# FAIR Data Initiatives at NIST

Robert Hanisch
Director, Office of Data and Informatics
Material Measurement Laboratory
National Institute of Standards and Technology

Wednesday September 11, 2019

MATERIAL
MEASUREMENT
LABORATORY

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

# Office of Data and Informatics (ODI)

## Standard Reference Data

- Distribution
- Sales
- Infrastructure
- Usage analysis and impact
- Improve web sites and user interfaces
- Provide APIs
- Register with data.gov

## Research Data

- Improve data management practices
- Data management planning tools
- Laboratory automation
- ELNs
- Open data policy implementation and guidance
- NIST open data repository
- NIST data portal

## Data Science

- Informatics and analytics resource
- Liaison with NIST Information Technology Laboratory
- Big data
- Cloud computing
- National Strategic Computing Initiative

## Community

- National Data Services Consortium
- Research Data Alliance
- DoC and other federal agencies (NIH, DOE, NSF)
- CENDI
- NMIs and BIPM
- CODATA, WDS

# ODI in Context

- National Institute of Standards and Technology (~5,000 people), total budget of ~$1B
- US National Metrology Institute
- Within US Department of Commerce
  - Material Measurement Laboratory (~1,000)
    - *ODI (16)*
  - Physical Measurement Laboratory (~1,000)
  - Engineering Laboratory
  - Information Technology Laboratory
  - Communications Technology Laboratory
  - NIST Center for Neutron Research

# Discover, Access, Interoperate

- Data challenges start with discovery, access, and interoperability
  - Support FAIR* principles: Findable, Accessible, Interoperable, Re-usable
  - Assure maximum return on national investment in basic research
  - Demonstrate best practices
  - Address reproducibility "crisis"
  - Cannot exploit data without sound data management
- OMB, OSTP directives; Open Data Act 2019
- Challenges: Data diversity, formats, proprietary representations, inadequate metadata, lack of credit for data curation and sharing

*Wilkinson et al. 2016, *Nature Scientific Data*, DOI: 10.1038/sdata.2016.18

# Policies, Directives, etc.

NIST

Search NIST | NIST MENU

**Public Access to NIST Research**

- Policies, Directives, and NIST's Public Access Plan
- Information for Awardees
- Information for Developers
- Copyright, Fair Use, and Licensing Statements for SRD, Data, and Software
- Open Government Performance Metrics

## Policies, Directives, and NIST's Public Access Plan

### NIST Directives related to public access

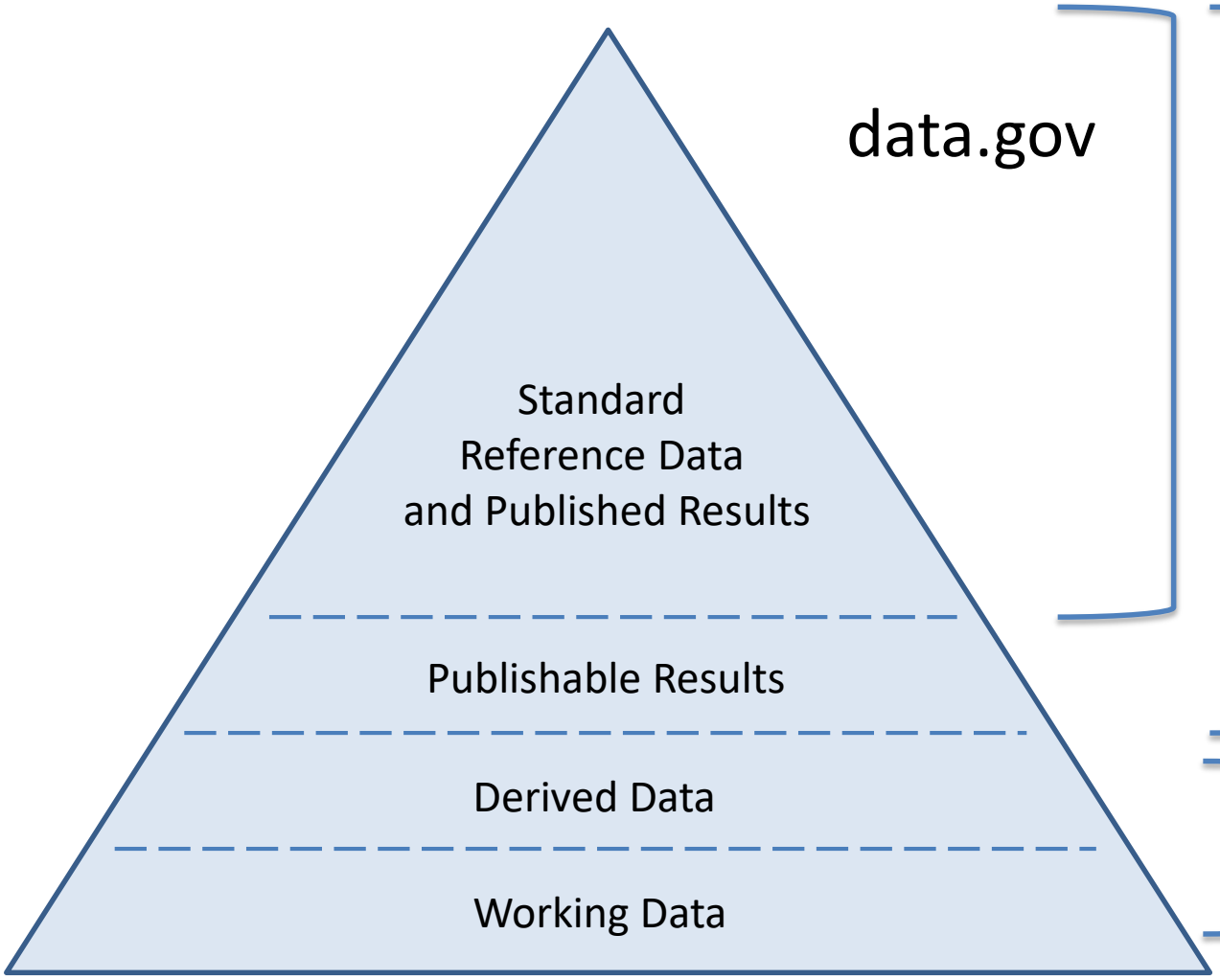
**Policy:** NIST P 5700.00 Managing Public Access to Results of Federally Funded Research
**Order:** NIST O 5701.00 Managing Public Access to Results of Federally Funded Research
**Guidance:** NIST G 5701.00 Guidance for Review of Data Management Plans Submitted with Applications and Proposals for NIST Funding for Scientific Research
**Order:** NIST O 5702.00 Preservation and Maintenance of Published Research Data
**Procedure:** NIST PR 5702.01 Managing Public Access to External Research Funded by NIST
**Guidance:** NIST G5702.01 Guidance for Making Non-NIST Data Available Through NIST's Data Management Infrastructure

