THE MATERIALS DATA CURATION SYSTEM

A PRACTICAL APPROACH TO THE LONG TAIL OF MATERIALS DATA AND METADATA

National Institute of Standards and Technology

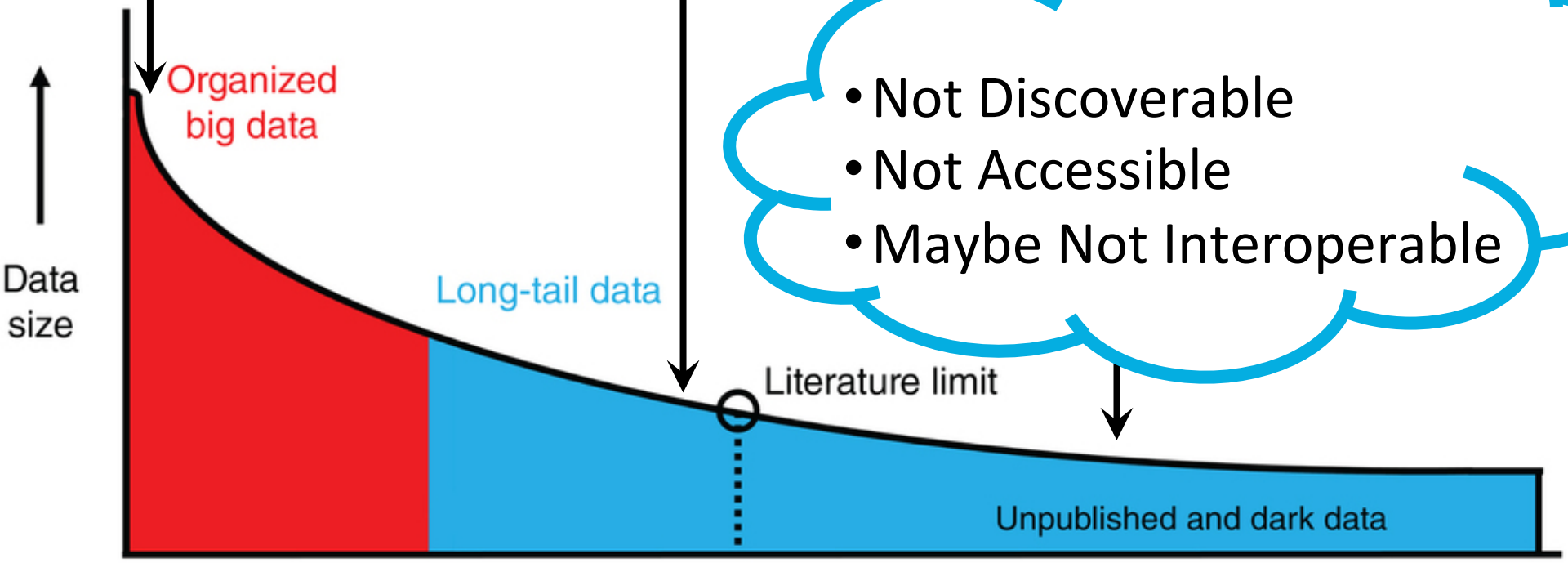
Alden Dima, Sunil Bhaskarla, Chandler Becker, Mary Brady, Carelyn Campbell, Philippe Dessauw, Lucas Hale, Robert Hanisch, Ursula Kattner, Kenneth Kroenlein, Chris Muzny, Marcus Newrock, Adele Peskin, Raymond Plante, Sheng-Yen Li, Pierre-François Rigodiat, Guillaume Sousa Amaral, **Zachary Trautt**, Xavier Schmitt, James Warren, Sharief Youssef

Materials Data

- Discoverable
- Accessible
- Interoperable

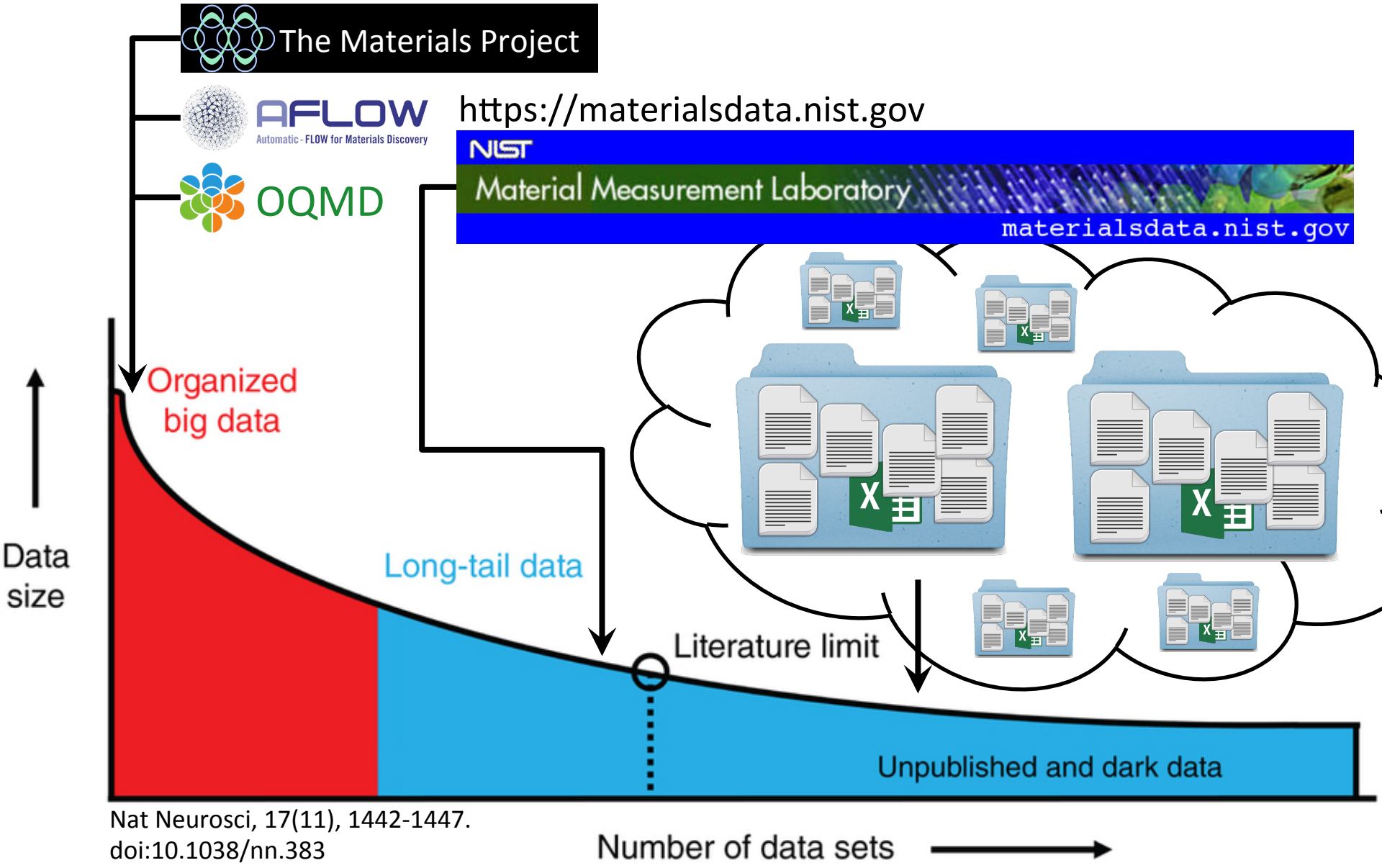
- Discoverable
- Accessible
- Maybe Not Interoperable

- Not Discoverable
- Not Accessible
- Maybe Not Interoperable



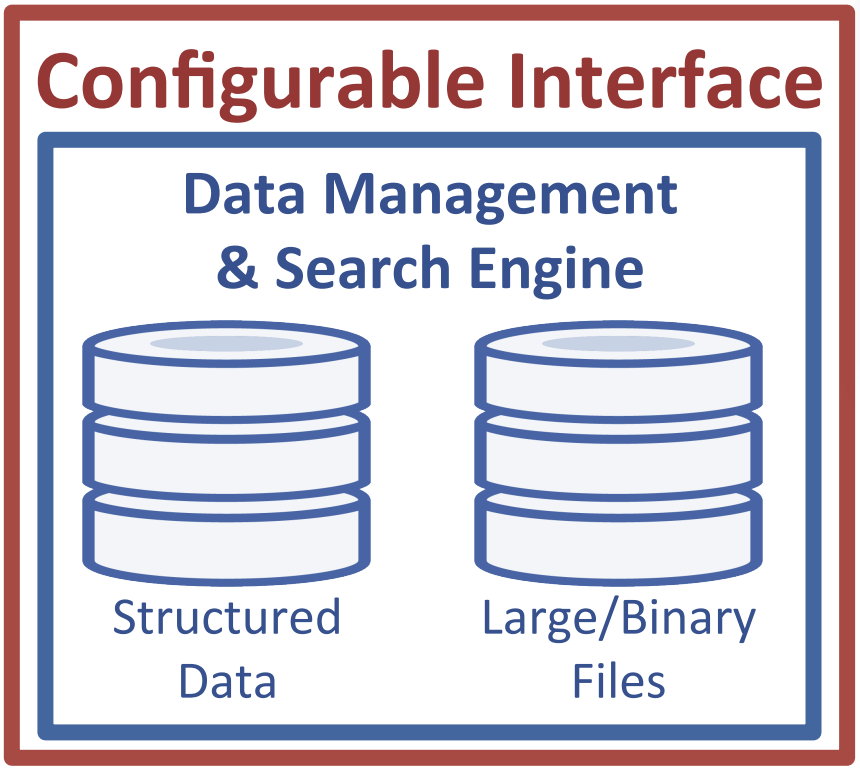
Nat Neurosci, 17(11), 1442-1447.
doi:10.1038/nn.383

