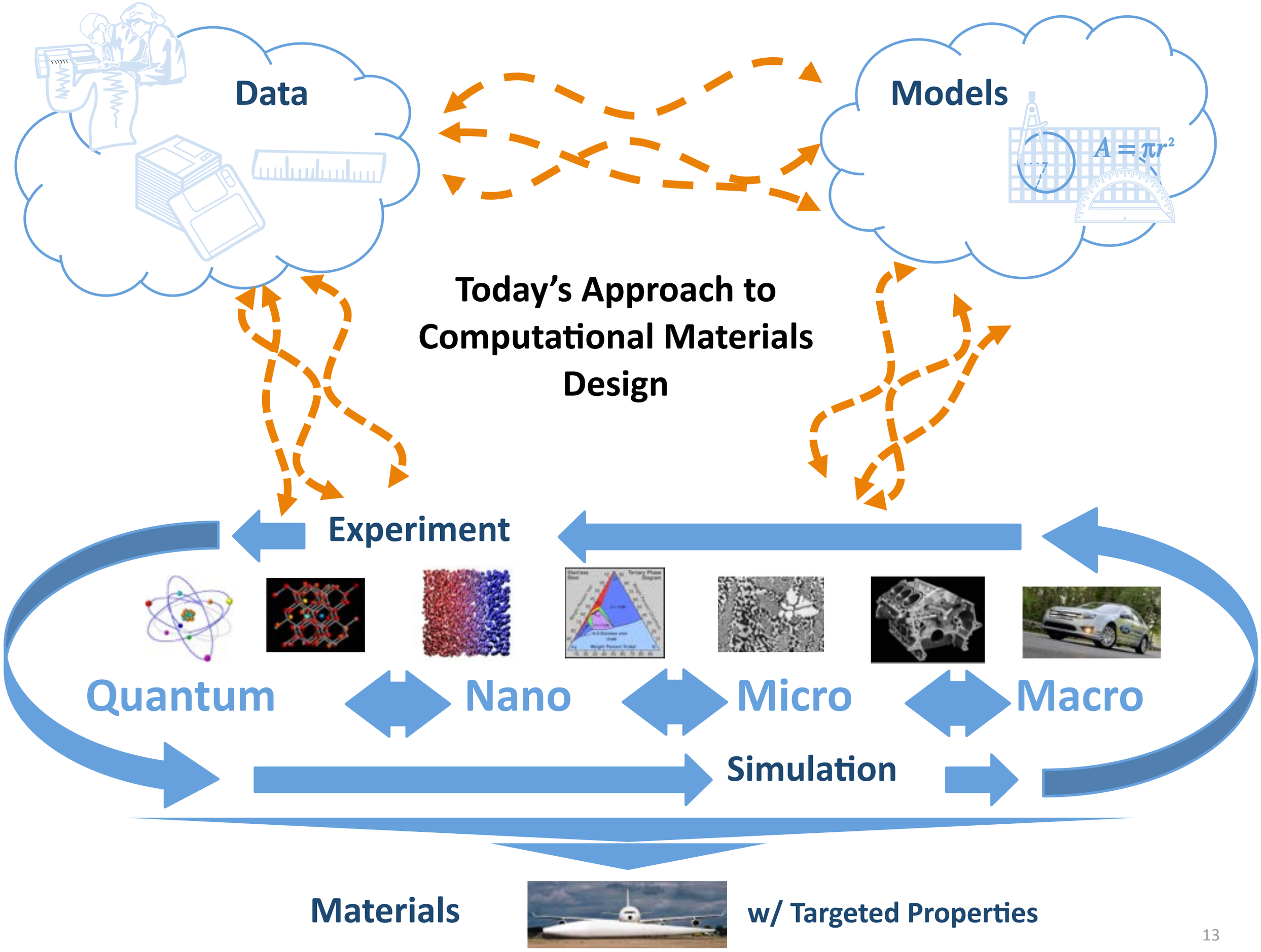
This Act may be cited as the "Digital Accountability and Transparency Act of 2014" or the "DATA Act".