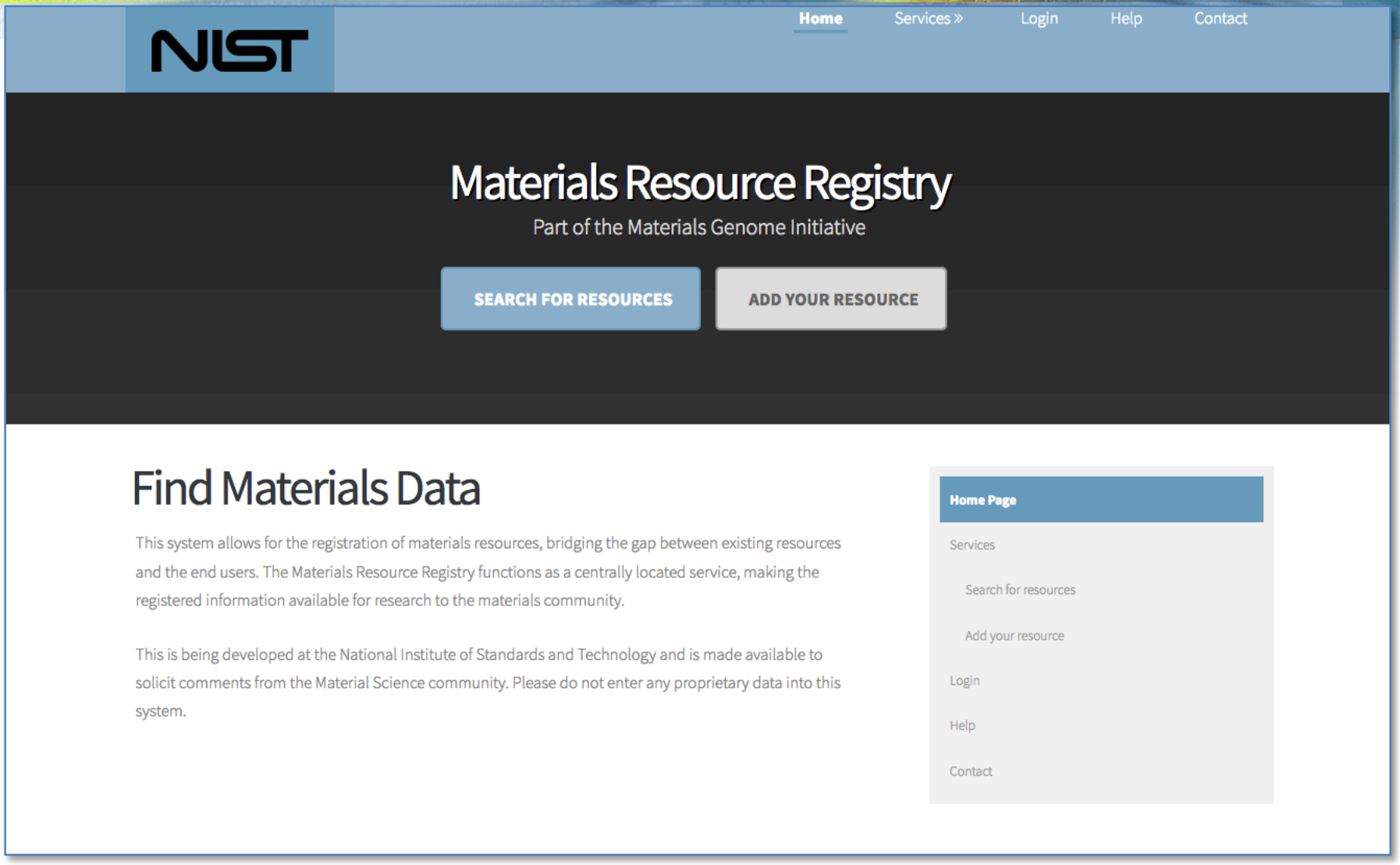
### NIST's Public Access Plan

The NIST Plan for Providing Public Access to the Results of Federally Funded Research was replaced by the directives above. The Plan and comments that were received in response to the *Federal Register Notice* announcing its availability for public comment, Docket Number 150414364-5364-01 are provided in NISTIR 8084.

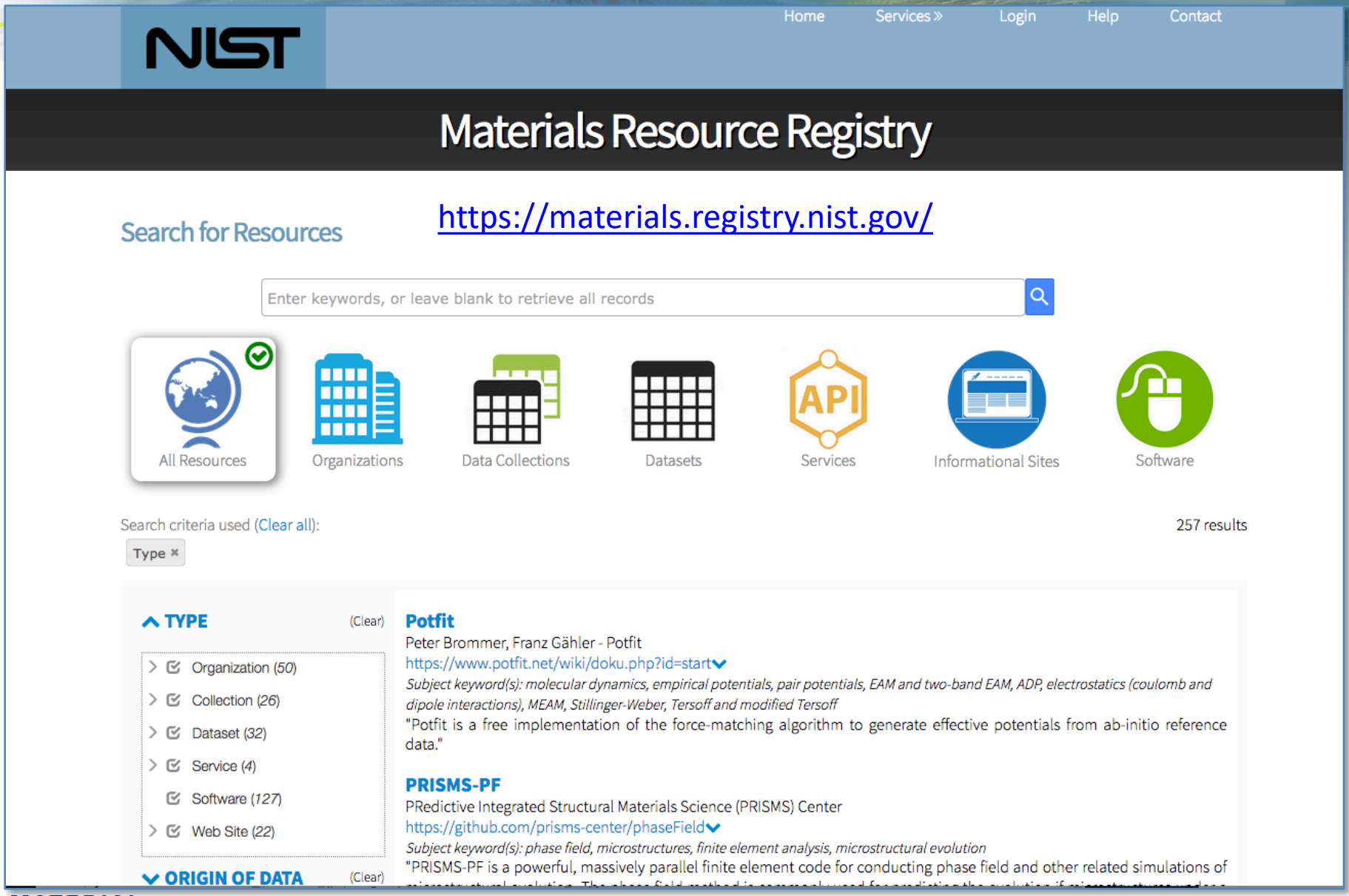
https://www.nist.gov/open/policies-directives-and-nists-public-access-plan

