
# DOE Data ID Service

CODATA/ICSTI Data Citation Workshop
July 12, 2016

Carly Robinson, PhD
OSTI Senior Science Advisor/ Product Strategist


Office of Science | Office of Scientific and Technical Information

# DOE Invests $12B per Year in R&D


**R&D Funding**

**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**

**SCIENTIFIC & TECHNICAL INFO**

- Journal articles/accepted manuscripts
- Technical reports
- Conference papers
- Theses/Dissertations
- Scientific and technical computer software
- Datasets
- Patents
- Workshop reports
- Videos

**≈ 50,000 STI "products" per year**

# DOE's Office of Scientific and Technical Information (OSTI)

**DOE-wide responsibility for ensuring access to DOE R&D results.**

- Public access to unclassified, unlimited
- Restricted access to classified and sensitive

## Mission

Advance science and sustain technological creativity by making R&D findings available and useful to Department of Energy (DOE) researchers and the public.

## Core Functions

1. Collect
2. Preserve
3. Disseminate

# DOE Funded Scientific Datasets

Handle datasets differently from other forms of STI – collect and preserve the metadata, disseminate the metadata with links to the datasets hosted elsewhere.

DOE's scientific research datasets are those resulting from the R&D activities of the Department. They may be large or small and may consist of any of several types of information that is not traditionally text-based. This could include:

- Numerical data
- Data plots or data figures
- Still images or photographs
- Genomic data (gene sequences, etc.)
- Computer simulations
- GIS data and interactive data maps/interfaces
- Multimedia (videos documenting experiments or animations captured from instruments)

# Data Citation

**Many DOE Data Catalogs** already existed, but OSTI saw benefits of assigning persistent identifiers to datasets.

**Benefits of obtaining persistent identifiers for data**

- Persistent identifiers enable easier citation which encourages more citation of datasets in scholarly publications
- Cited datasets support validation of scientific results and reuse of data
- Persistent identifiers help to ensure long-term discovery and access through persistent linkage and identification
- Cited datasets present additional attribution and impact-tracking opportunities for researchers, projects, data centers or repositories, and for funding sponsors
- Searchers are more likely to find datasets and also to make cross-disciplinary connections when both papers and datasets are treated as indexed, retrievable, linked objects
- Persistent identifiers support better version and historical control for data made available on the Web

# DOE Data ID Service Background

- **2011 OSTI become DataCite member** – allows OSTI ability to assign digital object identifiers (DOIs) to datasets
- **DataCite** – international organization that supports data visibility, ease of data citation in scholarly publications, data preservation and future re-use, and data access and retrievability by assigning DOIs to datasets
- **DOE Data ID Service** – OSTI provides service of assigning DOIs to datasets free for DOE funded research. Will provide the service to other federal agencies through cost recovery model.

# OSTP Memo

EXECUTIVE OFFICE OF THE PRESIDENT
**OFFICE OF SCIENCE AND TECHNOLOGY POLICY**
WASHINGTON, D.C. 20502

February 22, 2013

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: John P. Holdren
Director

SUBJECT: Increasing Access to the Results of Federally Funded Scientific Research

**1. Policy Principles**

The Administration is committed to ensuring that, to the greatest extent and with the fewest constraints possible and consistent with law and the objectives set out below, the direct results of federally funded scientific research are made available to and useful for the public, industry, and the scientific community. Such results include peer-reviewed publications and digital data.

Scientific research supported by the Federal Government catalyzes innovative breakthroughs that drive our economy. The results of that research become the grist for new insights and are assets for progress in areas such as health, energy, the environment, agriculture, and national security.

Access to digital data sets resulting from federally funded research allows companies to focus resources and efforts on understanding and exploiting discoveries. For example, open weather

# DOE Policy for Digital Research Data Management

**Requirement 2:** DMPs should provide a plan for making all research data displayed in publications resulting from the proposed research open, machine-readable, and digitally accessible to the public at the time of publication. This includes data that are displayed in charts, figures, images, etc. In addition, the underlying digital research data used to generate the displayed data should be made as accessible as possible to the public in accordance with the Principles stated above. The published article should indicate how these data can be accessed.