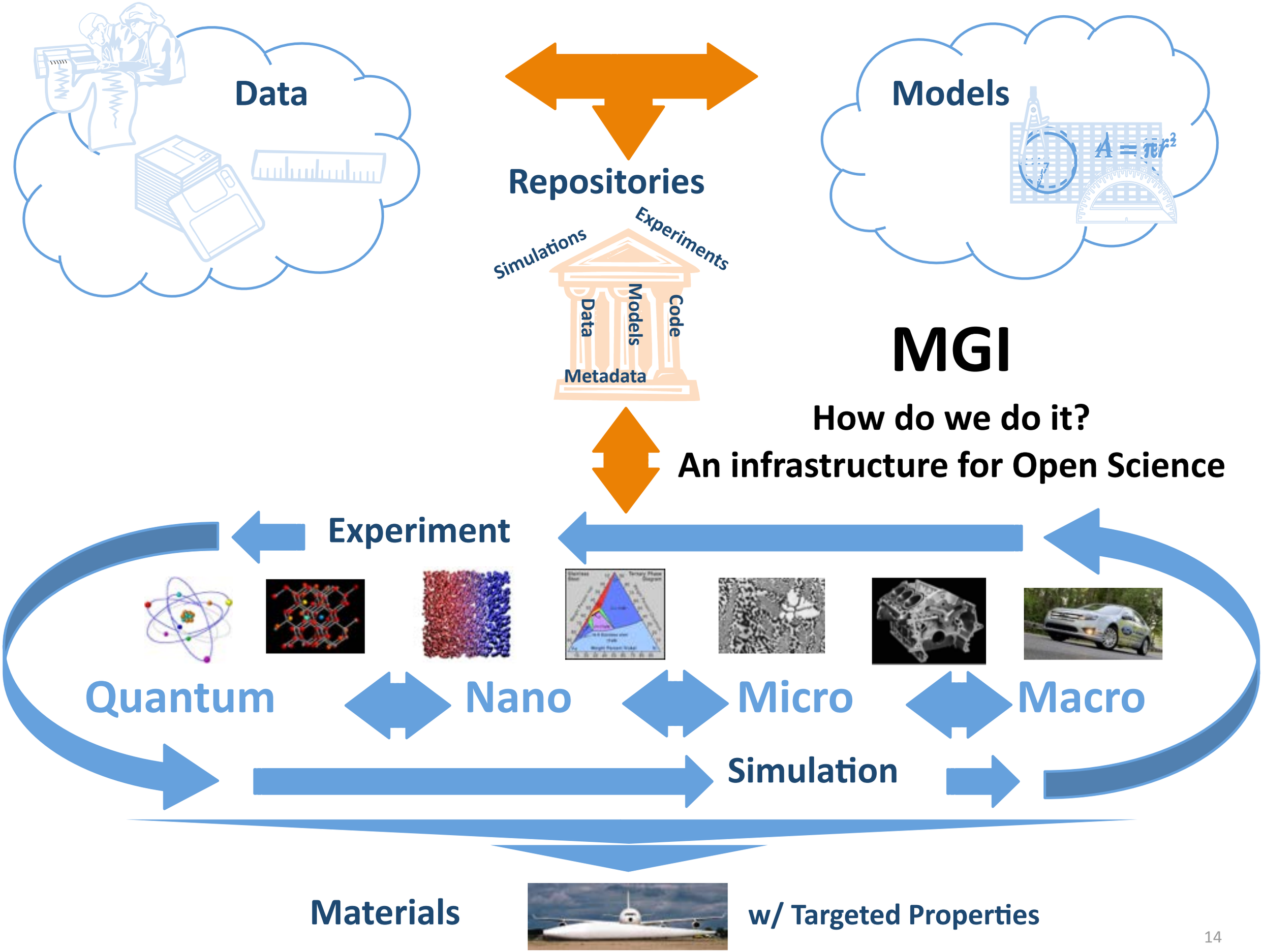
**SEC. 2. PURPOSES.**

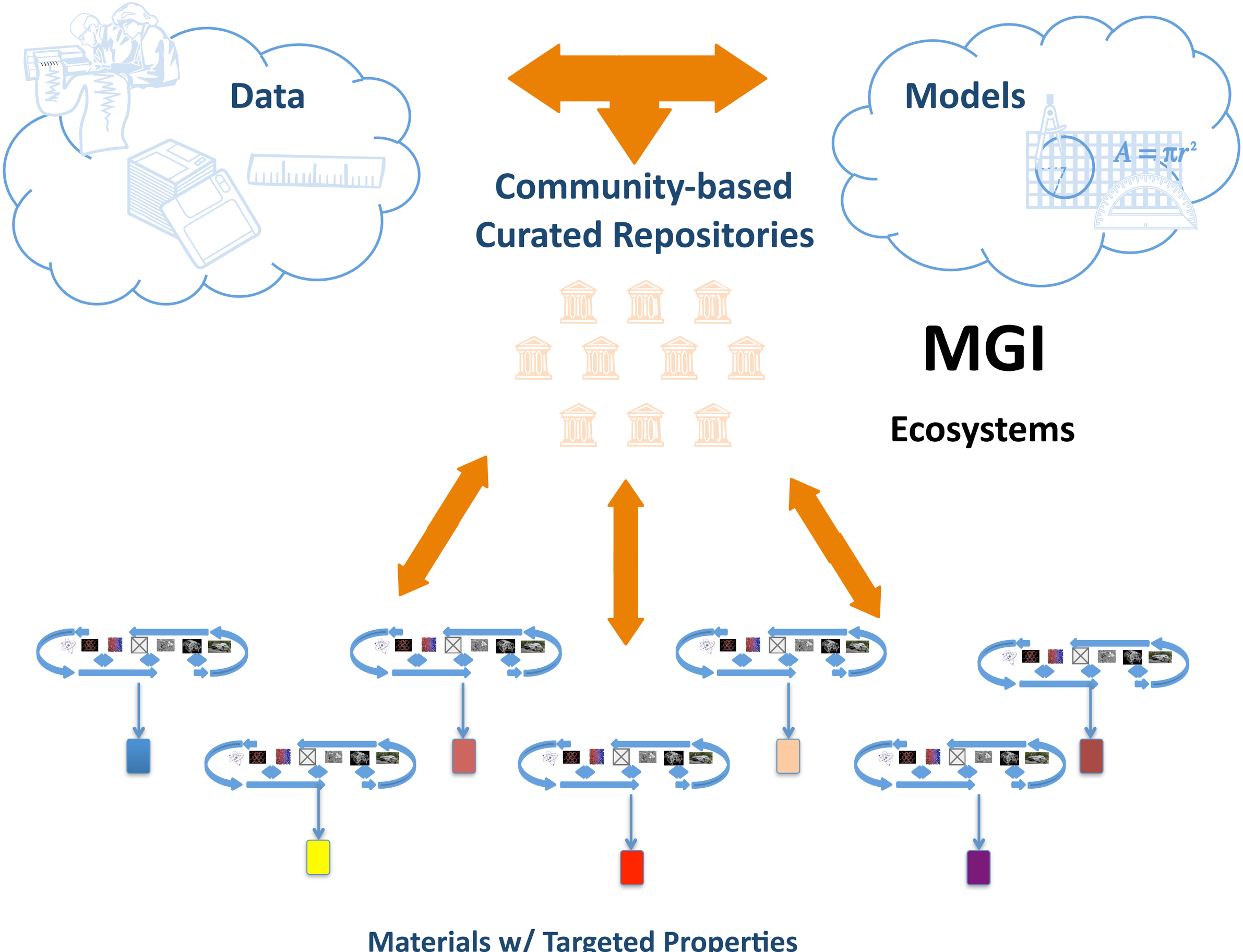
The purposes of this Act are to—

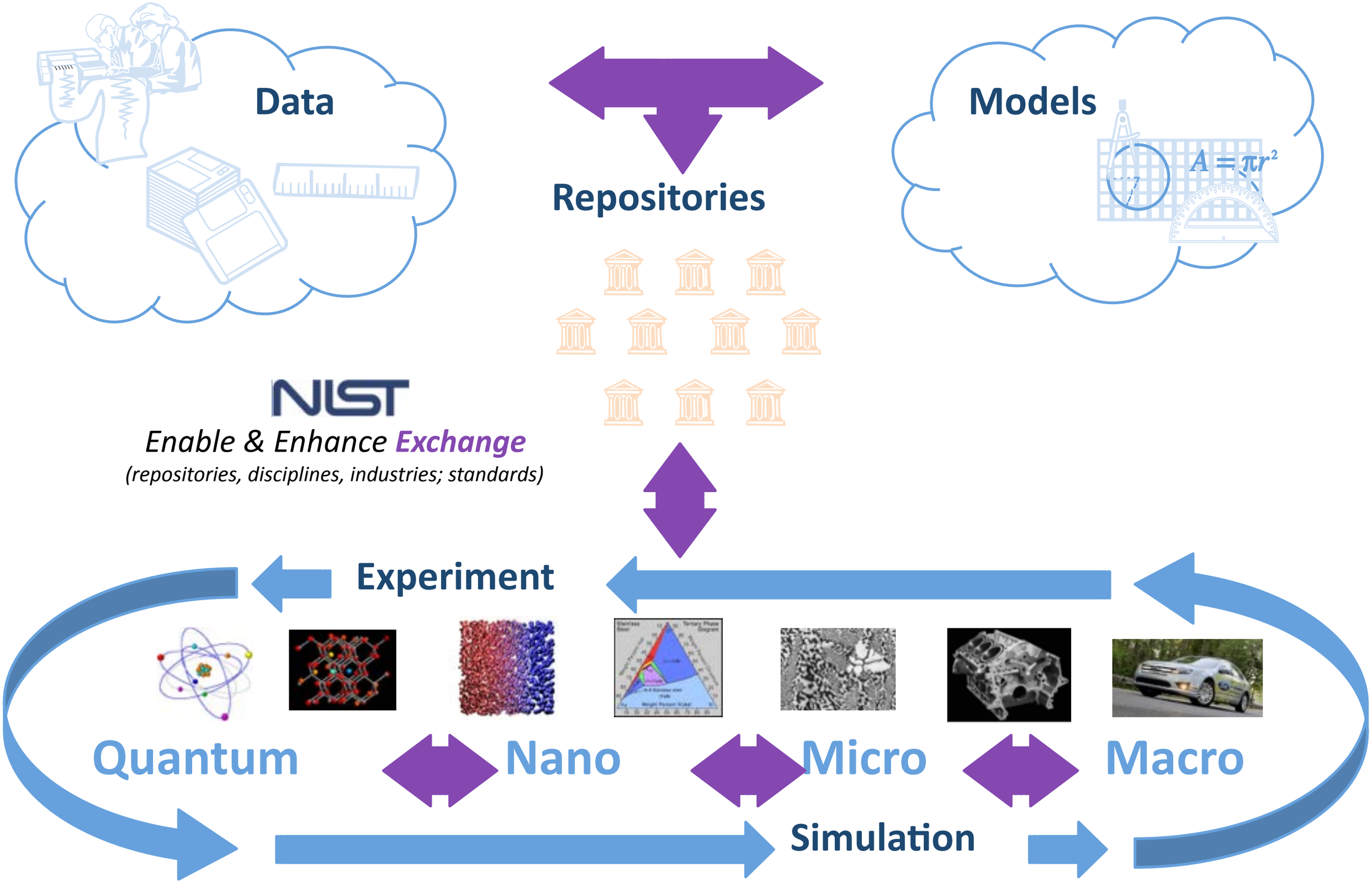
(1) expand the Federal Funding Accountability and Transparency Act of 2006 (31 U.S.C. 6101 note) by disclosing direct Federal agency expenditures and linking Federal contract, loan, and grant spending information to programs of Federal agencies to enable taxpayers and policy makers to track Federal spending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

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

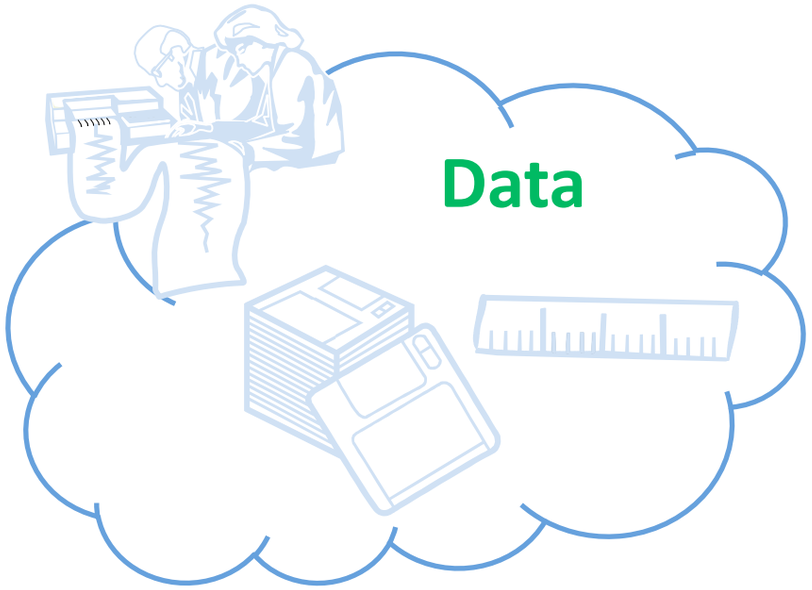

Data
Models
$A = \pi r^2$
Today's Approach to Computational Materials Design
Experiment
Quantum
Nano
Micro
Macro
Simulation
Materials
w/ Targeted Properties

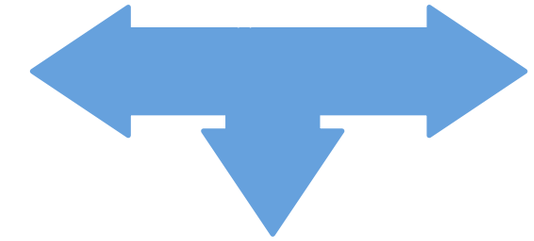

Data
Repositories
Models
$A = \pi r^2$
Simulations
Experiments
Data
Models
Code
Metadata
MGI
How do we do it?
An infrastructure for Open Science
Experiment
Quantum
Nano
Micro
Macro
Simulation
Materials
w/ Targeted Properties


Data
Models
$A = \pi r^2$
Community-based
Curated Repositories
MGI
Ecosystems
Materials w/ Targeted Properties