# NIST Data Pyramid






https://materials.registry.nist.gov/


NIST
Home
Services »
Login
Help
Contact
Materials Resource Registry
Search for Resources
https://materials.registry.nist.gov/
Enter keywords, or leave blank to retrieve all records
All Resources
Organizations
Data Collections
Datasets
API
Services
Informational Sites
Software
Search criteria used (Clear all):
257 results
Type ×
TYPE
(Clear)
Organization (50)
Collection (26)
Dataset (32)
Service (4)
Software (127)
Web Site (22)
ORIGIN OF DATA
(Clear)
Potfit
Peter Brommer, Franz Gähler - Potfit
https://www.potfit.net/wiki/doku.php?id=start
Subject keyword(s): molecular dynamics, empirical potentials, pair potentials, EAM and two-band EAM, ADP, electrostatics (coulomb and dipole interactions), MEAM, Stillinger-Weber, Tersoff and modified Tersoff
"Potfit is a free implementation of the force-matching algorithm to generate effective potentials from ab-initio reference data."
PRISMS-PF
PRedictive Integrated Structural Materials Science (PRISMS) Center
https://github.com/prisms-center/phaseField
Subject keyword(s): phase field, microstructures, finite element analysis, microstructural evolution
"PRISMS-PF is a powerful, massively parallel finite element code for conducting phase field and other related simulations of

Home | Services » | Dashboard | Logout | Help | Contact

Bureau International des Poids et Mesures

http://imrr.bipm.org/

IMRR

# International Metrology Resource Registry Delta

SEARCH FOR RESOURCES | ADD YOUR RESOURCE

## Find Resources

This system allows for the registration of resources, bridging the gap between existing resources and the end users. The International Metrology Resource Registry functions as a centrally located service, making the registered information available for research to the global community.

This is being developed at the Bureau International des Poids et Mesures and is made available to solicit comments from the global community. Please do not enter any proprietary data into this system.

- Home Page
- Services
  - Search for resources
  - Add your resource
- Dashboard
- Logout
- Help
- Contact

# International Metrology Resource Registry

## Search for Resources

Enter keywords, or leave blank to retrieve all records

All Resources | Participating Institutes | Databases | Datasets | Standards Documents | Data Services | Data Portals | Metrology-related Software and Tools

Search criteria used (Clear all):

Resource Type: ×

58 results

**RESOURCE TYPE:** (Clear)

- Database (20)
- Dataset (7)
- Document (13)
- Metrology Institute (9)
- Portal (6)
- Service (1)
- Software (2)

**STATUS OF THE DATA:** (Clear)

**STATUS OF THE DOCUMENT:** (Clear)

**APPLICATION AREA:** (Clear)

**TYPE OF PROPERTIES MEASURED:** (Clear)

### NIST/EPA Gas Phase Infrared Library (NIST-GPID)

US National Institute of Standards and Technology (NIST) (United States of America)

Home: 10.18434/T45K5N

*Subject keyword(s): infrared spectra, vapor phase*

This data collection contains 5,228 infrared spectra of different compounds along with chemical structures for most of them. Spectra are provided on a CD-ROM in the JCAMP-DX (Joint Committee for Atomic and Molecular Physical Data "Data Exchange") format. Chemical structures are provided in the MOL-file format. The IR data origin... show more

### NIST Standard Reference Data Gateway

US National Institute of Standards and Technology (NIST) (United States of America)

Home: http://srdata.nist.gov/gateway/

*Subject keyword(s): standards, standard reference data, metrology, materials, physics, computing, cybersecurity, law enforcement*

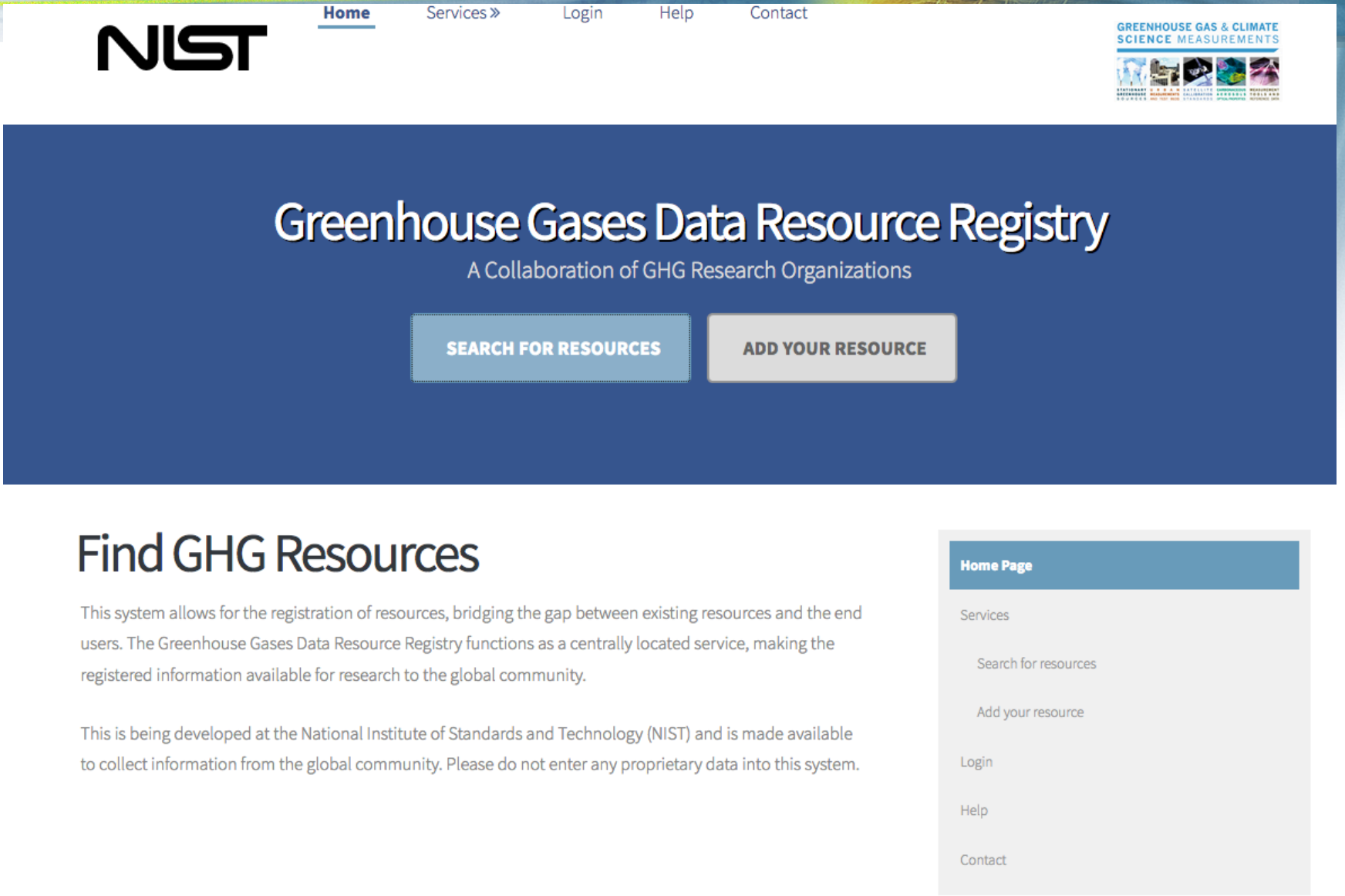
The NIST Data Gateway is an online system designed to provide users with easy access to NIST scientific and technical data. The Gateway provides (1) direct links to free NIST online systems and full text online versions of many Journal of Physical and Chemical Reference Data (JPCRD) articles, (2) links to descriptions and orderi... show more

