

How the Virtual Observatory Has Influenced Data Discovery, Access, and Re-Use in Other Disciplines

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

Tuesday November 21, 2017



Common Goals

Discover



- Materials Resource Registry (data, code)
- International Metrology Resource Registry
- NIST Enterprise Data Inventory
- data.gov
- NIST Public Data Repository and Search Portal

- Standard Reference Data
- Materials Data Repository
- Materials Data Facility
- Persistent identifiers (DOIs, handles)



Access

Interoperate

- Materials Data Curator
- Data type registry
- Schema repository
- Lab info mgmt systems



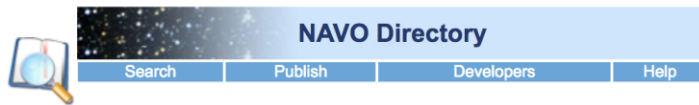
Discover, Access, Interoperate

2

- Why?
 - Support FAIR* principles: Findable, Accessible, Interoperable, Re-usable
 - Assure maximum return on national investment in basic research
 - Demonstrate best practices
 - Address reproducibility “crisis”
- US OMB, OSTP directives; FASTR legislation
- But, Astronomy was here ~20 years ago!

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

Registries



Find Astronomical Data Resources

Enter terms in the text box that describe the type of data you are looking for, or use the filters below to browse data by facet. Results will show catalogs and data collections that have these terms as part of their descriptions. Note that the search interface ignores slashes by default, so when searching for full or partial IDs you may want to put the search string in quotes (e.g. "ivo://mast.stsci/ssap/befs").

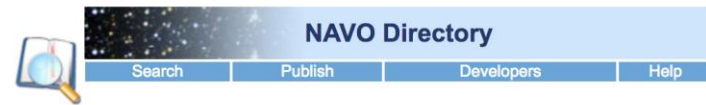
Examples: quasar, AGN, binary stars, Chandra, GALEX, far ultraviolet

Use the [Data Discovery Tool](#) to search and view data for a particular object or position.

Limit your search

- Subject** >
- Publisher** >
- Resource Type** >
- Content Level** >
- Instrument** >
- Facility** >
- Validation Level** >
- Rights** >
- Capability Type** >
- Result Type** >

**MEASUREMENT
LABORATORY**



Find Astronomical Data Resources

Enter terms in the text box that describe the type of data you are looking for, or use the filters below to browse data by facet. Results will show catalogs and data collections that have these terms as part of their descriptions. Note that the search interface ignores slashes by default, so when searching for full or partial IDs you may want to put the search string in quotes (e.g. "ivo://mast.stsci/ssap/befs").

Examples: quasar, AGN, binary stars, Chandra, GALEX, far ultraviolet

Use the [Data Discovery Tool](#) to search and view data for a particular object or position.

Limit your search

- Subject** >
 - qsos 460
 - redshifts 243
 - spectroscopy 123
 - agn 109
 - photometry:wide-band 101
 - galaxies 91
 - photometry 70
 - equivalent_widths 57
 - survey source 38
 - bl_lac_objects 34
 - more »
- Publisher** >
- Resource Type** >
- Content Level** >
- Instrument** >

You searched for:

« Previous | 1 - 10 of 671 | Next »

1. Liverpool Quasar Lens Monitoring

ID: ivo://org.gavo.dc/liverpool/res/rawframes/q
 Title: Liverpool Quasar Lens Monitoring
 Short Name: LivQLensMon
 Date: 08 May 2015 00:28:49
 Publisher: The GAVO DC team
 Description: This collection includes optical monitorings of gravitationally lensed quasars. The frames can be used to make light curves of quasar images and field objects. From quasar light curves, one may measure time delays and flux ratios, analyse variability and chromaticity, etc. These direct analyses/measurements are basic tools for different astrophysical studies, e.g., expansion rate of the Universe, mechanism of intrinsic variability in quasars, accretion disk structure, supermassive black holes, dark halos of galaxies (dust, collapsed dark matter, smoothly distributed dark matter,...)

2. Quasar polarization (Huterer et al. 1999)

standards and technology
U.S. Department of Commerce

Materials Resource Registry

NIST

[Home](#) [Services »](#) [Login](#) [Help](#) [Contact](#)

Materials Resource Registry

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

Search for Resources

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

Enter keywords, or leave blank to retrieve all records



Search criteria used (Clear all):

257 results

Type x

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 microstructural evolution. The phase field method is commonly used for predicting the evolution of microstructures under various conditions."

<http://imrr.bipm.org/>



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

7

Search for Resources



All Resources

Participating
Institutes

Databases



Datasets

Standards
Documents

Data Services



Data Portals

Metrology-related
Software and Tools

Search criteria used (Clear all):

58 results

Resource Type: ✕

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](https://srdata.nist.gov/gateway/) ✓*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*

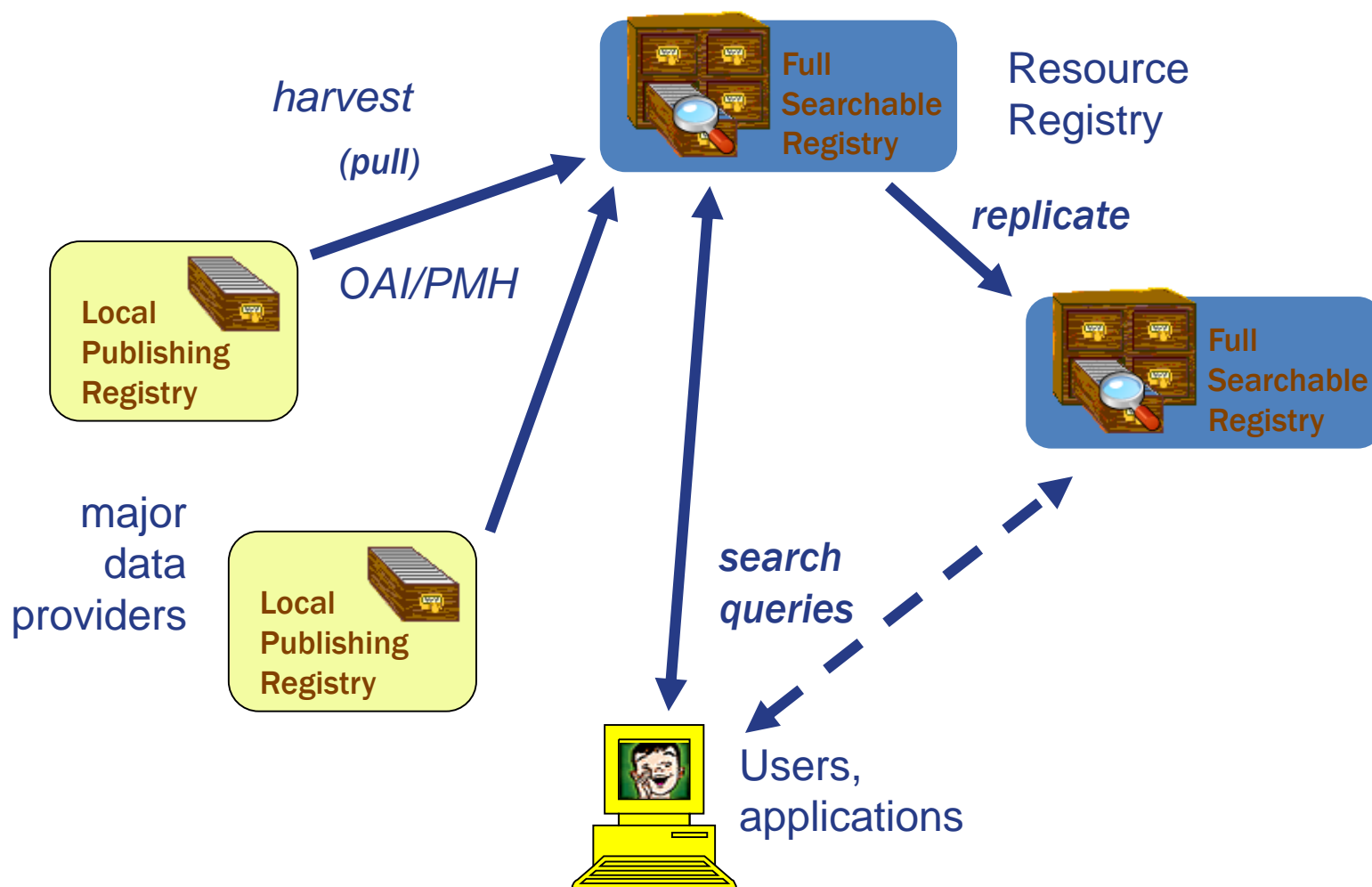
The NIST Data Gateway is an online system designed to provide users with easy access to NIST scientific and technical data. The Gateway provides (1) direct links to free NIST online systems and full text online versions of many Journal of Physical and Chemical Reference Data (JPCRD) articles, (2) links to descriptions and order... [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](https://srdata.nist.gov/gateway/) ✓*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*

