Informatics Infrastructure for Data-Driven Materials Research at NIST

John Henry J. Scott

Senior Advisor, Office of Data and Informatics &
Physicist, Materials Measurement Science Division
National Institute of Standards and Technology
Gaithersburg, MD, USA

Shared Metadata and Data Formats for Big-Data Driven Materials Science
Erwin-Schrödinger-Zentrum, Berlin
Monday July 8, 2019

NIST

NIST Disclaimer

Certain commercial equipment, instruments, materials, or software are identified in this talk to foster understanding. Such identification does not imply recommendation or endorsement by NIST, nor does it imply that the materials, equipment, or software identified are necessarily the best available for the purpose.

Acknowledgments

- ODI Staff and Detailees: Bob Hanisch, Gretchen Greene, Ray Plante, Adam Morey, Chandler Becker, Peter Linstrom, Arlin Stoltzfus, Andrea Medina-Smith, Zachary Trautt, June Lau, et al.
- NIST Information Technology Lab Staff: Mary Brady, Alden Dima, Peter Bajcsy, Sharief Youssef, et al.
- Others: Kathy Sharpless, Nicholas Ritchie, Gerhard Goldbeck, Larry Lannom, Giridhar Manepali

Research Data Alliance, Project FREYA

Office of Data and Informatics

Standard Reference Data

- Distribution
- Sales
- Infrastructure
- Usage analysis and impact
- Improve web sites and user interfaces
- Provide APIs

Research Data

- Improve data management practices
- Data management planning tools
- Laboratory automation
- Electronic Lab Notebooks
- NIST open data repository
- NIST data portal

Data Science

- Informatics and analytics resource
- Big data
- HPC and Cloud computing

Community

- National Data Services Consortium
- Research Data Alliance
- Other US federal agencies (NIH, DOE, NSF)
- CENDI
- National Metrology Institutes (NMIs)
- BIPM (Paris)
- CODATA, WDS

Robert Hanisch
Director, Office of Data and Informatics

Research Data Management at NIST

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)