### NIST Thermophysical Properties of Hydrocarbon Mixtures Database (SRD#4)

US National Institute of Standards and Technology (NIST) (United States of America)

Home: 10.18434/T4CC76

*Subject keyword(s): Chemical engineering, chemical manufacturing, chemical property, chemistry, dielectric constants, equation of state, fluids, heat capacity, hydrocarbons, LNG, mixtures, natural gas, petrochemical, phase equilibria, thermal conductivity, thermodynamics, thermophysics, transport property, viscosity*


NIST
Home
Services »
Login
Help
Contact
GREENHOUSE GAS & CLIMATE SCIENCE MEASUREMENTS
Greenhouse Gases Data Resource Registry
A Collaboration of GHG Research Organizations
SEARCH FOR RESOURCES
ADD YOUR RESOURCE
Find GHG Resources
This system allows for the registration of resources, bridging the gap between existing resources and the end users. The Greenhouse Gases Data Resource Registry functions as a centrally located service, making the registered information available for research to the global community.
This is being developed at the National Institute of Standards and Technology (NIST) and is made available to collect information from the global community. Please do not enter any proprietary data into this system.
Home Page
Services
Search for resources
Add your resource
Login
Help
Contact

# Greenhouse Gases Data Resource Registry

## Search for Resources

Enter keywords, or leave blank to retrieve all records

All Resources | Participating Organizations | Databases | Datasets | Data Services | Data Portals | Software and Tools

Search criteria used (Clear all):

Resource Type: ×

34 results

**RESOURCE TYPE:** (Clear)
- Database (17)
- Dataset (10)
- Document (0)
- Organization (0)
- Portal (4)
- Service (0)
- Software (3)

**SECTOR AREA:** (Clear)

**DATA COLLECTION METHOD:** (Clear)

**GREENHOUSE GAS:** (Clear)

**SOURCE CATEGORY** (Clear)

**STATE NAME:** (Clear)

**AVAILABILITY:** (Clear)

**TYPE OF MEDIA AVAILABLE:** (Clear)

### World Council on City Data Open Data Portal (WCCD Open Data Portal)
World Council on City Data
Home: http://open.dataforcities.org/
*Subject keyword(s): city indicators, emissions, pollution*
This portal provides data for WCCD cities on a variety of indicators, including GHGs, NO2, SO2, O3, PM10, PM2.5. Provides access to inventories for pollution data.

### CAIT Climate Data Explorer
World Resources Institute
Home: http://cait.wri.org/
Aggregation of reported emissions data reported to UNFCCC and others and targets from countries and states (cities coming soon)

### Carbon Disclosure Project (CDP)
Carbon Disclosure Project
Home: https://data.cdp.net/browse
*Subject keyword(s): emissions, targets, emissions reduction actions, cities, companies, maps*
City and company emissions reporting

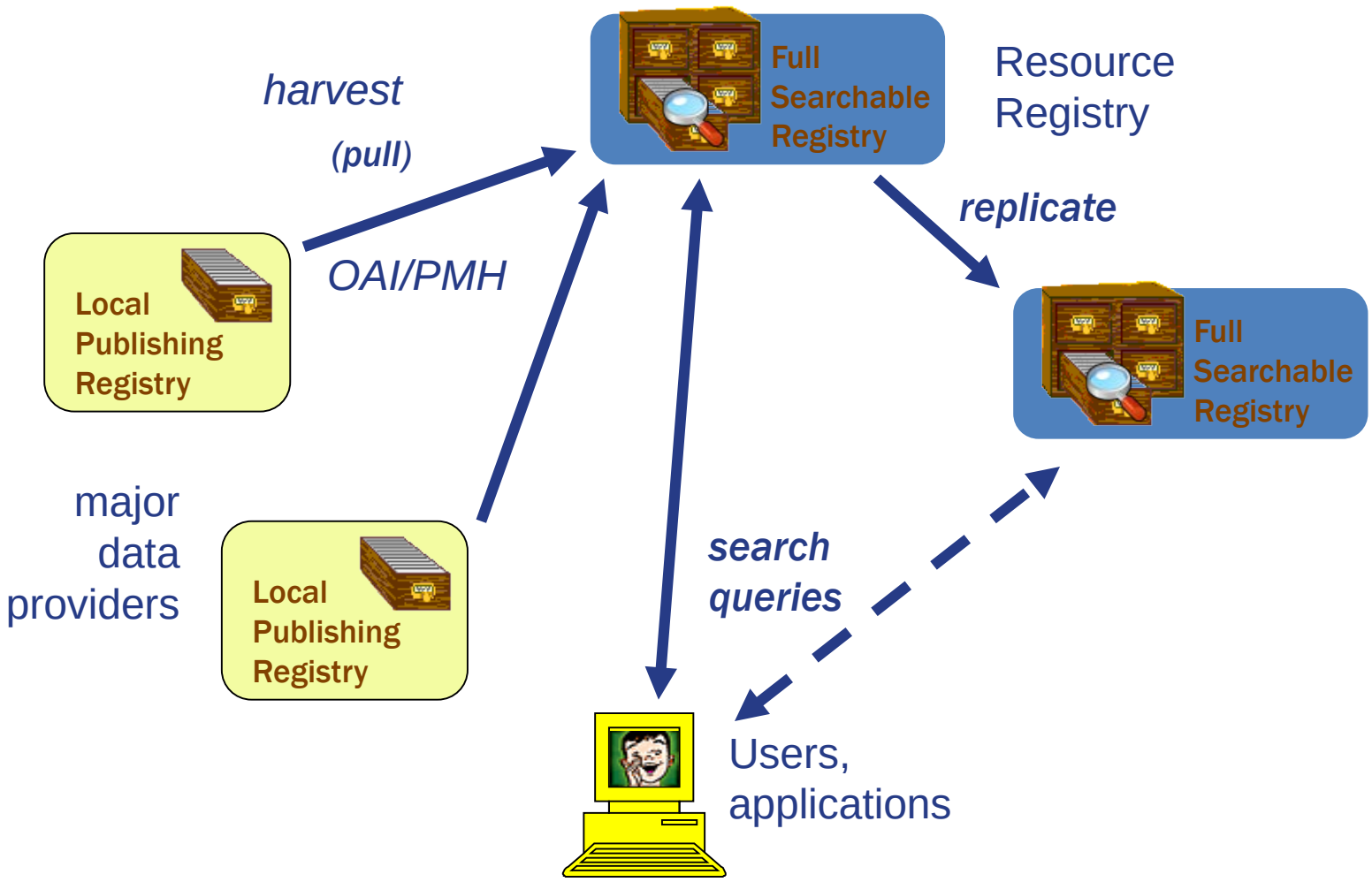
### Megacities Carbon Project
Jet Propulsion Laboratory (NASA)
Home: https://megacities.jpl.nasa.gov/portal/
*Subject keyword(s): CO2, CH4, CO, 14C, atmospheric*
The Megacities Carbon Project is developing, testing, and improving robust methods for assessing carbon emissions and monitoring the atmospheric trends of carbon attributed to the world's largest cities.