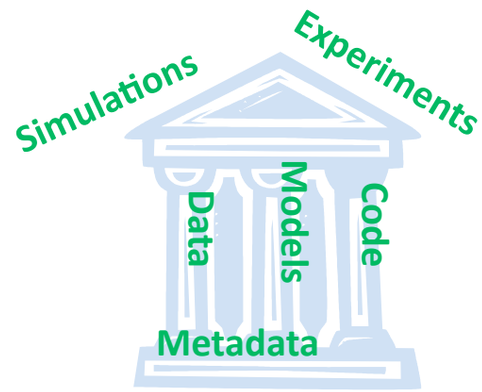


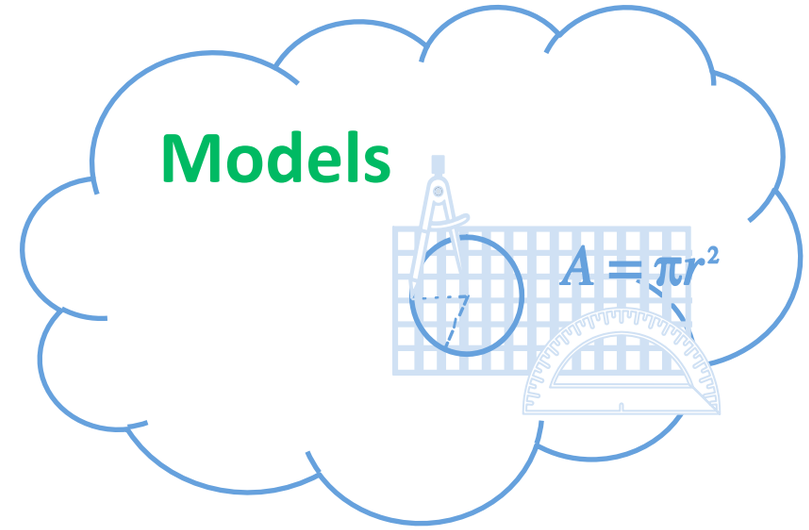

**Materials w/ Targeted Properties**



**Repositories**

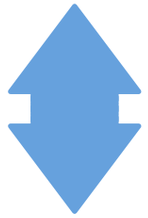


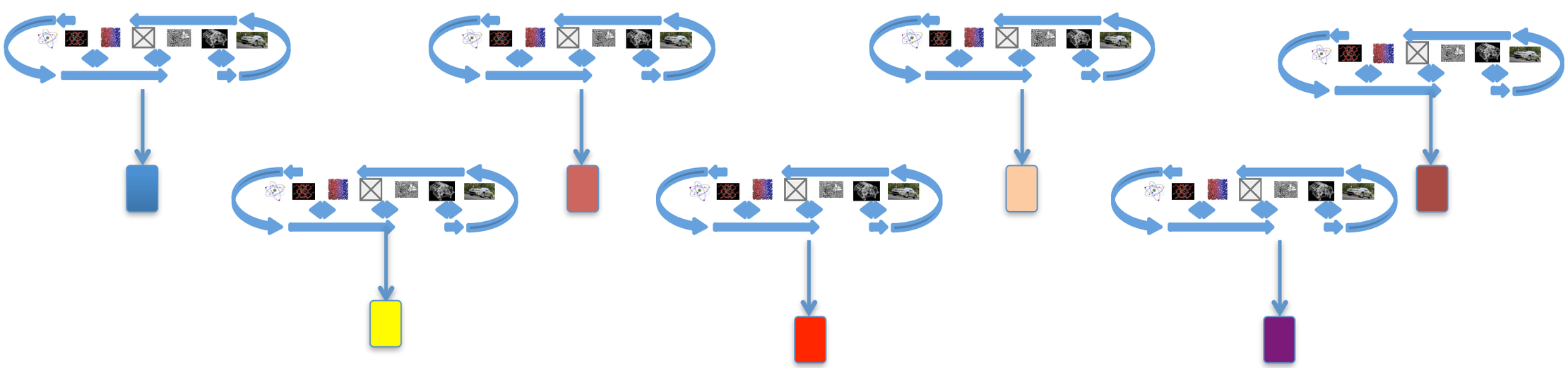




NIST

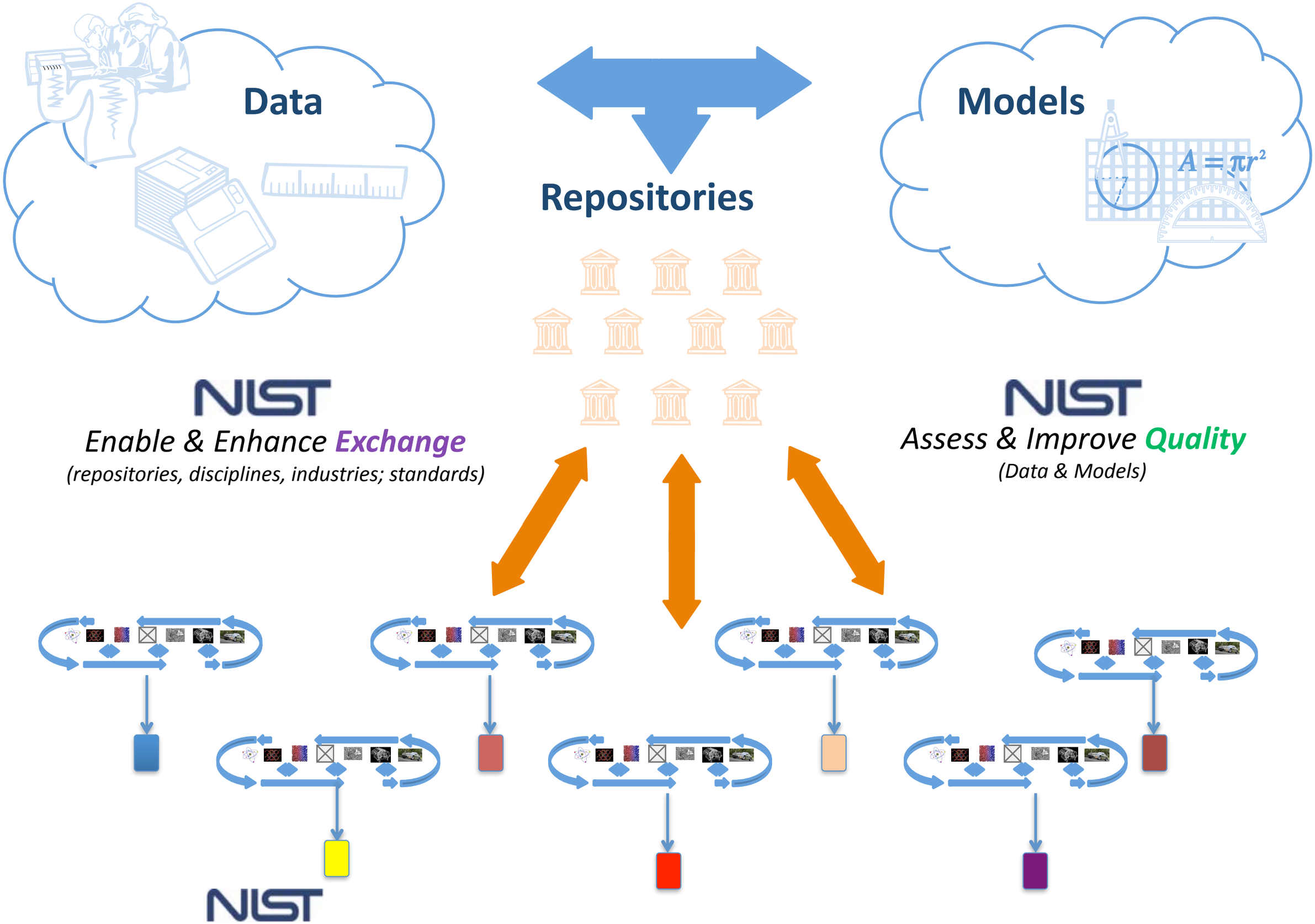
*Assess & Improve **Quality***

*(Data & Models)*

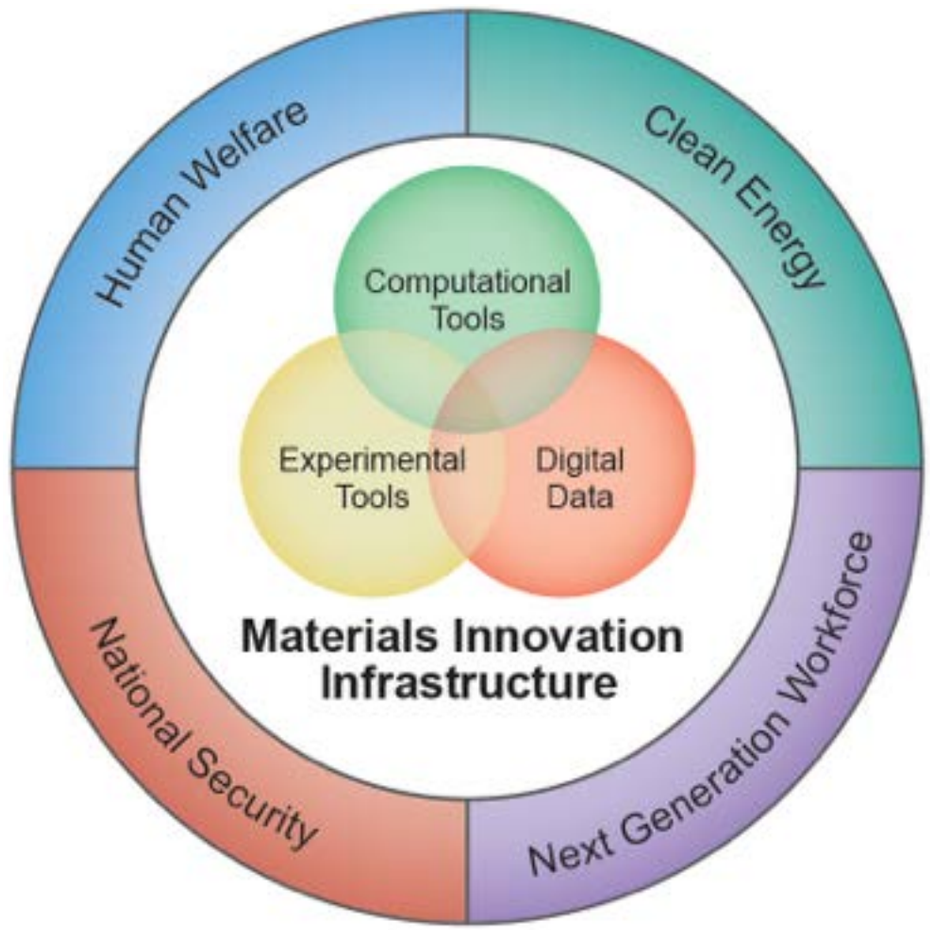




**Materials w/ Targeted Properties**

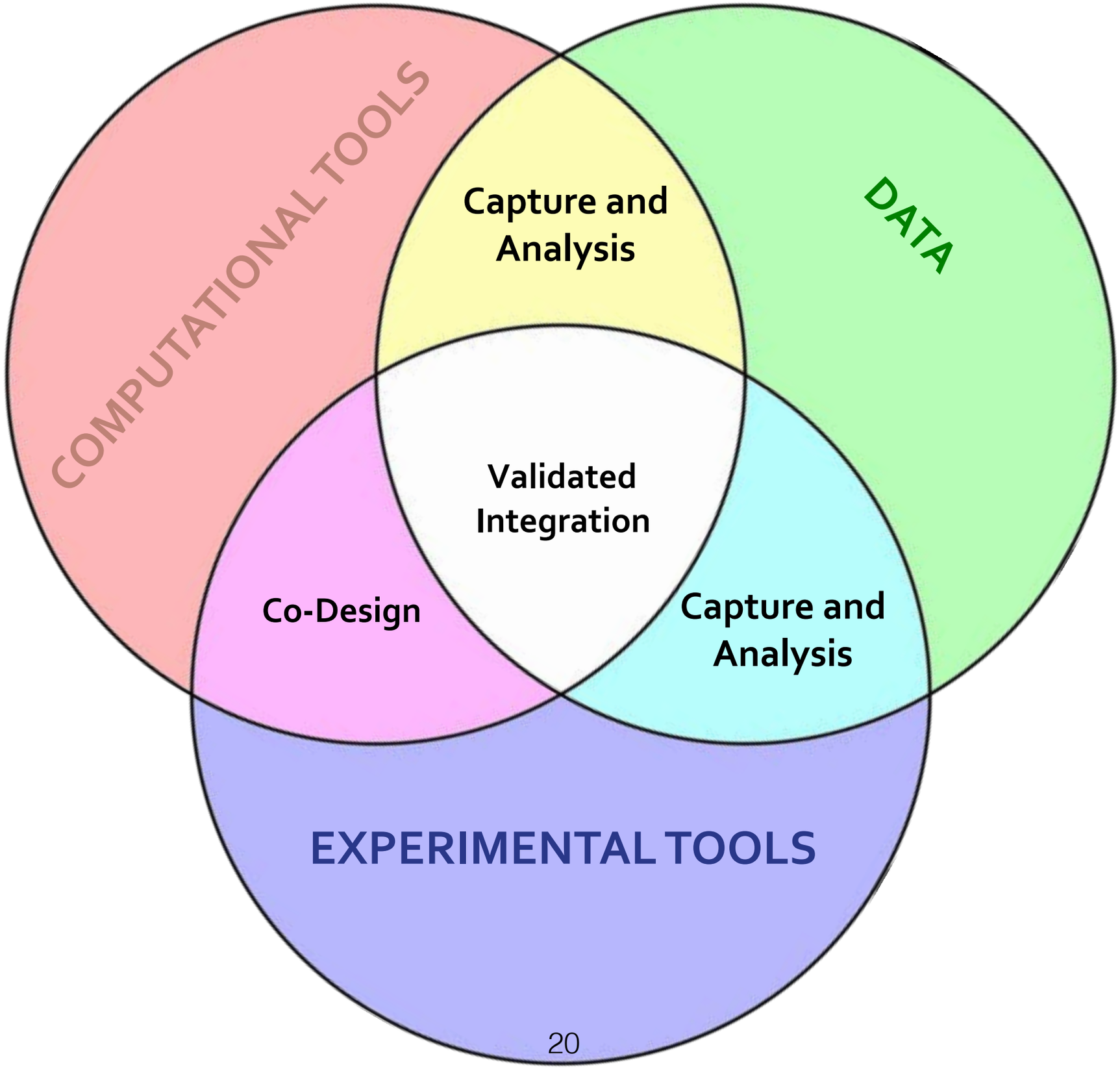

Data
Repositories
Models
$A = \pi r^2$
NIST
Enable & Enhance Exchange
(repositories, disciplines, industries; standards)
NIST
Assess & Improve Quality
(Data & Models)
NIST
New Methods and Metrologies
(data driven analysis and models)
Materials w/ Targeted Properties

