NDS7: Chicago APRIL 13, 2017
Scientific Data Platforms: from Silos to Ecosystems

Perspectives on DATA SERVICE DELIVERY VIA ECOSYSTEMS

Materials Data Federation: Enabling a Global Ecosystem for Materials Science Research

Ray Plante, Chandler Becker, and Sharief Youssef
National Institute of Standards and Technology (NIST)

Laura Bartolo
Northwestern University

NIST Science Team:
Chandler Becker, Bob Hanisch, Ray Plante, Zachary Trautt, Jim Warren

NIST Software Development Team:
Mary Brady, Alden Dima, Guillaume Sousa Amaral, Ben Long, Pierre Francois Rigodiat, Xavier Schmitt, Sharief Youssef

Motivation for Materials Resource Registry Federation

- Many materials resources exist (datasets, websites, repositories, registries, etc.), and the number is growing.

- How can we link them in a way that makes it easier to find and share relevant information and data?

What is a Resource Registry?

- A resource registry is a catalog containing descriptions of *resources** that are useful for (materials science) data-driven research
  - * Mainly datasets, databases, and data services
  - * Can also be portals, software, organizations, …

- A <u>starting</u> point for *discovering* useful data and tools
  - Make high level metadata descriptions searchable
  - Direct users to the web sites that host the data

Why Build a Registry Federation: Think Global, Act Local

- What does federation mean?
  - Network of registries; there is no single Registry
    Any registry can collect a globally-comprehensive collection of resource descriptions and make it searchable
  - Resource metadata exchange
    A common mechanism(s) for sharing descriptions of available data resources
  - Local metadata curation
    Any organization can run registry or register at remote registry their own data resources and share it with the world
- Why federate?
  - Distribute metadata curation
    Experts who provide data resources manage how they are described, update descriptions as they evolve
  - No single point of failure (including funding failure)
  - Allow innovation in providing search capabilities
- How do we federate?
  - Common metadata exchange mechanism
  - Common metadata schema

Metadata – Key to Federating

Federate registries by exchanging resource description records

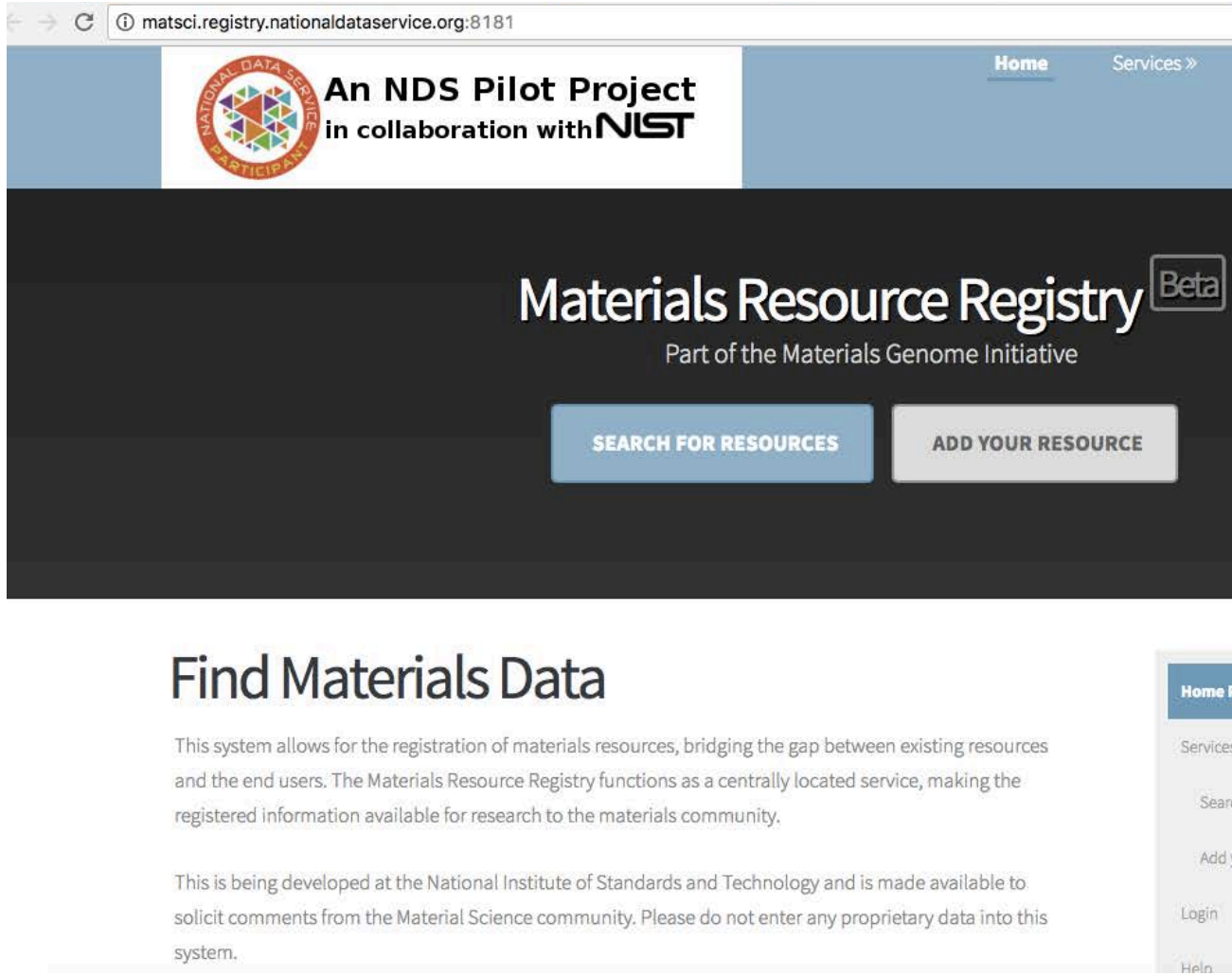
Through RDA Mat'ls Interest & Working Groups:

