

DOE Office of Scientific and Technical Information (OSTI) Artificial Intelligence and Machine Learning

DOE Data Days (D3)
June 2, 2022

Mary Beth West



U.S. DEPARTMENT OF
ENERGY

Office of
Science

Office of Scientific and
Technical Information



DOE Invests ~\$15B per Year in R&D



R&D Funding

- Advanced Research Projects Agency – Energy (ARPA-E)
- Office of Electricity Delivery & Energy Reliability
- Office of Energy Efficiency & Renewable Energy
- Office of Environmental Management
- Office of Fossil Energy
- Office of Legacy Management
- Office of Nuclear Energy
- Office of Science
- Office of Environment, Health, Safety & Security

NATIONAL LABS

Ames
Argonne
Brookhaven
Fermi
Idaho
Los Alamos
Lawrence Berkeley
Lawrence Livermore
NETL
NREL
Oak Ridge
Pacific Northwest
Princeton
SLAC
Sandia
Savannah River
Thomas Jefferson

GRANTEES
TECHNOLOGY CENTERS
SITES

R&D Results

- Journal articles/accepted manuscripts
- Technical reports
- Conference papers
- Theses/dissertations
- Scientific and technical software
- Data objects
- Patents
- Workshop reports
- Videos

≈ 50,000 R&D results per year



Main Drivers

OSTI's Mission: Advance science and sustain technological creativity by making R&D findings available and useful to Department of Energy researchers and the public.

The [Office of Scientific and Technical Information](#) (OSTI) collects, preserves, and disseminates DOE-funded R&D results, both through OSTI search tools and through other commonly-used search engines.

OSTI.GOV

Primary search tool for all DOE-funded R&D results

- Over 3.3 M metadata records; 800K full-text/resource available
- Semantic search “keyword to concept” mapping; AI potential

DEPARTMENT OF ENERGY
DOE PAGES
Public Access Gateway for Energy & Science

DOE Data Explorer

DOE **CODE**

DOE Patents

DOE ScienceCinema



AI / Machine Learning for R&D Results

R&D Results Data: Opportunities

- > 3.3M scientific and R&D results emanating from DOE-funded research and development (R&D) activities.
- Many types of research products represented, including journal articles, datasets, patents, software, technical reports.
- Data spans DOE research interests, research organizations, and sponsoring organizations.
- Metadata is searchable and available electronically, API / export use cases supported.

OSTI mission data is increasingly targeted as a valuable source of bulk metadata for analysis, machine learning exercises, and text mining.



AI / Machine Learning for R&D Results

R&D Results Data: Challenges

- R&D results are being generated more rapidly and in more varying forms.
- Access patterns are changing, conventional users / searches are no longer the only first-class use cases.
- New usage requirements for search results / formats are emerging.
- Existing data, structures, and formats don't always lend well to analysis at scale.
- Association of related data is increasingly critical for research and reproducibility.

OSTI data and procedures need to support comprehensive, FAIR collections.



AI / Machine Learning for R&D Results

R&D Results Data: Addressing Challenges

- Improve ingest processes of R&D results to handle scale of direct submission (i.e., other alternatives for submission, such as Web Services and APIs for ingest).
- Data integrity for R&D results and their products.
- Mechanisms to support the reuse of data.
- Procedural improvements to augment tools that standardize reporting and analysis (i.e., additional metadata dimensions will allow for better topic / trend / funding analysis).



Data Opportunity: DOE OSTI Organization Authority

- DOE OSTI ingests and curates organization information provided about DOE-funded research outputs – including author affiliations, research organizations, and funding organizations.
- We developed an internal organization authority to standardize organization naming during metadata submission and within in the research output records in our search tools.
- Organization authority allows for enhancement of user submitted metadata. Providing consistent naming structures, enhanced organization labeling, and hierarchical mapping.

United States Department of Energy
Energy Link System (E-Link)
DOE STI Management System

Authorities

These authority lists are established and used by DOE to standardize information in bibliographic records. Language, sponsoring organizations, originating research organizations, subject categories, and country of publication must be entered in required formats to optimize search and retrieval. The DOE Order 241.1B requires that DOE-funded STI include attribution to the sponsoring program with wording including DOE, the DOE program office, and the program office's subprogram(s) funding the research. Using the sponsor listings as shown in this Authority will meet that requirement.