# The NIST Portfolio and Wider Questions about Data

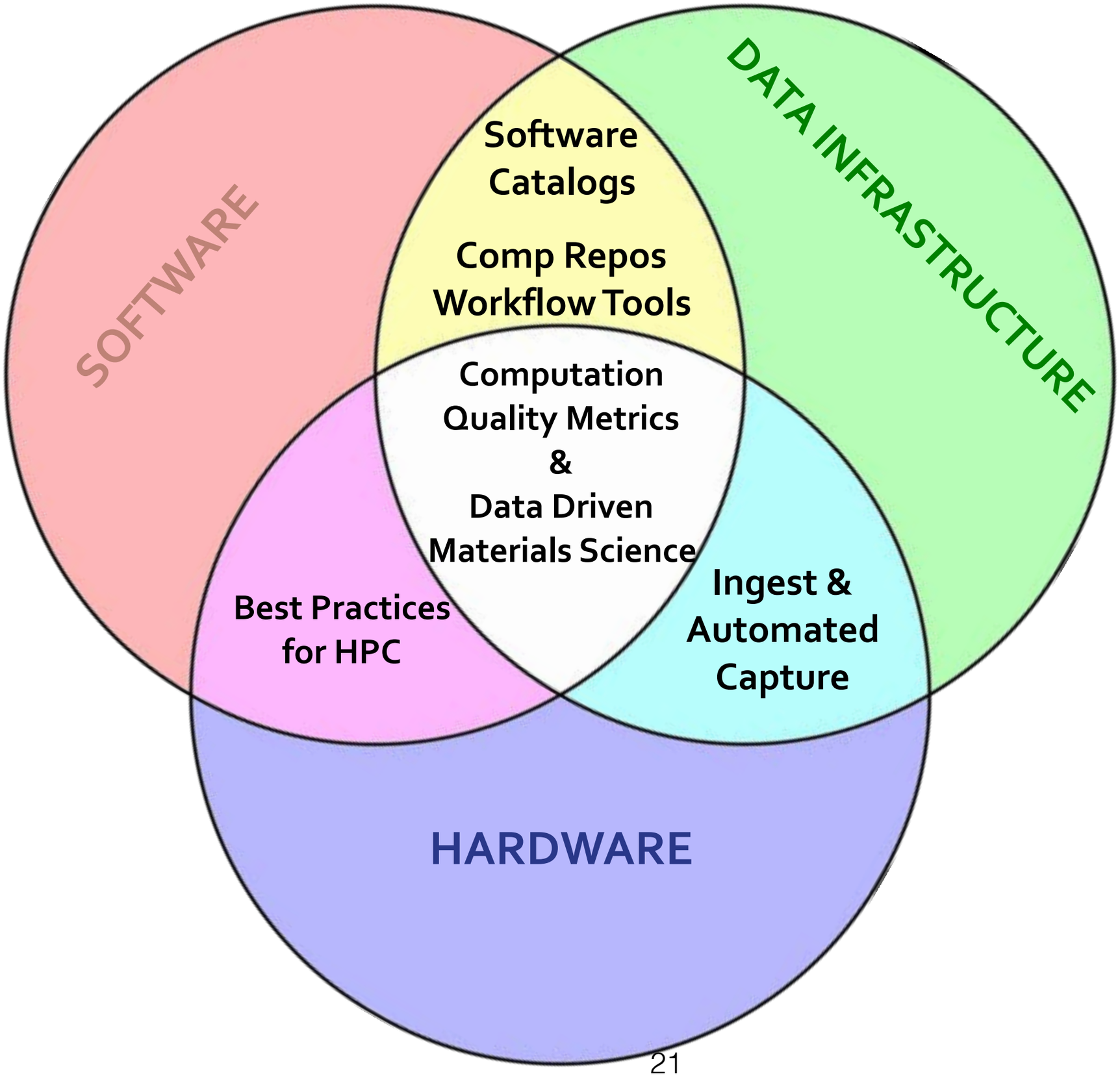



MGI.NIST.GOV

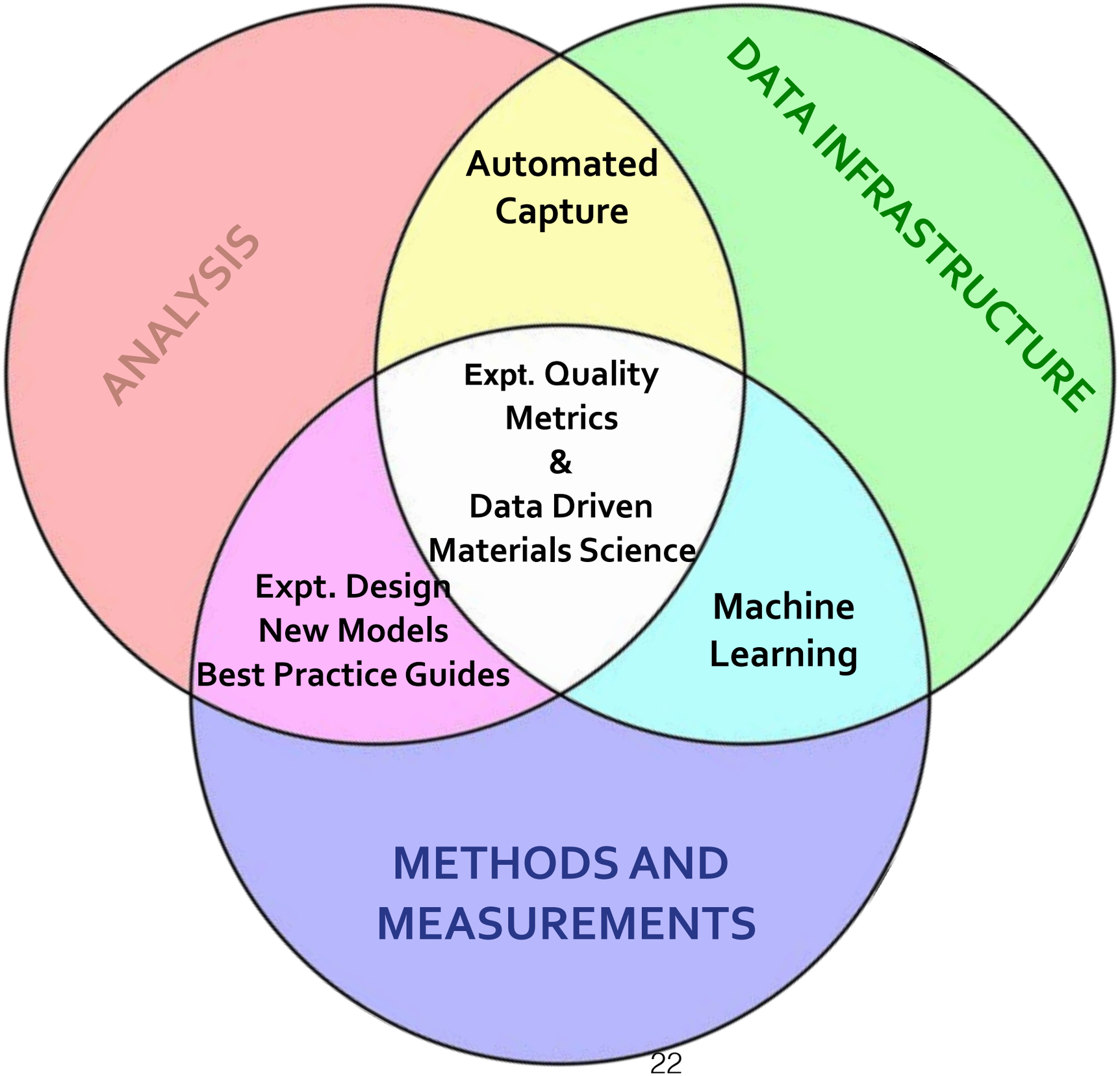
# MGI


COMPUTATIONAL TOOLS
DATA
EXPERIMENTAL TOOLS
Capture and Analysis
Validated Integration
Co-Design
Capture and Analysis

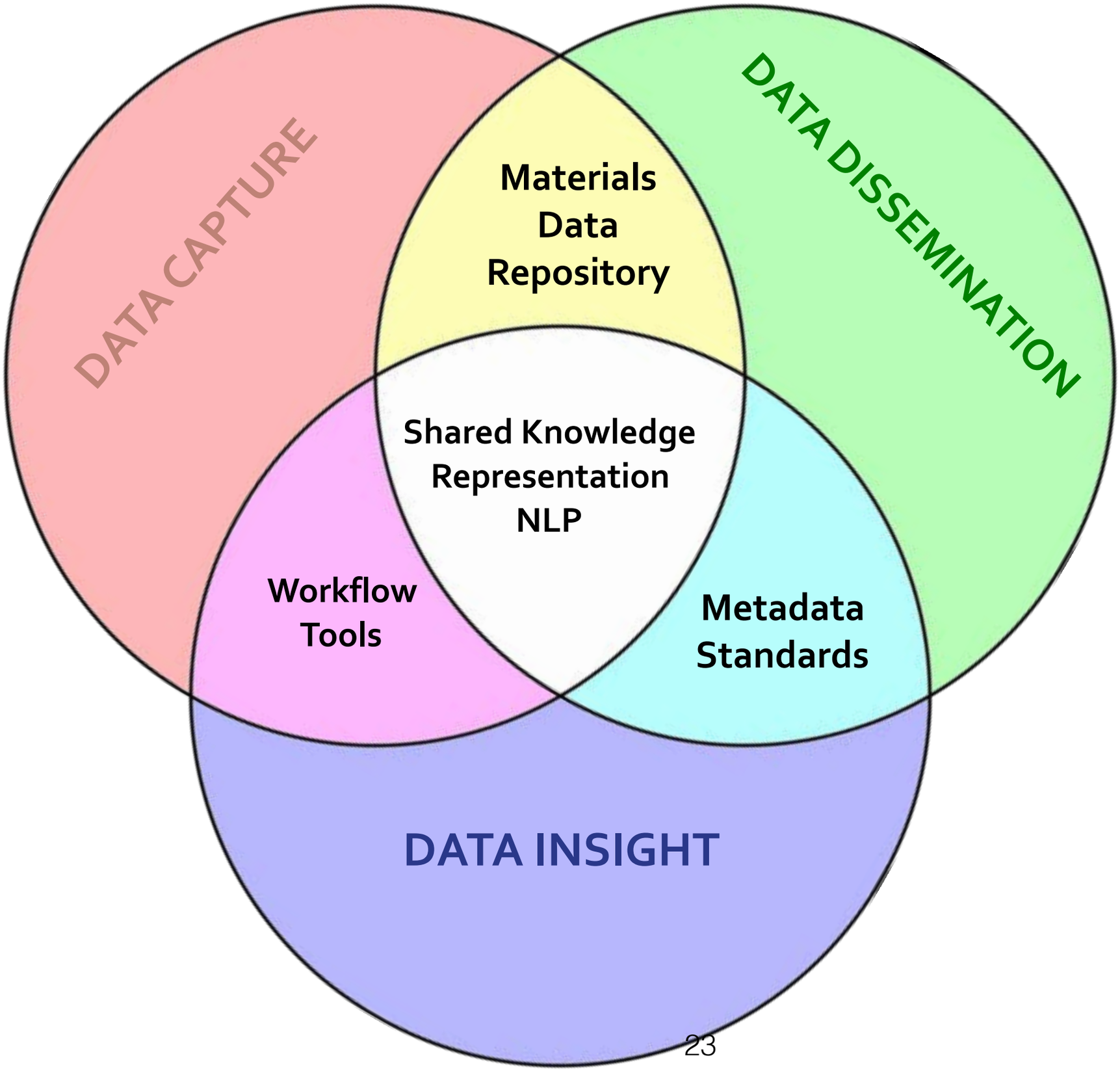
# Computational Tools



# Experimental Tools



# Data


DATA CAPTURE
DATA DISSEMINATION
Materials
Data
Repository
Shared Knowledge
Representation
NLP
Workflow
Tools
Metadata
Standards
DATA INSIGHT