- Establish a common description schema to enable record exchange
  - Latest version incorporating feedback at RDA WG site:
    https://www.rd-alliance.org/system/files/documents/Materials_Registry_vocab_draft_170321.pdf
- Adopt an XML-based approach
  - OAI-PMH for exchanging records
  - XML Schema for defining record format/syntax
- Build evolving extensible metadata
  - Define concepts, vocabularies in format-free way
  - Encode into XML with techniques that allow for future extension and evolution

Pilot NIST Materials Resource Registry

Through the National Data Service:

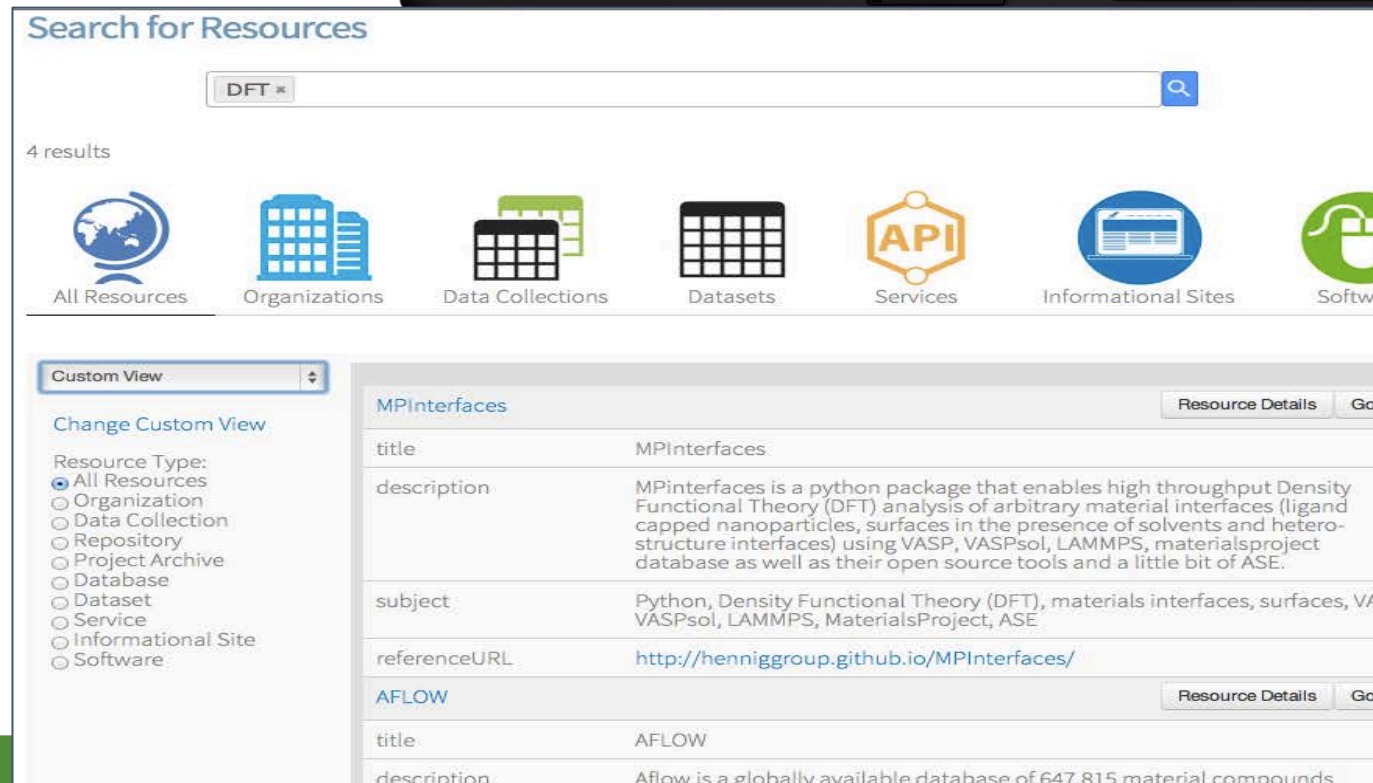
- NDS Pilot Project in Materials Science
- NDS sponsors pilot projects to help build the NDS community and prototype the NDS infrastructure
- The NDS Pilot Project enabled NIST to prototype and roll out working registry