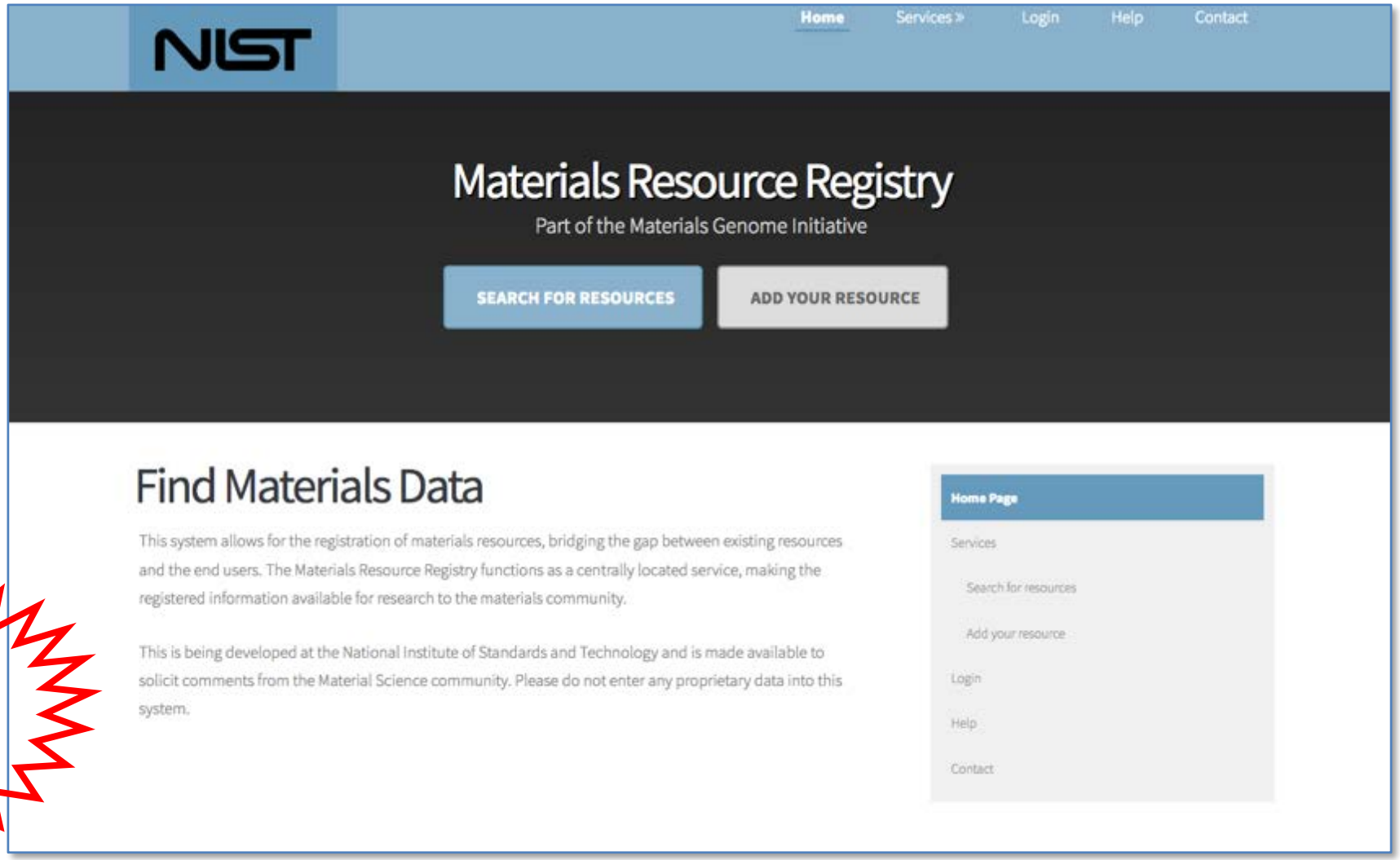
Access

Interoperate

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

Discovery

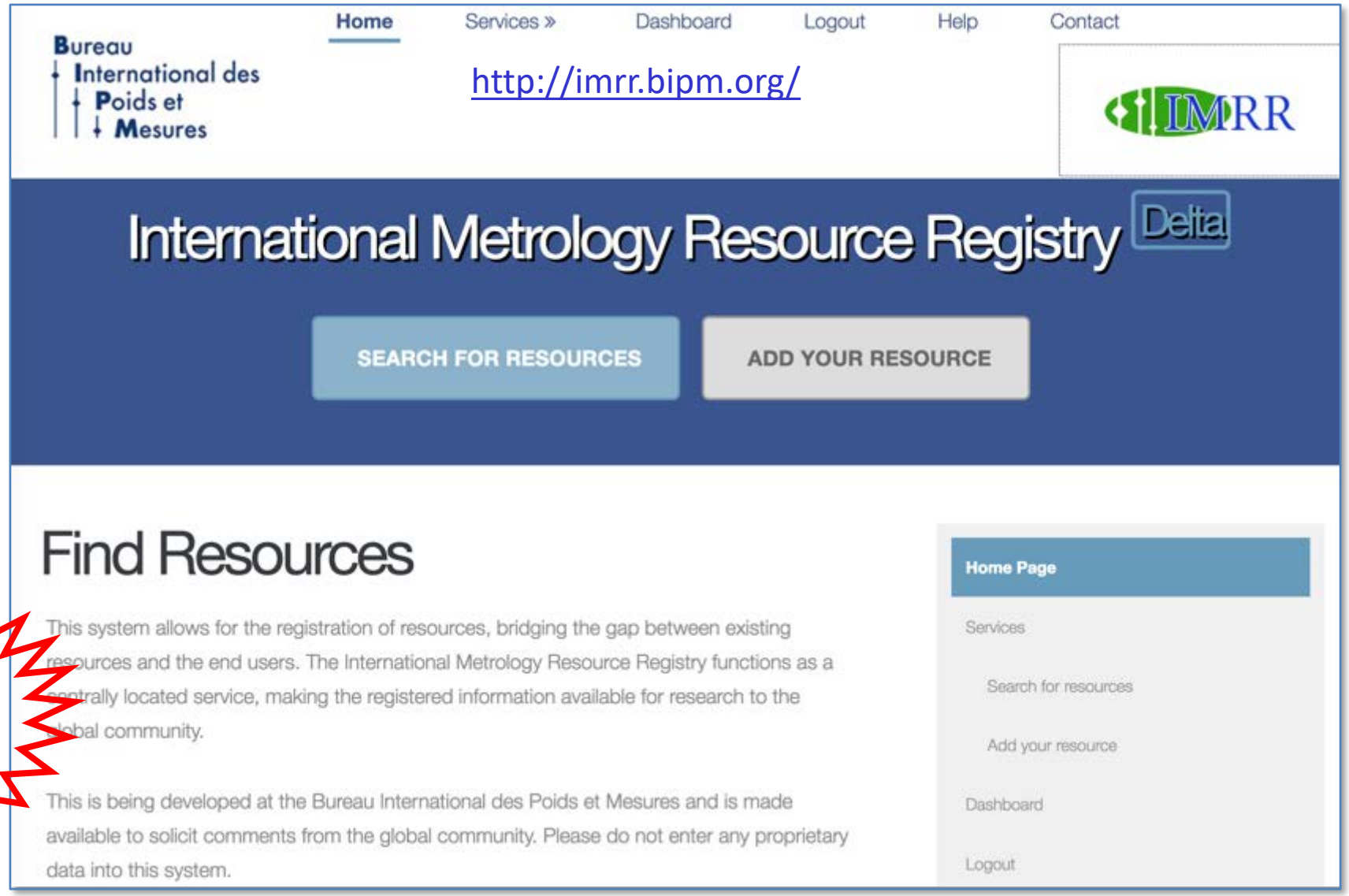
NIST Materials Resource Registry



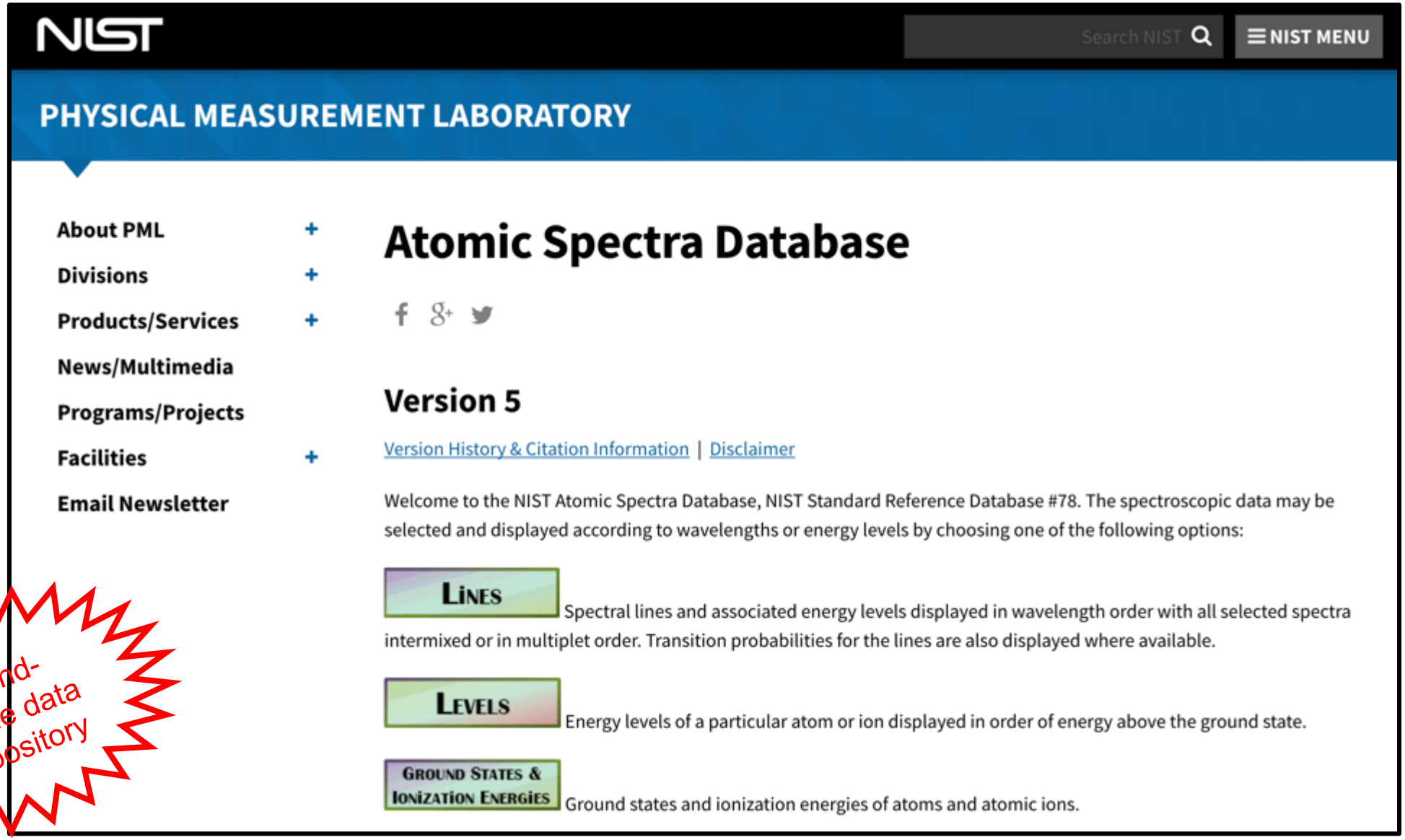
stand-alone data registry

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

International Metrology Resource Registry



stand-alone data registry

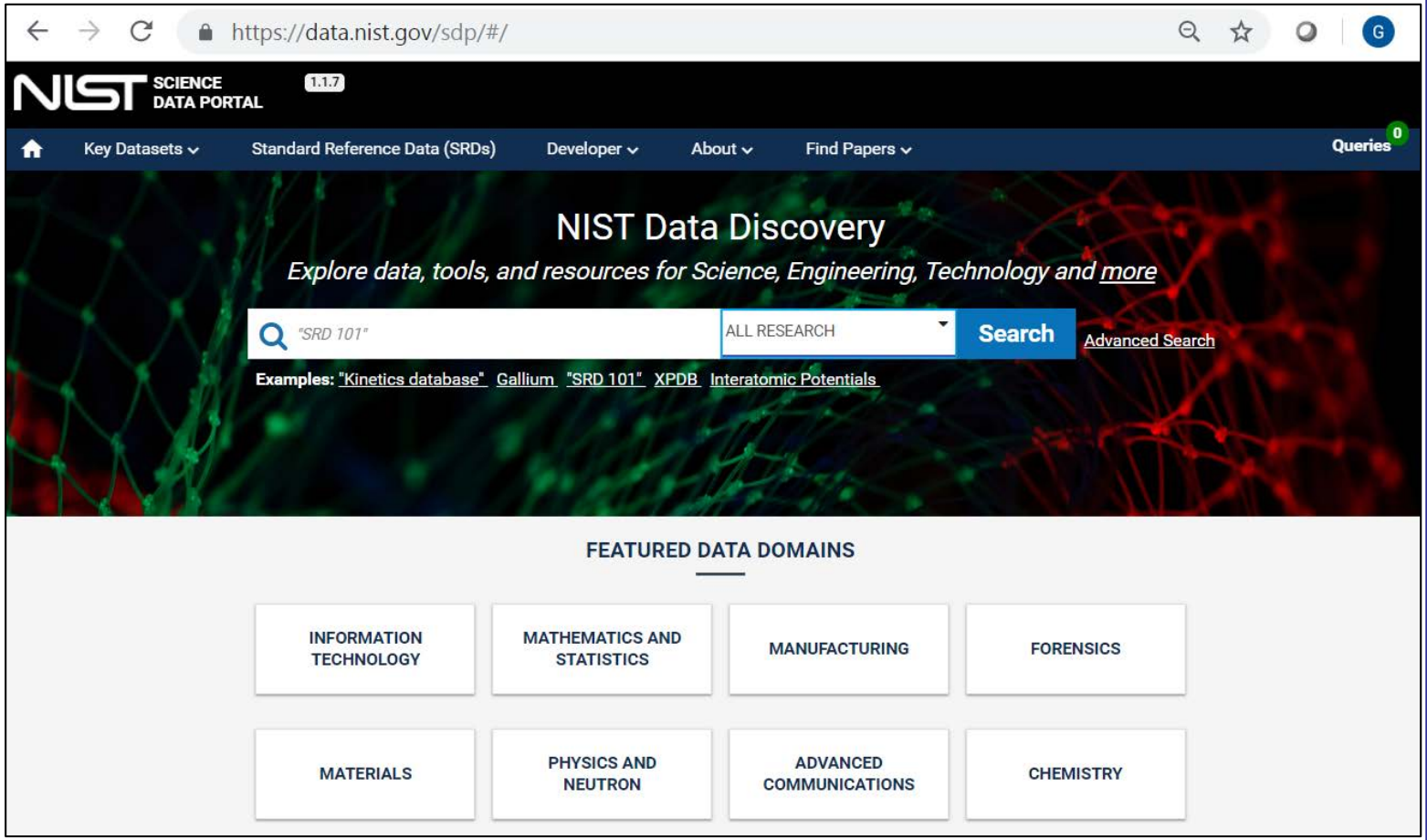


stand-alone data repository