# Office of Data and Informatics

## SRD

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

## Research Data

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

## Lead/Liaison

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

## Data Science

*The 4th paradigm?*

- will it stand next to
  - theoretical
  - experimental
  - computational
- Cloud
- Statistical Learning
- Big Data
- Knowledge Discovery
- very fast growing
- *many* new jobs
- new degrees/depts

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

# 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

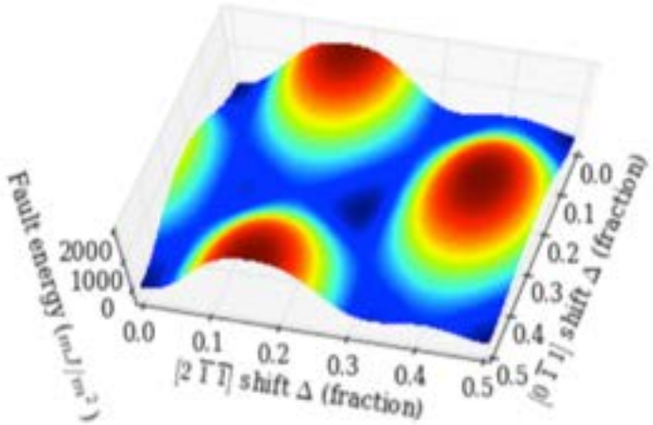
Properties | Methods | References | FAQ | Resources | People | Contact

## 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.

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

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

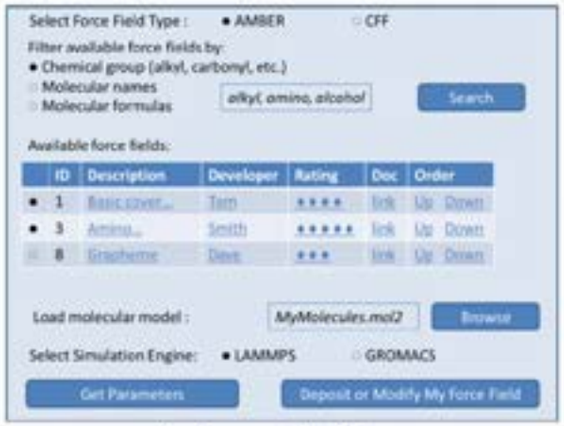
Frederick R. Phelan Jr. , *Material Science & Engineering Division, Polymers & Complex Fluids Group NIST , Gaithersburg, MD*

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

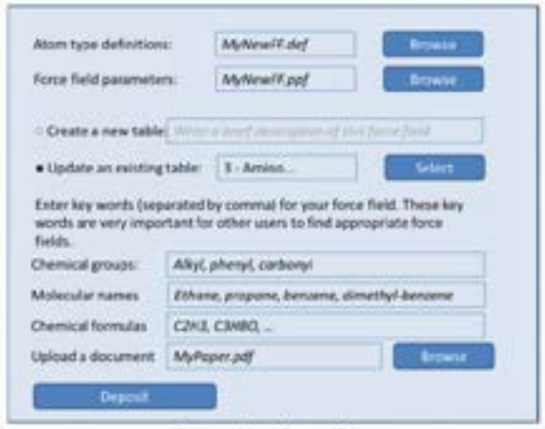
## WebFF

### *Smart force-field repository*

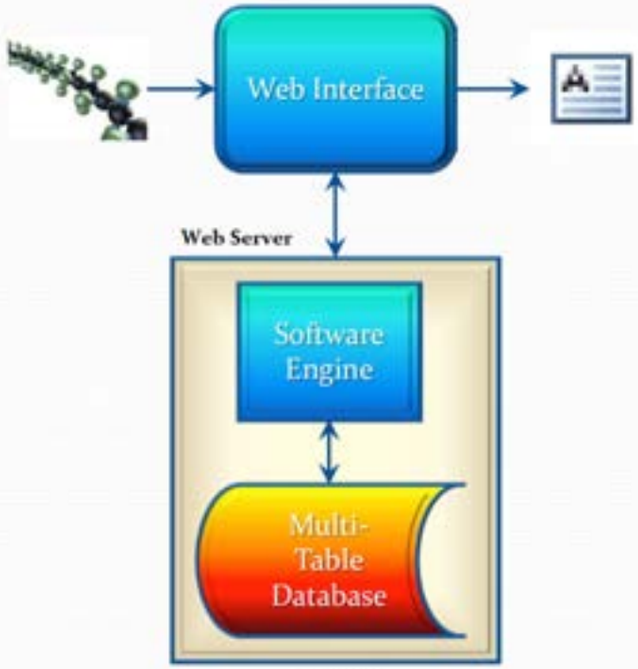
- Assignment Mode
  - Integrated database with software assignment engine, creates workflow
    - Input: Single molecule or molecular system
    - Output: LAMMPS/Gromacs data files
  - Multiple input and output formats
  - Fills missing parameters
- Insertion Mode
  - Extensible, multiple table format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




Assignment Interface




Insertion Interface




**WebFF**
**Web hosted, extensible repository and assignment engine**

# 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**

Go

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

Login

NIST
Material Measurement Laboratory
materialsdata.nist.gov

NIST Repositories → Experimental Data Repository → Diffusion Data → View Item

**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) 135–146
http://dx.doi.org/10.1016/j.msea.2005.07.01

**Related Publications by Author:** → **Related Work**

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

Campbell CE, Zhao JC, Henry MF (2004) Comparison of experimental and simulated multicomponent Ni-base superalloy diffusion couples. J Phase Equil Dif 25 (1):6-15. DOI: http://dx.doi.org/10.1361/10549710417966

**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.

**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[illegible]luso, Rebecca C[illegible], Mitch Hammond and Karen Denike for their experimental support.

**Data files**

**Files in this item**

**Name:** R95-R88-expsimul.txt — View/Open
**Size:** 25.32Kb
**Format:** Text file
**Description:**
Experimental and simulate composition profiles for the R88/R95 diffusion couple at 1150 C for 1000 h

**Name:** r95r88-1000h-labe ... — View/Open
**Size:** 58.76Mb
**Format:** TIFF image
**Description:**
Micrograph of R95/R88 diffusion couple after 1000 h at 1150 C

The following license files are associated with this item:
- Creative Commons

**Offer licenses with attribution 3.0**

creative commons

Except where otherwise noted, [illegible] Universal

**This item appears in the following Collection(s)**
- Diffusion Data

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

Further Studies on the Nickel-Aluminum System. I. The β-Ni2Al3 Phase Fields
Taylor, A; Doyle, N.J. (1972-01-31)
New lattice parameter and density results have been obtained for alloys in the fl-NiA1 and 6-Ni2A13 phase fields of the nickel-aluminum system. The lattice parameter of the fl-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

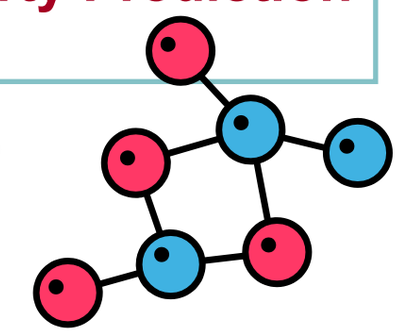
# THERMODYNAMICS RESEARCH CENTER

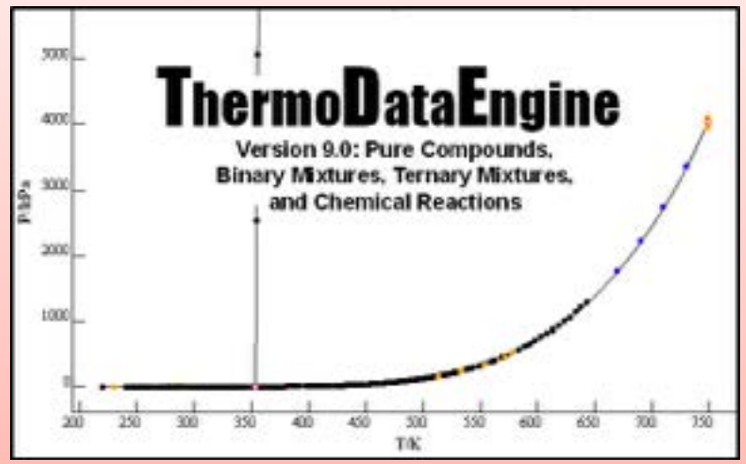
**Open Literature**

Data Capture

Model Development

**Molecular Modeling & Property Prediction**





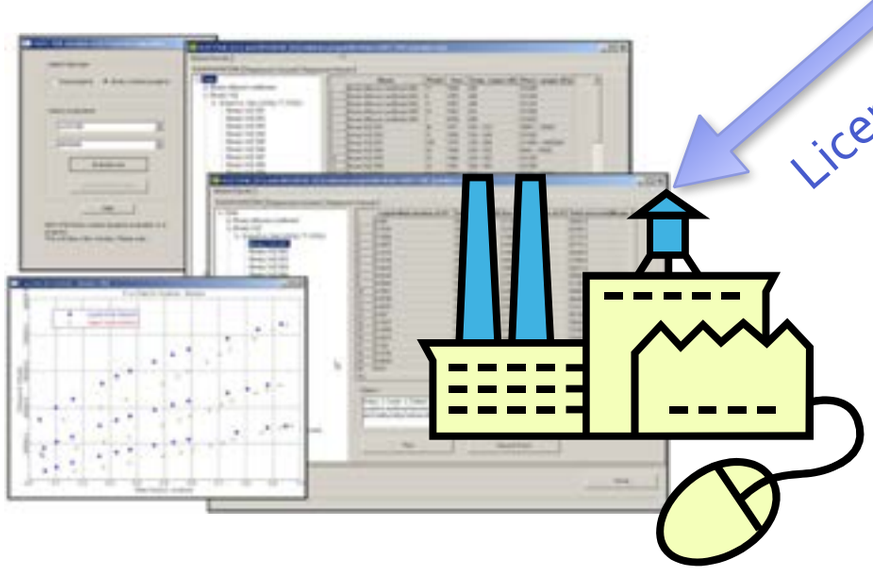

ThermoData Engine (SRD 103b)