NIST Materials Resource Registry

- General materials science resources
- Intended to interact with other registries that are more focused and/or housed at other institutions
- OAI-PMH protocol enabled, built on the Materials Data Curation System platform
  - Code on GitHub

https://github.com/usnistgov/MaterialsResourceRegistry



Browse Registered Resources

NIST

Home Services » Login Help Contact

Materials Resource Registry Beta

Search for Resources

Enter keywords, or leave blank to retrieve all records

66 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

Primary Audience:
- Education
- Public Outreach
- Research

Rights:
- Fee-Required
- Open-Login
- Proprietary
- Public

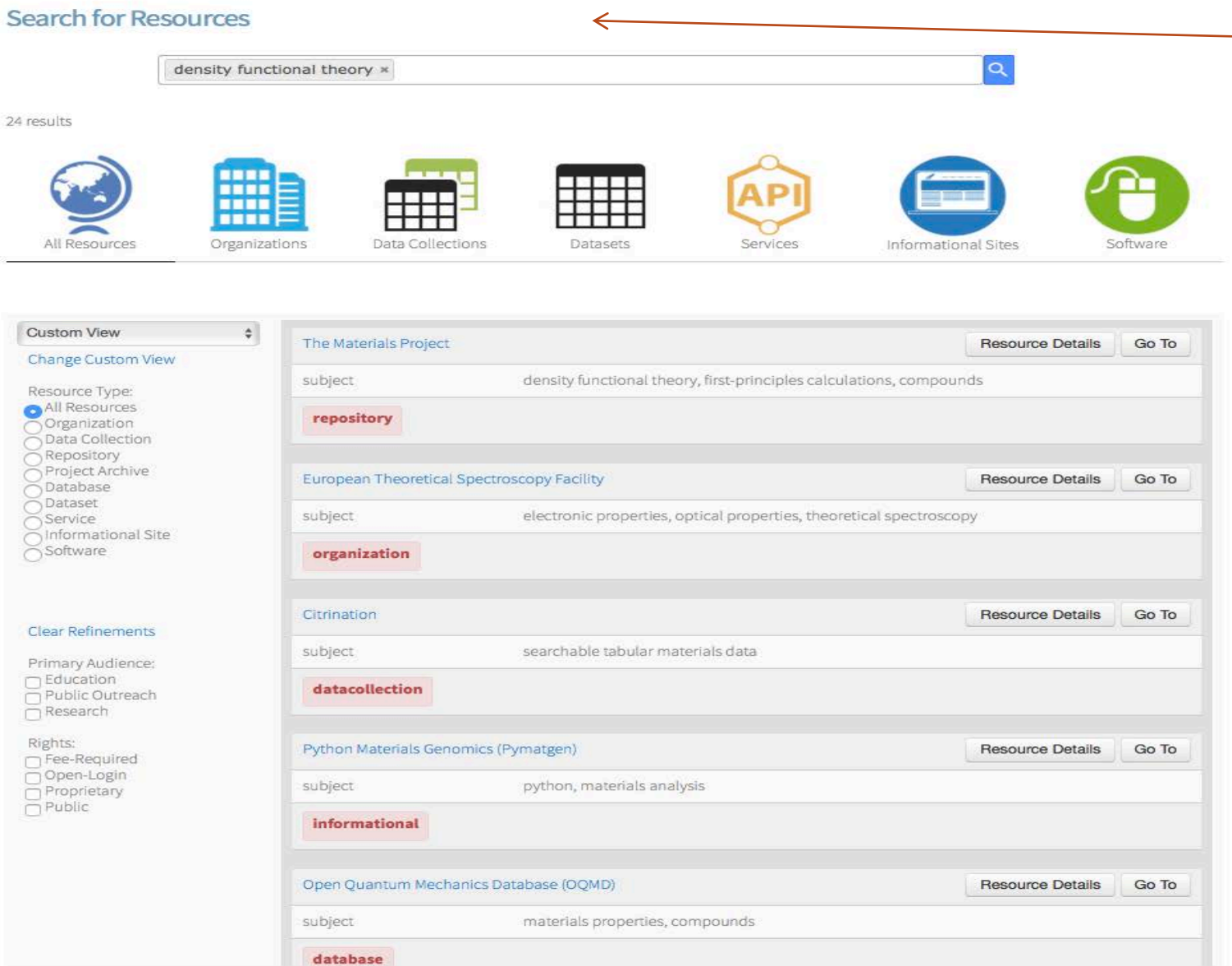
European Theoretical Spectroscopy Facility	Resource Details	Go To
publisher	European Theoretical Spectroscopy Facility	
subject	electronic properties, optical properties, theoretical spectroscopy	
organization		
MATIN: Materials e-Collaboration, Data Sciences and Informatics	Resource Details	Go To
publisher	Georgia Institute of Technology	
subject	materials discovery, open source software platform, materials informatics	
service		
Citrination	Resource Details	Go To
publisher	Citrine Informatics	
subject	searchable tabular materials data	
datacollection		
Cloud of Reproducible Records	Resource Details	Go To
publisher	CoRR team, National Institute of Standards and Technology	