Federated Architecture



Data Discovery for Public Research Data

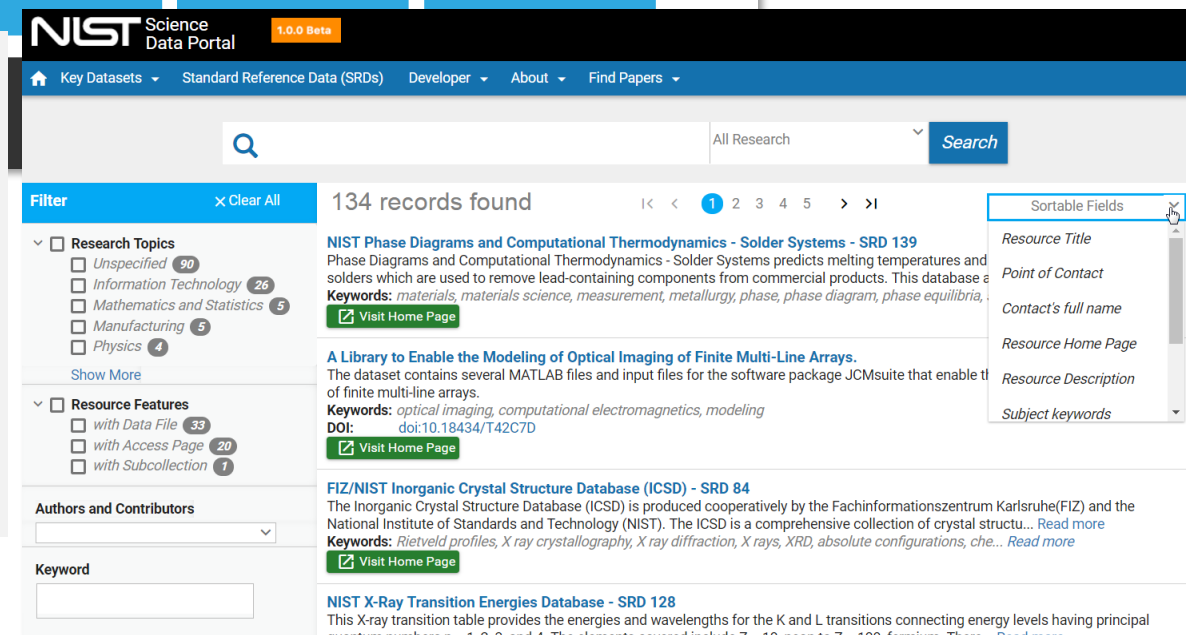
9




Search NIST public data records

- View metadata
- Filter results
- Access data files, metadata
- APIs allow interoperability with client tools
- Records link to Public Data Repository

**MATERIAL
MEASUREMENT
LABORATORY**



NIST Public Data Repository – Basic Landing Page


Public Data Repository
1.0.0 Beta

About | Search

Public Data Resource

NIST Phase Diagrams and Computational Thermodynamics - Solder Systems - SRD 139

Contact: [Ursula Kattner](#) ..
 Identifier: [doi:10.18434/T4759N](#)
 Described in article:
 Last modified: **1990-01-01**

Description

Phase Diagrams and Computational Thermodynamics - Solder Systems predicts melting temperatures and freezing ranges of lead-free solders which are used to remove lead-containing components from commercial products. This database also shows the effects of non-equilibrium solidification. Also included is a collection of calculated binary and ternary systems that are relevant to solders.

Research Topics:
Subject Keywords: materials, materials science, measurement, metallurgy, phase, phase diagram, phase equilibria, science, solder, standards, technical activities, thermodynamic

Files

This resource contains:

Files

solder.tdb

Access

- [Visit Home Page](#)
- [Download all data](#)
- [Export Metadata](#)

Use

- [Cite this resource](#)
- [License Statement](#)