Discover NIST Data at data.nist.gov

Web portal Discovery Tool for NIST public data

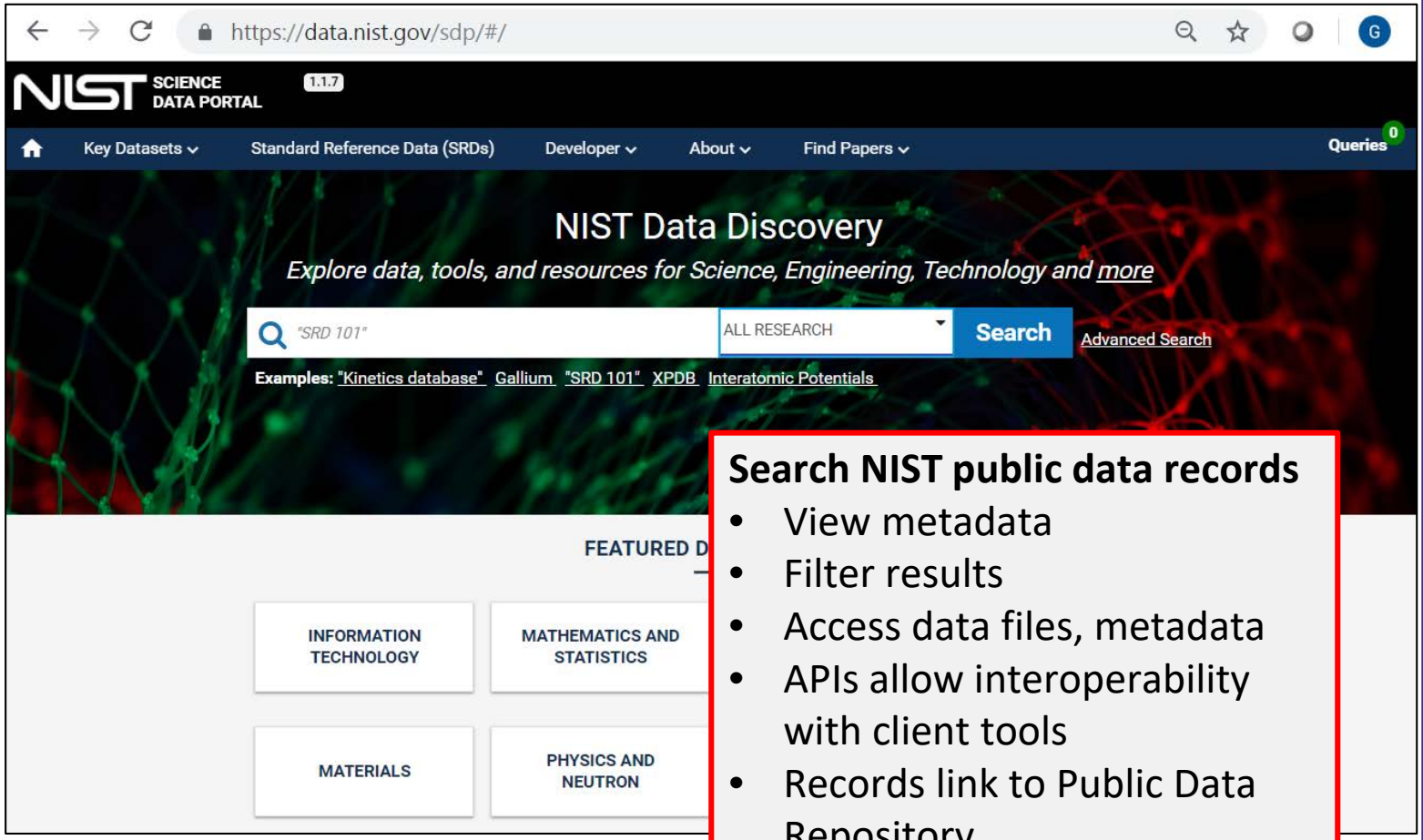
- Search All NIST inventory of public datasets
- Access to the NIST Public Data Repository
- Navigation to Data Publications, Custom Repositories



Discover NIST Data at data.nist.gov

Web portal Discovery Tool for NIST public data

- Search All NIST inventory of public datasets
- Access to the NIST Public Data Repository
- Navigation to Data Publications, Custom Repositories