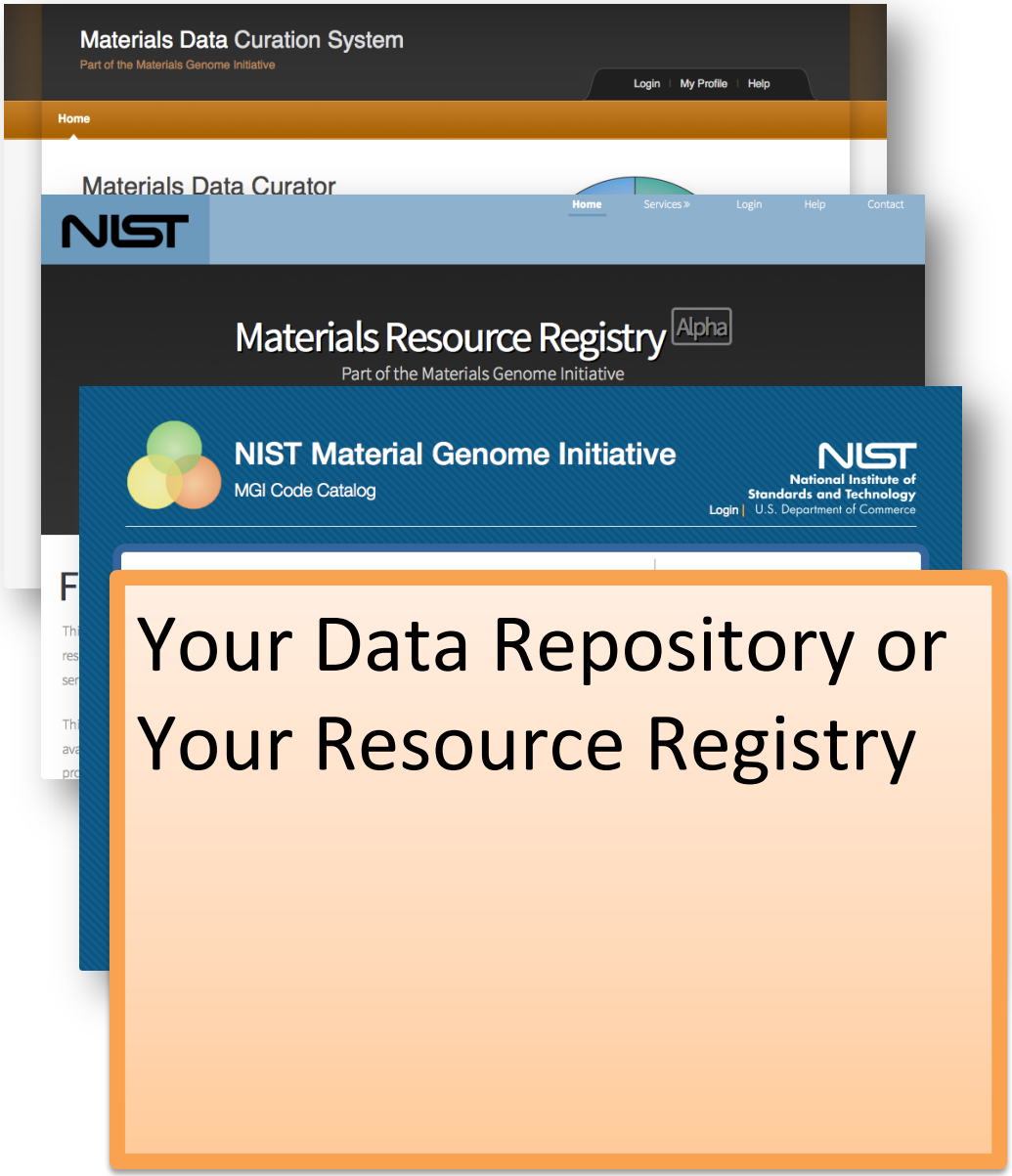
Materials Data



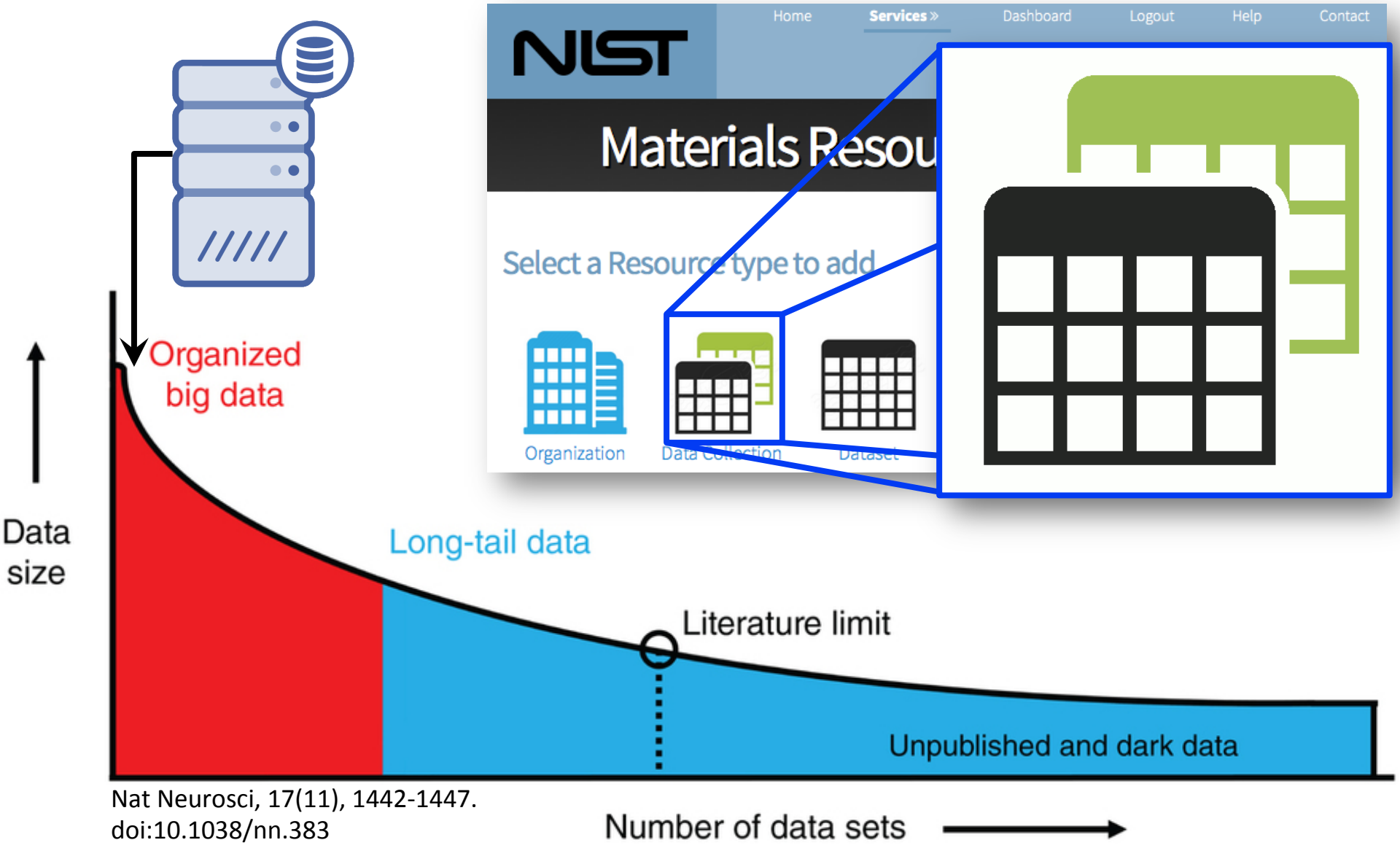
Nat Neurosci, 17(11), 1442-1447.
doi:10.1038/nn.383