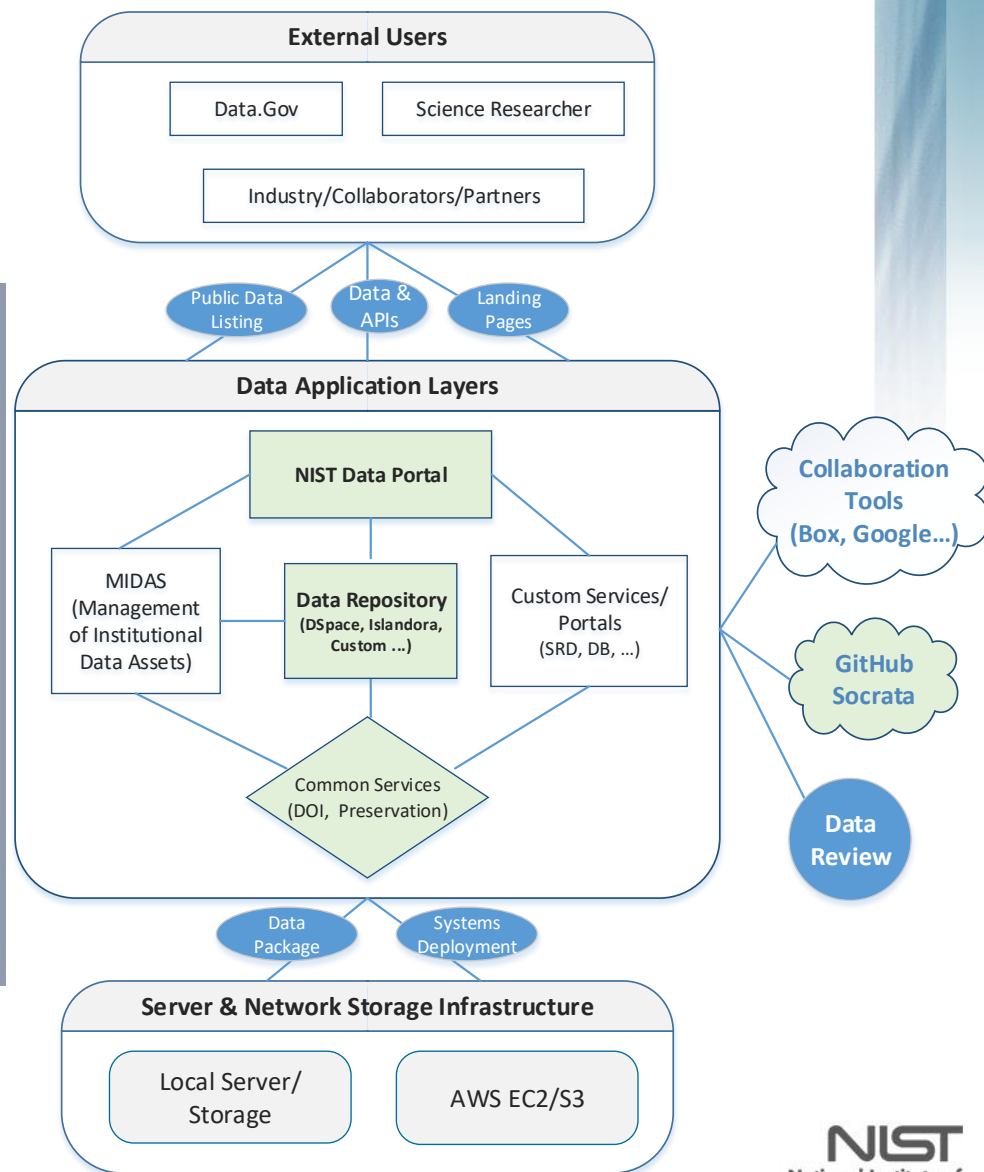
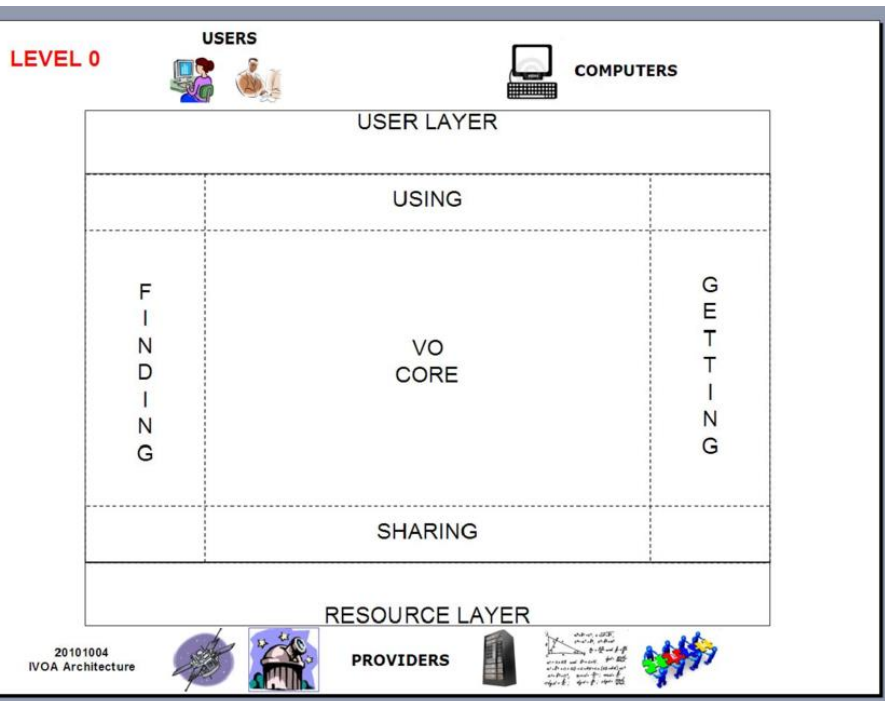
Find

- [Similar Resources](#)
- [Resources by Authors](#)

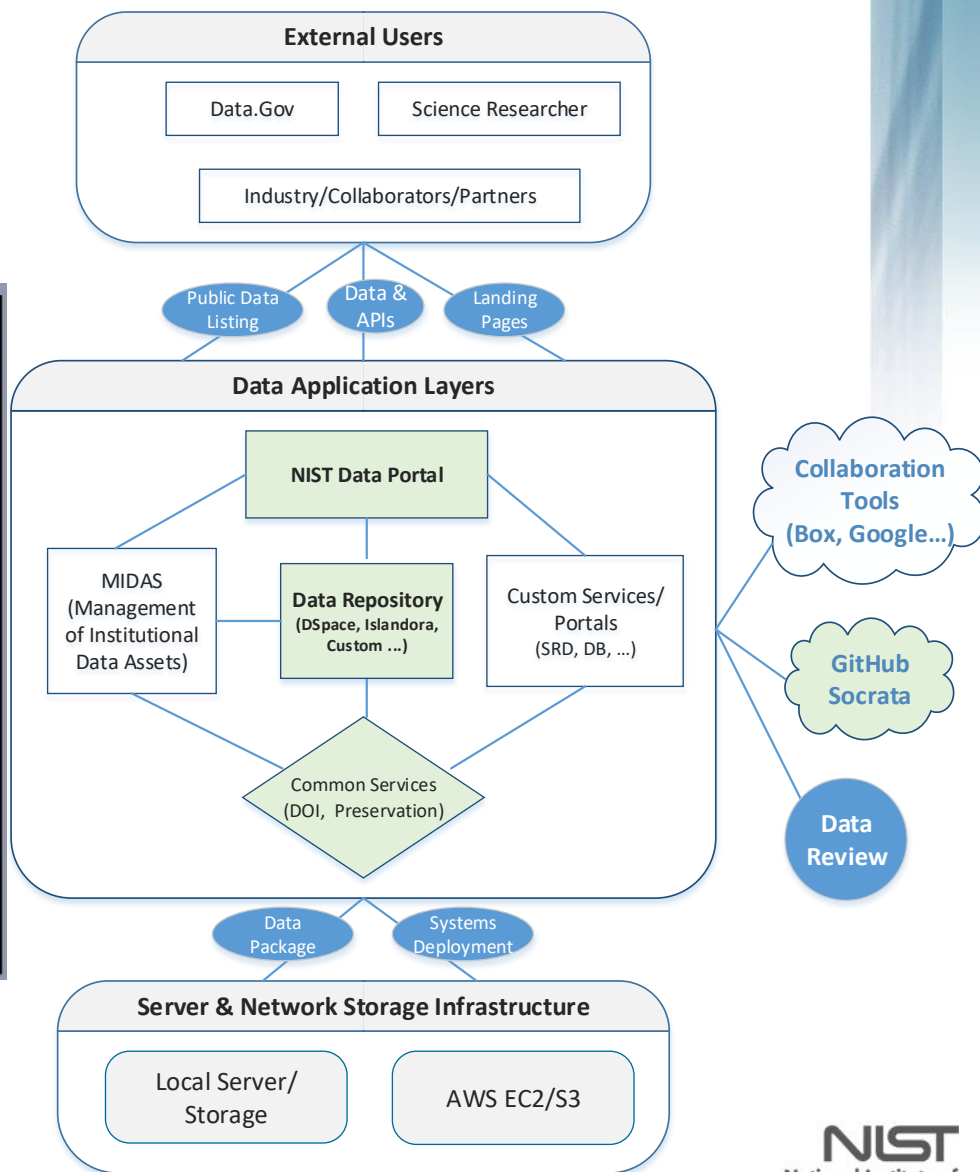
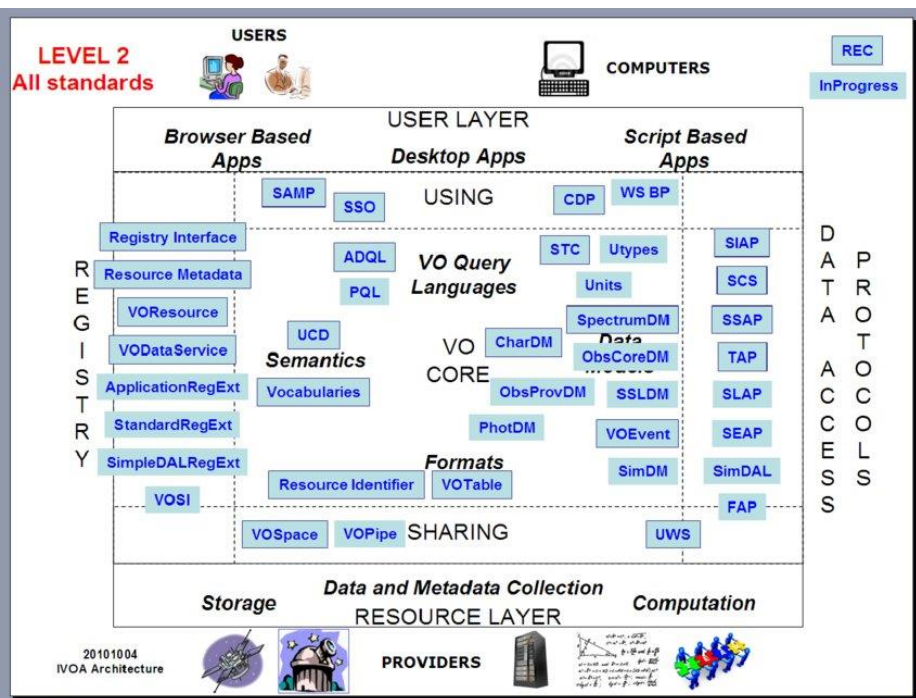
Table of Contents