### The Berkeley High-Resolution NO2 Product

MATE
MEAS
LABO

NIST
al Institute of
d Technology
nt of Commerce

# Federated Architecture



# materialsdata.nist.gov

Data Communities API MGI Login

Materials Data Repository Home

## NIST Materials Data Repository

The National Institute of Standards and Technology has created a materials science data repository as part of an effort in coordination with the Materials Genome Initiative (MGI) to establish data exchange protocols and mechanisms that will foster data sharing and reuse across a wide community of researchers, with the goal of enhancing the quality of materials data and models. Data present on this system are varied and may originate from within NIST or from the worldwide materials community. Data present on this system may or may not be critically reviewed.

NIST provides this data as a public service and assumes no liability whatsoever for use of the data. NIST provides the data "AS IS" and makes NO WARRANTY OF ANY TYPE, EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION NO WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR USE, AND NO WARRANTY OF NON-INFRINGEMENT OF THIRD PARTY INTELLECTUAL PROPERTY RIGHTS. The individual data providers define how the data may be used and the terms of use associated with each data entry.

All data in this repository are intended to be public; some data may be a part of invitation-only collections while researchers continue to analyze the data prior to publication, with the goal of making the public data presentable in a meaningful way. Publishable data derived from working data will be made public at the appropriate time (this may coincide with end of a grant or when the data has been accepted for publication).

The NIST Materials Data Repository is open to the research community interested in developing best practices in the management of materials data. We are currently looking into use cases for expansion of the site including, but not limited to, partnerships with other projects focusing on discoverability.

Anyone having an interest in an invitation-only collection should contact the site administrator for more information.

Search

RELATED PRODUCTS

Root and Rules Search

BROWSE

All of Materials Data Repository

Communities & Collections

By Issue Date

Authors

Titles

Subjects

MY ACCOUNT

Login

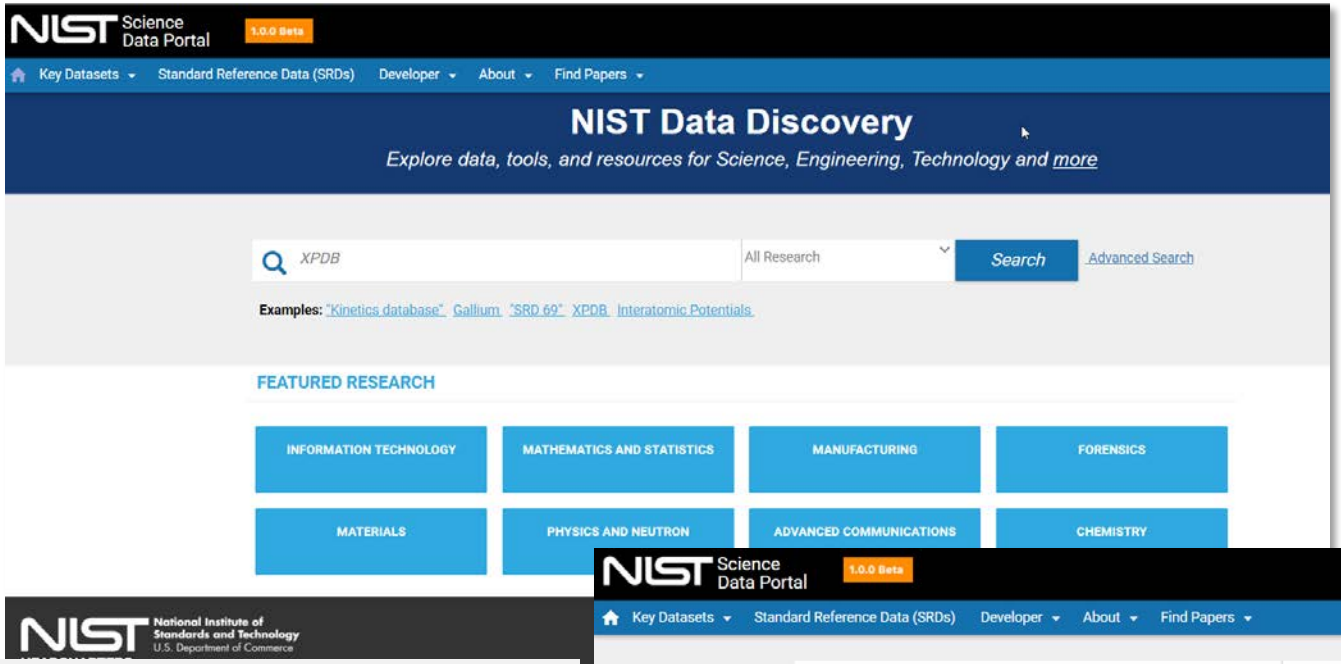
# Based on DSpace

## Communities in Materials Data Repository

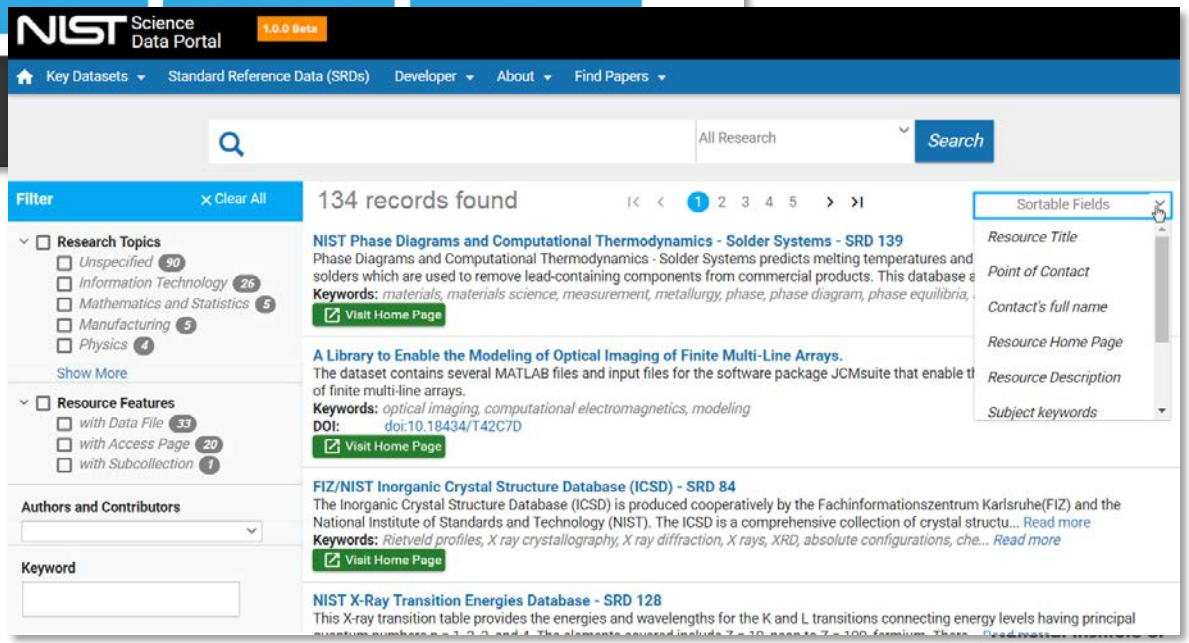
Select a community to browse its collections.