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

Enterprise Data Inventory

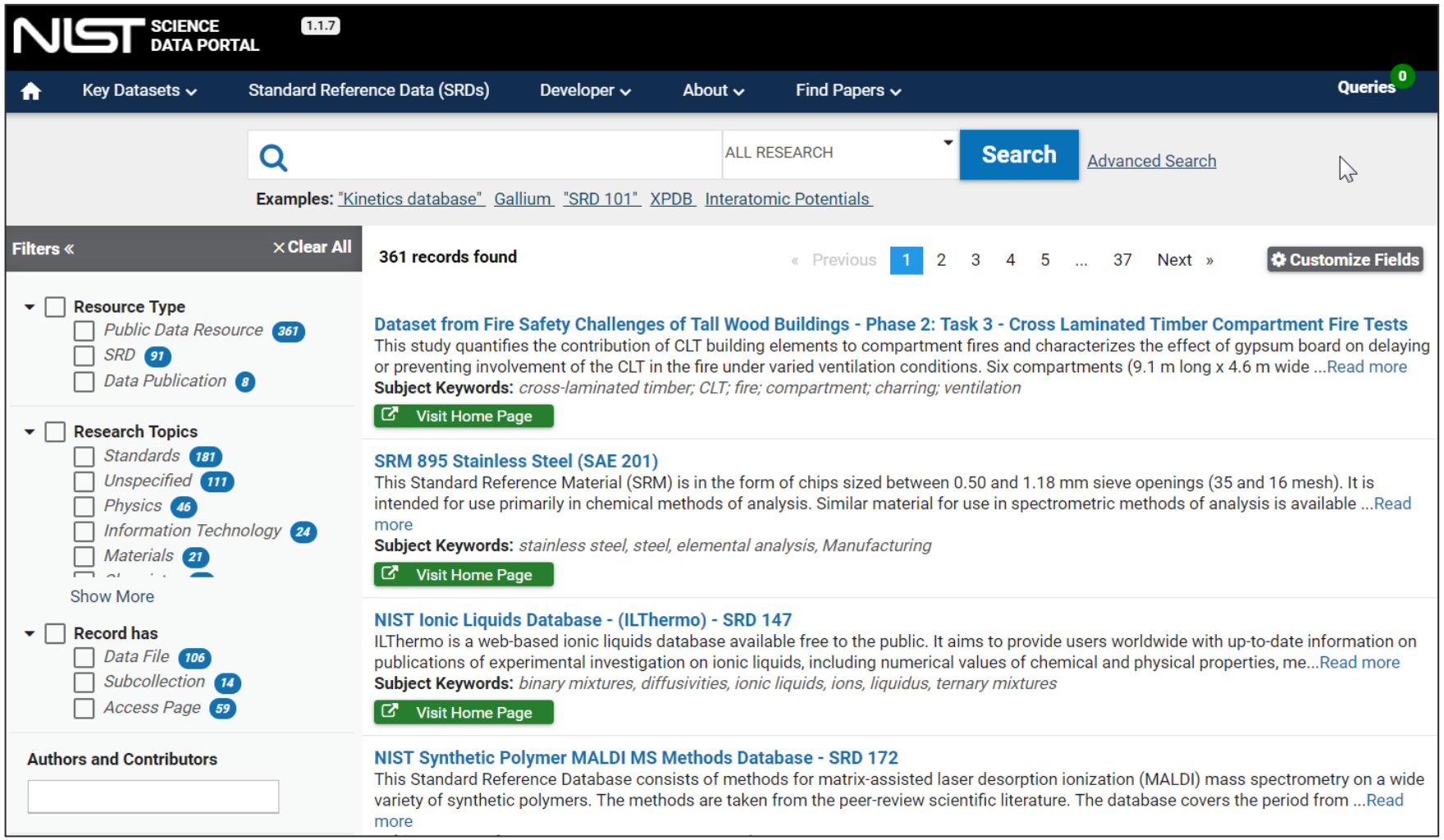
- *Enhanced for **Discoverability***

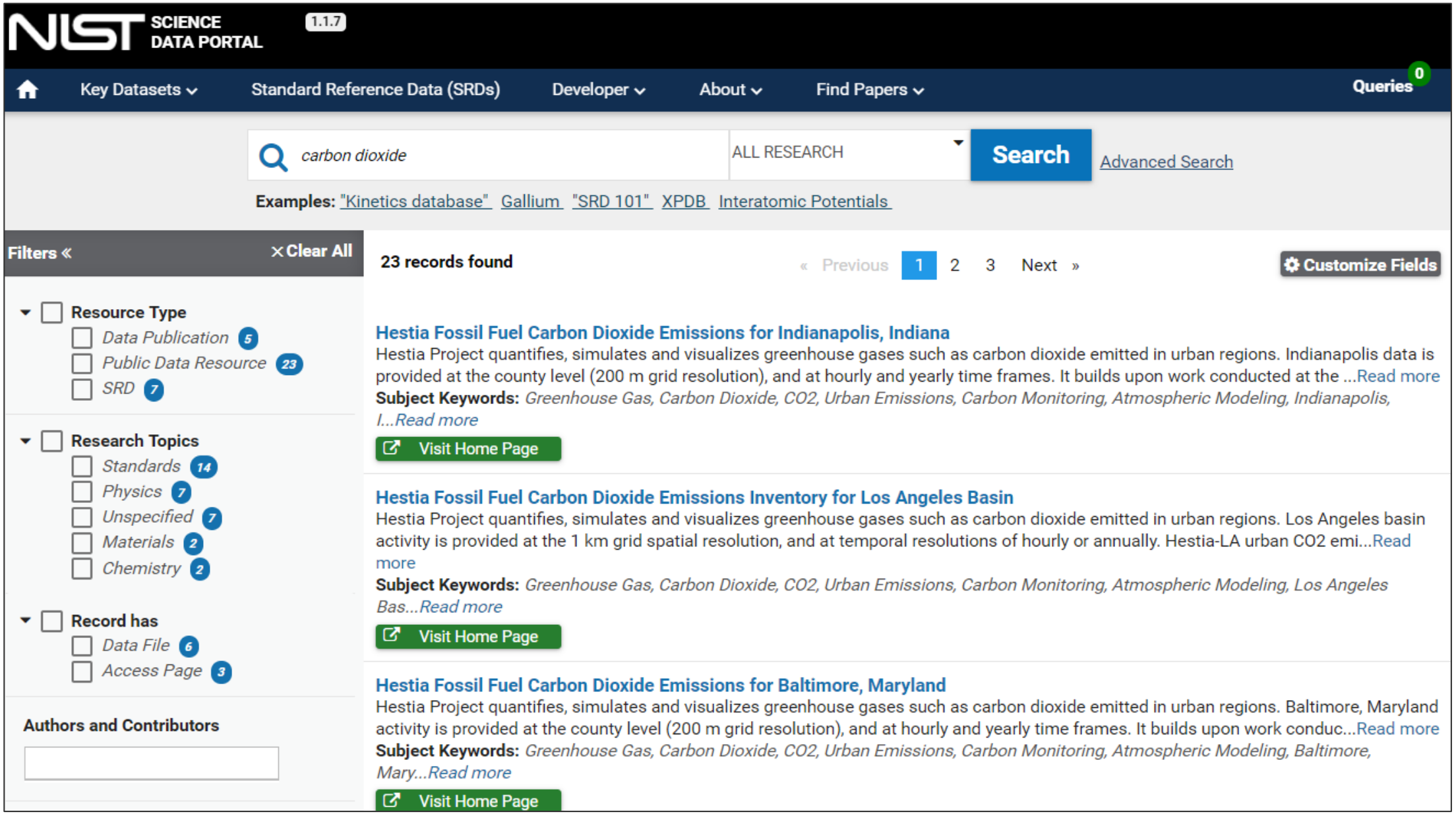
Faceted Filters for Research Topic

- *Based on the **NIST Taxonomy***

New Data Publications via MIDAS

- **“Management of Institutional Data Assets”**




NIST SCIENCE DATA PORTAL
1.1.7
Key Datasets
Standard Reference Data (SRDs)
Developer
About
Find Papers
Queries 0
carbon dioxide
ALL RESEARCH
Search
Advanced Search
Examples: "Kinetics database" Gallium "SRD 101" XPDB Interatomic Potentials
Filters «
Clear All
23 records found
« Previous 1 2 3 Next »
Customize Fields
Resource Type
Data Publication 5
Public Data Resource 23
SRD 7
Research Topics
Standards 14
Physics 7
Unspecified 7
Materials 2
Chemistry 2
Record has
Data File 6
Access Page 3
Authors and Contributors
Hestia Fossil Fuel Carbon Dioxide Emissions for Indianapolis, Indiana
Hestia Project quantifies, simulates and visualizes greenhouse gases such as carbon dioxide emitted in urban regions. Indianapolis data is provided at the county level (200 m grid resolution), and at hourly and yearly time frames. It builds upon work conducted at the ...Read more
Subject Keywords: Greenhouse Gas, Carbon Dioxide, CO2, Urban Emissions, Carbon Monitoring, Atmospheric Modeling, Indianapolis, I...Read more
Visit Home Page
Hestia Fossil Fuel Carbon Dioxide Emissions Inventory for Los Angeles Basin
Hestia Project quantifies, simulates and visualizes greenhouse gases such as carbon dioxide emitted in urban regions. Los Angeles basin activity is provided at the 1 km grid spatial resolution, and at temporal resolutions of hourly or annually. Hestia-LA urban CO2 emi...Read more
Subject Keywords: Greenhouse Gas, Carbon Dioxide, CO2, Urban Emissions, Carbon Monitoring, Atmospheric Modeling, Los Angeles Bas...Read more
Visit Home Page
Hestia Fossil Fuel Carbon Dioxide Emissions for Baltimore, Maryland
Hestia Project quantifies, simulates and visualizes greenhouse gases such as carbon dioxide emitted in urban regions. Baltimore, Maryland activity is provided at the county level (200 m grid resolution), and at hourly and yearly time frames. It builds upon work conduc...Read more
Subject Keywords: Greenhouse Gas, Carbon Dioxide, CO2, Urban Emissions, Carbon Monitoring, Atmospheric Modeling, Baltimore, Mary...Read more
Visit Home Page

Metadata and Vocabulary Standards for NIST Data

This page provides pointers to the various metadata and vocabulary standards used to serve and manage NIST data as part of NIST's open data initiative (Open Access to Research).

Table of Contents

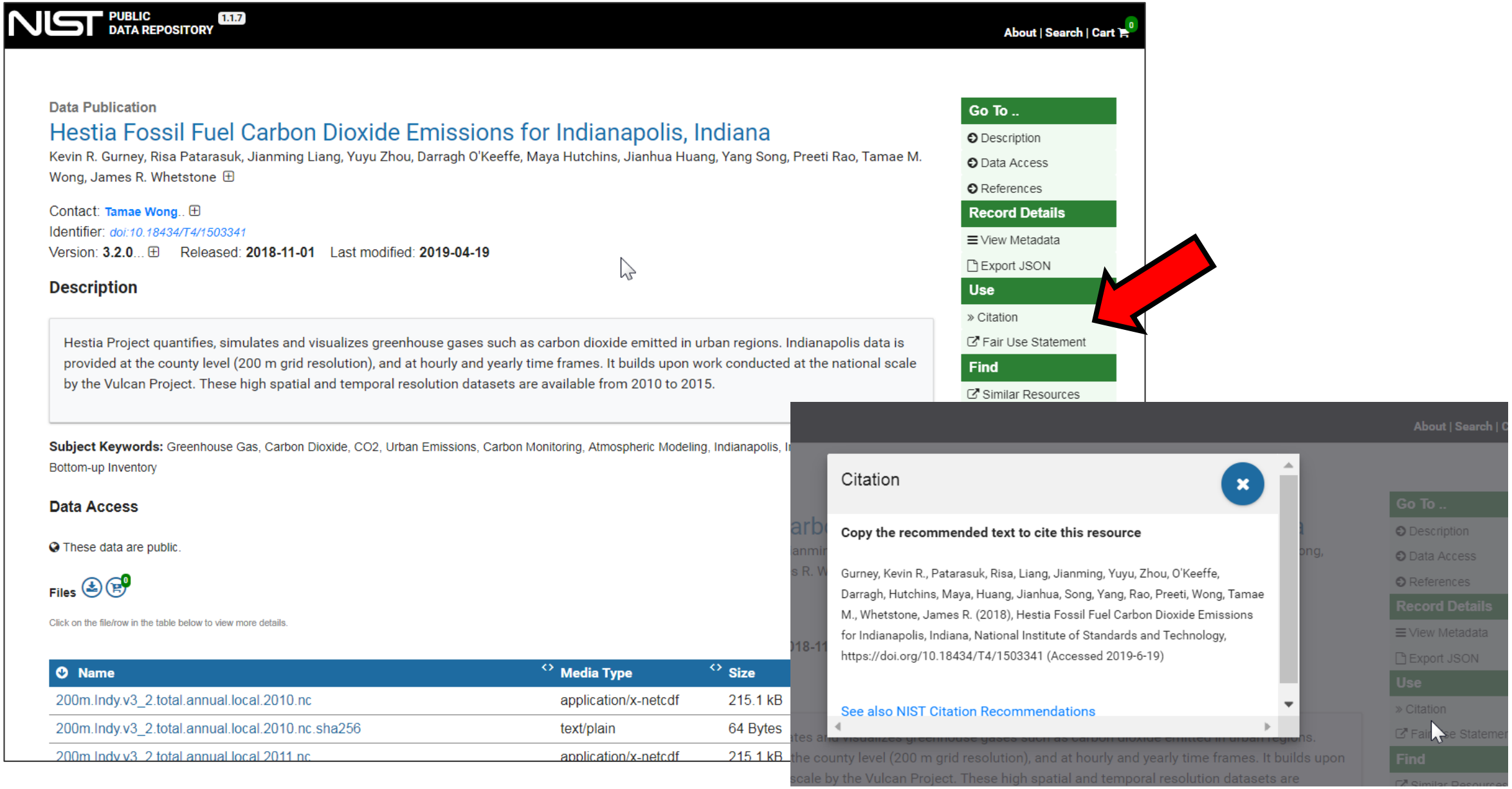
- NIST Extensible Resource Data Model (NERDm)
- Project Open Data (POD)

NIST Extensible Resource Data Model (NERDm)

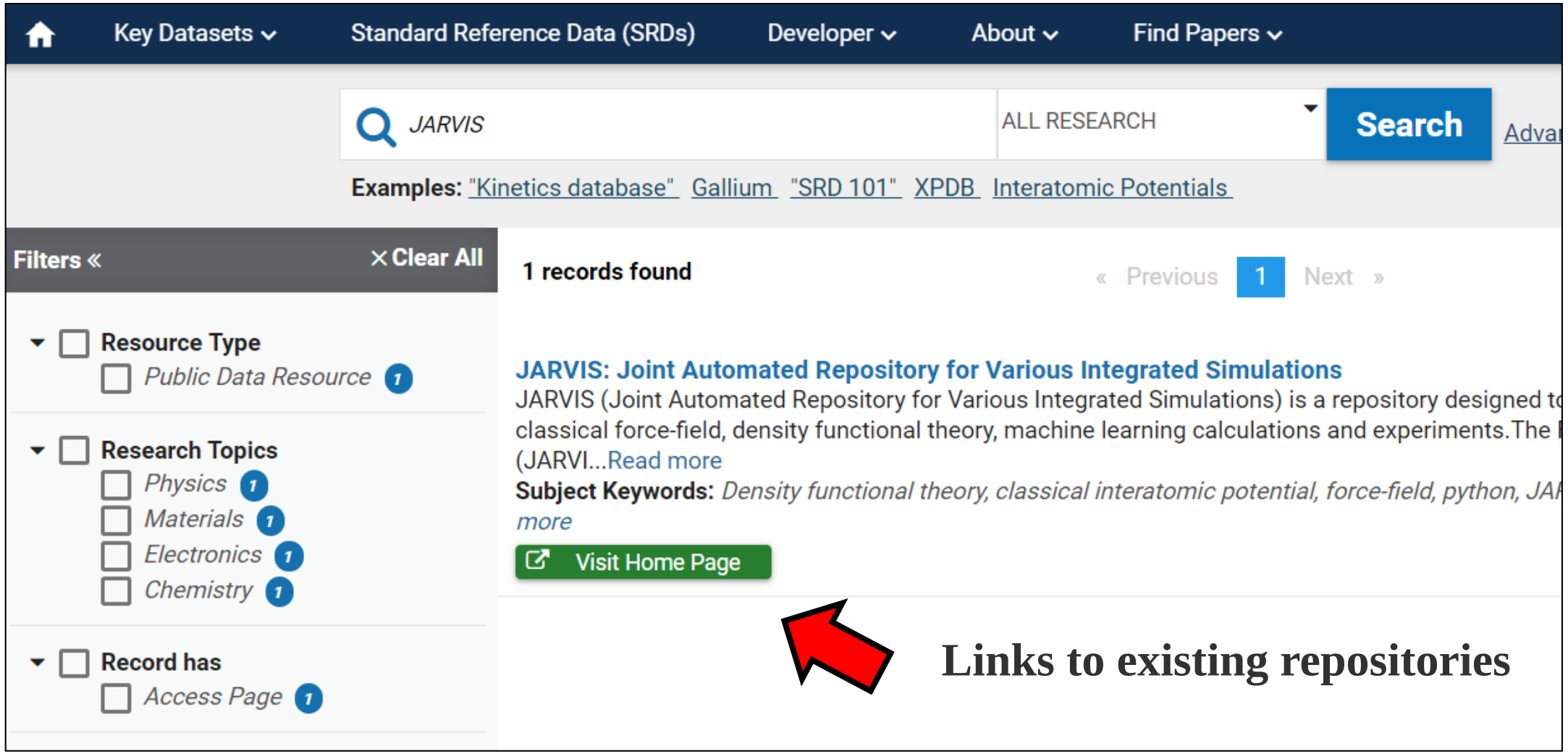
The NIST Extensible Resource Data Model (NERDm) refers to the JSON-LD-formatted metadata schema used by the NIST Public Data Repository (PDR) and Science Data Portal internally to describe data resources available from NIST. It is defined using JSON Schema.