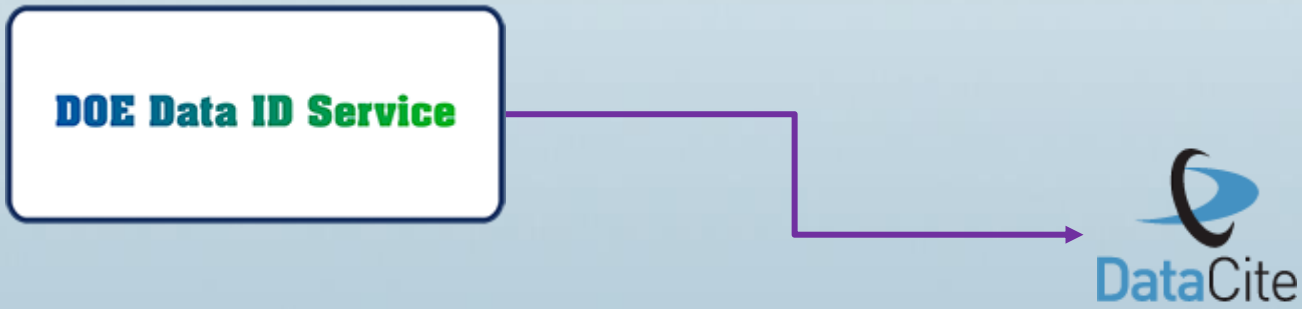
- Policy encouraging the linkage between publications and data
- Support and encourage the use of DOIs for the linkage – citation of dataset DOIs in publication references

# DOE Data ID Service

**DOE Data ID Service is an "umbrella" name for the submission tools, the processing steps, and the DOI assignment system that interacts with DataCite.**

**Service includes:**

- Dataset metadata submission process
- DOI assignment process
- DOI registration process with DataCite



# DOE Data ID Service Process

1. Client contact OSTI – DOEDataID@osti.gov
2. Planning Phase – initial conversations about data researcher is producing and how the service works
   - Conversation about commitment and requirements of making data publicly available indefinitely - manage, update, curate, etc.
   - Understand client data - static or dynamic data (need API), data granularity for DOI
   - Discussion about landing pages (required by DataCite)
   - Offer unique infix for clients -

     DOI prefix/**DOI infix**/DOI suffix

     prefix from DataCite/**infix from submitter**/suffix from OSTI ID

     10.19597/**myprojectname**/1105143
3. Testing Phase for API clients – clients work within our test environment
   - XML formatting tags - send dummy records to make sure correct tag formatting
   - Send real test records to review metadata coming into E-Link (OSTI's ingest tool that communicates with DataCite)
4. Production Phase – clients work within the production environment
   - Begin sending in actual dataset records
   - Metadata records submitted to E-Link → OSTI checks metadata, assigns DOI, and each night sends records DataCite → DataCite does their own validate, officially mints DOI, and makes DOI live globally. If a record fails at DataCite, E-Link is notified and we work with client to correct the record.
5. Datasets discoverable through DOE Data Explorer and SciTech Connect – OSTI discovery/dissemination tools