- Description
- References
- Files
- Metadata

Research Data Infrastructure



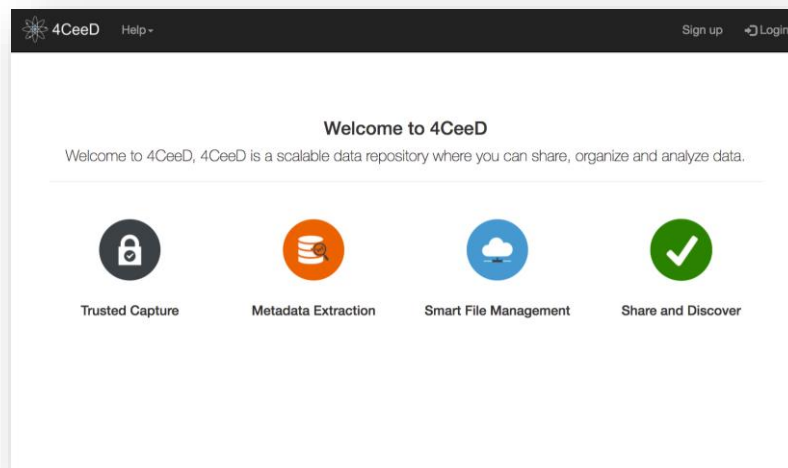
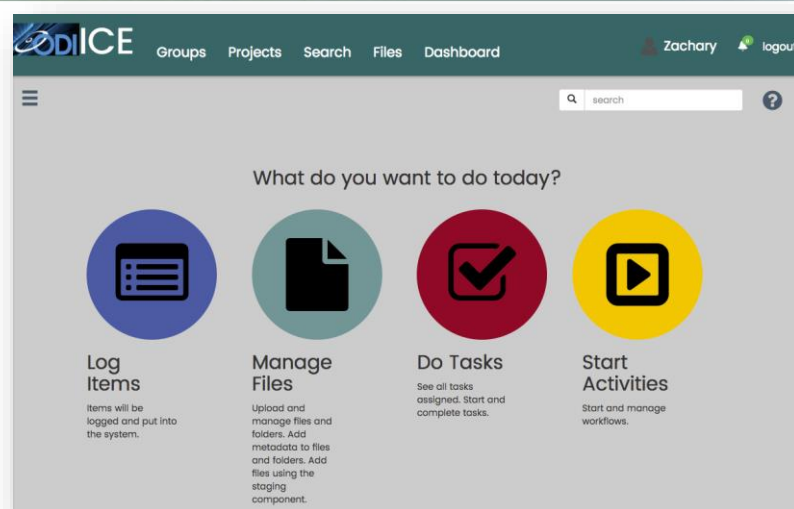
Research Data Infrastructure



Interoperability

Laboratory Information Management Systems

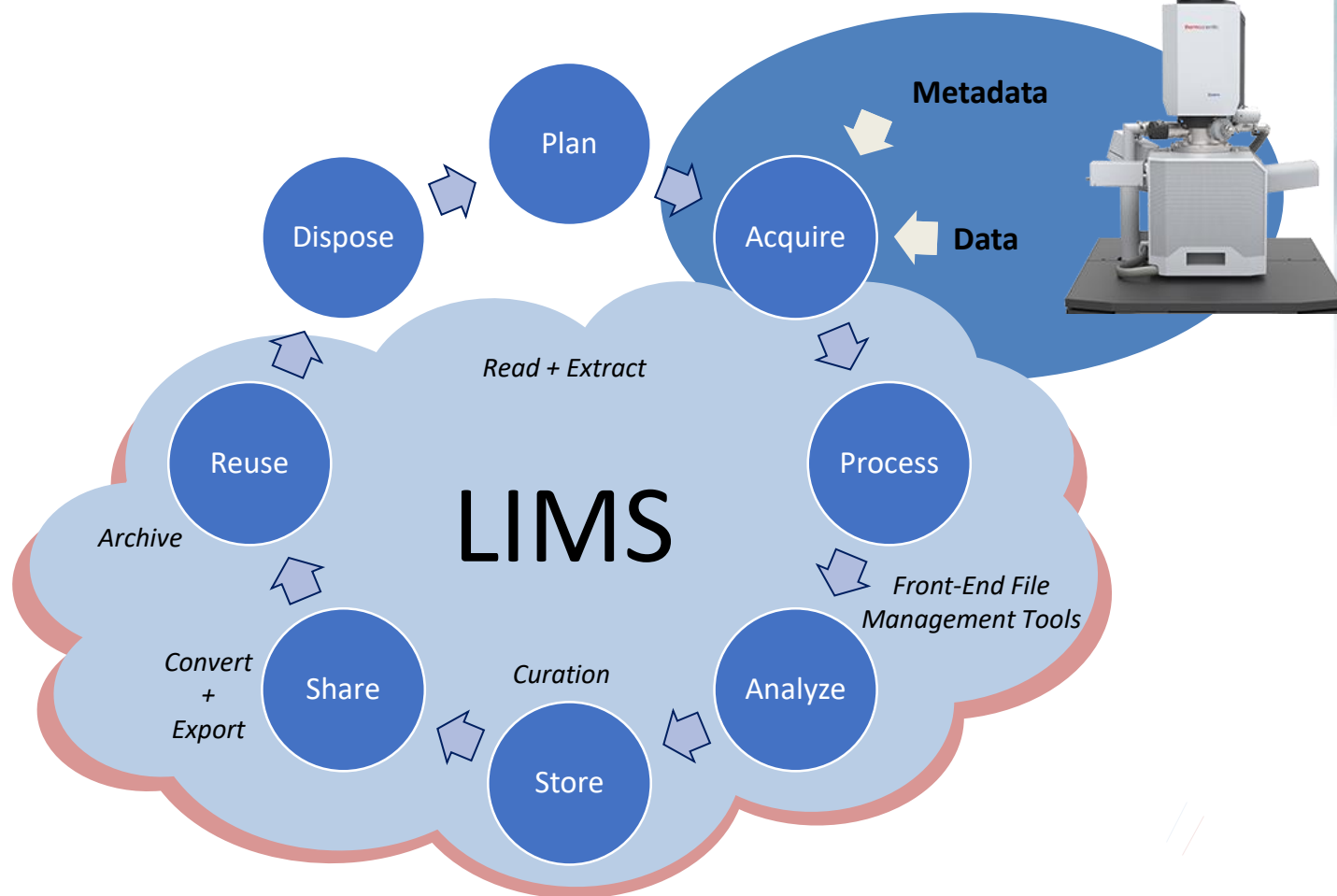
- Integrated Collaborative Environment (ICE)
 - Running now at <http://ice.nist.gov>
 - Developed by Air Force Research Laboratory
- Timely and Trustworthy Curating and Coordinating Data Framework (T2C2) 4CeeD system
 - Running now at <http://t2c2.nist.gov:32500/>
 - Developed by University of Illinois at Urbana-Champaign
- Also considering Discovery Environment for Relational Information and Versioned Assets (DERIVA) from USC



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

LIMS Help Manage the Data Lifecycle



Materials Data Curation System

Materials Data Curation System

Part of the Materials Genome Initiative

[Login](#) | [My Profile](#) | [Help](#)

Home

Materials Data Curator

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

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.