Resources

- A Reader's Guide to NERDm Metadata
- JSON Schema definition files:
  - NERDm Core (latest): nerdm-schema-0.2.json (v0.1)
  - Extension: Publications (latest): nerdm-schema-0.2.json (v0.1)
- JSON-LD context files:
  - nerdm-schema-context.jsonld


NIST PUBLIC DATA REPOSITORY 1.1.7
About | Search | Cart 0
Data Publication
Hestia Fossil Fuel Carbon Dioxide Emissions for Indianapolis, Indiana
Kevin R. Gurney, Risa Patarasuk, Jianming Liang, Yuyu Zhou, Darragh O'Keeffe, Maya Hutchins, Jianhua Huang, Yang Song, Preeti Rao, Tamae M. Wong, James R. Whetstone
Contact: Tamae Wong..
Identifier: doi:10.18434/T4/1503341
Version: 3.2.0... Released: 2018-11-01 Last modified: 2019-04-19
Description
Hestia Project quantifies, simulates and visualizes greenhouse gases such as carbon dioxide emitted in urban regions. Indianapolis data is provided at the county level (200 m grid resolution), and at hourly and yearly time frames. It builds upon work conducted at the national scale by the Vulcan Project. These high spatial and temporal resolution datasets are available from 2010 to 2015.
Subject Keywords: Greenhouse Gas, Carbon Dioxide, CO2, Urban Emissions, Carbon Monitoring, Atmospheric Modeling, Indianapolis, I
Bottom-up Inventory
Data Access
These data are public.
Files
Click on the file/row in the table below to view more details.
Name
Media Type
Size
200m.Indy.v3_2.total.annual.local.2010.nc
application/x-netcdf
215.1 kB
200m.Indy.v3_2.total.annual.local.2010.nc.sha256
text/plain
64 Bytes
200m.Indy.v3_2.total.annual.local.2011.nc
application/x-netcdf
215.1 kB
Go To ..
Description
Data Access
References
Record Details
View Metadata
Export JSON
Use
» Citation
Fair Use Statement
Find
Similar Resources
Citation
Copy the recommended text to cite this resource
Gurney, Kevin R., Patarasuk, Risa, Liang, Jianming, Yuyu, Zhou, O'Keeffe, Darragh, Hutchins, Maya, Huang, Jianhua, Song, Yang, Rao, Preeti, Wong, Tamae M., Whetstone, James R. (2018), Hestia Fossil Fuel Carbon Dioxide Emissions for Indianapolis, Indiana, National Institute of Standards and Technology, https://doi.org/10.18434/T4/1503341 (Accessed 2019-6-19)
See also NIST Citation Recommendations

Example Search: “JARVIS”



Public Data Resource

JARVIS: Joint Automated Repository for Various Integrated Simulations

Contact: Kamal Choudhary..

Identifier: doi:10.18434/M3HQ1W

Last modified: **2017-10-18**

Visit Home Page

Go To ..
Description
Data Acces
References
Record De
View Metad
Export JSO
Use
» Citation
Fair Use St
Find
Similar Res
Resources

Description

JARVIS (Joint Automated Repository for Various Integrated Simulations) is a repository designed to automate materials discovery using classical force-field, density functional theory, machin rning calculations and experiments.The Force-field section of JARVIS (JARVIS-FF) consists of thousands of automated LAMMPS based force-field calculations on DFT geometries. Some of the pr ties included in JARVIS-FF are energetics, elastic constants, surface energies, defect formations energies and phonon frequencies of materials.The Density functional theory section of JARVIS (VIS-DFT) consists of thousands of VASP based calculations for 3D-bulk, single layer (2D), nanowire (1D) and molecular (0D) systems. Most of the calculations are carried out with optB88vDW functi JARVIS-DFT includes materials data such as: energetics, diffraction pattern, radial distribution function, band-structure, density of states, carrier effective mass, temperature and carrier concentrat ependent thermoelectric properties, elastic constants and gamma-point phonons.The Machine-learning section of JARVIS (JARVIS-ML) consists of machine learning prediction tools, trained on JAR DFT data. Some of the ML-predictions focus on energetics, heat of formation, GGA/METAGGA bandgaps, bulk and shear modulus.

Research Topics: Physics: Condensed matter , Materials: Modeling and computational material science , Electronics: Thin-film electronics , Electronics: Optoelectronics , Chemistry: Molecular characterization , istry: Theoretical chemistry and modeling , Chemistry: Chemical thermodynamics and chemical properties , Electronics: Semiconductors , Materials: Materials characterization , Physics: Optical physics

Subject Keywords: Density functional theory, classical interatomic potential, force-field, python, JARVIS, MGI, MDCS, RESTAPI, automation

Data Access

These data are public. For more information, please visit the home page.

Data is available via the following locations:

- JARVIS for DFT
- JARVIS for Force-fields

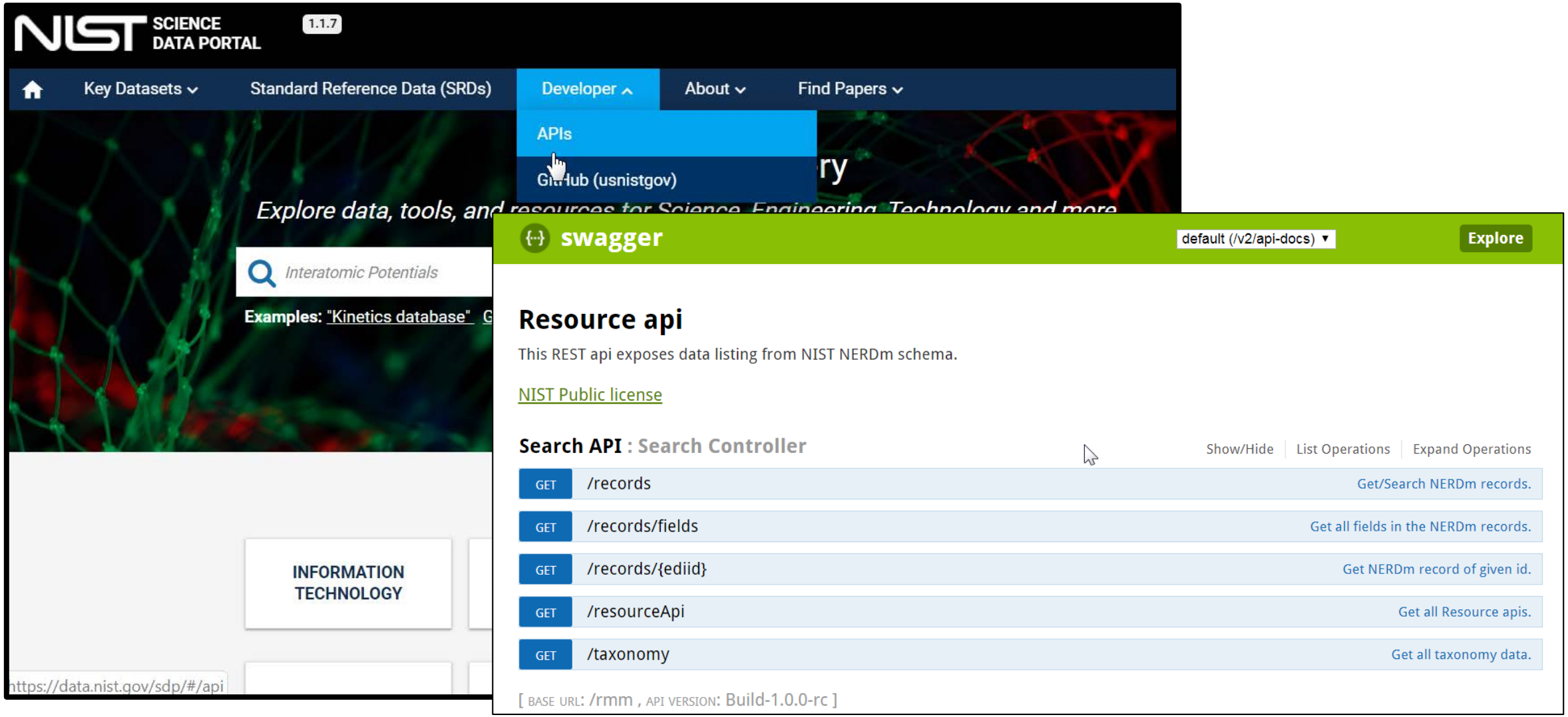
References

This data is referenced in :

