



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 starting 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

The screenshot shows a web browser window with the URL "matsci.registry.nationaldataservice.org:8181". The page header includes the NDS logo and the text "An NDS Pilot Project in collaboration with NIST". Navigation links for "Home" and "Services" are in the top right. The main heading is "Materials Resource Registry" with a "Beta" badge, followed by the subtitle "Part of the Materials Genome Initiative". Two large buttons are present: "SEARCH FOR RESOURCES" and "ADD YOUR RESOURCE". Below this, a section titled "Find Materials Data" contains two paragraphs of text explaining the system's purpose and development. A sidebar on the right contains links for "Home", "Services", "Search", "Add", "Login", and "Help".

matsci.registry.nationaldataservice.org:8181

An NDS Pilot Project
in collaboration with NIST

Home Services »

Materials Resource Registry Beta
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 Services Search Add Login Help

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)
<https://github.com/usnistgov/MaterialsResourceRegistry>



Search for Resources

DFT

4 results

 All Resources
  Organizations
  Data Collections
  Datasets
  API Services
  Informational Sites
  Software

Custom View

Change Custom View

Resource Type:

- ☒ All Resources
- ☐ Organization
- ☐ Data Collection
- ☐ Repository
- ☐ Project Archive
- ☐ Database
- ☐ Dataset
- ☐ Service
- ☐ Informational Site
- ☐ Software

MPInterfaces		Resource Details
title	MPInterfaces	
description	MPInterfaces is a python package that enables high throughput Density Functional Theory (DFT) analysis of arbitrary material interfaces (ligand capped nanoparticles, surfaces in the presence of solvents and hetero-structure interfaces) using VASP, VASPsol, LAMMPS, materialsproject database as well as their open source tools and a little bit of ASE.	
subject	Python, Density Functional Theory (DFT), materials interfaces, surfaces, VASPsol, LAMMPS, MaterialsProject, ASE	
referenceURL	http://henniggroup.github.io/MPInterfaces/	

AFLOW		Resource Details
title	AFLOW	
description	Aflow is a globally available database of 647,815 material compounds	

Browse Registered Resources

Different types of resources, including:

- Organizations
- Collections
- Services
- Software

Links to registered resources and more information

Change which fields are displayed

The screenshot shows the NIST Materials Resource Registry Beta website. The header includes the NIST logo and navigation links: Home, Services, Login, Help, and Contact. Below the header is a search bar with the text "Search for Resources" and a placeholder "Enter keywords, or leave blank to retrieve all records". Below the search bar, it indicates "66 results". A row of icons represents different resource types: All Resources, Organizations, Data Collections, Datasets, Services, Informational Sites, and Software. On the left side, there is a sidebar with filters for "Resource Type" (All Resources, Organization, Data Collection, Repository, Project Archive, Database, Dataset) and "Primary Audience" (Education, Public Outreach, Research). Below these are filters for "Rights" (Fee-Required, Open-Login, Proprietary, Public). The main content area displays a list of resources, each with a title, publisher, subject, and a resource type label (organization, service, datacollection). Each resource entry has a "Resource Details" button and a "Go To" button.








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

Search for resources

Search for Resources

density functional theory

24 results

 All Resources
  Organizations
  Data Collections
  Datasets
  Services
  Informational Sites
  Software

Custom View

Change Custom View

Resource Type:

- ☒ All Resources
- ☐ Organization
- ☐ Data Collection
- ☐ Repository
- ☐ Project Archive
- ☐ Database
- ☐ Dataset
- ☐ Service
- ☐ Informational Site
- ☐ Software

Clear Refinements

Primary Audience:

- ☐ Education
- ☐ Public Outreach
- ☐ Research

Rights:

- ☐ Fee-Required
- ☐ Open-Login
- ☐ Proprietary
- ☐ Public

The Materials Project

subject: density functional theory, first-principles calculations, compounds

repository

European Theoretical Spectroscopy Facility

subject: electronic properties, optical properties, theoretical spectroscopy

organization

Citration

subject: searchable tabular materials data

datacollection

Python Materials Genomics (Pymatgen)

subject: python, materials analysis

informational

Open Quantum Mechanics Database (OQMD)

subject: materials properties, compounds

database

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):

73 results

Type *

^ 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 Li National 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

Project REST API

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

- | | | |
|---|---|---|
| <input type="checkbox"/> polymer | <input checked="" type="checkbox"/> metal | <input checked="" type="checkbox"/> nanomaterials |
| <input type="checkbox"/> organic | <input checked="" type="checkbox"/> oxide | <input checked="" type="checkbox"/> inorganic |
| <input type="checkbox"/> composite | <input checked="" type="checkbox"/> semiconductor | <input checked="" type="checkbox"/> non-specific |
| <input checked="" type="checkbox"/> ceramic | <input type="checkbox"/> biomaterial | <input type="checkbox"/> superconductor |

Morphology/Structures

- | | | |
|---|--|--|
| <input checked="" type="checkbox"/> 2D | <input type="checkbox"/> amorphous | <input checked="" type="checkbox"/> non-specific |
| <input type="checkbox"/> fiber | <input type="checkbox"/> fluid | <input checked="" type="checkbox"/> crystalline |
| <input checked="" type="checkbox"/> interfacial | <input checked="" type="checkbox"/> bulk | <input type="checkbox"/> quasi-periodic |
| <input checked="" type="checkbox"/> interphase | <input type="checkbox"/> 1D | <input checked="" type="checkbox"/> film |
| <input checked="" type="checkbox"/> composite | <input type="checkbox"/> line defect | |
| <input type="checkbox"/> nanotube | <input type="checkbox"/> point defect | |

Material Property Classes

- | | | |
|---|--|--|
| <input checked="" type="checkbox"/> thermodynamic | <input type="checkbox"/> optical | <input type="checkbox"/> transport |
| <input checked="" type="checkbox"/> mechanical | <input checked="" type="checkbox"/> simulated | <input checked="" type="checkbox"/> structural |
| <input checked="" type="checkbox"/> defect | <input checked="" type="checkbox"/> non-specific | |

Experimental Data Acquisition Methods

- | | | |
|--|---|---|
| <input type="checkbox"/> electron microscopy | <input type="checkbox"/> impact testing | <input type="checkbox"/> scattering-diffraction |
| <input type="checkbox"/> indentation | <input type="checkbox"/> other | <input type="checkbox"/> load frame testing |
| <input type="checkbox"/> calorimetry | <input type="checkbox"/> spectroscopy | <input type="checkbox"/> optical microscopy |
| <input type="checkbox"/> atom probe microscopy | <input type="checkbox"/> dilatometry | <input type="checkbox"/> non-specific |

Computational Data Acquisition Methods

- | | | |
|---|---|---|
| <input type="checkbox"/> computational thermodynamics | <input type="checkbox"/> boundary tracking/level set | <input type="checkbox"/> non-specific |
| <input type="checkbox"/> numerical simulations | <input checked="" type="checkbox"/> density functional theory calculation | <input type="checkbox"/> crystal plasticity calculation |
| <input checked="" type="checkbox"/> molecular dynamics simulation | <input type="checkbox"/> multiscale simulations | <input type="checkbox"/> dislocation dynamics |
| <input type="checkbox"/> phase field calculation | <input type="checkbox"/> finite element analysis | |
| <input type="checkbox"/> statistical mechanics | <input type="checkbox"/> monte-carlo simulation | |

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