The diagram illustrates the Materials Innovation Infrastructure. It features a central Venn diagram with three overlapping circles: 'Computational Tools' (green), 'Experimental Tools' (yellow), and 'Digital Data' (orange). These circles are contained within a larger black circle labeled 'Materials Innovation Infrastructure'. Surrounding this central circle are four colored segments representing different domains: 'Human Welfare' (blue), 'Clean Energy' (teal), 'National Security' (red), and 'Next Generation Workforce' (purple).

Available Options

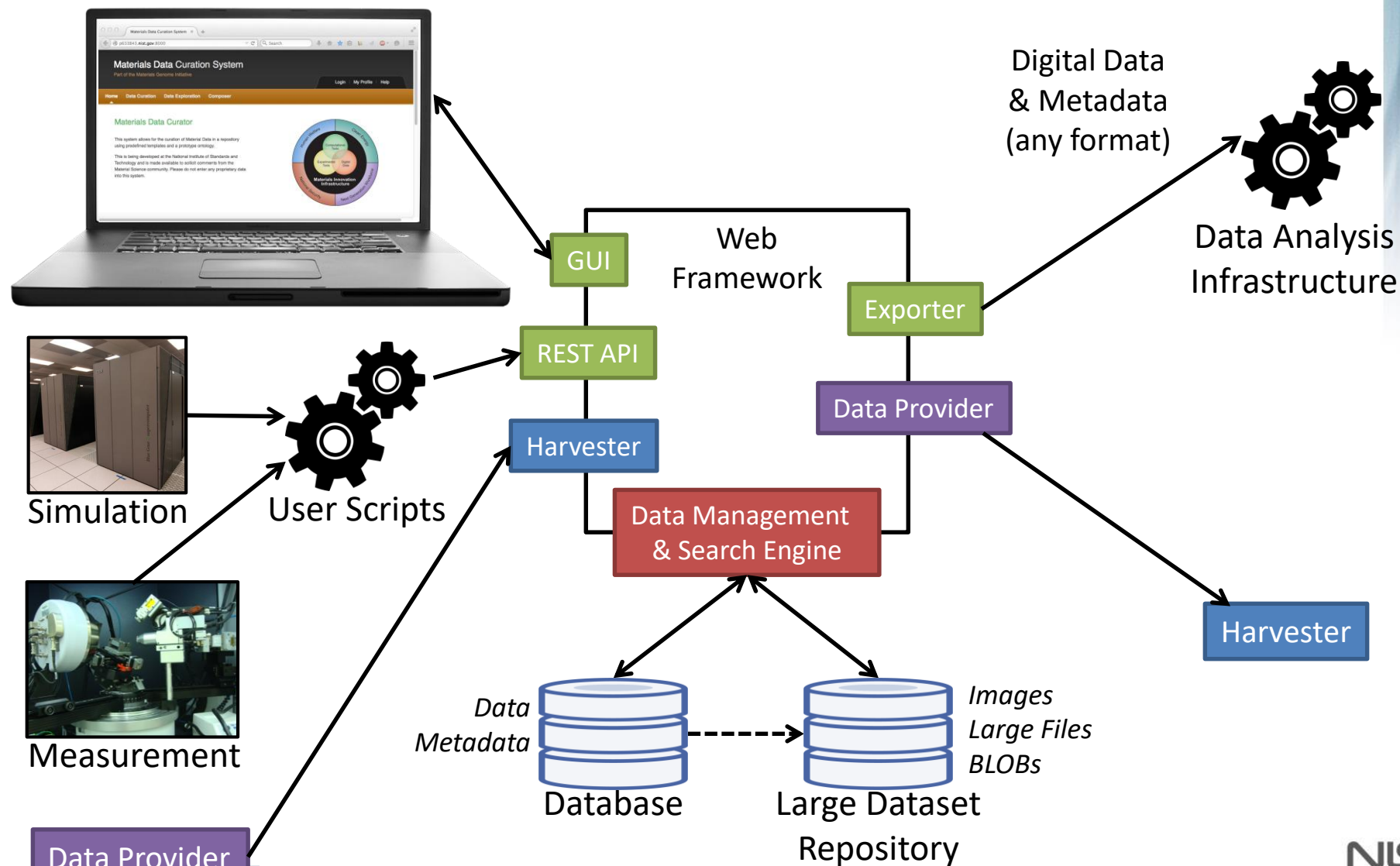
[All Options »](#)

Most Recent Templates

[Browse All »](#)

- species-qs | species.xsd.xml
- TEM-Tutorial-Tamu | workshop-TEM.xsd
- TEM-Tut | workshop-TEM.xsd

Materials Data Curation System



Data Provider
LABORATORY

Challenges with Experimental Data

**Undefined
Structure**

SampleID	CPD	RR	S
BANK	1	7251	
1	153	1	161
1	129	1	129
1	129	1	134

**Different
Formats**

SampleID	CPD	RR	Sample 1B	DataFileName
DiffType	PW3710	GeneratorVoltage	40	TubeCurrent
Anode	Cu	Alpha1	1.54056	Alpha2
MonochromatorUsed	YES	DivergenceSlit	1	ReceivingSlit
5.000	0.020	150.000	MeasureDate	Time
184	171	182	184	StepTime
171	163	146	158	3.00
156	158	148	153	

Only an expert human can understand this number.