- https://www.nature.com/articles/sdata2016125
- https://www.nature.com/articles/s41598-017-05402-0

Customized Links to existing repositories

Links to Scholarly Publications auto-generated


NIST SCIENCE DATA PORTAL
1.1.7
Key Datasets
Standard Reference Data (SRDs)
Developer
About
Find Papers
APIs
GitHub (usnistgov)
Explore data, tools, and resources for Science, Engineering, Technology and more
Interatomic Potentials
Examples: "Kinetics database"
INFORMATION TECHNOLOGY
https://data.nist.gov/sdp/#/api
swagger
default (/v2/api-docs)
Explore
Resource api
This REST api exposes data listing from NIST NERDm schema.
NIST Public license
Search API : Search Controller
Show/Hide
List Operations
Expand Operations
GET /records Get/Search NERDm records.
GET /records/fields Get all fields in the NERDm records.
GET /records/{ediid} Get NERDm record of given id.
GET /resourceApi Get all Resource apis.
GET /taxonomy Get all taxonomy data.
[BASE URL: /rmm , API VERSION: Build-1.0.0-rc]

Access

Options for Access include Direct Download

Or via Data Cart

Data Access

These data are public.

Files

Click on the file/row in the table below to view more details.

Total No. files: 56

Name	Media Type	Size	Status
200m.Indy.v3_2.total.annual.local.2010.nc	application/x-netcdf	215.1 kB	Added
200m.Indy.v3_2.total.annual.local.2010.nc.sha256	text/plain	64 Bytes	Added
200m.Indy.v3_2.total.annual.local.2011.nc			
200m.Indy.v3_2.total.annual.local.2011.nc.sha256			
200m.Indy.v3_2.total.annual.local.2012.nc			
200m.Indy.v3_2.total.annual.local.2012.nc.sha256			
200m.Indy.v3_2.total.annual.local.2013.nc			
200m.Indy.v3_2.total.annual.local.2013.nc.sha256			
200m.Indy.v3_2.total.annual.local.2014.nc			
200m.Indy.v3_2.total.annual.local.2014.nc.sha256			
200m.Indy.v3_2.total.annual.local.2015.nc			
200m.Indy.v3_2.total.annual.local.2015.nc.sha256			
200m.Indy.v3_2.total.annual.UTC.2010.nc			
200m.Indy.v3_2.total.annual.UTC.2010.nc.sha256			

Go To ..

- Description
- Data Access
- References

Record Details

- View Metadata
- Export JSON

Use

- » Citation

NIST PUBLIC DATA REPOSITORY 1.1.7

About | Search | Cart 3

Download Manager

Zip File Name	Download Status
download56035-1.zip	✔ downloaded

Download Selected 3 | Remove downloaded 3 | Remove selected 3

Name	Media Type	Size	Status
Hestia Fossil Fuel Carbon Dioxide Emissions for Indianapolis, Indiana			
200m.Indy.v3_2.total.annual.local.2010.nc *download56035-1.zip*	application/x-netcdf	215.1 kB	downloaded
200m.Indy.v3_2.total.annual.local.2010.nc.sha256 *download56035-1.zip*	text/plain	64 Bytes	downloaded
200m.Indy.v3_2.total.annual.local.2011.nc *download56035-1.zip*	application/x-netcdf	215.1 kB	downloaded

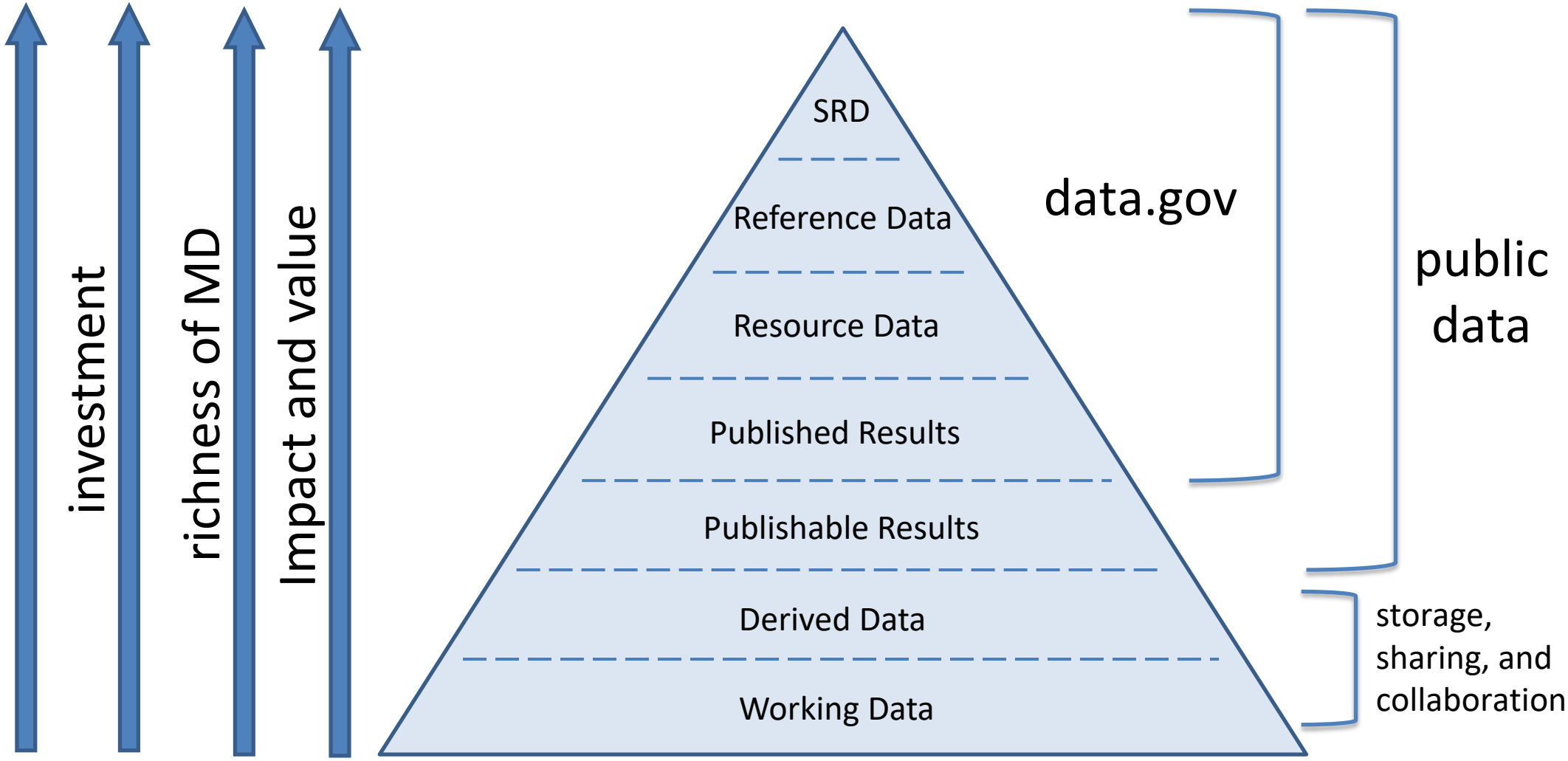
NIST National Institute of Standards and Technology U.S. Department of Commerce