Different types of resources, including:
- Organizations
- Collections
- Services
- Software

Links to registered resources and more information

Change which fields are displayed

Search for resources



All metadata text is searchable

Moving toward resources connected by metadata harvesting protocols such as OAI-PMH

For example:

- Materials Data Facility
- Instances hosted by universities or professional societies
- Other implementations that use OAI-PMH but different code

Search criteria used (Clear all):

Type ×

73 results

TYPE (Clear)

- Organization (21)
- Collection (25)
- Dataset (12)
- Service (0)
- Software (2)
- Web Site (14)

MATERIAL TYPE (Clear)

STRUCTURAL FEATURE (Clear)

- composites (6)
- defects (2)
  - unspecified defects (0)
  - cracks (0)
  - dislocations (2)
  - inclusions (0)
  - interstitials (0)

Materials Design Toolkit

Shengyen LiNational Institute of Standards and Technology (NIST) - National Institute of Standards and Technology (NIST)

https://mgi.nist.gov/generic-materials-design-toolkit

Subject keyword(s): structural materials, superalloys, Co alloys, Ni-based superalloys, two-phase microstructures, Materials Data Curation System (MDCS)

This framework provides an environment for materials design. The ICME (Integrated Computational Materials Engineering) approach is implemented for processing-structure-property correlation with a optimizer for material selection. Microstructure evolution can be simulated using the CALPHAD approach and phase based models using thermodynamics softwar... show more

Exascale Co-design Center for Materials in Extreme Environments

Tim Germann, Jim Belak, David Richards, Allen McPherson

http://www.exmatex.org/

The objective of the Exascale Co-design Center for Materials in Extreme Environments (ExMatEx) is to establish the interrelationship among algorithms, system software, and hardware required to develop a multiphysics exascale simulation framework for modeling materials subjected to extreme mechanical and radiation environments. Such a simulation cap... show more

Bilbao Crystallographic Server

Walter de Gruyter

http://cryst.ehu.es/

Subject keyword(s): Crystallography

Initiated in 1997, at the Materials Laboratory of the University of the Basque Country, Spain, the Bilbao Crystallographic Server has since been offering its crystallographic and solid state programs and utilities, free of charge. The programs hosted on the server are on a wide variety of topics, and are grouped units called "shells".