Materials Data Curation System



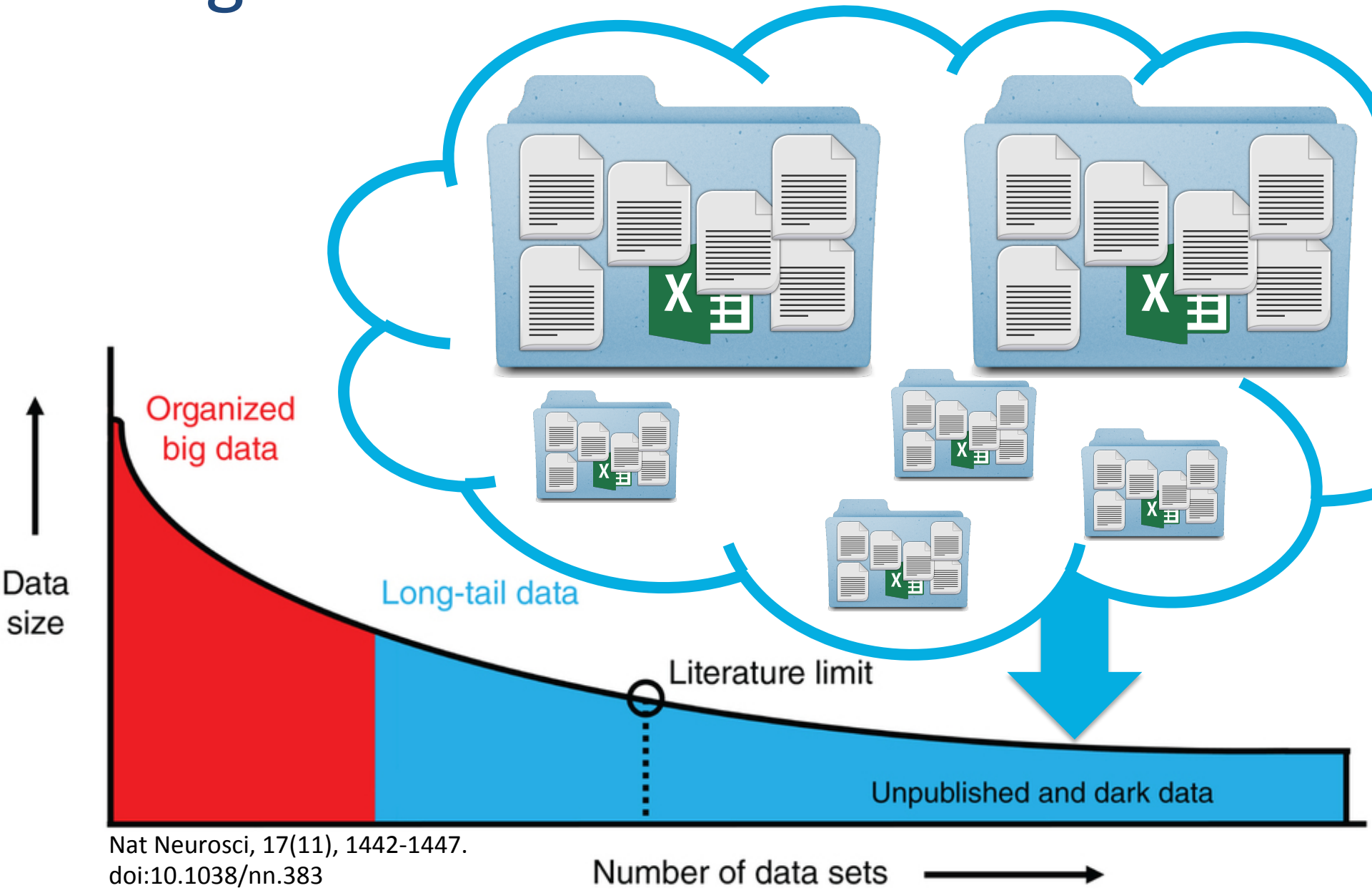


Head Data



Nat Neurosci, 17(11), 1442-1447.
doi:10.1038/nn.383

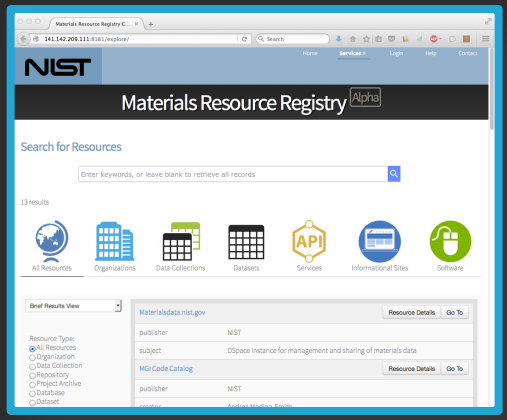
Long Tail



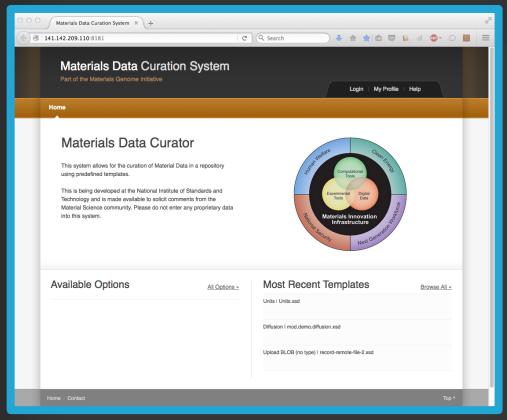
Nat Neurosci, 17(11), 1442-1447.
doi:10.1038/nn.383

NIST MGI APPROACH TO LONG TAIL DATA

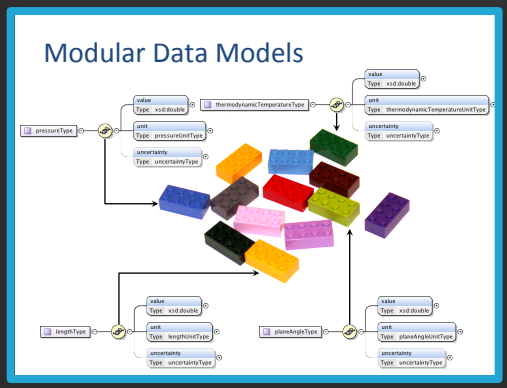
- Discoverable
  (via the Registry)
  https://mgi.nist.gov/Zkp



- Accessible
  (via the Curator)
  https://mgi.nist.gov/ZkS

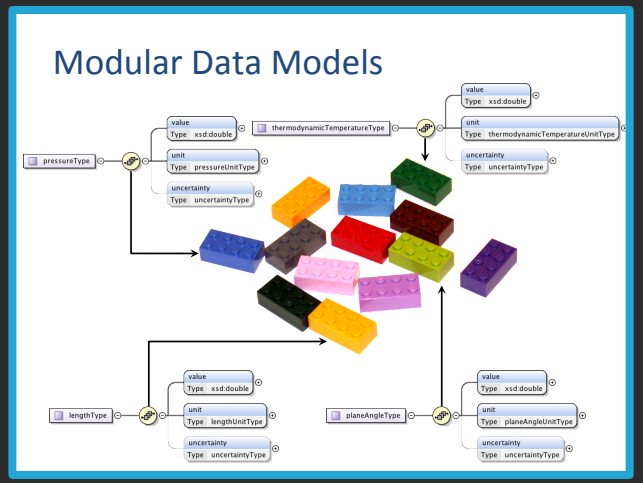


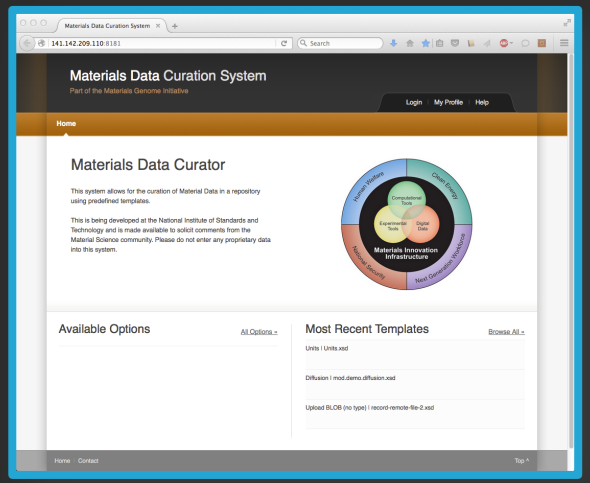
- Interoperable
  (via Community Data Standards)
  https://mgi.nist.gov/ZkG

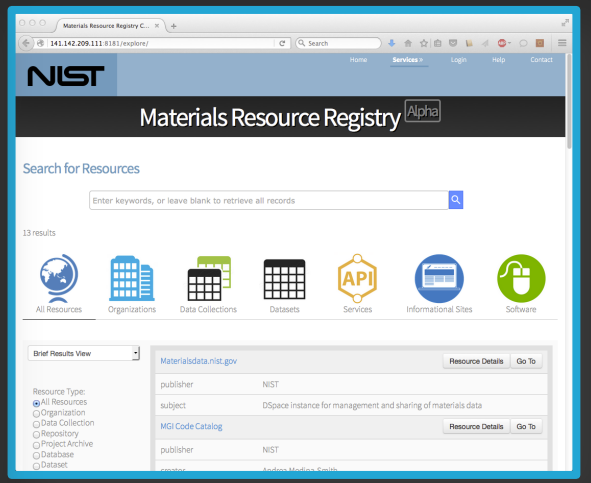


TAKE HOME MESSAGE

- Come Talk to Us
- Organize a Domain Workshops:
  1. Develop Modular Community Data Standards
  2. Hack-A-Thon Style Curator Adoption
  3. Hack-A-Thon Style Registry Adoption





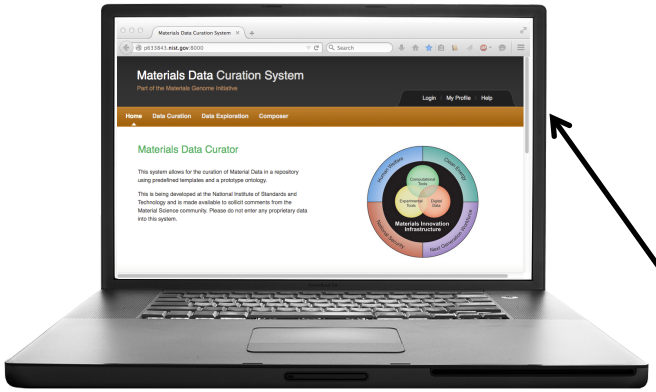


DESIGN OF THE CURATOR

REQUIREMENT #1

Materials researchers require a platform for interoperable exchange of materials data and metadata, which supports an approach of modular community-developed data standards.

Overall Design



Digital Data
& Metadata
(any format)