HEADQUARTERS

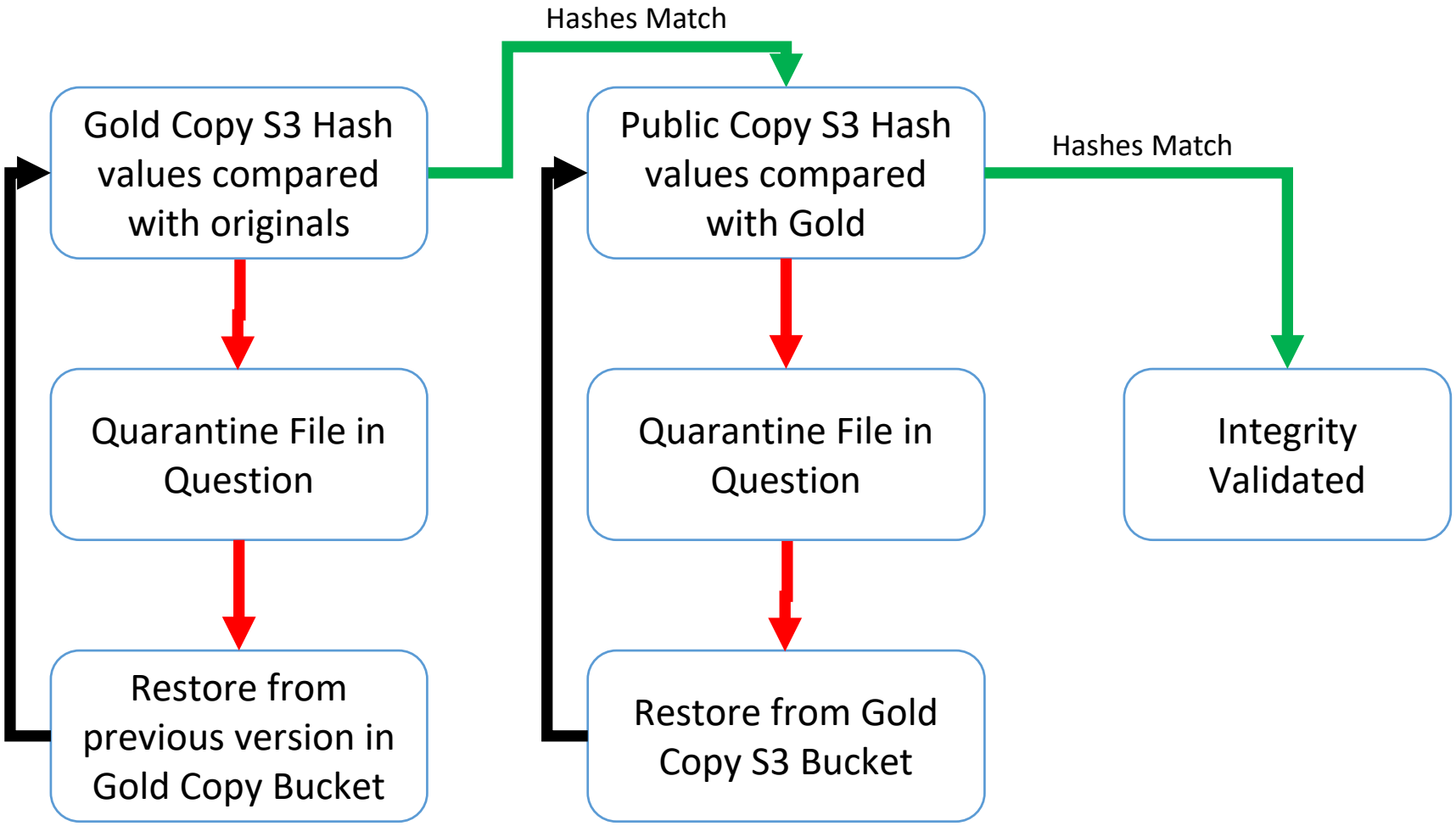
download56035-1.zip

Show all

NIST Public Data Access Policy



Data Integrity



- Integrity checked once per day
- Files are quarantined and reverted back to the previous version if an issue is detected

Future Enhancements Planned

- Metrics for the NIST Public Data patterns of use
  - “Data as a First-Class Research Output” (https://makedatacount.org/)
  - Researchers can see how many Page Views or Downloads their data
- Enhance machine readability of datasets for AI
- Public software & analysis tool linked integration
- Large-scale datasets – they already exist as a challenge!
  - Evaluating Globus (community demonstrated with archive systems for public access)
- Search – auto suggest enhance with Data Driven index
  - E.g. natural language processing of the metadata (e.g. rules-based ontology)

Interoperability & Reusability

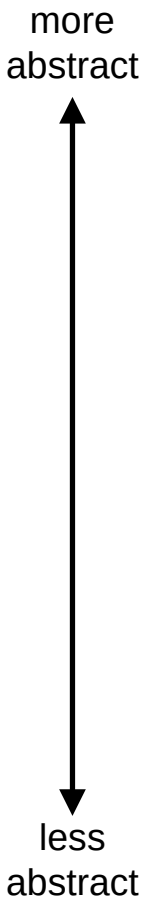
Part of a Larger Interoperability Problem

Long-tail research data is limited by lack of :

- Interoperability
- Provenance
- Automated reasoning

Experimentalists, instrument operators, and modeling/simulation researchers face interoperability challenges at many levels of abstraction

As you all know, fixing these problems is *hard.*



level	meaning	examples
ontology	concepts, relationships	samples, observations
metadata semantics	concepts, taxonomy	spatial scale, affine xforms
metadata syntax	terms, controlled vocabulary	magnification *vs.* horizontal field width (HFW)
data shapes	dimensions, data types	img focus series *vs.* SEM-EDS map
on-disk serialization	dataset file formats	Gatan .dm3 TIFF 6.0

Research Data Alliance (RDA) Activities

https://rd-alliance.org

RDA/CODATA Materials Data, Infrastructure & Interoperability IG

“Materials IG”

Chairs

James Warren (NIST)
Laura Bartolo (Northwestern)
Takuya Kadohira (NIMS)
Adham Hashibon (Fraunhofer IWM)
Alysia Garmulewicz (Universidad de Santiago de Chile)

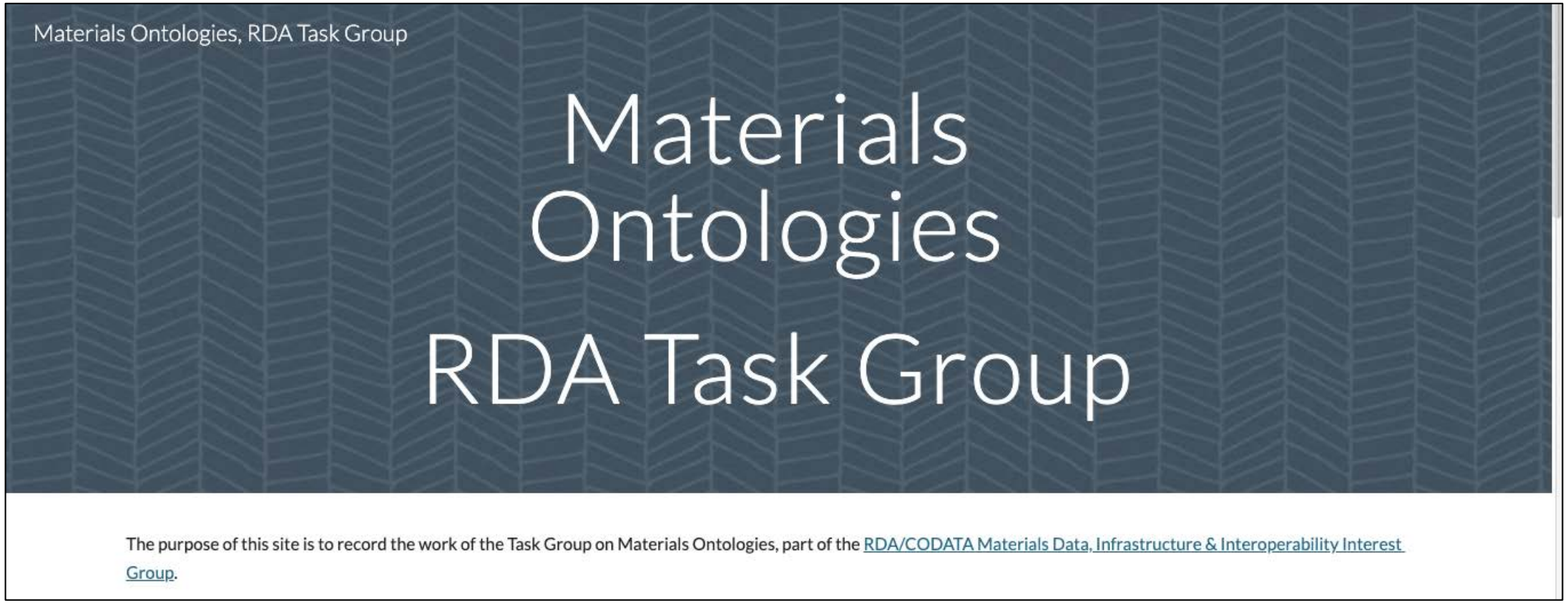
Materials Ontologies Task Group

Chairs

Gerhard Goldbeck (Goldbeck Consulting)
Clare Paul (Air Force Research Lab)
John Henry Scott (NIST)