# Current Data Clients


Atmospheric Radiation Measurement Program
ARM Archive at ORNL

CXIDB

Coherent X-ray Imaging Data Bank
LBNL and Sweden

US Lattice Quantum Chromodynamics

USQCD

GDR

Geothermal Data Repository

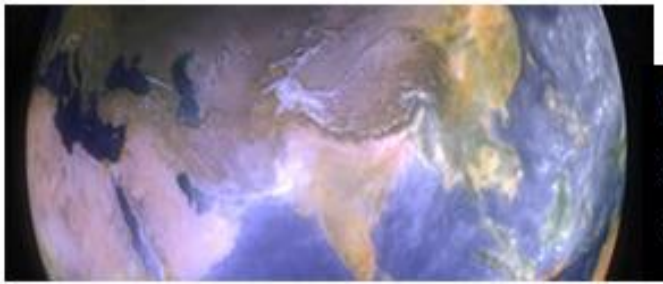
The Materials Project

TCCON

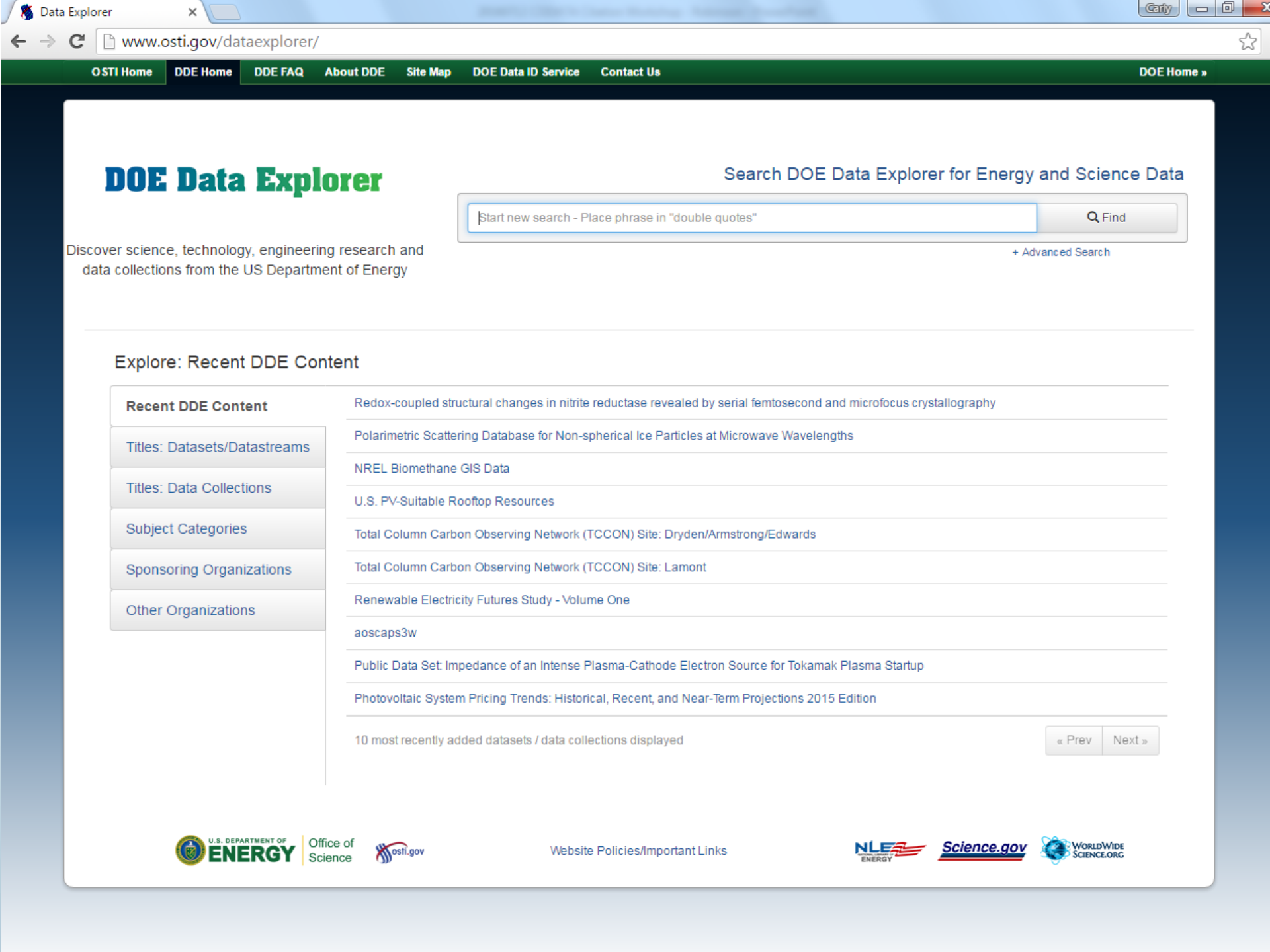
Total Carbon Column Observing Network, CDIAC