Authority:

Current Sponsoring Organizations	
AITO	USDOE Office of Artificial Intelligence and Technology (AITO)
ARPA-E	USDOE Advanced Research Projects Agency - Energy (ARPA-E)
AU	USDOE Office of Environment, Health, Safety and Security (AU)

1. Product Description

2. Product Type Info

3. Authors

4. Content

5. Related Documents

Content

Contributor Organization(s) ?
Enter names of organizations that provided significant assistance but not direct funding

Report/Product Number ?
None

* Sponsoring DOE Program Office ?
[Select Primary First]

(Select Primary First)
Current Sponsoring Organizations

- USDOE
- USDOE Office of Artificial Intelligence and Technology (AITO)
- USDOE Advanced Research Projects Agency - Energy (ARPA-E)
- USDOE Office of Environment, Health, Safety and Security (AU)
- USDOE Office of Environment, Health, Safety and Security (AU), Office of Corporate Security Strategy
- USDOE Office of Environment, Health, Safety and Security (AU), Office of Health and Safety
- USDOE Office of Environment, Health, Safety and Security (AU), Office of Environmental Protection and E&SH Reporting
- USDOE Office of Environment, Health, Safety and Security (AU), Office of Nuclear Safety
- USDOE Office of Environment, Health, Safety and Security (AU), Office of Security
- USDOE Office of Environment, Health, Safety and Security (AU), Office of Classification
- USDOE Office of Cybersecurity, Energy Security and Emergency Response (CESER)
- USDOE Office of Cybersecurity, Energy Security and Emergency Response (CESER), Infrastructure Security and Energy Restoration (ISER)
- USDOE Office of Congressional and Intergovernmental Affairs (CI)
- USDOE Office of Congressional and Intergovernmental Affairs (CI), Energy Policy (CI-30)
- USDOE Office of Enterprise Assessments (EA)
- USDOE Office of Energy Efficiency and Renewable Energy (EERE)
- USDOE Office of Energy Efficiency and Renewable Energy (EERE), Transportation Office
- USDOE Office of Energy Efficiency and Renewable Energy (EERE), Transportation Office, Bioenergy Technologies Office

* Required



Organization Disambiguation

Turn these ...

- Oak Ridge National Lab. (ORNL), Oak Ridge, TN (United States)
- Oak Ridge National Laboratory, Oak Ridge, TN (United States)
- ORNL
- Oak Ridge Nat. Lab
- Oak Ridge National Lab. (ORNL)
- Oak Ridge National Lab., Tenn.
- etc. ...



... into this

Name
Oak Ridge National Laboratory

Type
Facility

Latitude
-84.31700134277344

Longitude
35.93299865722656

Parent(s)
United States
Department of Energy

Congressional District
TN 3

Funder Registry ID
100006228

GRID ID
grid.135519.a
etc...

Drivers

- ✓ Improves the findability of research outputs by enhancing our search tools through normalization of metadata.
- ✓ Enables organization and sector-level analysis of research outputs.
- ✓ Supports enhancements to organization data that enable meaningful analytics (e.g., geolocation, congressional district).
- ✓ Makes data more interoperable by supporting organization data with persistent identifiers.
- ✓ Supports additional metadata for other research objects (e.g., award DOIs, ORCID iDs, etc.).

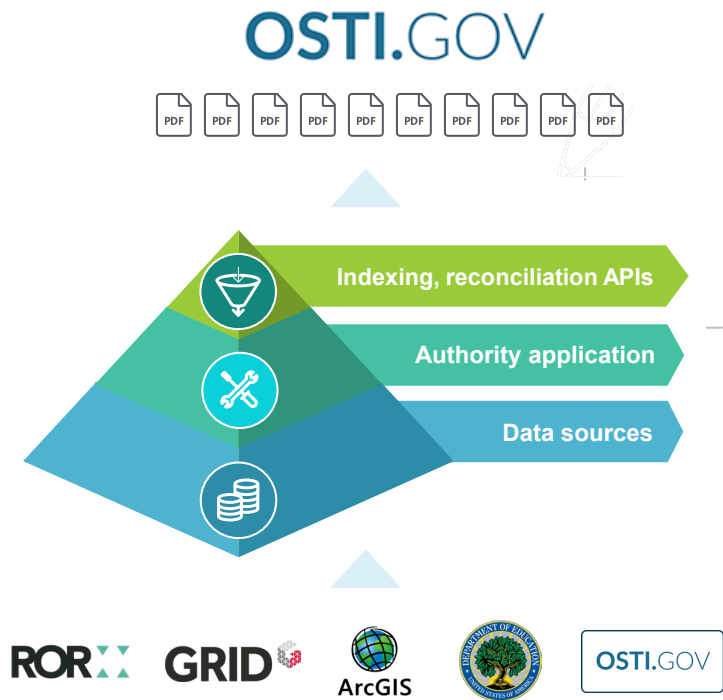


U.S. DEPARTMENT OF
ENERGY

Office of
Science

Office of Scientific and
Technical Information

Organization Disambiguation



How it works ...