Data Analysis
Infrastructure

GUI

Web
Framework

Exporter

REST API

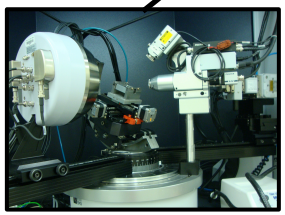
Data Provider

Harvester

Simulation

User Scripts

Data Management
& Search Engine



Measurement

Data
Metadata

Images
Large Files
BLOBs

Database

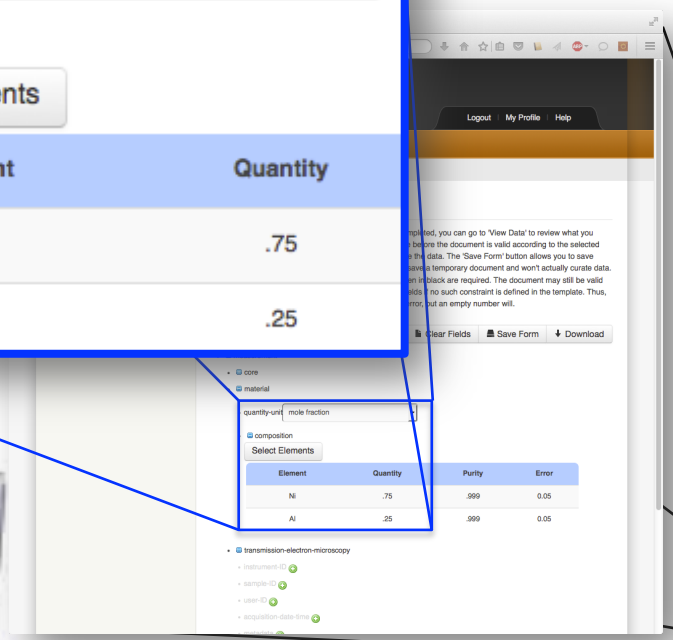
Large Dataset
Repository

Data Provider

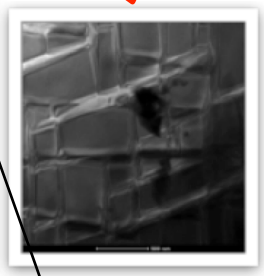
Harvester

Web Interface

Summer Undergraduate Researcher

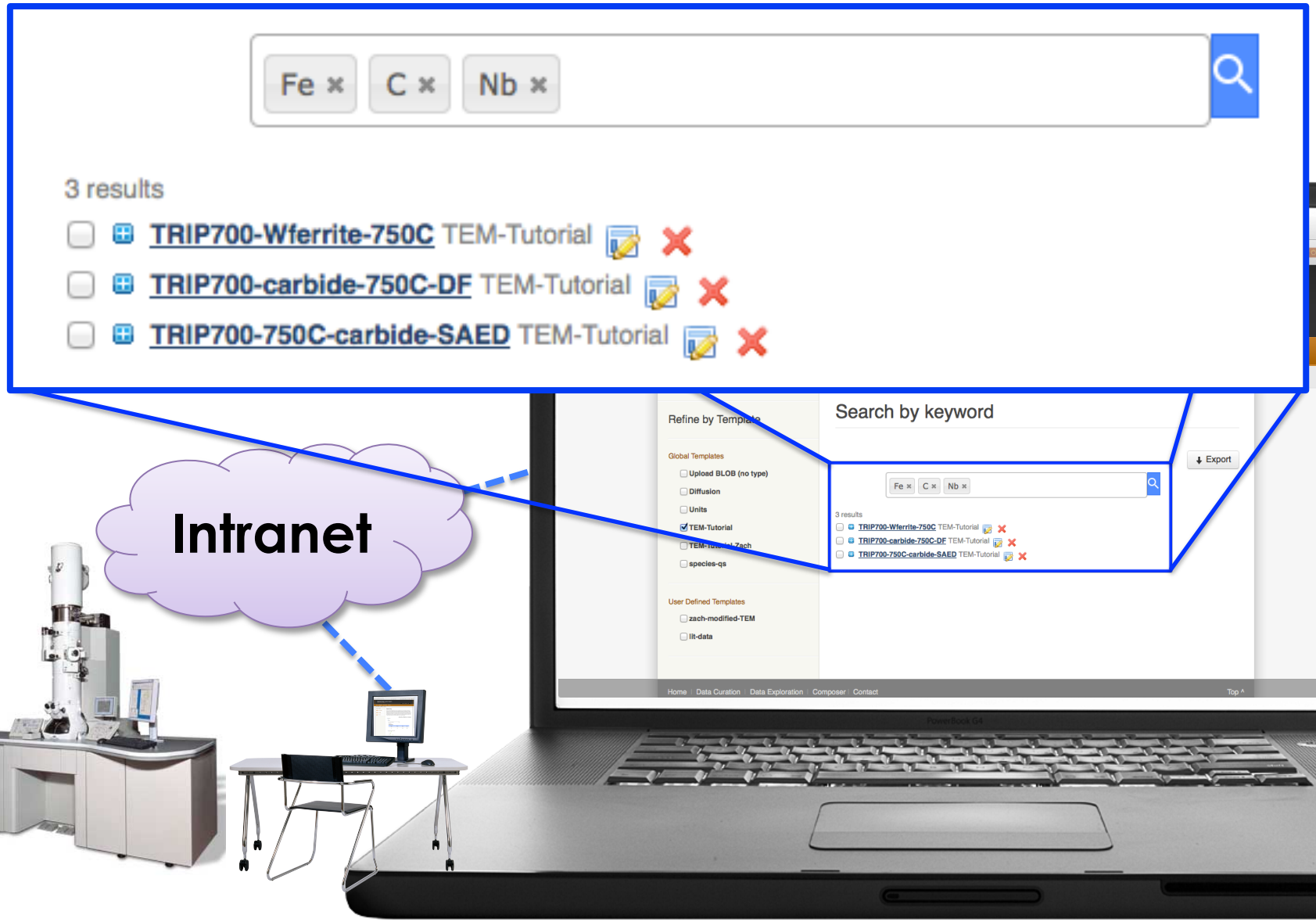


Sample 3?

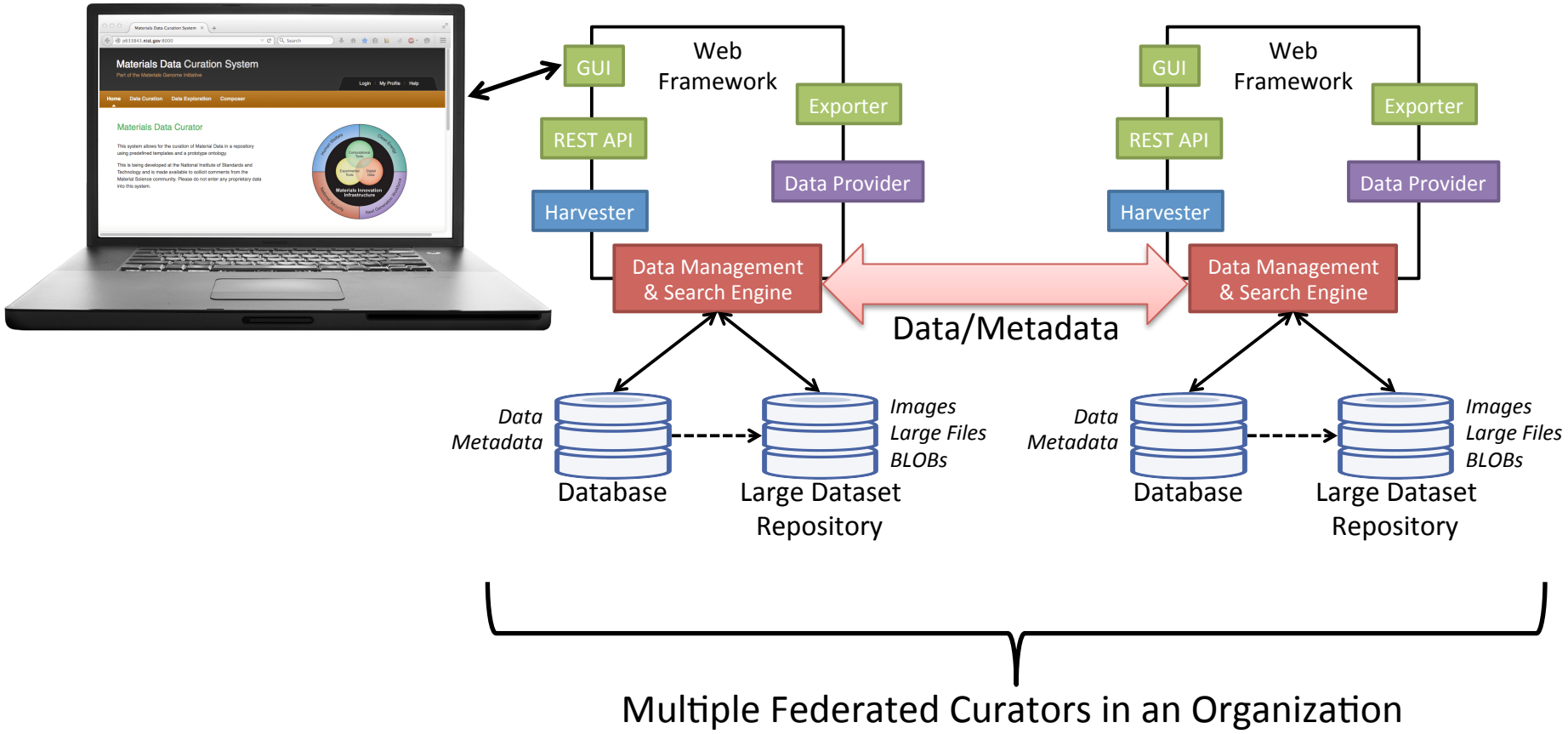


Sample-3.jpg

Web Interface

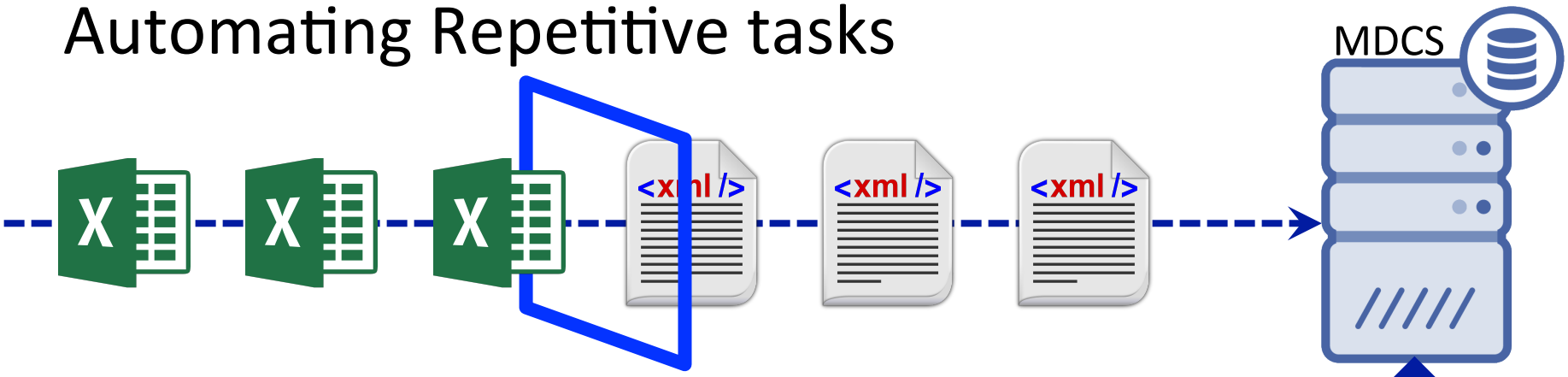


Federated Search

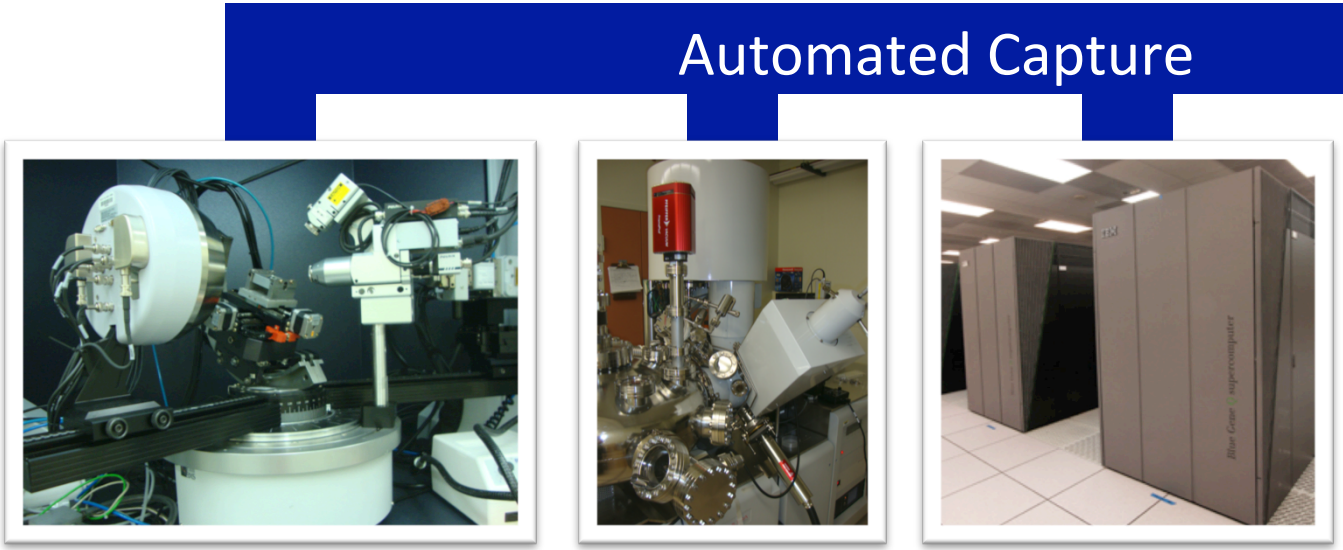


Application Programming Interface

Automating Repetitive tasks



Integration with equipment

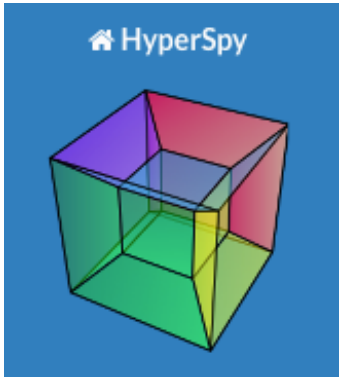


Complement Community Standards

Community could develop a schema and deploy:

- Repository for microstructure HDF5 Files

- Repository for electron microscopy HDF5 Files

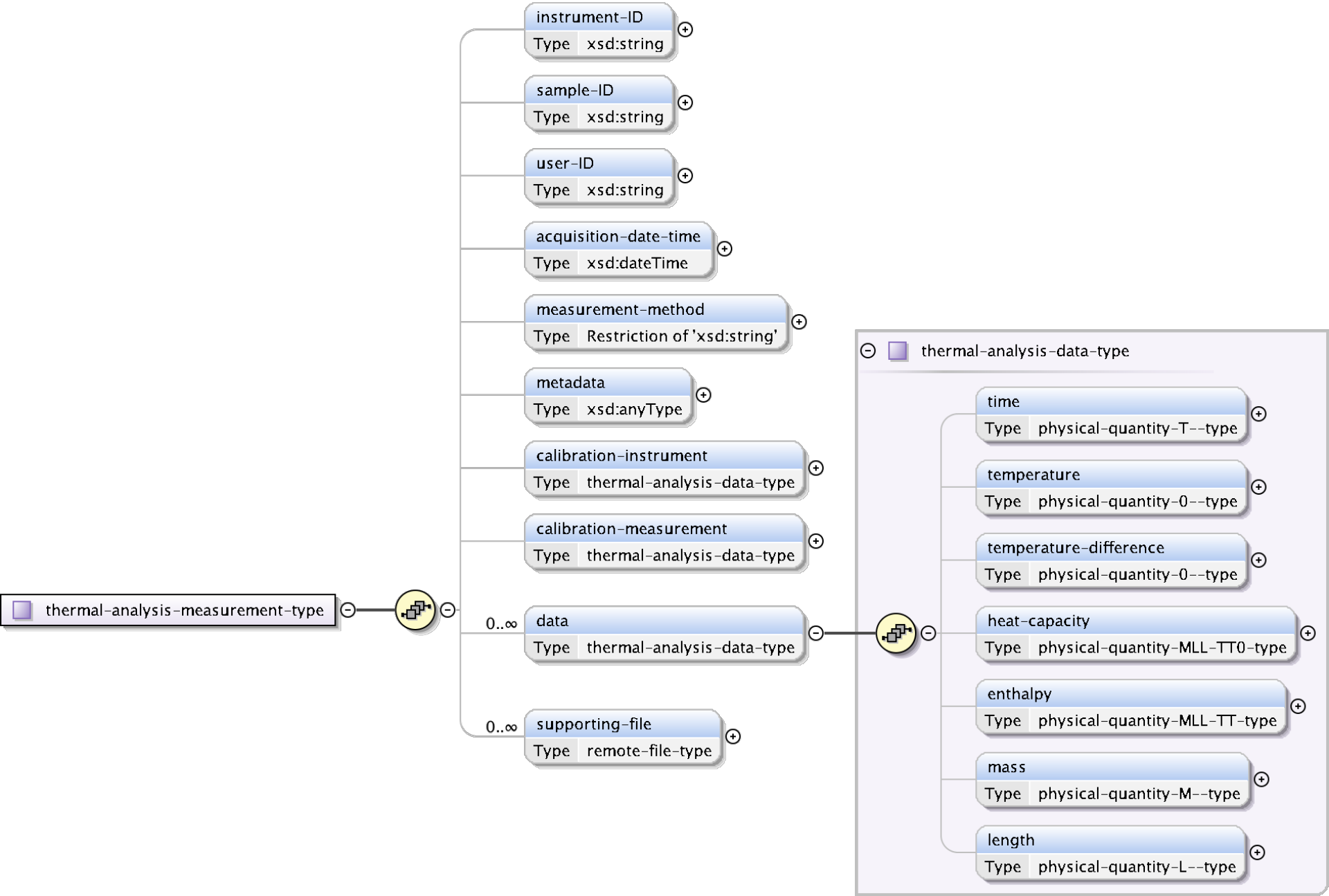


- Repository for diffraction files HDF5 Files

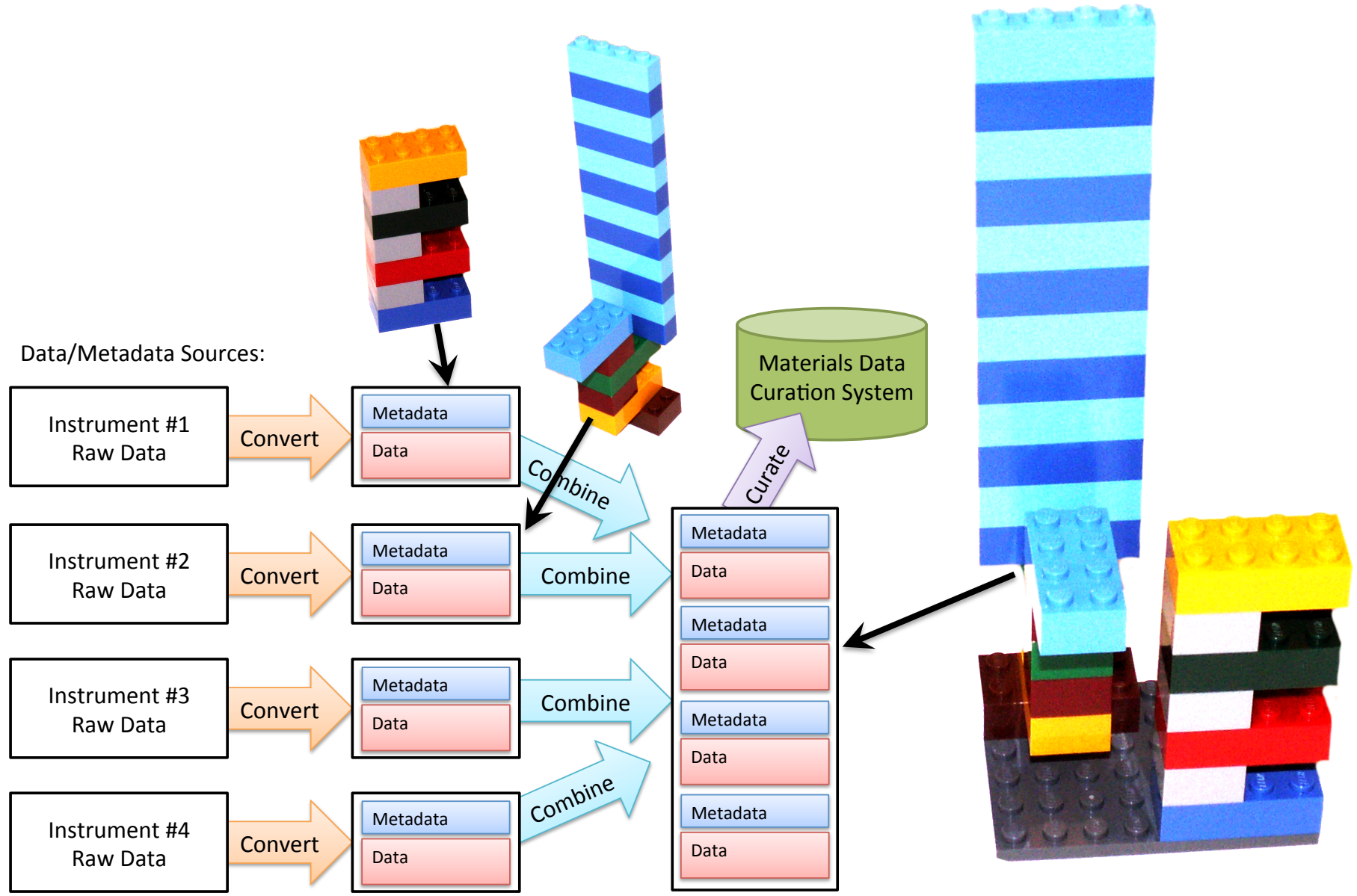
NeXus

Enable Discovery and Access of Data in Existing Standardized Formats

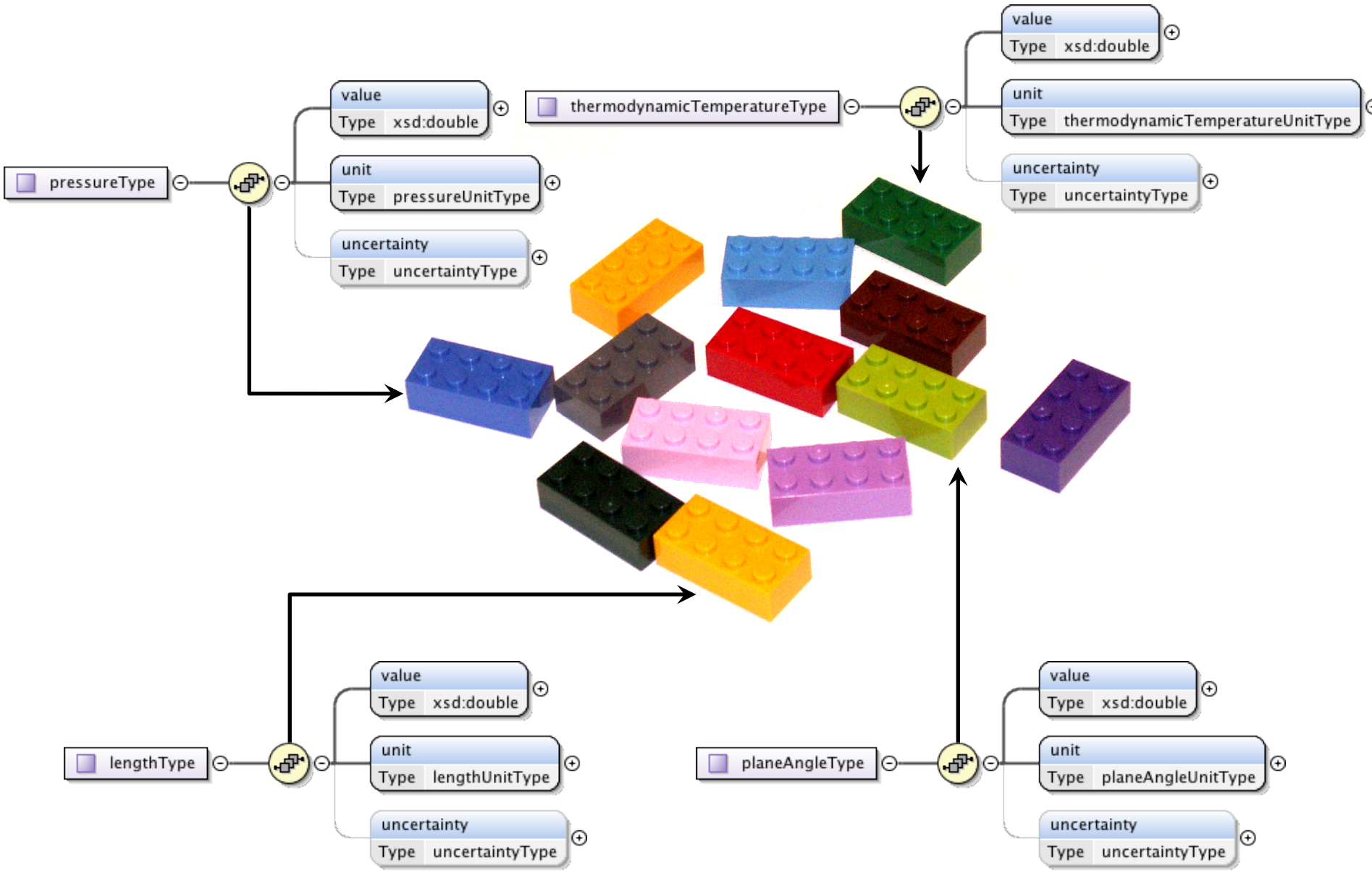
Development of New Community Standards



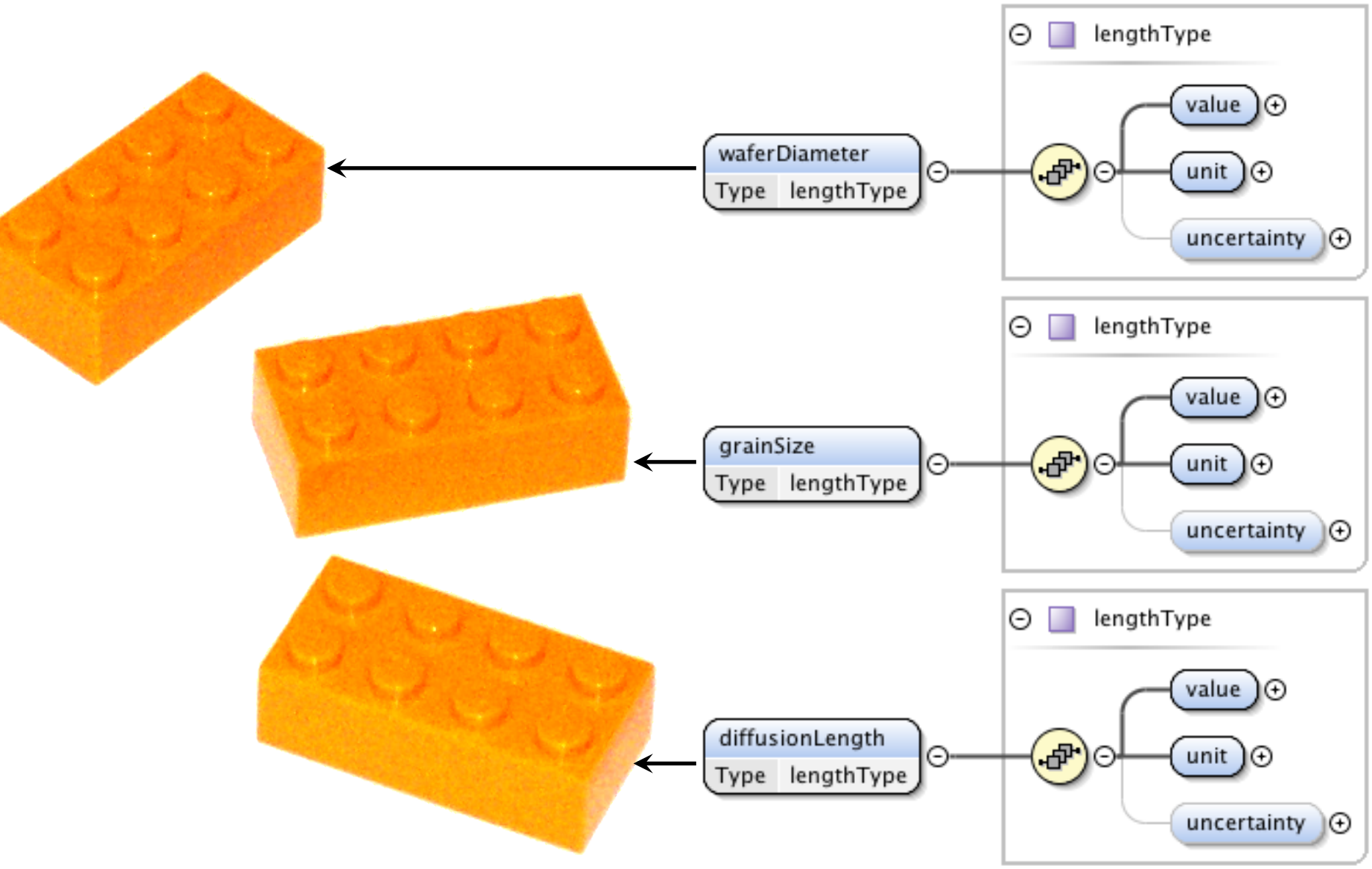
Modular Data Models



Modularity: Foundational Types



Modularity: Foundational Types



Coming Soon: Schema Repository and Registry

- Discover existing schemas
- Register and describe currently accessible schemas
- Upload, register, and describe new schemas
- Scope?
  - XSD, JSON-LD, JSON Schema
  - How to do this with HDF5 formats?
  - Capture and link to vocabularies and ontologies?

Free and Open-Source Software

README.md

Materials Data Curation System

The NIST Materials Data Curation System (MDCS) provides a means for capturing, sharing, and transforming materials data into a structured format that is XML based amenable to transformation to other formats. The data are organized using user-selected templates encoded in XML Schema. These templates are used to create data entry forms. The documents are saved in a non-relational (NoSQL) database, namely MongoDB. The data can be searched and retrieved via several means: by a