- ASM Journal of Phase Equilibria and Diffusion
  Datasets submitted in support of articles published as JPED may be archived at the NIST data repository.
- ASM Structural Materials Data Demonstration Project
- Chemical Sciences
- CHiMaD Data Collections [R]
- Community for Greenhouse Gases
  SPO Greenhouse Gas and Climate Science Measurements Program
- Computational File Repository
- Experimental Data Repository
- Heusler Phases: First Principles Simulations [R]
- ICME Approach to Development of Lightweight 3GAHSS Vehicle Assembly [R]
- ICME of Carbon Fiber Composites for Lightweight Vehicles [R]
- MGI Catalogs
  A community for collections of useful resources, within the general scope of the MGI.
- MICCoM Collections [R]
  Midwest Integrated Center for Computational Materials
- NanoRelease
- NIST/DOE-EERE Advanced Automotive Cast Magnesium Alloys [R]
  Addressed the need for fundamental understanding of Mg phase transformation, diffusion, and corrosion behavior
- NIST Thermodynamics and Kinetics Test Space [R]
  Repository for developmental purposes restricted the NIST Materials Informatics Effort.
- Porous Metals and Ceramics: Freeze-casting under microgravity and terrestrial conditions
- RDA Demonstration Project: DTR/PID & MGI Infrastructure [R]
- State Variable Model for QP980
- Synchrotron Studies of Slot Die Coated Films [R]
- Thermal Conductivity of CVD Diamond - DARPA Round Robin [R]
- TMS Springer Integrating Materials and Manufacturing Innovation (IMMI)
  This community holds the data associated with articles published in Integrating Materials and Manufacturing Innovation (IMMI)

# Data Discovery for Public Research Data




*Data and s/w reside in AWS*




**Search NIST public data records**

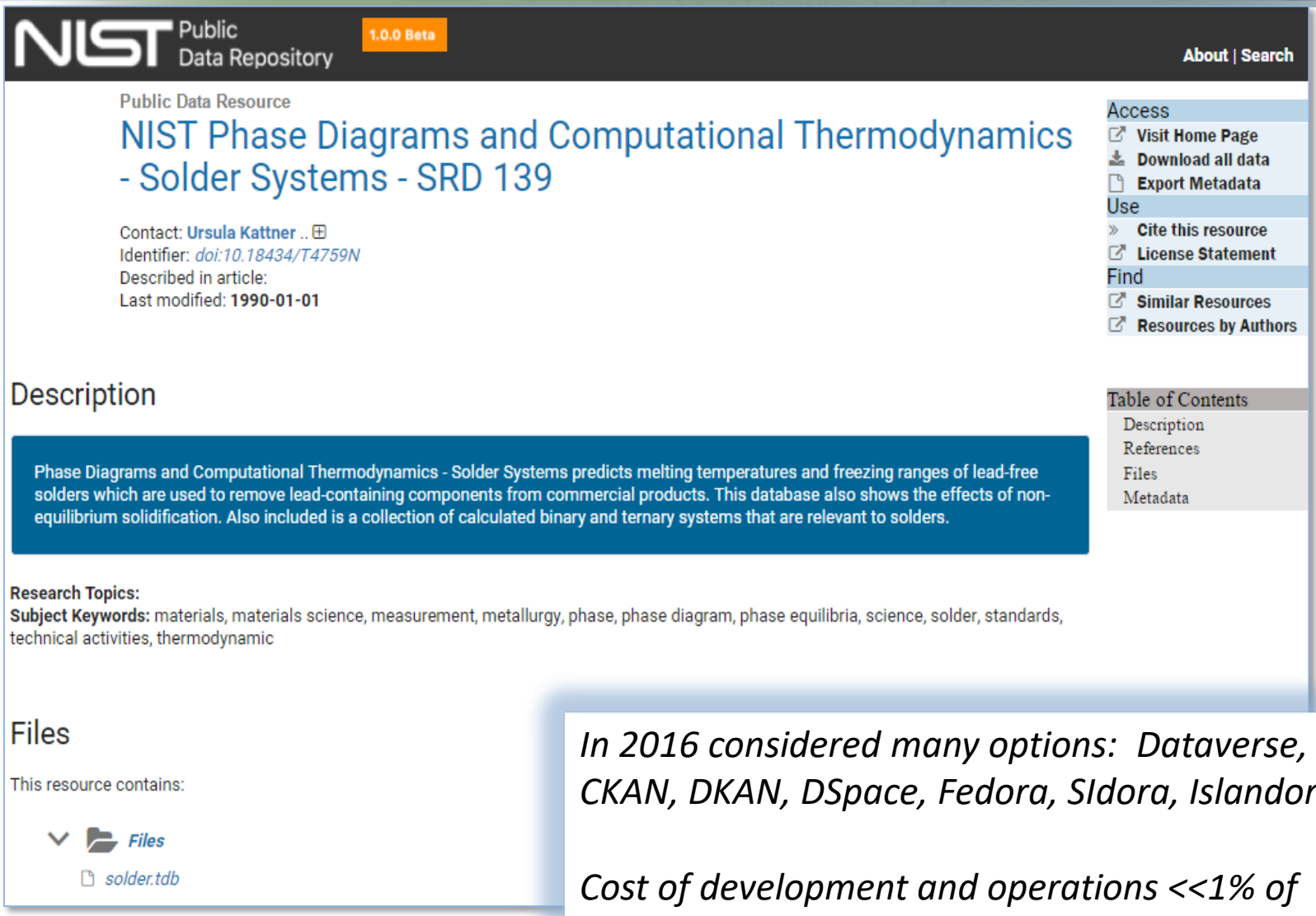
- View metadata, filter results
- Access data files, metadata
- APIs allow interoperability with client tools
- Records uploaded and synced with data.gov