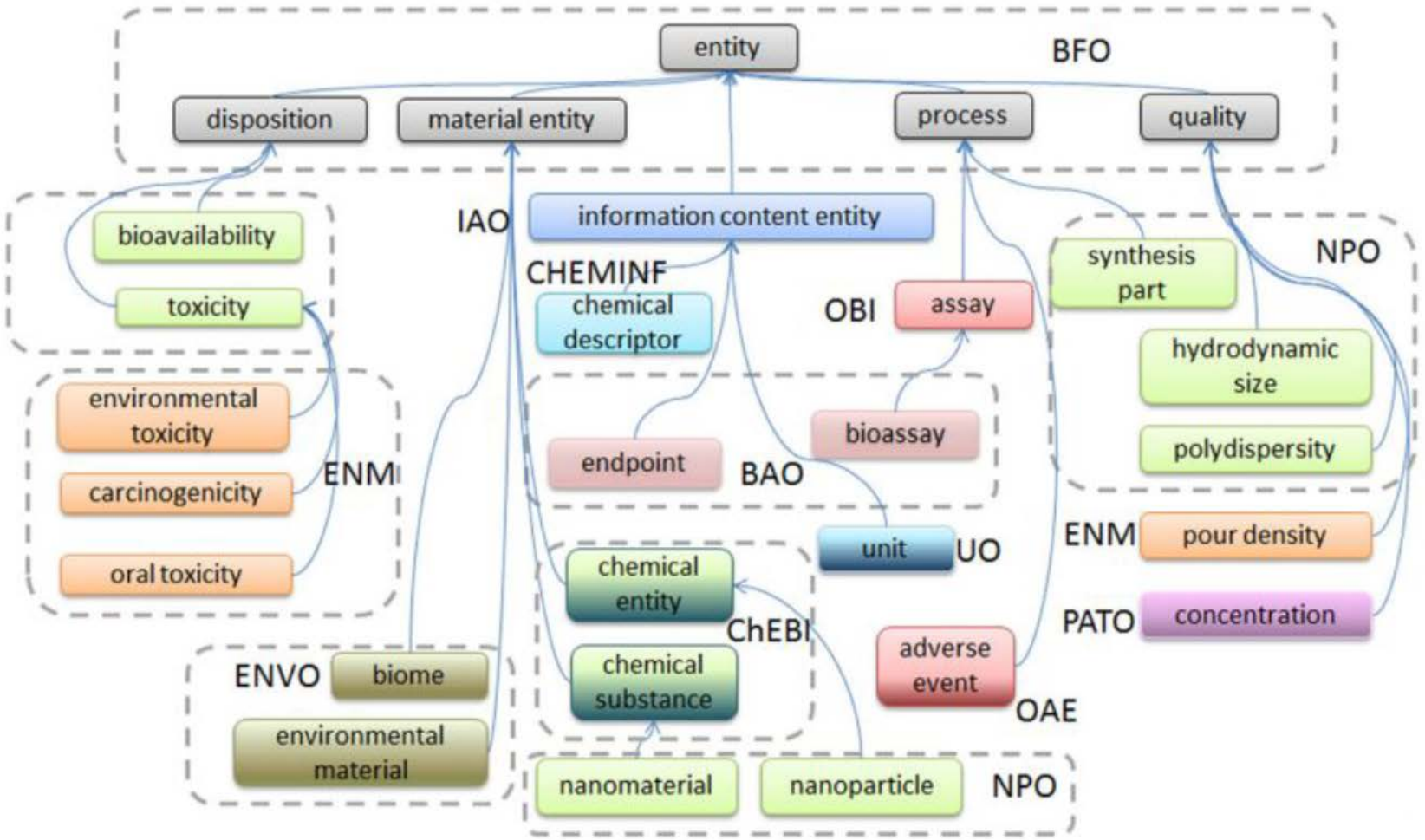
1. **Initial Call: 13 March 2019**
2. **Task Group Approved: 04 Apr 2019 @ P13**
3. **1st Telecon: 25 May 2019**

RDA Materials Ontologies Task Group



https://sites.google.com/view/rda-materials-ontologies-tg/home

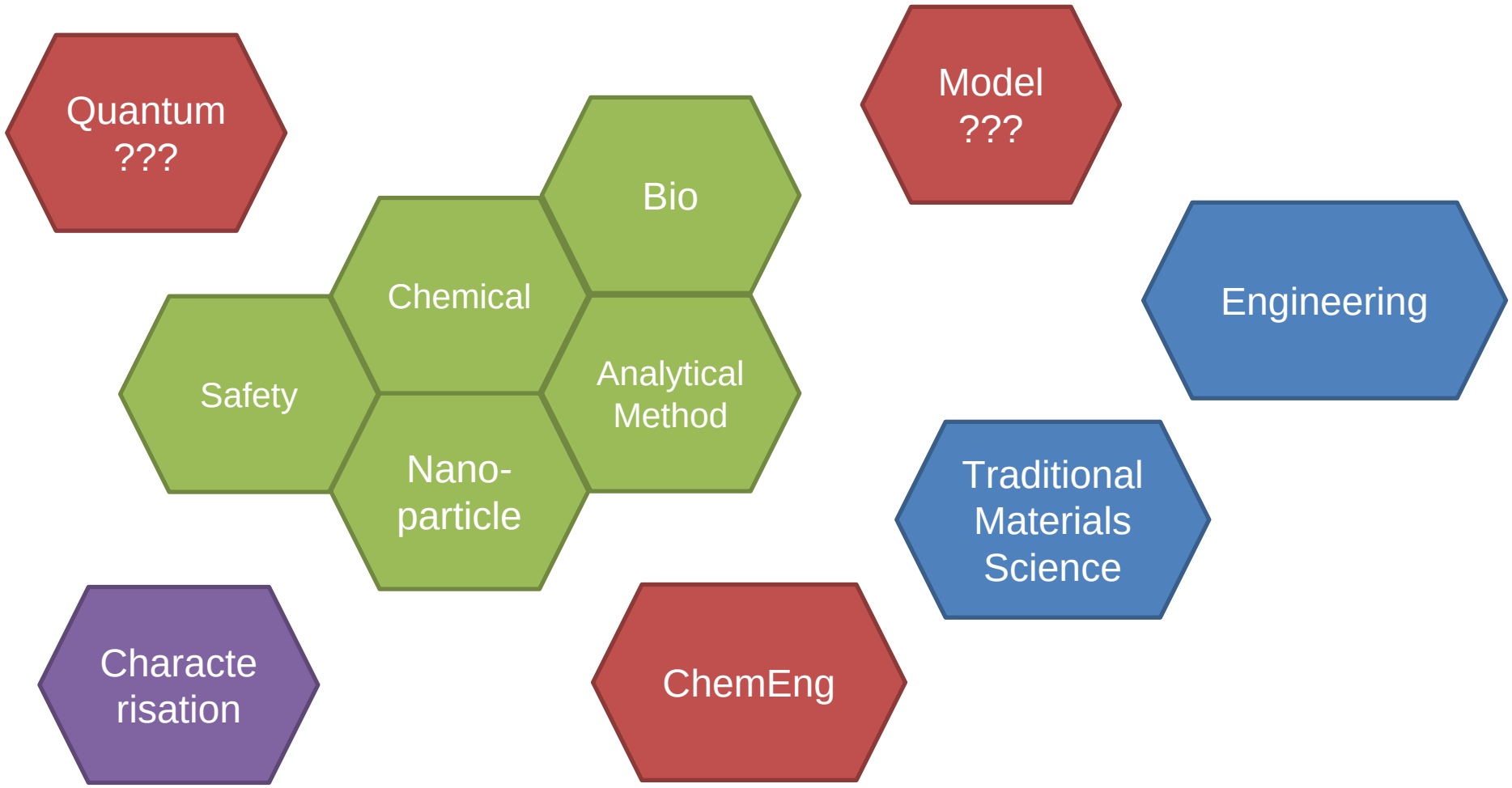
Balkanization of materials ontologies



Courtesy of

Slide courtesy: Gerhard Goldbeck, after Egon Willighagen, Univ Maastricht

Current Status



Slide courtesy: Gerhard Goldbeck

Where do we want to get to?

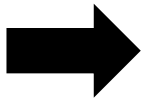
Slide courtesy: Gerhard Goldbeck

Research Data Alliance (RDA) Activities

RDA Vocabulary Services IG

Chairs

Adam Shepherd
Simon Cox
Yann Le Franc
John Graybeal



Semantic Assets for Materials Science Task Group

Chairs

Vasily Bunakov (STFC)

Zach Trautt (NIST)

RDA Heritage Science @ P13

Heritage Science Research Data: Challenges of Linked Open Data Authoritative Sources

Fenella France (LoC)
Andrew Forsberg (LoC)
John Henry Scott (NIST)

Research Data Alliance (RDA) Activities

Schema Repository and Registry

Part of the Materials Genome Initiative

https://schemas.nist.gov

Zach Trautt

The EMMC Taxonda Registry

EMMC

THE EUROPEAN MATERIALS MODELLING COUNCIL

Recent NIST Metadata Workshops

CHiMaD

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

WORKSHOP 5

2018 NIST/CHiMaD Materials Microscopy Data Conference

October 25-26, 2018 (Northwestern University)

Program, videos, all talks online at:

http://chimad.northwestern.edu/news-events/CHiMaD_Data_Database_Efforts.html

NIST

Recent NIST Metadata Workshops

CHiMaD

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

WORKSHOP 6

2019 NIST/CHiMaD Video Conference

***Workflow*, June Lau, April 9, 1:30-3:00PM ET**

To learn more about the conference click here.

***Vocabulary/Metadata*, John Henry Scott, March 19, 1:30-3:00PM ET**

To learn more about the conference click here.

***Sample History*, Carelyn Campbell, March 5, 1:30-3:00PM ET**

All content online at:

http://chimad.northwestern.edu/news-events/CHiMaD_Data_Database_Efforts.html

NIST

Recent NIST Metadata Workshops

CHiMaD

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

WORKSHOP 7

2019 NIST/CHiMaD Materials Microscopy Data Conference

May 15-16, 2019 (NIST National Cybersecurity Center of Excellence)

All content online at:

http://chimad.northwestern.edu/news-events/CHiMaD_Data_Database_Efforts.html

NIST

Recent NIST Metadata Workshops

Microstructure Repository and Metadata Workshop: May 13-14

Lehigh University, the National Institute for Standards and Technology (NIST), and the National Science Foundation will be sponsoring a workshop from mid-day May 13, 2019 – May 14, 2019 to address key issues involved in materials data curation, assess current best practices, and develop a plan for establishing future protocols for widely accessible data repositories. This workshop will serve to create a new microstructures data repository that will enable researchers to share, access, and analyze images and associated data from multiple platforms. In preparation for this workshop, we held a preliminary meeting last November to discuss potential schema and the curation of metadata as related to electron micrographs and associated data. As a result of this November meeting, we have started construction of the repository as a testbed to evaluate proposed schema.

Program online at:

https://www.nist.gov/document/agenda-microstructure-workshop-may-13-14-2019

Thank You

please contact me if you want these slides,
or if you have questions (or suggestions!)

John Henry Scott
NIST
+1 301-975-4981

johnhenry.scott@nist.gov
@js0b
ORCID: https://orcid.org/0000-0002-3796-2110
ResearcherID: B-4593-2015