https://github.com/usnistgov/MDCS

Installation

To install and run the MDCS on your machine:

- Pick the instruction notes for your operating system, inside the docs folder,
- Follow the installation instructions,
- Make sure that the python packages and software that you are installing, match the versions listed in the document Required Python Packages and Required Software,
- Recommended Internet Browser: Mozilla Firefox.

REQUIREMENT #2

Materials researchers need a decentralized infrastructure to enable finding and sharing of materials resources.

Materials Resource Registry Alpha

Part of the Materials Genome Initiative

SEARCH FOR RESOURCES ADD YOUR RESOURCE

Find Materials Data

This system allows for the registration of materials resources, bridging the gap between existing resources and the end users. The 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 available to solicit comments from the Material Science community. Please do not enter any proprietary data into this system.

Home Page

Services

Search for resources

Add your resource

Login

Help

Contact

Materials Resource Registry Alpha

Search for Resources

Enter keywords, or leave blank to retrieve all records

13 results

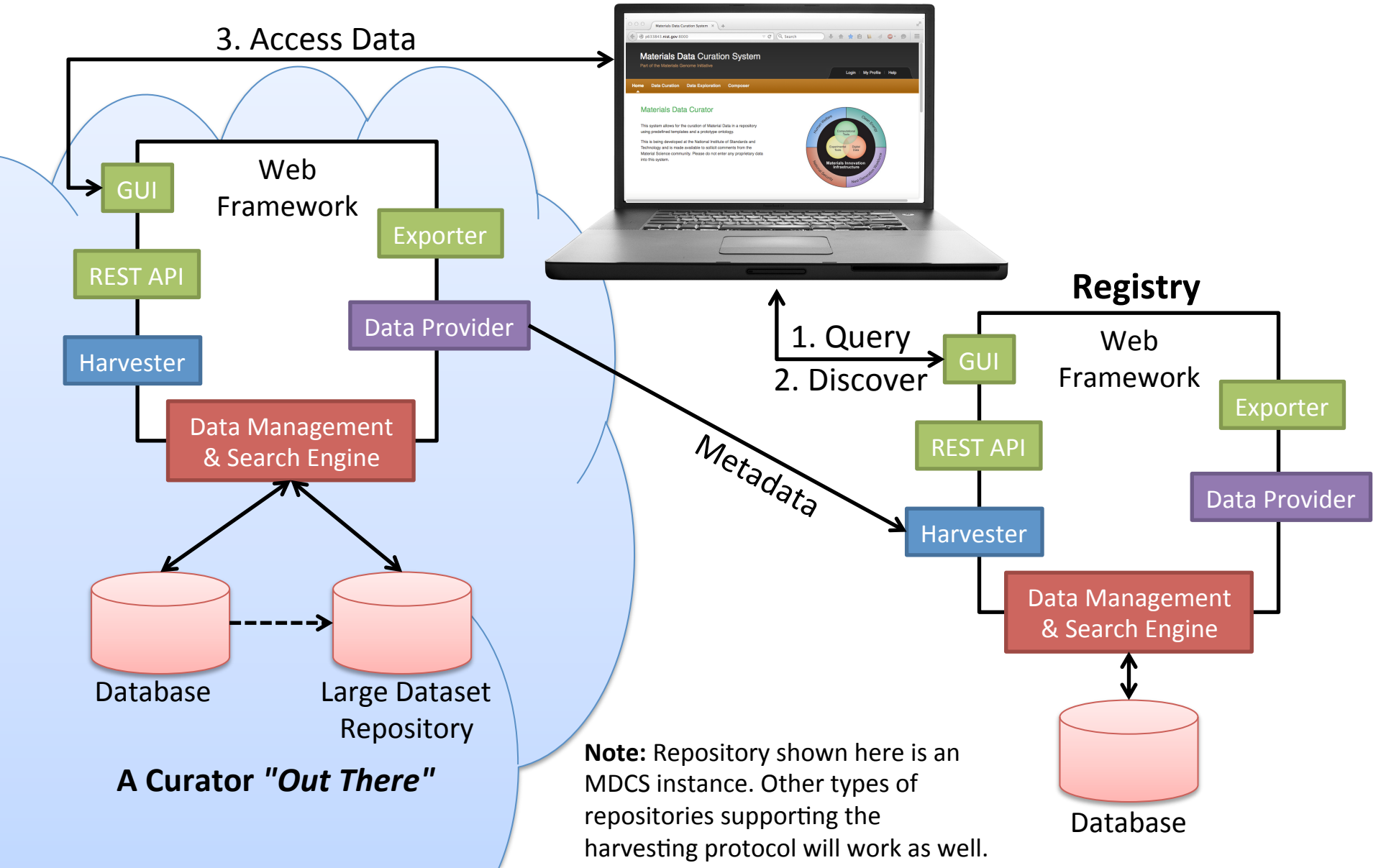
All Resources | Organizations | Data Collections | Datasets | Services | Informational Sites | Software