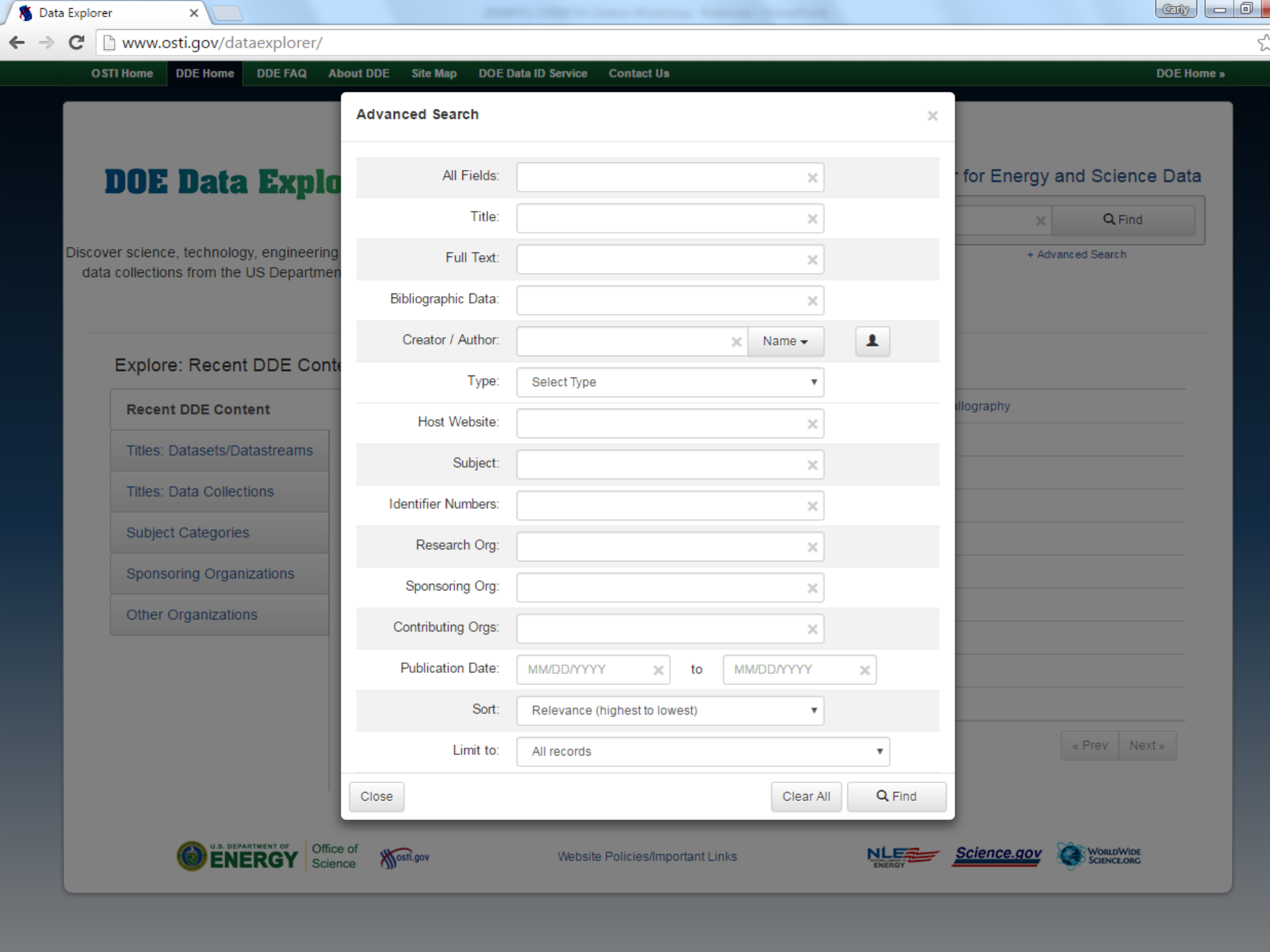
NGEE Arctic

Next-Generation Ecosystem Experiments



Climate Change Science Institute
ORNL


Data Explorer
www.osti.gov/dataexplorer/
OSTI Home
DDE Home
DDE FAQ
About DDE
Site Map
DOE Data ID Service
Contact Us
DOE Home »
DOE Data Explorer
Search DOE Data Explorer for Energy and Science Data
Start new search - Place phrase in "double quotes"
Find
+ Advanced Search
Discover science, technology, engineering research and data collections from the US Department of Energy
Explore: Recent DDE Content
Recent DDE Content
Titles: Datasets/Datastreams
Titles: Data Collections
Subject Categories
Sponsoring Organizations
Other Organizations
Redox-coupled structural changes in nitrite reductase revealed by serial femtosecond and microfocus crystallography
Polarimetric Scattering Database for Non-spherical Ice Particles at Microwave Wavelengths
NREL Biomethane GIS Data
U.S. PV-Suitable Rooftop Resources
Total Column Carbon Observing Network (TCCON) Site: Dryden/Armstrong/Edwards
Total Column Carbon Observing Network (TCCON) Site: Lamont
Renewable Electricity Futures Study - Volume One
aoscaps3w
Public Data Set: Impedance of an Intense Plasma-Cathode Electron Source for Tokamak Plasma Startup
Photovoltaic System Pricing Trends: Historical, Recent, and Near-Term Projections 2015 Edition
10 most recently added datasets / data collections displayed
« Prev
Next »
U.S. DEPARTMENT OF ENERGY
Office of Science
osti.gov
Website Policies/Important Links
NLE
Science.gov
WorldWide Science.org


