ENABLING DATA DISCOVERY IN THE U.S. DEPARTMENT OF ENERGY

SEPTEMBER 12, 2018, FDLP ACADEMY WEBINAR SERIES
DOE INVESTS $12B PER YEAR IN R&D

U.S. DEPARTMENT OF
ENERGY
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 INFORMATION
(STI/R&D Results)
Text
• Journal articles/accepted manuscripts
• Technical reports
• Conference papers
• Patents
Data
• Large and small datasets
• Images
• Visualizations
Software/Code

≥ 50,000 STI “products” annually

OSTI works to improve discoverability and use of all of this
Vision

The Office of Scientific and Technical Information (OSTI) will fulfill a critical U.S. Department of Energy (DOE) mission to ensure long-term preservation of and access to the results of DOE research and development (R&D) investments. Across the full spectrum of DOE R&D programs, OSTI will provide accountability for all DOE scientific and technical information (STI)—in its many forms—through electronic, efficient, and user-friendly tools and technology.

Mission

Advance science and sustain technological creativity by making R&D findings available and useful to Department of Energy (DOE) researchers and the public.
How Do We Make R&D Results Accessible and Useful?

a) Partnership
b) Technology
c) Services
d) Search tools

STIP

ECOLINK

OSTI.GOV

DOIs for data, technical reports, and software. Software repository services.

PAGES

DOE Data Explorer

DOECODE

DOEpatents

SCIENCECINEMA

Science.gov

WORLDWIDE SCIENCE.ORG

(Federated Search)
OSTI’S DATA TOOLS

DOE Data Explorer

- The DOE Data Explorer (DDE) launched in June 2008 as a tool to help users discover publicly available, DOE-sponsored data and other non-text information.
- DDE is the search tool for finding DOE-funded, publicly available, scientific data submitted by data centers, repositories, and other organizations funded by the Department. DDE includes data Project, data Collection, and individual Dataset records.

Data ID Service

- OSTI became DataCite member in 2011 – Allows OSTI to assign digital object identifiers (DOIs) to datasets.
- DOE Data ID Service – OSTI provides service of assigning DOIs to datasets for DOE-funded research. Also assigns DOIs for other agencies on a cost-reimbursable basis.
WHY REORGANIZE OSTI’S DATA DISCOVERY TOOL?

DOE Data Explorer – search tool for discovering DOE research data

- Feedback from lab workshops and meetings with data researchers
- Provide a more intuitive way to browse/find data
  - Hierarchical
  - Interrelated
  - Customizable
NEED FOR A NEW DATA PRODUCT “TYPE”

- Need for three “types” to organize data
  - data Project
  - data Collection
  - Dataset

- Thought experiment → published in Data Science Journal → now must execute!
A “NEW” DDE
VISUALLY REPRESENTING THE NEW ORGANIZATIONAL STRUCTURE

- Toggle between Projects, Collections, and Datasets using tabs at the top of the results list.
- Under each Project result is the number of associated Collections and Datasets.
- Refine options change based on results product type.
SEARCH AND NAVIGATION USING THE NEW RELATIONS

Can toggle among the Project and associated Collections and Datasets by tabs or hyperlinks.

Includes:

- List of related Projects, Collections, or Datasets
- Link out to data
- Suggested “similar” records
- Geolocation info for Projects
The Materials Project

Harnessing the power of supercomputing and state of the art electronic structure methods, the Materials Project provides open web-based access to computed information on known and predicted materials. Experimental research can be targeted to these sets. Supercomputing clusters at national laboratories provide the infrastructure and algorithms to run at unparalleled speed. Researchers are able to run calculations in a fraction of the time, providing materials researchers with the information they need to design and create new materials.