Brief Results View

Resource Type:
- (•) All Resources
- ( ) Organization
- ( ) Data Collection
- ( ) Repository
- ( ) Project Archive
- ( ) Database
- ( ) Dataset
- ( ) Service

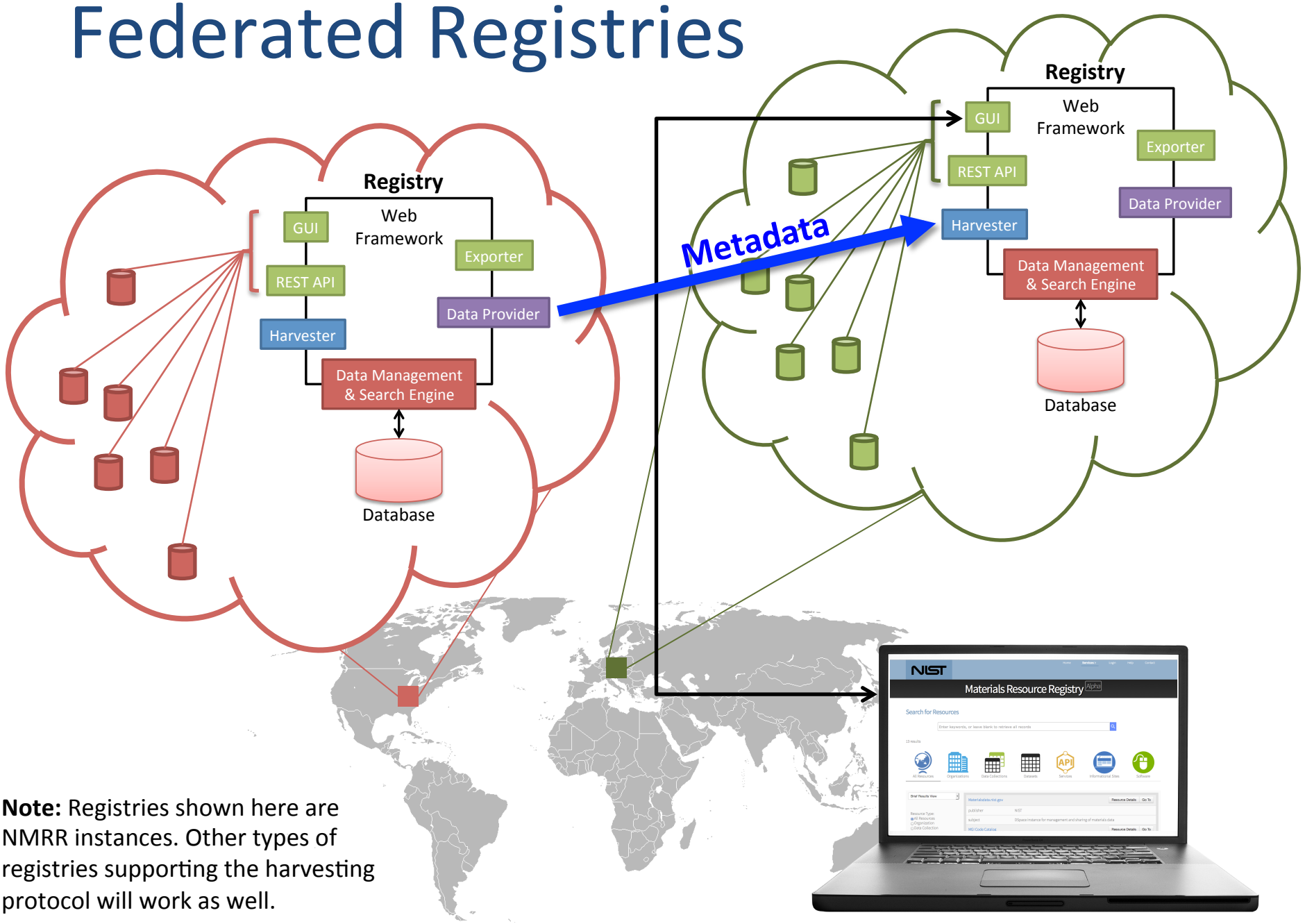
Materialsdata.nist.gov		Resource Details / Go To
publisher	NIST	
subject	DSpace instance for management and sharing of materials data	
MGI Code Catalog		Resource Details / Go To
publisher	NIST	
creator	Andrea Medina-Smith	

Registry-Enabled Data Discovery



Note: Repository shown here is an MDCS instance. Other types of repositories supporting the harvesting protocol will work as well.

Federated Registries



Note: Registries shown here are NMRR instances. Other types of registries supporting the harvesting protocol will work as well.

Free and Open-Source Software

README.md

Materials Resource Registry

Access to scientific data and resources across the materials community is limited and fragmented. Local resource providers are frequently populated with new resources but the larger community is often unaware that these potentially interesting resources are available. The NIST Materials Resource Registry (MRR) bridges the gap between existing resources and the end users by registering the resources and their metadata for search and discovery.

https://github.com/usnistgov/MaterialsResourceRegistry

The MRR software was developed by the National Institute of Standards and Technology (NIST).

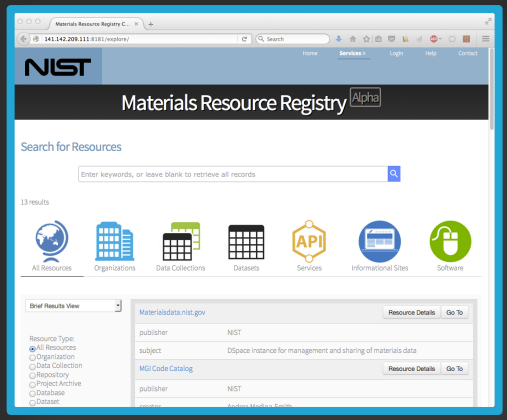
Installation

To install and run the Registry on your machine:

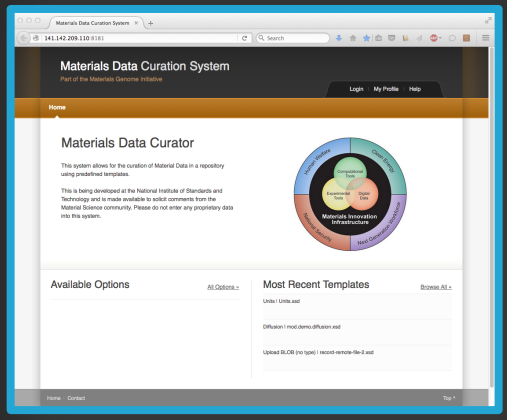
- Pick the instruction notes for your operating system, inside the docs folder,
- Follow the installation instructions,
- Make sure that the python packages and software that you are installing, match the versions listed in the document Required Python Packages and Required Software,
- Recommended Internet Browser: Google Chrome/Safari.

NIST MGI APPROACH TO LONG TAIL DATA

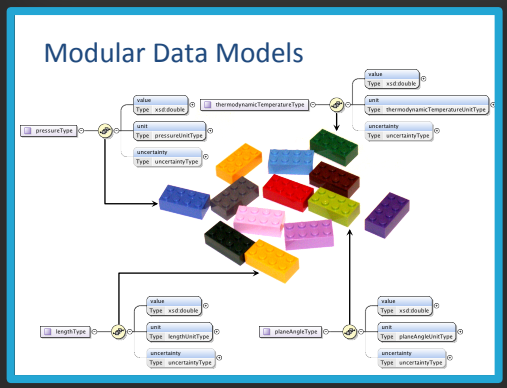
- Discoverable
  (via the Registry)
  https://mgi.nist.gov/Zkp



- Accessible
  (via the Curator)
  https://mgi.nist.gov/ZkS

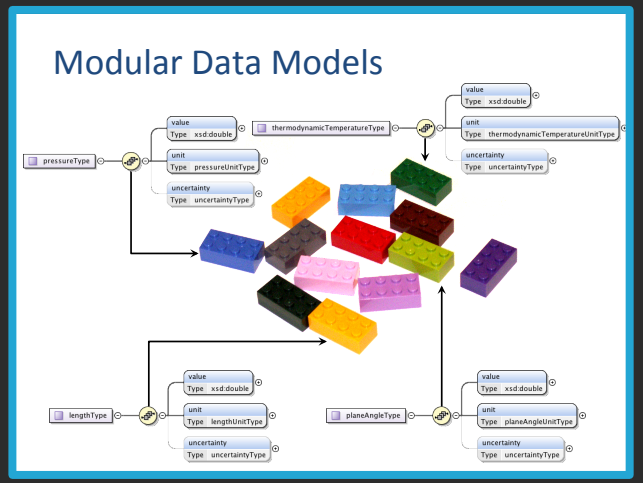


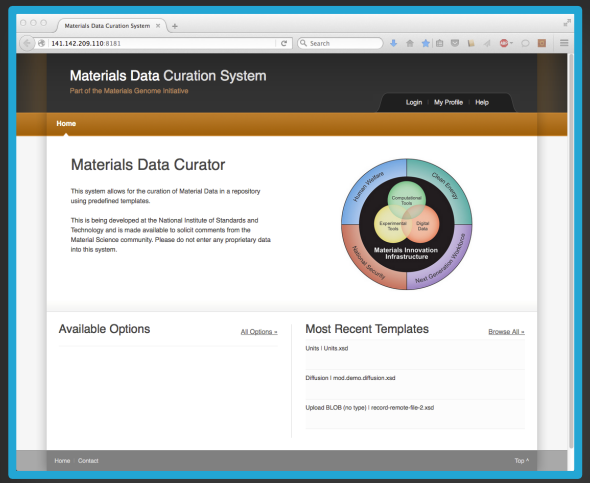
- Interoperable
  (via Community Data Standards)
  https://mgi.nist.gov/ZkG