Data Explorer
www.osti.gov/dataexplorer/
OSTI Home
DDE Home
DDE FAQ
About DDE
Site Map
DOE Data ID Service
Contact Us
DOE Home »
DOE Data Explo
for Energy and Science Data
Discover science, technology, engineering
data collections from the US Departmen
Find
+ Advanced Search
Explore: Recent DDE Conte
Recent DDE Content
Titles: Datasets/Datastreams
Titles: Data Collections
Subject Categories
Sponsoring Organizations
Other Organizations
« Prev
Next »
Advanced Search
All Fields:
Title:
Full Text:
Bibliographic Data:
Creator / Author:
Name
Type:
Select Type
Host Website:
Subject:
Identifier Numbers:
Research Org:
Sponsoring Org:
Contributing Orgs:
Publication Date:
MM/DD/YYYY
to
MM/DD/YYYY
Sort:
Relevance (highest to lowest)
Limit to:
All records
Close
Clear All
Find
U.S. DEPARTMENT OF ENERGY
Office of Science
osti.gov
Website Policies/Important Links
Science.gov
WorldWide Science.org

Search DOE Data Explorer for Energy and Science Data

# DOE Data Explorer

Start new search - Place phrase in "double quotes" | Find

+ Advanced Search



## Search for: All records

Limit to: Datasets

Sort by Relevance | « Prev | Next »

| Total Results | 17,785 |
|---|---|
| Numeric Data | 960 |
| Genome / Genetics Data | 0 |
| Interactive Data Maps | 1 |
| Multimedia | 0 |
| Animations / Simulations | 0 |
| Figures / Data Plots | 12 |
| Specialized Mix | 16,753 |
| Still Images or Photos | 55 |
| Software | 4 |

FILTER RESULTS

Filter by Author

SAVE RESULTS

Excel
CSV
XML

Have feedback or suggestions for a

1. 3D Model of the San Emidio Geothermal Area
James E. Faulds

> The San Emidio geothermal system is characterized by a left-step in a west-dipping normal fault system that bounds the western side of the Lake Range. The 3D geologic model consists of 5 geologic units and 55 faults. Overlying Jurrassic-Triassic metasedimentary basement is a ~500 m-1000 m thick section of the Miocene lower Pyramid sequence, pre- syn-extensional Quaternary sedimentary rocks and post-extensional Quaternary rocks. 15-30° eastward dip of the stratigraphy is controlled by the predominant west-dipping fault set. Both geothermal production and injection are concentrated north of the step over in an area of closely spaced west dipping normal faults.

December 2013 — View Dataset

2. Seismic Line Location Map Hot Pot Project, Humboldt County, Nevada 2010
Michael Lane

> Seismic Line Location Map Hot Pot Project, Humboldt County, Nevada 2010. ArcGIS map package containing topographic base map, Township and Range layer, Oski BLM and private leases at time of survey, and locations, with selected shot points, of the five seismic lines.

January 2010 — View Dataset

3. Nitrogen concentration and isotope dataset for environmental samples from 2012 and 2013, Barrow, Alaska
Jeff Heikoop ; Heather Throckmorton

> Dataset includes nitrate concentrations for polygonal active layer samples, snowmelt; ammonium concentrations for active layer samples; nitrate isotopes for active layer samples, snowmelt, permafrost; ammonium isotopes for active layer samples; and nitrogen isotopes for soils and dissolved organic nitrogen extracted from soil pore waters.