MATERIAL MEASUREMENT LABORATORY



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

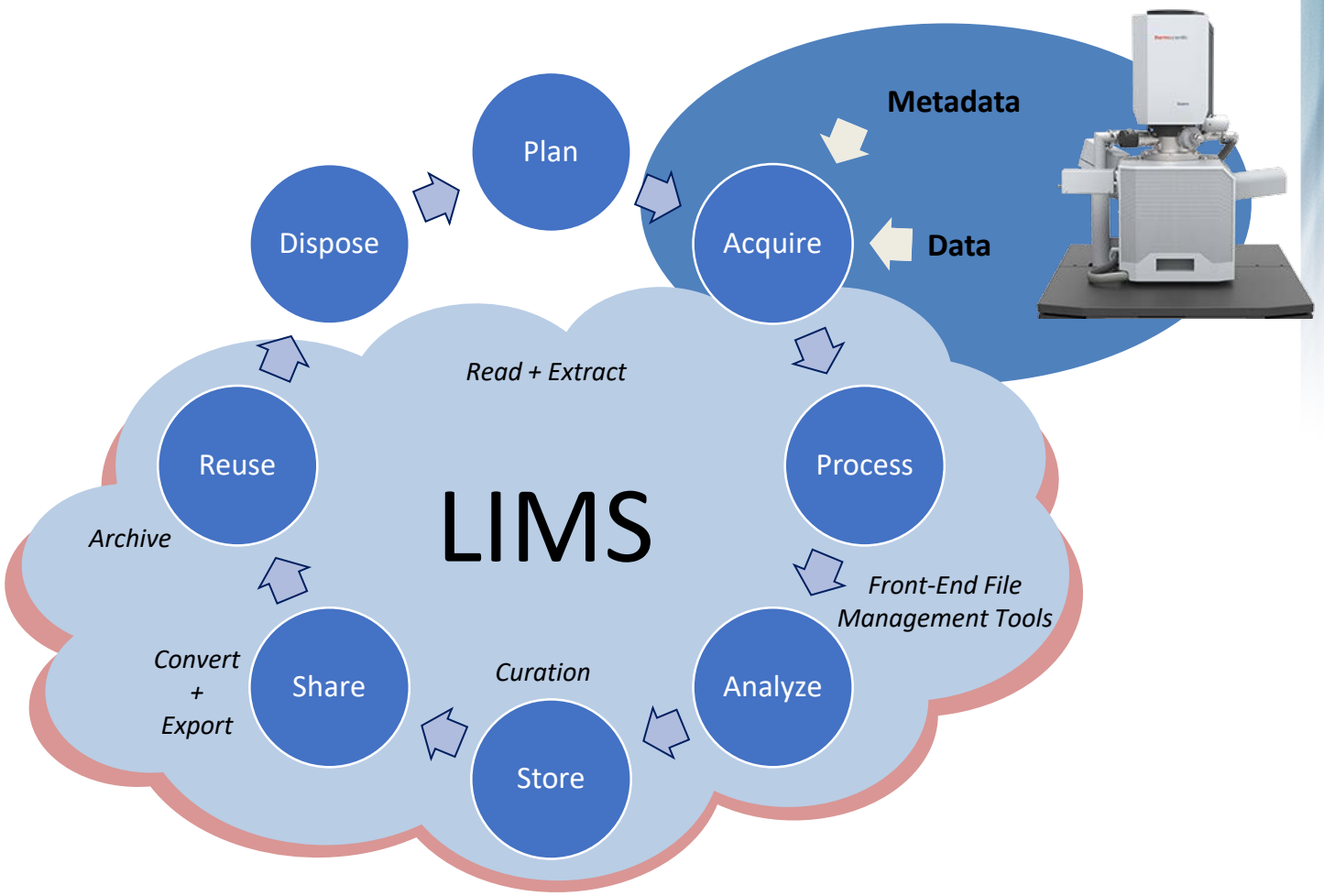
# NIST Public Data Repository – Basic Landing Page




*In 2016 considered many options: Dataverse, CKAN, DKAN, DSpace, Fedora, SIdora, Islandora*

*Cost of development and operations <<1% of NIST research budget*

# LIMS Help Manage the Data Lifecycle

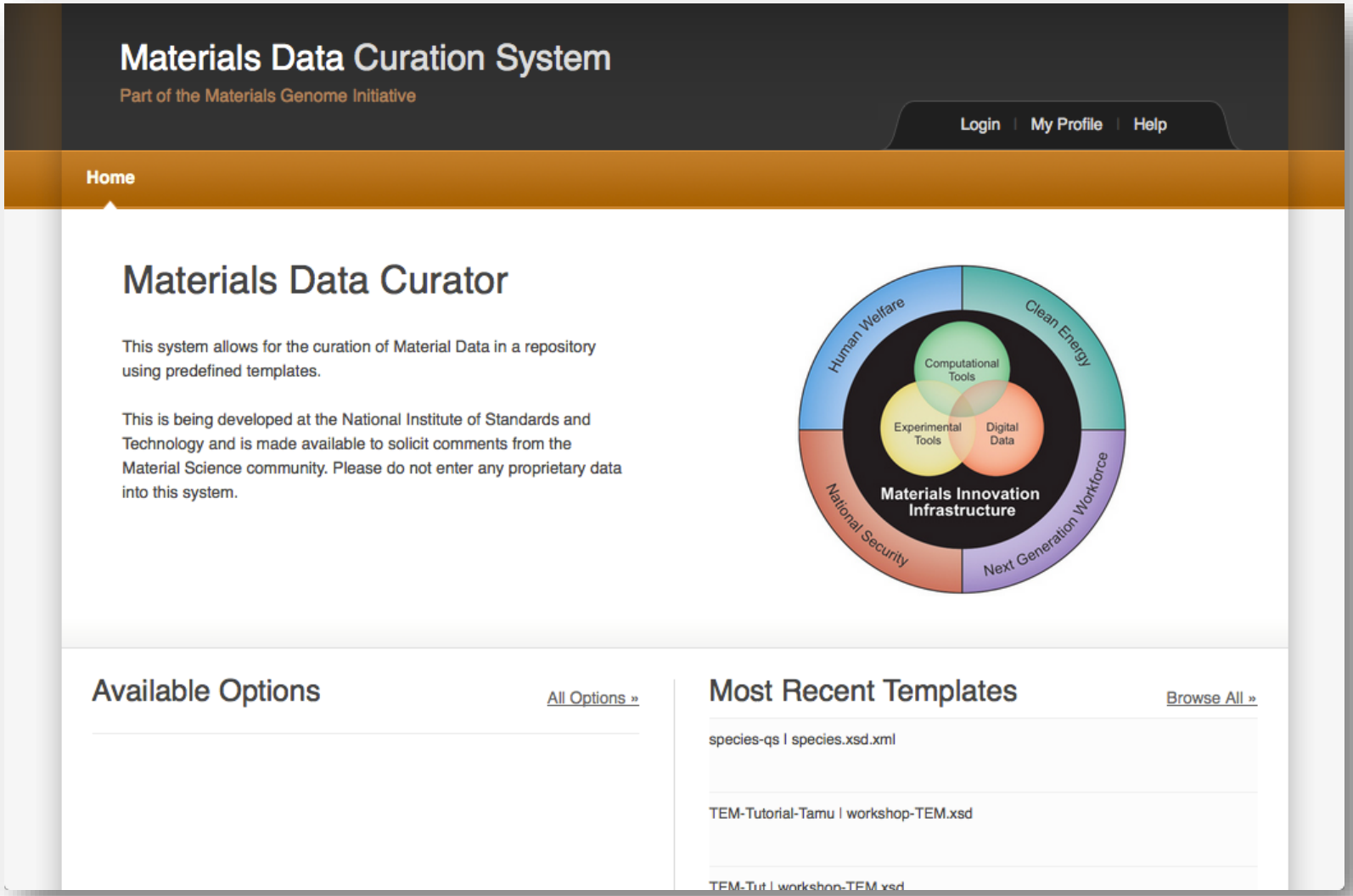



Credit: Rachel Devers, SURF program

# Laboratory Information Management Systems

- Capture instrument metadata at the source
  - Metadata extractors
  - Often must reverse engineer proprietary binary formats
- Move experiment metadata into database
  - Enable search across many experiments
  - Do not use filenames/file system for metadata storage
- Enable scripted data processing, calibration, feature extraction
- Support data management from acquisition to publication; improve reproducibility

# Configurable Data Curation System



# Challenges with Experimental Data

**Undefined Structure**

**Different Formats**

```
SampleIdent CPD RR S
BANK    1       7251                                    0.000     0.000
 1   153 1   161 1                                  139 1   129 1   132
 1   129 1   129 1   151 1                                      146
 1   129 1   134 1   125 1   114 1   129     127 1
```

```
SampleIdent CPD RR Sample 1B        DataFileNam
DiffrType PW3710   GeneratorVoltage 40   TubeCurr
Anode Cu    Alpha1  1.54056    Alpha2  1.54439    Ratio
MonochromatorUsed YES   DivergenceSlit 1      ReceivingSlit 0.3
   5.000   0.020 150.000    MeasureDateTime 20/12/1997   17:18   StepTime   3.00
     184     171     182     184     158     169     182     151     171     136
     171     163     146     158     158     169     182     151     171     136
     156     158     148     153     151     156     139     158     125     163
```

Only an expert human can understand this number.