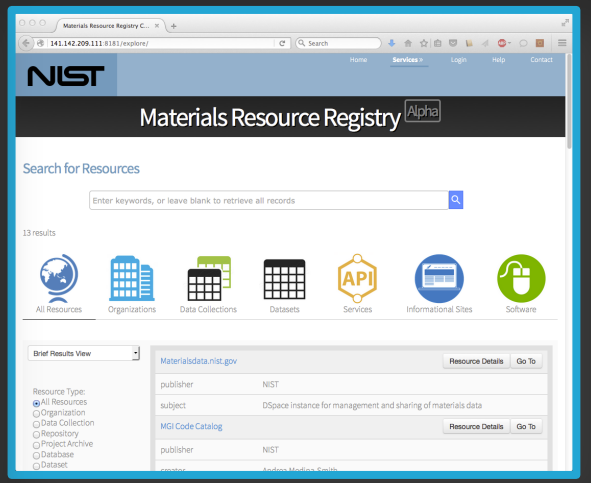


TAKE HOME MESSAGE

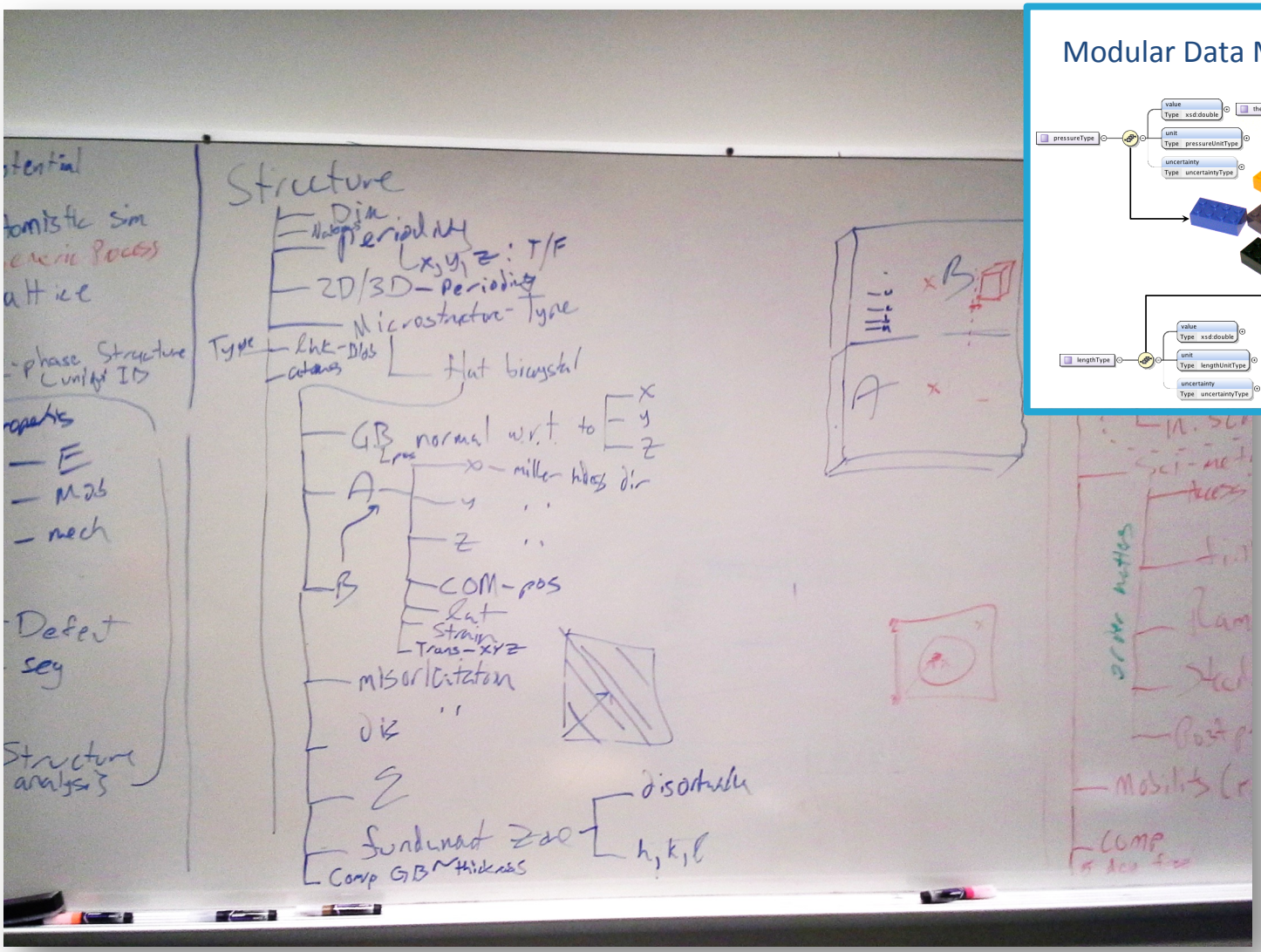
- Come Talk to Us
- Organize a Domain Workshops:
  1. Develop Modular Community Data Standards
  2. Hack-A-Thon Style Curator Adoption
  3. Hack-A-Thon Style Registry Adoption

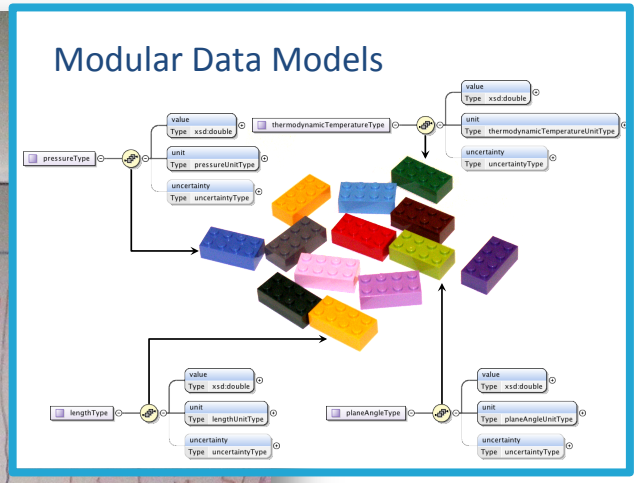




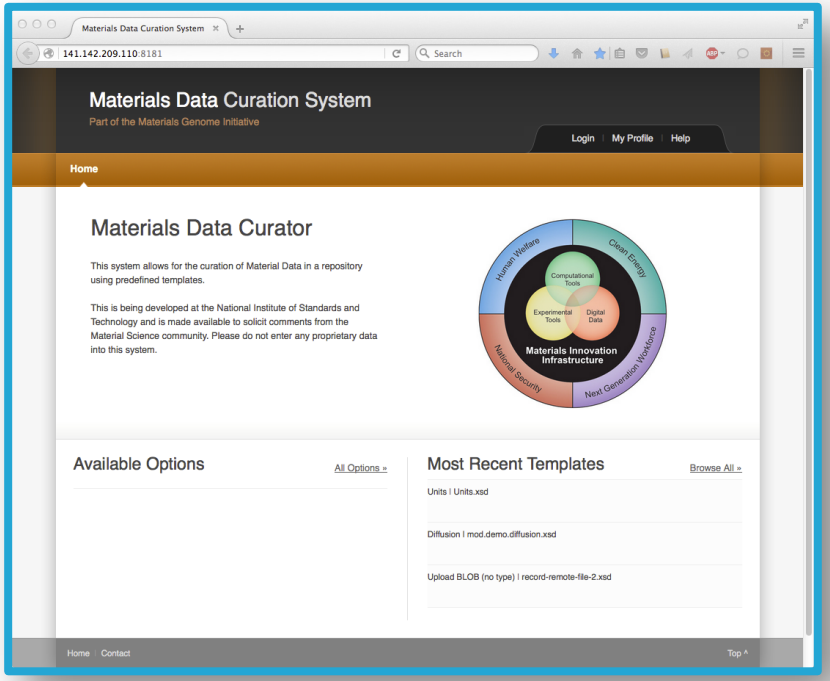


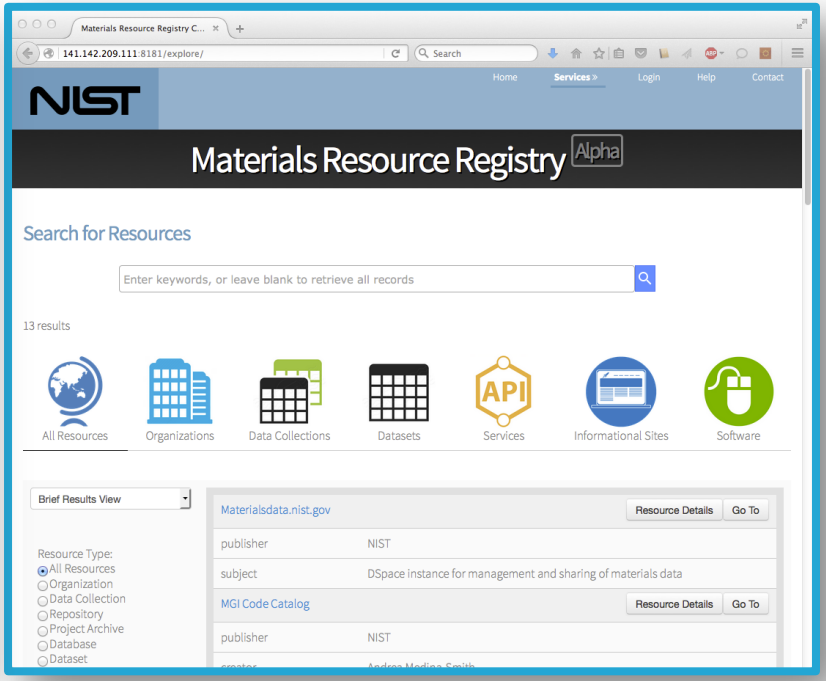
Workshop Part 1: Data Models





Workshop Part 2: Integration





THANK YOU

Disclaimer: Certain commercial equipment or software are identified in this presentation to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.