May 2015 — View Dataset

4. Calibrated Hydrothermal Parameters, Barrow, Alaska, 2013
Atchley, Adam ; Painter, Scott ; Harp, Dylan ; Coon, Ethan ; Wilson, Cathy ; Liljedahl, Anna ; Romanovsky, Vladimir

# DOE Data Explorer



## Materials Data on LaH4C4NO8 (SG:14) by Materials Project

Computed materials data using density functional theory calculations. These calculations determine the electronic structure of bulk materials by solving approximations to the Schrodinger equation. For more information, see https://materialsproject.org/docs/calculations

| | |
|---|---|
| Authors: | Kristin Persson |
| Publication Date: | 2014-11-02 |
| OSTI Identifier: | 1183904 |
| Report Number(s): | mp-557512 |
| DOE Contract Number: | AC02-05CH11231; EDCBEE |
| Resource Type: | Dataset |
| Data Type: | Specialized Mix |
| Resource Relation: | Related Information: https://materialsproject.org/citing |
| Research Org: | LBNL Materials Project; Lawrence Berkeley National Laboratory (LBNL), Berkeley, CA (United States) |
| Sponsoring Org: | USDOE Office of Science (SC), Basic Energy Sciences (BES) (SC-22) |
| Country of Publication: | United States |
| Language: | English |
| Subject: | 36 MATERIALS SCIENCE crystal structure; C4 H4 La1 N1 O8; C-H-La-N-O; ICSD-109887; electronic bandstructure |

DATASET

View Dataset (-0.00 MB)

DOI: 10.17188/1183904
*Select the DOI to obtain a copy of this dataset from the publisher.*

Have feedback or suggestions for a way to improve these results? Let us know!

SAVE / SHARE THIS RECORD NO

Citation Formats
Export Metadata
Send to Email

GO TO NEXT/PREVIOUS RECORD

« Prev Next »

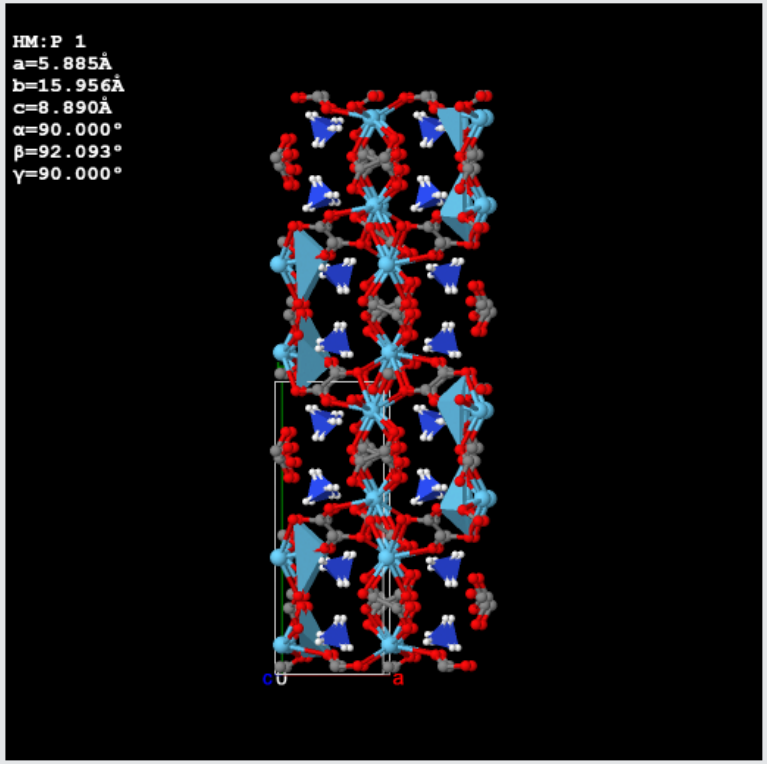
Back to Search Results

Materials Data on LaH4C4 | Material mp-557512

https://materialsproject.org/materials/mp-557512/?null

Home | About | Apps | Documentation | API | Login

| MATERIAL | ID: | DOI: |
| --- | --- | --- |
| $LaH_4C_4NO_8$ | mp-557512 | 10.17188/1183904 |




Structure Type: Conventional Standard | Primitive | Refined

CIF

Space Filling ☐ Polyhedra ☑

| | |
| --- | --- |
| **Zoom in/out** | Shift + Drag cursor |
| **Rotate along the center axis** | Option + Drag cursor |

Registered users see material details, including calculations for band structure, density of states, x-ray diffraction, elasticity, and user-submitted calculated and experimental data. Registration is free.

Learn more | Sign In or Register

Show JSON History | Show BibTex Citation | Download BibTex Citation

# Thank you!

# Questions?