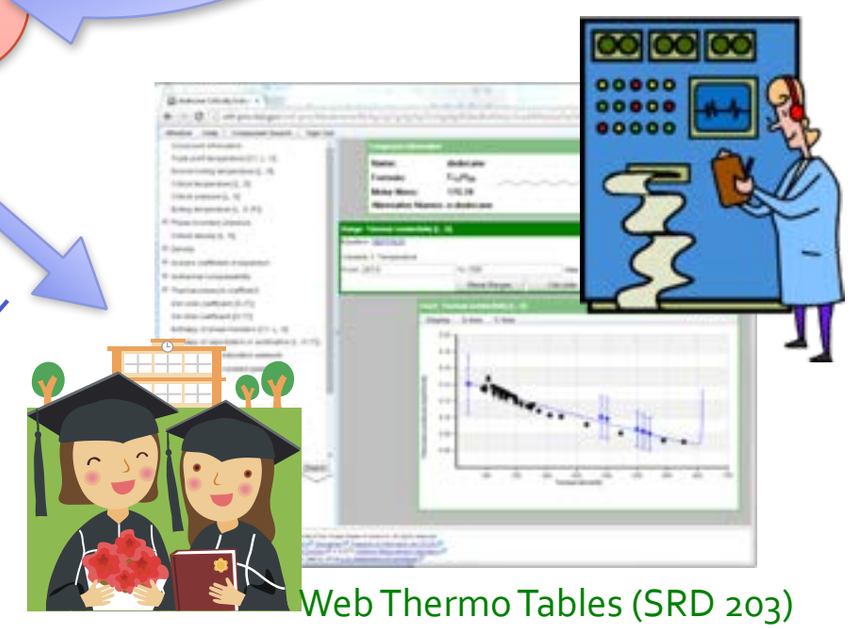
Quality Checks in Peer Review

Algorithmic Enhancement

Licensing

Web-based Dissemination



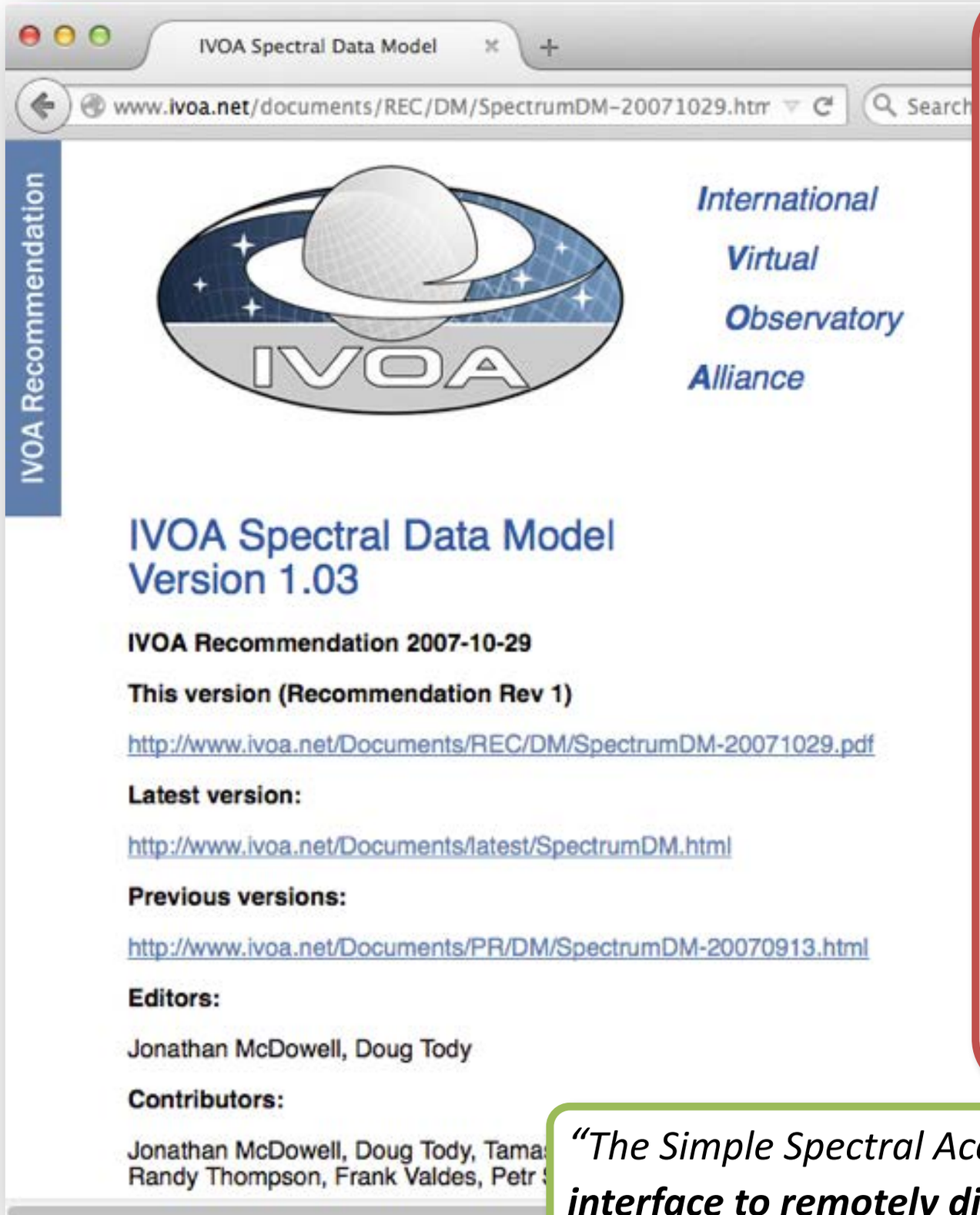


Web Thermo Tables (SRD 203)
ThermoLit (SRD 171)

http://trc.nist.gov/

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

# A Successful Example



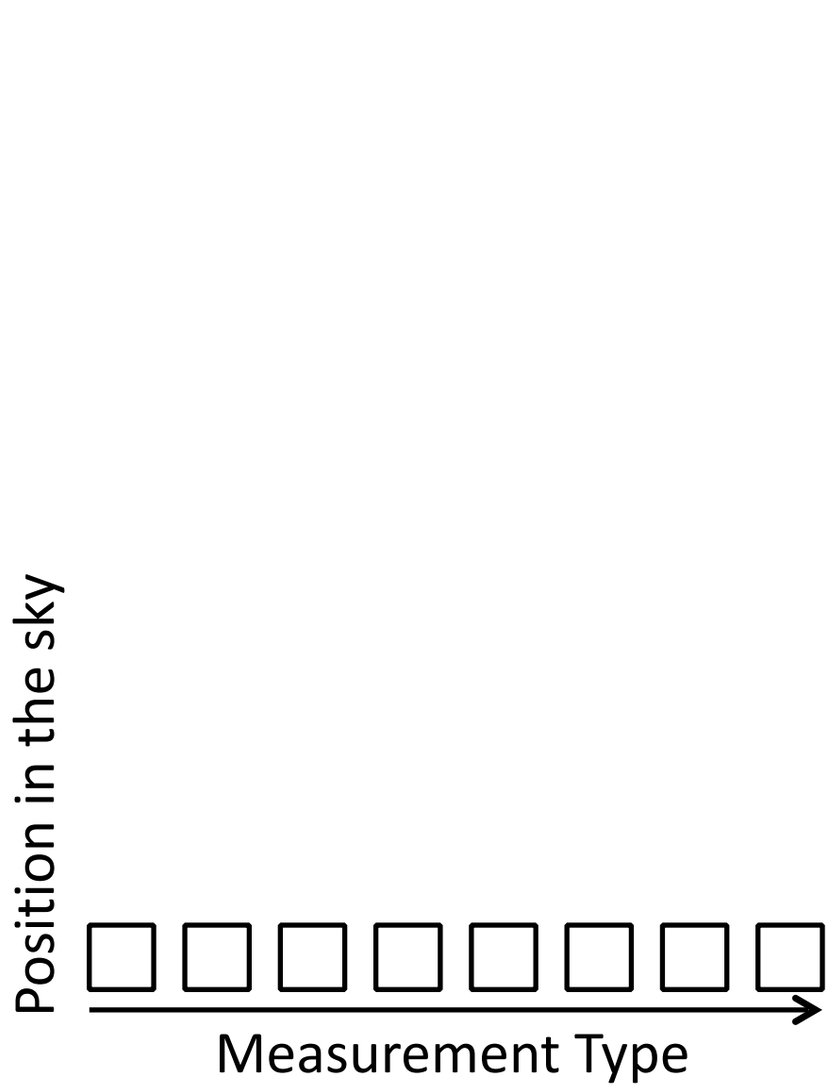

*"We present a **data model describing the structure of** spectrophotometric **datasets** with spectral and temporal coordinates and **associated metadata**. This data model may be used to represent spectra, time series data, segments of SED (Spectral Energy Distributions) and other spectral or temporal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

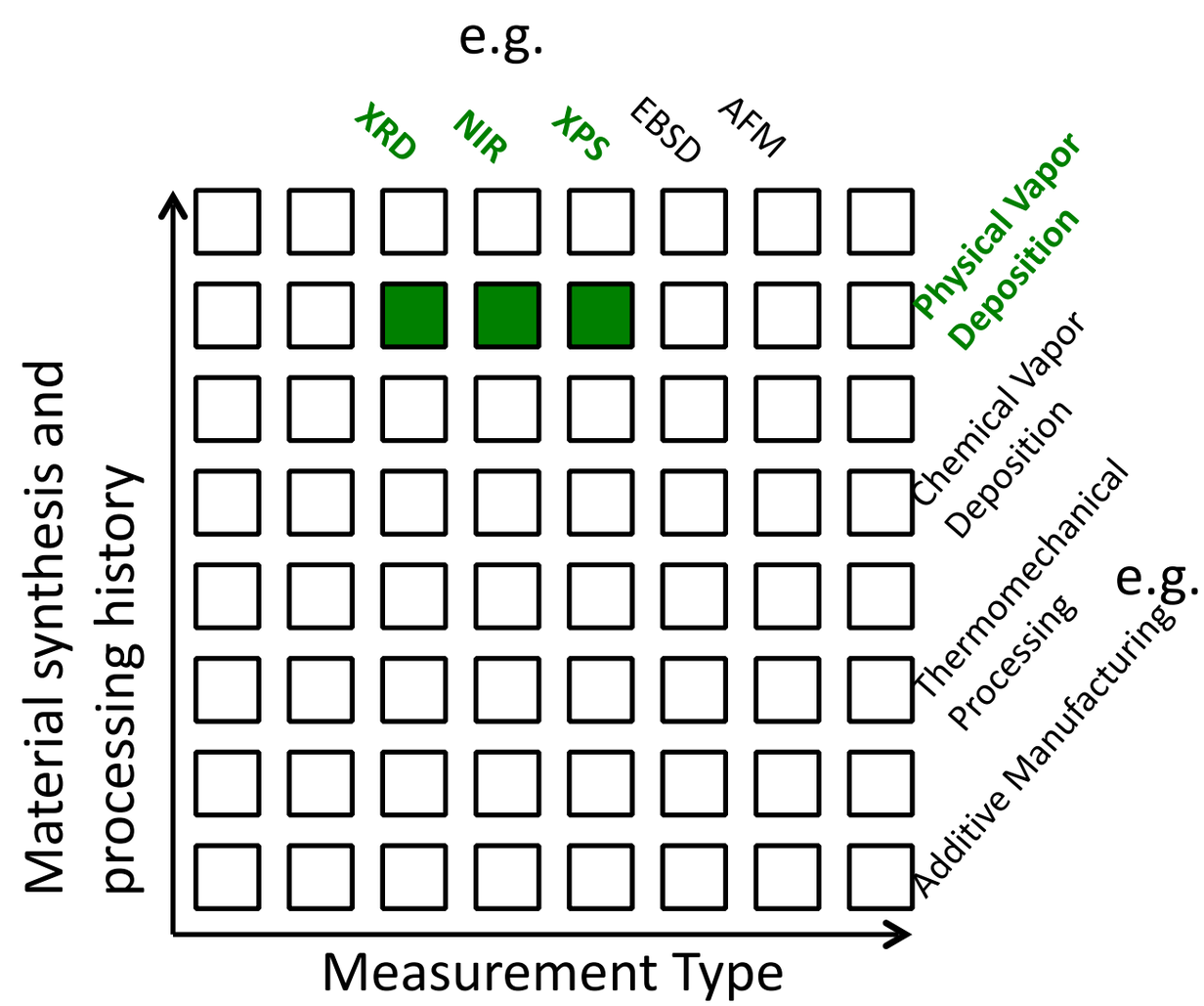
*"This is a Proposed Recommendation, developed with the intention to **support the Simple Spectral Access Protocol**."*

*"The Simple Spectral Access (SSA) Protocol (SSAP) defines a uniform **interface to remotely discover and access** one-dimensional spectra."*

# Astronomy vs. Material Measurement




Singular data models




Modular data models

# Materials Data Curation System

Welcome, cecamp. Thanks for logging in.