DOE Data Explorer

Department of Energy

http://www.osti.gov/dataexplorer/

Subject keyword(s): DOE sponsored public data collections

Get More Information

Detailed information about resources, including

- who created them
- who maintains them
- what they contain
- how to access them

Plus links to the resources themselves

Resource Details

National Renewable Energy Laboratory Materials Database	
localid	DYJM46H37MU0FHYBE8NE
status	active
title	National Renewable Energy Laboratory Materials Database
shortName	NRELMatDB
publisher	National Renewable Energy Laboratory
contributor	Ann Deml, Stephan Lany, Haowei Peng, Vladan Stevanovic, Jun Yan, Pawel Zawadzki
name	
description	NRELMatDB is a computational materials database with the specific focus on materials for renewable energy applications including, but not limited to, photovoltaic materials, materials for photo-electrochemical water splitting, thermoelectrics, etc. The main goal of NRELMatDB is to enable and facilitate the access and exchange of computational data between different research groups following the guidelines outlined in the Presidential Materials Genome Initiative (http://www.whitehouse.gov/mgi).
subject	renewable energy, enthalpies of formation, enthalpies of decomposition, band gaps, dielectric functions
referenceURL	http://materials.nrel.gov/
referenceCitation	"Correcting density functional theory for accurate predictions of compound enthalpies of formation: Fitted elemental-phase reference energies". V. Stevanovic, S. Lany, X. Zhang, A. Zunger, Physical Review B 85, 115104 (2012), http://dx.doi.org/10.1103/PhysRevB.85.115104, "Band-structure calculations for the 3d transition metal oxides in GW". S. Lany, Physical Review B 87, 085112 (2013), http://dx.doi.org/10.1103/PhysRevB.87.085112, "Semiconducting transition metal oxides". S. Lany, J. Phys.: Cond. Matter 27, 283203 (2015), http://dx.doi.org/10.1088/0953-8984/27/28/283203
primaryAudience	research
materialType	metal, semiconductor
structuralMorphology	bulk, crystalline
propertyClass	thermodynamic, optical, simulated
computationalDataAcquisitionMethod	density functional theory calculation
rights	public
database	

Ok

Add a Resource

Resource

Local ID

UKK8LPL6G0TAK9DEWC6W

Status

Active

Identity

Software Name: MPInterfaces

Short Name: MPInterfaces

Version

Identifier

Logo

Curation

Publisher: Hennig group, University of Florida

PID:

Creator: Arunima Singh

Built on the Materials Data Curation System software, but with a specialized schema and interface

Materials Science

Material Types

- [ ] polymer
- [ ] organic
- [ ] composite
- [x] ceramic
- [x] metal
- [x] oxide
- [x] semiconductor
- [ ] biomaterial
- [x] nanomaterials
- [x] inorganic
- [x] non-specific
- [ ] superconductor

Morphology/Structures

- [x] 2D
- [ ] fiber
- [x] interfacial
- [x] interphase
- [x] composite
- [ ] nanotube
- [ ] amorphous
- [ ] fluid
- [x] bulk
- [ ] 1D
- [ ] line defect
- [ ] point defect
- [x] non-specific
- [x] crystalline
- [ ] quasi-periodic
- [x] film

Material Property Classes

- [x] thermodynamic
- [x] mechanical
- [x] defect
- [ ] optical
- [x] simulated
- [x] non-specific
- [ ] transport
- [x] structural

Experimental Data Acquisition Methods

- [ ] electron microscopy
- [ ] indentation
- [ ] calorimetry
- [ ] atom probe microscopy
- [ ] impact testing
- [ ] other
- [ ] spectroscopy
- [ ] dilatometry
- [ ] scattering-diffraction
- [ ] load frame testing
- [ ] optical microscopy
- [ ] non-specific

Computational Data Acquisition Methods

- [ ] computational thermodynamics
- [x] numerical simulations
- [x] molecular dynamics simulation
- [ ] phase field calculation
- [ ] statistical mechanics
- [ ] boundary tracking/level set
- [x] density functional theory calculation
- [ ] multiscale simulations
- [ ] finite element analysis
- [ ] monte-carlo simulation
- [ ] non-specific
- [ ] crystal plasticity calculation
- [ ] dislocation dynamics

Sample Processing Methods

Resource Registry: https://github.com/usnistgov/MaterialsResourceRegistry Applicable for other communities & other uses?

- Instead of This…

Lots of incompatible resources

Confusion, frustration, data loss, and missed opportunities

- Try This…

Common protocols

Mappings between content and approaches of different projects