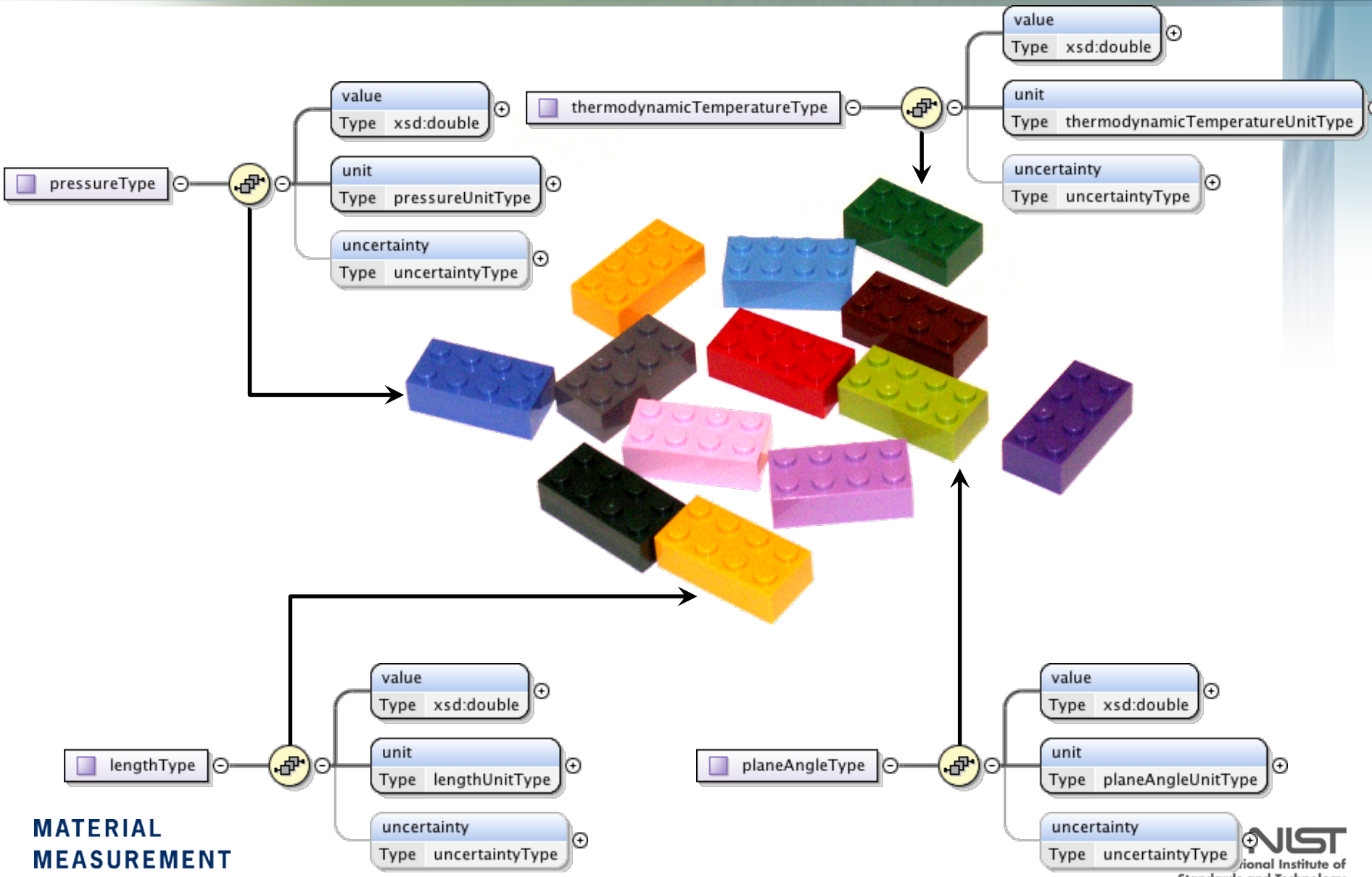
To a computer, this is a meaningless collection of numbers

```
 5.0000  150.0000    0.0200    1.0000
177.   182.   174.   154.   177.   156.   172.   161.   146.   169.
144.   154.   161.   156.   144.   166.   164.   119.   182.   135.
164.   128.   154.   114.   142.   121.   144.   154.   137.   137.
```

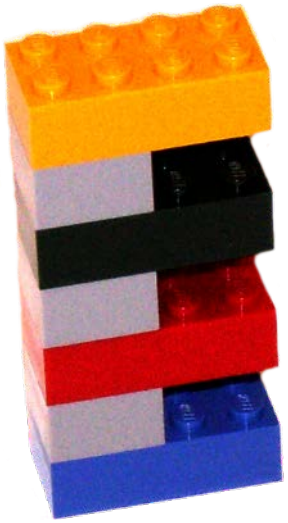
# Structured Data Based on Data Model

```
{"diffractogram": {
    "xray-source": {
        "tube": {"anode-material": "Cu",
            "spectra": {"emission-line": [
                    {"Siegbahn": "Kalpha",
                     "wavelength": {"value": 1.54184,"unit": "angstrom"}},
                    {"Siegbahn": "Kalpha1",
                     "wavelength": {"value": 1.54056,"unit": "angstrom"}},
                    {"Siegbahn": "Kalpha2",
                     "wavelength": {"value": 1.54439,"unit": "angstrom"}}
                    ]}}},
    "pattern-data": {
        "angle-2-theta": {
            "value": [9.3,9.32,9.34, ... 75.16,75.18,75.2],
            "unit": "degree"},
        "intensity": {
            "value": [681.02,687.34,703.49, ... 127.52,124.29,118.32],
            "unit": "arbitrary"}}}}
```

# Modularity: Foundational Types



# Data Models Re-Use Components



# Other NIST Activities

- Office of Science and Technology Policy, National Science and Technology Council
  - Committee on Science
    - Subcommittee on Open Science
    - Succeeds Interagency Working Group on Open Science
  - Committee on Technology
    - Materials Genome Initiative
    - Machine Learning / Artificial Intelligence
  - Committee on Science and Technology Enterprise
    - Subcommittee on Scientific Collections
    - Subcommittee on R&D Infrastructure
  - Joint Committee (CoS and CSTE)
    - Subcommittee on Research Integrity
  - Interagency Working Group on Bio-Data Sharing

# NIST Research Data Framework (RDaF)

- Pilot program to provide an overall guide to the actors and stakeholders in the research data space
- NIST Cybersecurity Framework is the model
- Community consensus, not NIST imposition
- Steering Committee met yesterday, anticipate kick-off event this fall

# Summary

- Core activities:  data management, informatics and machine learning, data sharing and dissemination, community engagement
- NIST aims to be an exemplar organization in the stewardship of research data
- But there are lots of players in the research data arena!!  To what extent can we coordinate and synthesize?

robert.hanisch@nist.gov