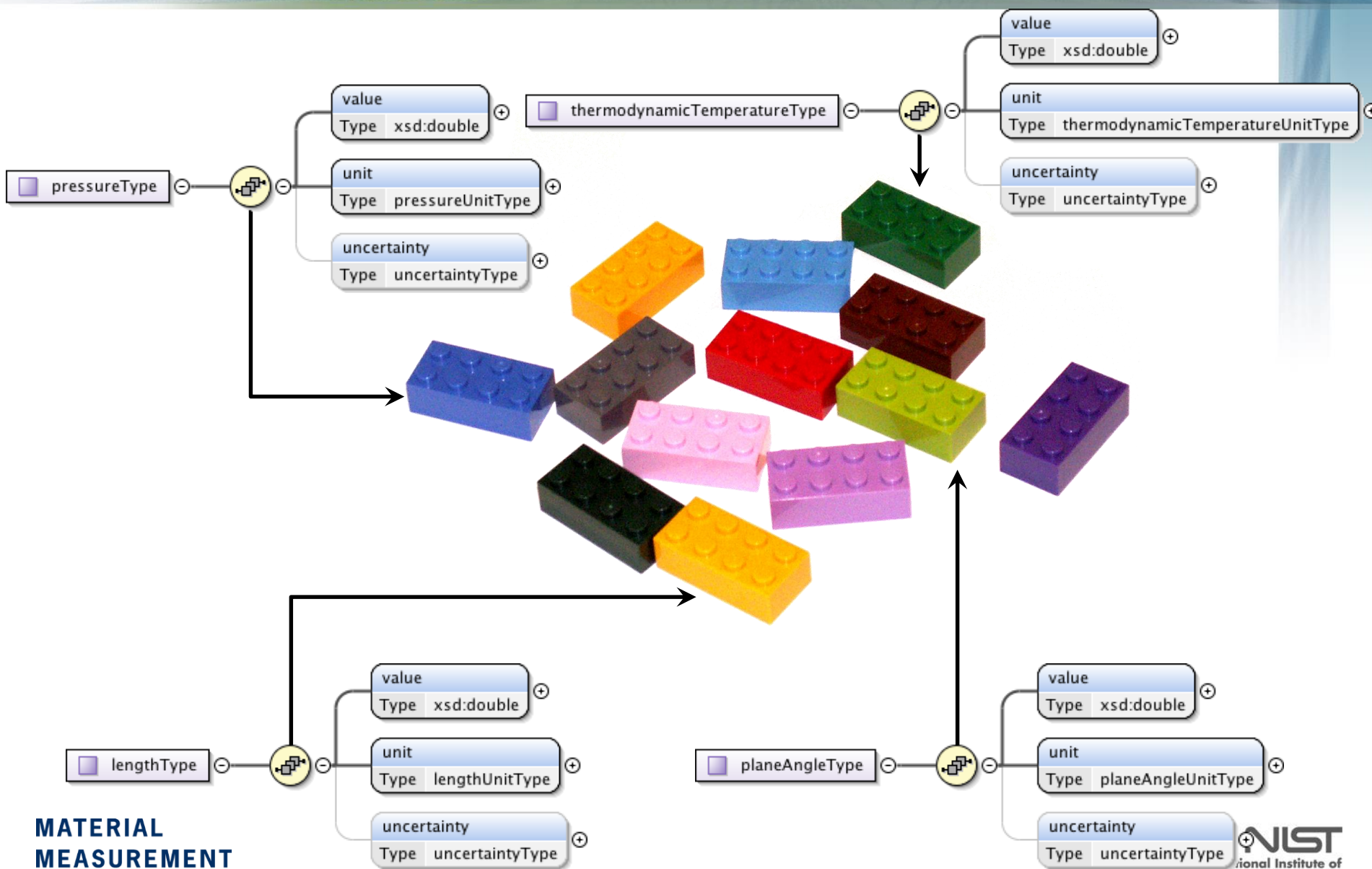
To a computer, this is a meaningless collection of numbers

5.0000	150.0000	0.0200	1.0000
177.	182.	174.	154.
144.	154.	161.	156.
164.	128.	154.	114.