Logout | My Profile | Help

Home | Data Curation | Data Exploration | Composer

https://github.com/usnistgov/MDCS

## 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 sollicit comments from the Material Science community. Please do not enter any proprietary data into this system.

## 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.

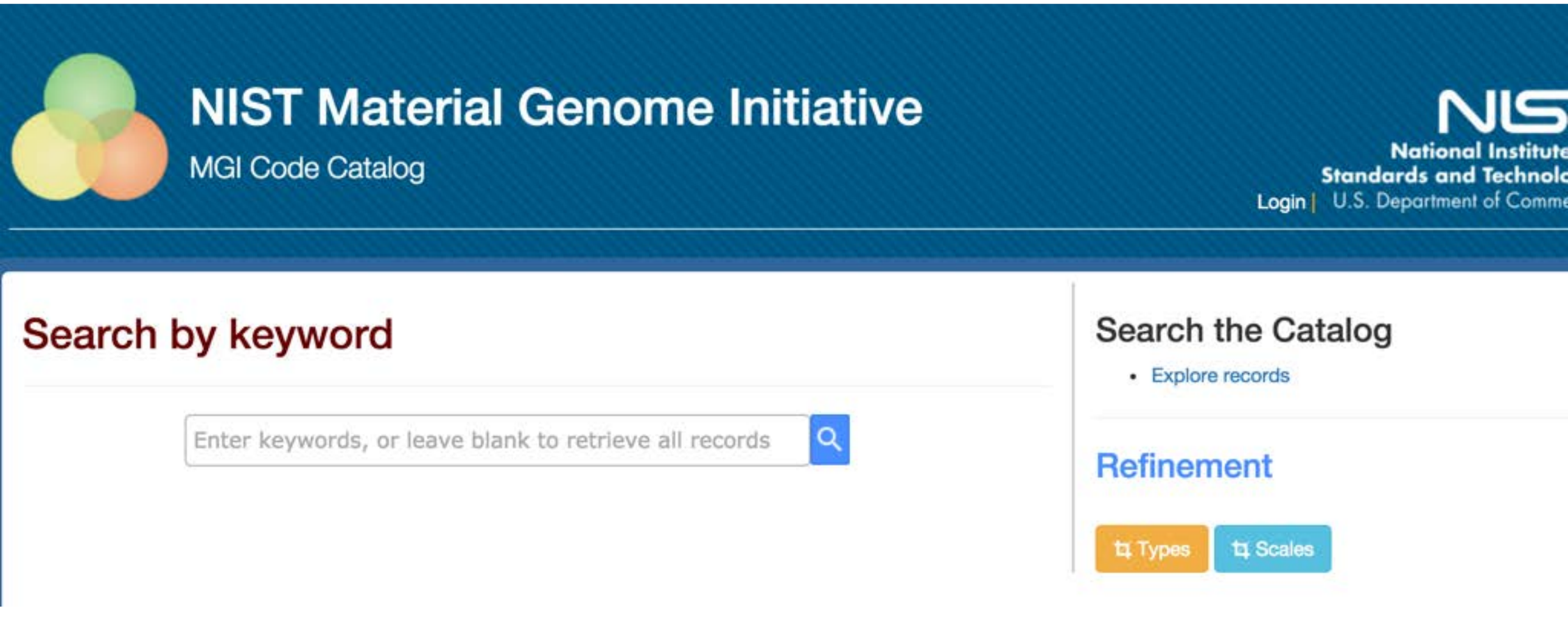
## Most Recent Templates

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

# A simple use for the curator



# Registries

See Poster by Chandler Becker!

National Institute of
Standards and Technology

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

# The case for a Materials Resource Registry

Register. Discover. Access.
Research Data Alliance

# Collaborators

National Institute of Standards and Technology (NIST)

National Data Service (NDS)

Center for Hierarchical Materials Design (CHiMaD)

Research Data Alliance (RDA)

# How do you use to describe your data resource?

**Chandler A. Becker**[1], Raymond L. Plante[1], Sharief Youssef[2], Alden Dima[2], Robert J. Hanisch[1], Laura M. Bartolo[3], James A. Warren[1]

[1] Material Measurement Laboratory, NIST
[2] Information Technology Laboratory, NIST
[3] Center for Hierarchical Materials Design, Northwestern Univ.

NIST MGI seminar, 1 Oct. 2015

# Everybody loves a challenge

## Challenge Details

Discussions 12

Solutions 0

Rules

Submit Solution

### About the Challenge

We are seeking novel uses of accessible digital data to advance Materials Science and Engineering knowledge to accelerate the transition to industrial applications

Posted By: Air Force Research Lab
Category: Scientific/Engineering
Partnership With: National Institute for Standards and Technology
National Science Foundation

Submission Dates : **Jul 01, 2015 - Mar 31, 2016**

Judging Dates : **Apr 01, 2016 - May 11, 2016**

Winners Announced: **May 11, 2016**

People have successfully applied advanced analysis techniques to 'big data' and have solved significant challenges in several areas of society including business, biology, and astronomy. Using these techniques has provided people with new and valuable information from the data that was not readily available by other means. However, materials science and engineering data has not yet been exploited to its full potential because of its complexity. This complexity

### Prizes

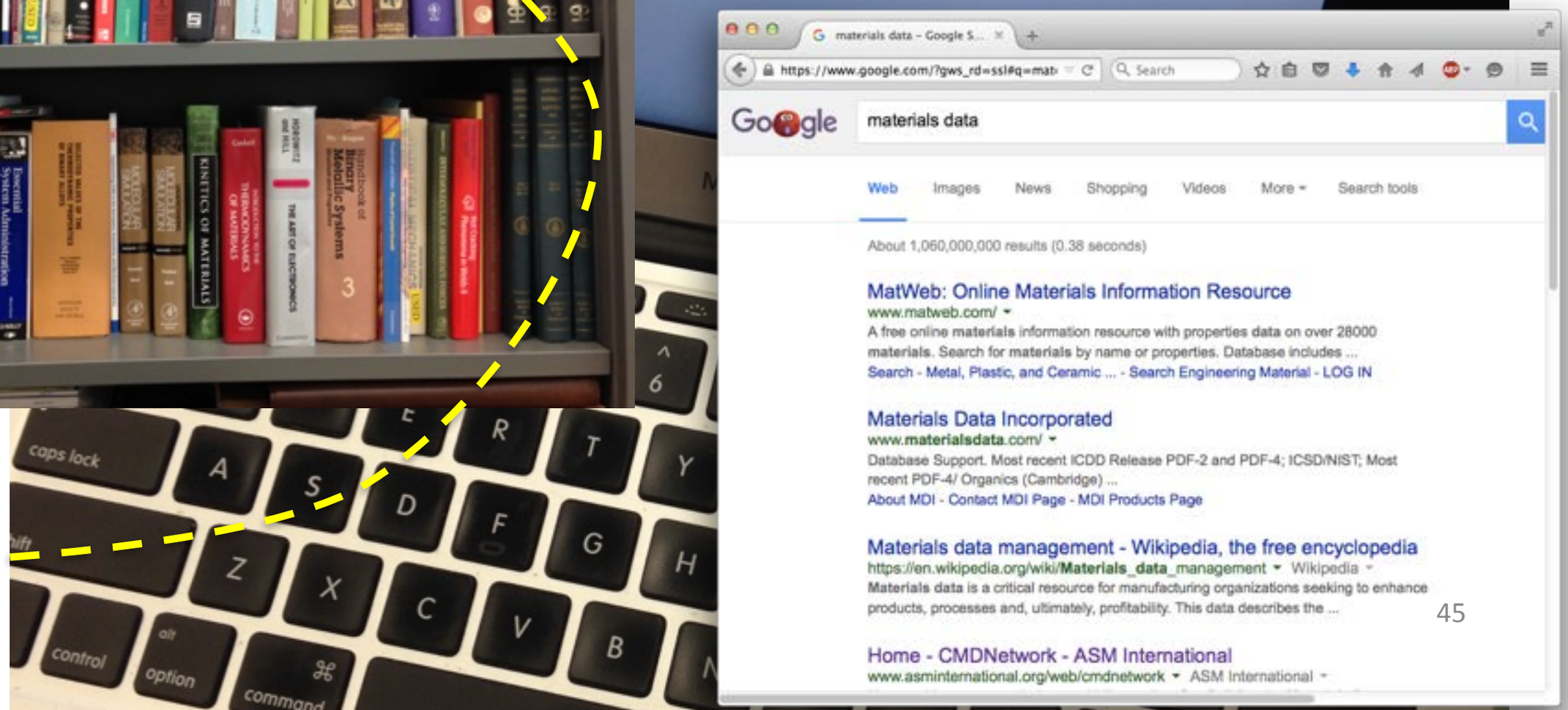
**Materals Genome Iniatiative Prize**

$50,000.00

*The awards will be paid to the best submission(s) as solely determined by the review committee. The US Government intends to make multiple awards with the top award being $25,000 and no award being smaller than $5,000. • In addition to the monetary awards, winning Solvers will be invited to participate (present their research and engage the audience in a question and answer session) in a conference session at Materials Science & Technology 2016, to be held October 24-27, 2016, Salt Lake City, Utah.*

# How do you find data?

- Search engine, library, browser bookmarks, walk down the hall, random walk




?!?

MATERIALS ACCELERATOR NETWORK

Search Go

HOME ABOUT INITIATIVES EVENTS EDUCATION NEWS CONTACT

# MATERIALS SCIENCE AND ENGINEERING DATA CHALLENGE RESOURCES

This page provides links to data, informatics tools, and related information to support participation in the Materials Science and Engineering Data Challenge. Thank you to the many supportive colleagues who have helped us develop this list – it continues to be a team effort. Please note that due to practical limitations these resources have not been reviewed by us and are provided without any assurances of accuracy, quality, performance, or value for your particular project.