- Organization data is aggregated from multiple authoritative sources, including ROR, GRID, OSTI authorities, and others.
- Authority application developed to maintain / query authoritative references and extended metadata.
- Disambiguation tools (indexing, reconciliation APIs) leverage NLP, similarity, and authority info to resolve variations and assign authoritative entity data to variant organization names.
- Normalized labels and enhanced organization metadata are made available to downstream processes (e.g. submission and search tools).



U.S. DEPARTMENT OF
ENERGY

Office of
Science

Office of Scientific and
Technical Information

Organization Authority and Reconciliation API

Organization Name
Aliases

Organization Type

Geographic Information

Organization PIDs

Details for: Oak Ridge National Laboratory

Organization ID: 2139

Status: Active

Aliases: Oak Ridge National Lab. (ORNL), Oak Ridge, TN (United States)
Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States)
Oak Ridge National Laboratory, Oak Ridge, TN (United States)
ORNL
Oak Ridge Nat. Lab
Oak Ridge National Lab. (ORNL)
Oak Ridge National Lab., Tenn.
Oak Ridge National Lab., TN (USA)
Oak Ridge National Lab., TN (United States)
Oak Ridge National Laboratory, Oak Ridge, TN (US)
Oak Ridge National Laboratory (ORNL)
Oak Ridge National Lab., TN
Oak Ridge National Lab., Tenn. (USA)
Oak Ridge National Lab., TN (US)
Oak Ridge National Laboratory (ORNL), Oak Ridge, TN
Oak Ridge National Laboratory
ORNL (Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States))
ORNL, Oak Ridge National Laboratory (US)
LOCKHEED MARTIN ENERGY SYST INC
Oak Ridge National Lab.
Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830
Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831
Oak Ridge National Lab., Oak Ridge, TN (United States)
Oak Ridge National Lab. (ORNL), TN (United States)
LOCKHEED MARTIN ENERGY RES COR
Lockheed Martin Energy Research Corporation
Lockheed Martin Energy Syst Inc

Acronyms: ORNL

Types: Facility

Wikipedia URL: http://en.wikipedia.org/wiki/Oak_Ridge_National_Laboratory

Year Established: 1943

Links: <http://www.ornl.gov/>

Longitude/Latitude: -84.3170013, 35.9329987

Address: Oak Ridge, Tennessee
United States

Parent(s): United States Department of Energy
United States

Direct Descendants: [Joint Institute for Computational Sciences](#)

External ID's

Type	Value	Preferred?
FundRef	100006228	Yes
GeoNames City ID	4548571	Yes
GRID	grid.135519.a	Yes
ISNI	0000 0004 0449 2859	Yes
ORGRRef	38147	Yes
ROR	https://ror.org/01qz5mb59	Yes
Wikidata	Q714439	Yes
Wikidata	Q5671313	No



```
{
  "id" : "2139",
  "name" : "Oak Ridge National Laboratory",
  "lat" : "-84.3170013",
  "lng" : "35.9329987",
  "city" : "Oak Ridge",
  "state" : "Tennessee",
  "state_code" : "TN",
  "country_code" : "US",
  "org_type" : "facility",
  "org_acronyms" : "ORNL",
  "parent_id" : "{2052}",
  "ext_ids" : {
    "fundref" : "100006228",
    "isni" : "0000 0004 0449 2659",
    "ror" : "https://ror.org/01qz5mb56",
    "grid" : "grid.135519.a",
    "geonames_city_id" : "4646571",
    "wikidata" : "Q714439",
    "orgref" : "38147"
  }
}
```



Organization Authority and Reconciliation API

Reconciliation API

- The API adheres to the draft reconciliation API spec for query behavior and result formats.
- Query results return confidence scores and match IDs, which can be used to fetch extended organization metadata (e.g., PIDs, sector, MSI status).
- Supports additional query fields (e.g., city, state, country) to improve match resolution and multiple queries per request to enable bulk operations.
- The API integrates with data cleaning and transformation tools like OpenRefine.
- Standard JSON is used for both input and output.

Application Status: Currently in development and testing; anticipate deploying late FY22.



Comments/Questions

Mary Beth West
westm@osti.gov



www.osti.gov
www.osti.gov/ai/



AI@osti.gov
comments@osti.gov



[@osti.gov](https://twitter.com/osti.gov)



U.S. DEPARTMENT OF
ENERGY

Office of
Science

Office of Scientific and
Technical Information