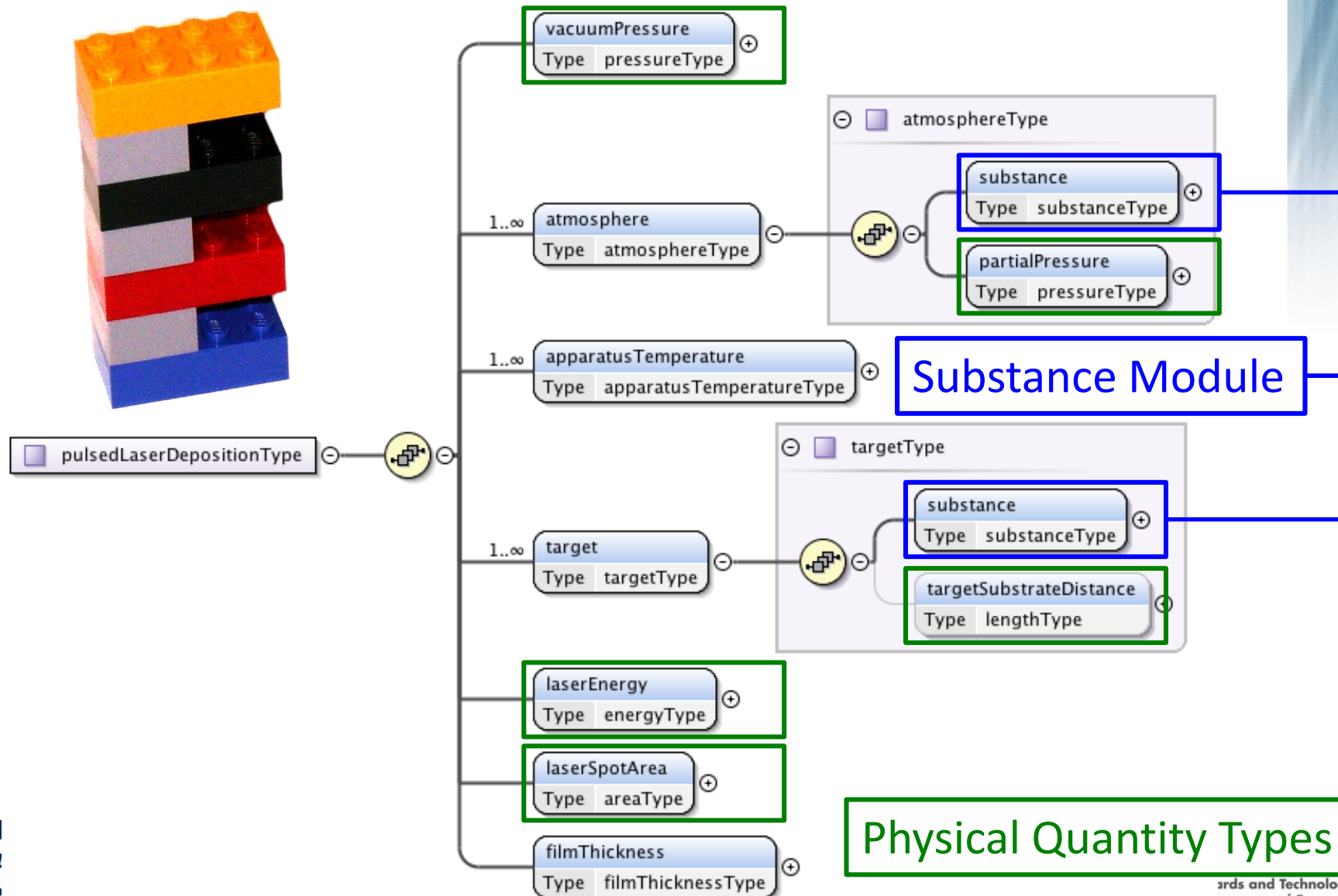
Structured Data Based on Data Model

```
{
  "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



Metadata Standards



Search all of FAIRsh:

Standards

Databases

Policies

Collections

Add/Claim Content

Stats

Log in or Register

Showing records 1 - 50 of 1068.

View as Table

View as Grid

Sort by

Best Match

Recommended Records

Recommended

Associated Publication?

No Publication

Has Publication

Claimed?

No Maintainer

Has Maintainer

Record Status

Uncertal

Deprecal

In develo

Ready

Standard Type

Terminology Artifact 708

Model/format 240

Reporting Guideline 120

Domains

Life Science 406

Biomedical Science 198

Ontology And Terminology 69

Deoxyribonucleic Acid (DNA) 59

« 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 »

Registry	Name	Abbreviation	Type	Domain	Taxonomy	Related Database	Related Standard	Related Policy
	CDISC Analysis Data Model	CDISC ADaM	Standard	Analysis Biomedical Science Data Model Data Transformation Preclinical And Clinical Studies	Homo sapiens	None	CDISC PRM CDISC SEND CDISC SDTM CDISC CDASH CDISC Define.xml Plus 6 more...	None
	mz peptide and protein Identification Markup Language	mzIdentML	Standard	Centrally Registered Identifier Life Science Peptide Identification Protein Protein Identification	All	None	mzQuantML PSI-MS CV PSI SpML QCML mzTab	None
	mz Quantitative Markup Language	mzQuantML	Standard	Experimental Measurement Life Science Protein Quantification	All	PRIDE	mzIdentML PSI-MS CV HUPO-PSI TraML mzML PSI SpML Plus 3 more...	None
	Anatomical Entity Ontology	AEO	Standard	Anatomy Life Science	Fungi Invertebrata Plantae Vertebrata	None	EHDAA2 CARO	None
	Amphibian Taxonomy Ontology	ATG	Standard	Taxonomic Classification	Amphibia	None	None	None
	Bilateria anatomy	BILA	Standard	Anatomy	Bilateria	4DXpress	None	None
	Biological Spatial Ontology	BSPO	Standard	Anatomy Life Science	Animalia Fungi Virdiplantae	ZFIN	CARO	None

The Open Microscopy Environment

OME develops open-source software and data format standards for the storage and manipulation of biological microscopy data. It is a joint project between universities, research establishments, industry and the software development community.

OMERO



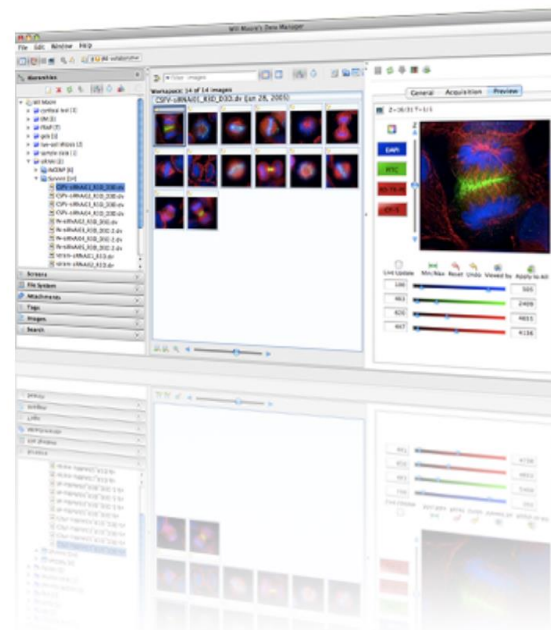
OMERO is client-server software for visualization, management and analysis of biological microscope images.

[About OMERO](#) | [Download](#) | [New Users](#) | [Projects](#) | [Shared Scripts](#)

Latest Release: OMERO 5.3.2

(May 2017)

Full details of this release can be found in the [Release announcement](#).



Bio-Formats



Bio-Formats is a Java library for reading and writing biological image files. It can be used as an ImageJ plugin, Matlab toolbox, or in your own software.

[About Bio-Formats](#) | [Download](#)

Latest Release: Bio-Formats 5.5.1

(May 2017)

Full details of this release can be found in the [Release announcement](#).

OME-TIFF Format

A TIFF-based image format that includes the OME-XML standard. [Read more](#)

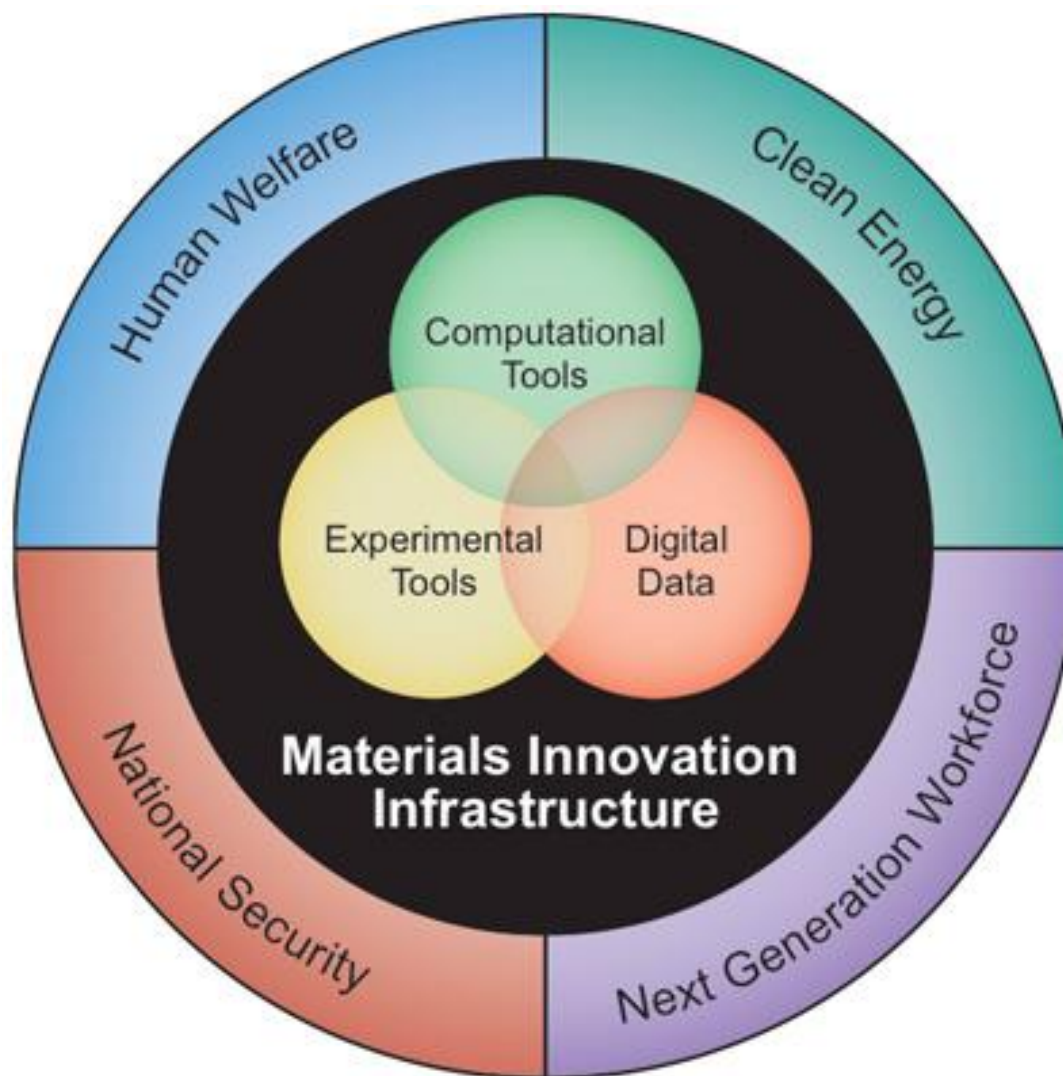
OME Data Model

A common specification for storing details of microscope set-up and image acquisition. See [the File Formats](#)

Inspiration

- Data management practices in astronomy, and the VO in particular, have inspired similar efforts in other fields
 - Space science/space physics VxOs (VSO, VHO, VMO, VITMO)
 - Materials science
 - Metrology
 - Ecology/environmental science
 - Life/bioscience
 - Neuroscience
- Other communities are envious of astronomy's global data format, FITS
 - Other fields must contend with myriad of formats, many proprietary
- Also similar challenges, such as interoperability and semantic standards
- Independent development of similar architecture

Materials Genome Initiative



Environmental Science



What is DataONE?

Data Observation Network for Earth (DataONE) is the foundation of new innovative environmental science through a distributed framework and sustainable cyberinfrastructure that meets the needs of science and society for open, persistent, robust, and secure access to well-described and easily discovered Earth observational data.

Life Sciences



VISION: TRANSFORMING SCIENCE THROUGH DATA-DRIVEN DISCOVERY.