## About the Materials Science and Engineering Data Challenge

The Air Force Research Lab (AFRL), National Institute of Standards and Technology (NIST), and National Science Foundation (NSF) are sponsoring a contest to pursue the goals of the Materials Genome Initiative. The focus is to seek novel uses of accessible digital data to advance Materials Science and Engineering knowledge, and accelerate the transition to industrial applications.

http://acceleratornetwork.org/mse-challenge/

## Data Resources

**Computed Data**

AFLOW database

Computational Materials Data (CMD) Network

Harvard Clean Energy Project

Materials Project

National Institute of Standards and Technology (NIST) Interatomic Potentials Repository Project

Open Knowledgebase of Interatomic Models (KIM) or OpenKIM

Open Quantum Materials Database (OQMD)

**Experimental (and possibly computed) Data**

3D Materials Atlas

American Mineralogist Crystal Structure

## Data Mining Tools

Best Data Mining Tools by Quora

Citrine. See also their blog posts on machine learning for the materials scientist part 1 and part 2.

Dream3D

Fiji (ImageJ)

Granta (Material Intelligence)

Massive Online Analysis (MOA)

Materials Knowledge Systems in Python (PyMKS)

Matlab

Matlab Toolbox for Dimensionality Reduction by Laurens van der Maaten

nanoHUB

Nutonian Eureqa

## Places to Publish, Share (and Find) Data

**Journals with Data Focus**

Data in Brief (DiB) (Elsevier) . See also Harvard Dataverse DiB section.

Harvard Dataverse

Integrating Materials and Manufacturing Innovation (IMMI) (see Data Descriptor article type)

Materials Discovery (Elsevier)

Open Data journals at Elsevier. Part of a number of projects at Elsevier supporting the Materials Genome Initiative. See also Elsevier's page on their resources for the MS&E Data Challenge.

Scientific Data (Nature Publishing Group)

**Data Repositories and Data Sharing Tools**

Citrine (see their blog for details on their support of datasets for the Challenge)

What does each resource represent? How do I find out?
What if I want to search only these resources or a subset of them?

NIST
Home
Services
Dashboard
Help
Contact
API
NIST Materials Resource Registry
Part of the Materials Genome Initiative
SEARCH FOR RESOURCES
ADD YOUR RESOURCE
Find Materials Resources
This system allows for the registration of materials resources, bridging the gap between existing resources and the end users. The NIST Materials Resource Registry functions as a centrally located service, making the registered information available for research to the materials community.
This is being developed at the National Institute of Standards and Technology and is made
Home Page
Services
Search for Resources
Add Your Resource
Logout

# Browse and refine

# Get more information

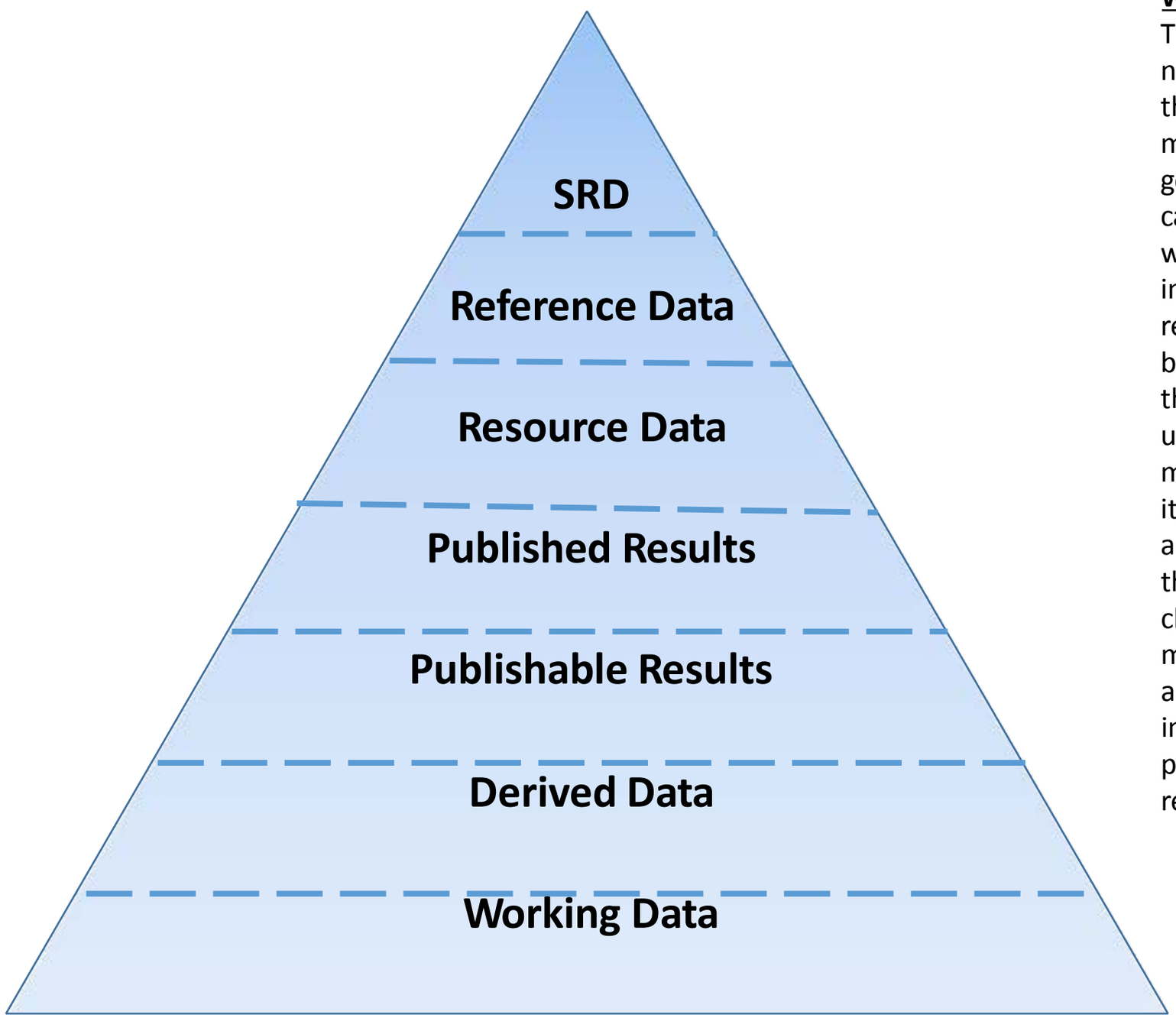
**Resource Details**

| | |
|---|---|
| Resource Name: | AFLOW |
| Short Name: | Aflow |
| Version: | v31014 |
| Publisher: | AFLOW Consortium |
| Creator: | |
| Contributor: | |
| Contact: | Stefano Curtarolo, stefano@duke.edu |
| Date: | |
| Resource Type: | Repository |
| Description: | Aflow is a globally available database of 647,815 material compounds (including metal alloys) - with over 58,303,350 calculated properties (and growing). The Aflow portal provides interactive search applications as well as a REST interface for programmatic access. |
| Subjects: | materials property |
| Reference URL: | http://aflowlib.org/ |
| Reference Citation: | |

# Data Policy

A NIST Approach (in progress)
We can help you with DMPs!

# 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: STRAWMAN-DRAFT

- 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[1]
  4. Inventory record flagged for public access

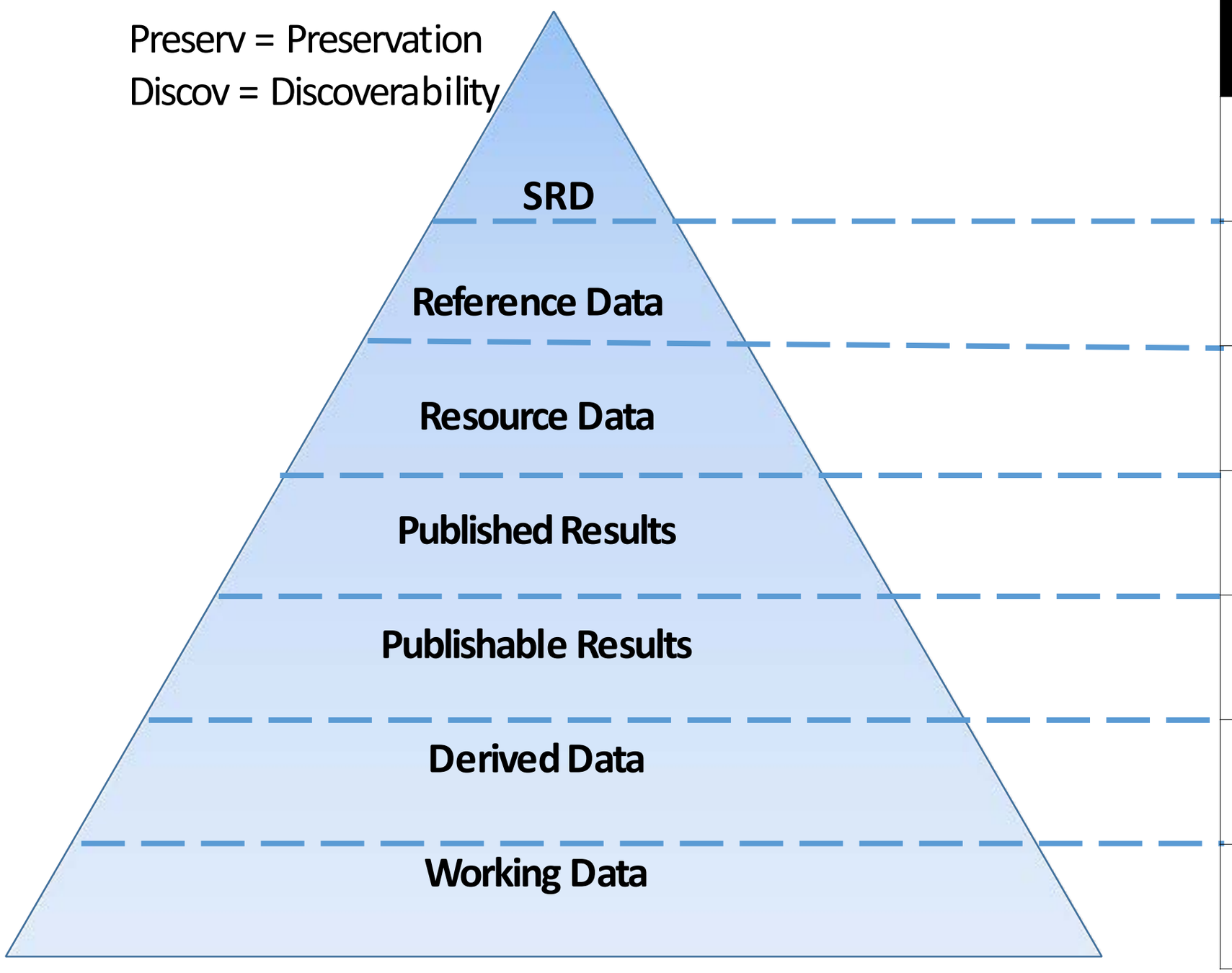
[1]PID assigned + NIST minimum metadata present

# 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 |

# NIST Center of Excellence: Advanced Materials

CHiMaD

# CENTER FOR HIERARCHICAL MATERIALS DESIGN

THE UNIVERSITY OF CHICAGO

NORTHWESTERN UNIVERSITY

Argonne NATIONAL LABORATORY

QUESTEK INNOVATIONS LLC

ASM INTERNATIONAL

FAYETTEVILLE STATE UNIVERSITY

STOP

MGI.NIST.GOV