Product Type: Project
Project Lead: Parson, Kristin
Research Org(s): LBNL Materials Project, Lawrence Berkeley National Lab (ORNL), Oak Ridge, TN (United States); Argonne National Lab, Argonne, IL (United States)
Sponsoring Org: USDOE Office of Science (SC), Basic Energy Sciences (BES), Office of Basic Energy Sciences (BER), Office of Energy Efficiency and Renewable Energy (EERE), and Office of Vehicle Technologies (FTO)

Collection of SG:189 Materials Data

1. Materials Data on CePd2 (SG:189) by Materials Project
   - Kristin Parson
   - 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/doi:10.17186/1351485
   - DOI: 10.17186/1351485
   - Details

2. Materials Data on SiP2 (SG:189) by Materials Project
   - Kristin Parson
   - 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/doi:10.17186/1356520
   - DOI: 10.17186/1356520
   - Details

3. Materials Data on CePd5 (SG:189) by Materials Project
   - Kristin Parson
   - 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/doi:10.17186/1355340
   - DOI: 10.17186/1355340
   - Details

Creators: Parson, Kristin
Publication Date: 2017-04-14
Report Number(s): mp-102467, mp-102544, mp-1124
DOE Contract Number: EDCBEE, AC02-05CH11231
Product Type: Collection
Research Org(s): Lawrence Berkeley National Lab (LBNL), Berkeley, CA (United States)
Subject: MATERIALS SCIENCE, ICSD-948698, Ce-Pd2, Ce(1) Se(2); Ce-9e, Pd-Si, ICSD-649614, crystal structure, Ce(3) Pd5, P21/
OSTI became DataCite member in 2011 – Allows OSTI to assign digital object identifiers (DOIs) to datasets.

DataCite – An international organization 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.

Data ID Service allows data creators to obtain DOIs for datasets, encouraging and increasing data citation in publications.
The Materials Project submits to EoLINK, which assigns a DOI “10.17188/1342472”. The Record with a live DOI becomes available in OSTI.GOV and DDE. DOI links back to MP data. Journal article cites MP data.
BENEFITS TO ASSIGNING DOIS

1. RDA WGDC: assigning a persistent identifier to data ensures preservation and access to data, allowing for citation and reuse
2. Data updates/versioning
3. Permanence of large data
4. Dynamic data
5. Data can be independently indexed and discovered
6. Visibility and use/reuse apart from associated article
USE CASES

- DOI granularity level - dynamic data, beamline experiments, etc.
- DOE researchers produce huge datasets, other people use portions of this data for other research
- Sharing data within a community
- Much data is not directly associated with a publication
- Researcher may submit slice of data to publisher
- Need a DOI for the data prior to submitting article to publisher
- Including DOI in reference list prior to publication
DOE CODE

CREATE
Create code inside or outside DOE CODE.

SUBMIT
Submit your DOE-funded code.

DISCOVER
Discover DOE-funded software and code.

DOE CODE is an open-source product on GitHub that other institutions can download and deploy for their own purposes.

DOE CODE issues digital object identifiers (DOIs) for code and software, making it easier to cite and discoverable. This allows for stronger connections between code, publications, and data through citation in reference sections.

DOE CODE is easy to use and requires minimal metadata, often auto-populated.

DOE CODE seamlessly interfaces with common development platforms (e.g., open web API).

DOE CODE is the DOE Software Catalog.

DOE CODE offers public and private repository services in GitHub and GitLab.
STRATEGIC PRIORITIES
INTERLINKING RESEARCH RESULTS

End goal: interlink all related research outcomes (e.g., from publication to related data to related software)
How can we better communicate/educate data producers of the value of assigning DOIs to data to increase linkages in the scholarly record?

How do we create additional hierarchical associations such as “lab rollups,” associating user facilities (like Argonne National Lab-Advanced Photon Source) to the overarching lab (Argonne National Lab) so that a user can find data related to an instrument instead of an individual project?

Some data can logically be related to more than one Project. How can we address these linkages and expose them in DDE?
QUESTIONS?

SARA STUDWELL
LIBRARIAN/PRODUCT MANAGER
DOE OFFICE OF SCIENTIFIC AND TECHNICAL INFORMATION
STUDWELLS@OSTI.GOV