MISSION: OUR MISSION IS TO DESIGN, DEPLOY, AND EXPAND A NATIONAL CYBERINFRASTRUCTURE FOR LIFE SCIENCES RESEARCH, AND TO TRAIN SCIENTISTS IN ITS USE.

THE PROJECT

CyVerse is funded by the National Science Foundation's Directorate for Biological Sciences. We are a dynamic virtual organization led by the University of Arizona to fulfill a broad mission that spans our partner institutions: Texas Advanced Computing Center, Cold Spring Harbor Laboratory, and the University of North Carolina at Wilmington.

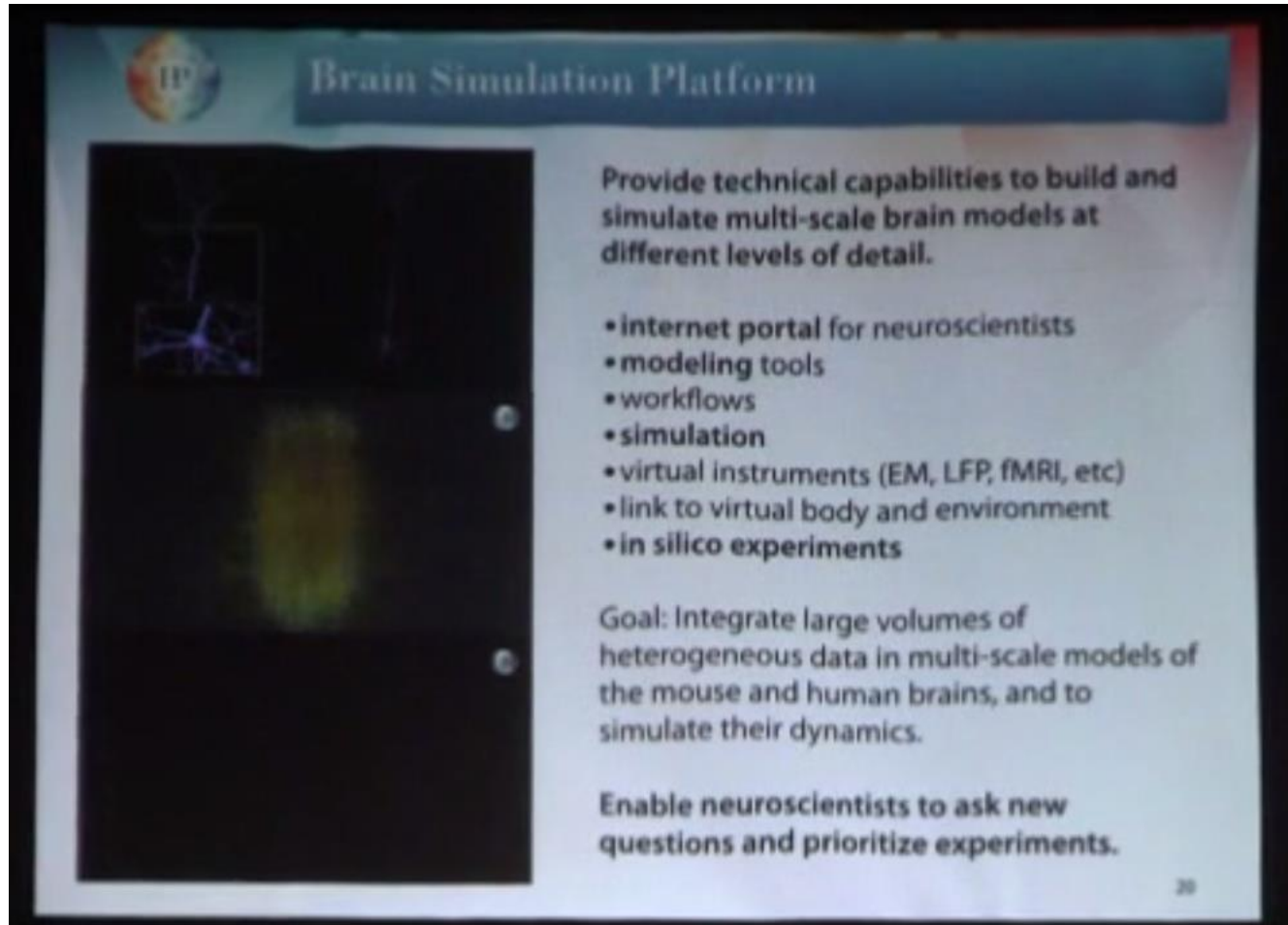
CyVerse fills a niche created by the computing epoch and a rapidly evolving world. Developing solutions to today's grand

scientific challenges means that we must understand how the organisms that contribute to our food, fuels, and ecosystem are shaped by interactions with their environment. CyVerse provides life scientists with powerful computational infrastructure to handle huge datasets and complex analyses, thus enabling data-driven discovery. Our powerful extensible platforms provide data storage, bioinformatics tools, image analyses, cloud services, APIs, and more.



Neuroscience

- Sean Hill, European Brain Initiative...



The screenshot shows a presentation slide titled "Brain Simulation Platform". On the left side of the slide, there is a dark rectangular area containing a glowing, yellowish-green 3D model of a brain structure. Above this image is a small circular logo with the letters "IP". To the right of the image, the text reads: "Provide technical capabilities to build and simulate multi-scale brain models at different levels of detail." Below this is a bulleted list of features: "• internet portal for neuroscientists", "• modeling tools", "• workflows", "• simulation", "• virtual instruments (EM, LFP, fMRI, etc)", "• link to virtual body and environment", and "• in silico experiments". Further down, it states: "Goal: Integrate large volumes of heterogeneous data in multi-scale models of the mouse and human brains, and to simulate their dynamics." At the bottom, it says: "Enable neuroscientists to ask new questions and prioritize experiments." A small number "20" is visible in the bottom right corner of the slide.

Brain Simulation Platform

Provide technical capabilities to build and simulate multi-scale brain models at different levels of detail.

- internet portal for neuroscientists
- modeling tools
- workflows
- simulation
- virtual instruments (EM, LFP, fMRI, etc)
- link to virtual body and environment
- in silico experiments

Goal: Integrate large volumes of heterogeneous data in multi-scale models of the mouse and human brains, and to simulate their dynamics.

Enable neuroscientists to ask new questions and prioritize experiments.

Neuroscience

NSF 16-047

Dear Colleague Letter: National Brain Observatory: A Phased Approach for Developing a National Research Infrastructure for Neuroscience

February 19, 2016

Dear Colleagues:

With this Dear Colleague Letter (DCL), the National Science Foundation (NSF) is announcing the intention to foster the development of a national research infrastructure for neuroscience (National Brain Observatory) to support collaborative and team science for achieving a comprehensive understanding of the brain in action and context.

- Facilitates the emergence of a coherent national infrastructure comprising the above shared and accessible tools, resources and networks that will allow rapid integration, analysis, and modeling of brain data associated with behaviors from multi-disciplinary projects and enable large-scale collaborative research efforts nationally and internationally that will advance our understanding of brain structure and function.

Conclusions

- Core VO design/architecture is strong and has been adopted by major data providing organizations in astronomy; now informing data system design in other fields
- Federated, distributed systems have many benefits...
 - Scalability
 - Flexibility
 - Distributed curation by experts
 - Communities can build on core infrastructure
- ...and challenges
 - Consensus must be reached on metadata standards and realistic goals for interoperability

<https://www.nist.gov/mml/odi>

