# 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





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**



#### **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

## OSTI'S VISION AND MISSION

#### **Vision**

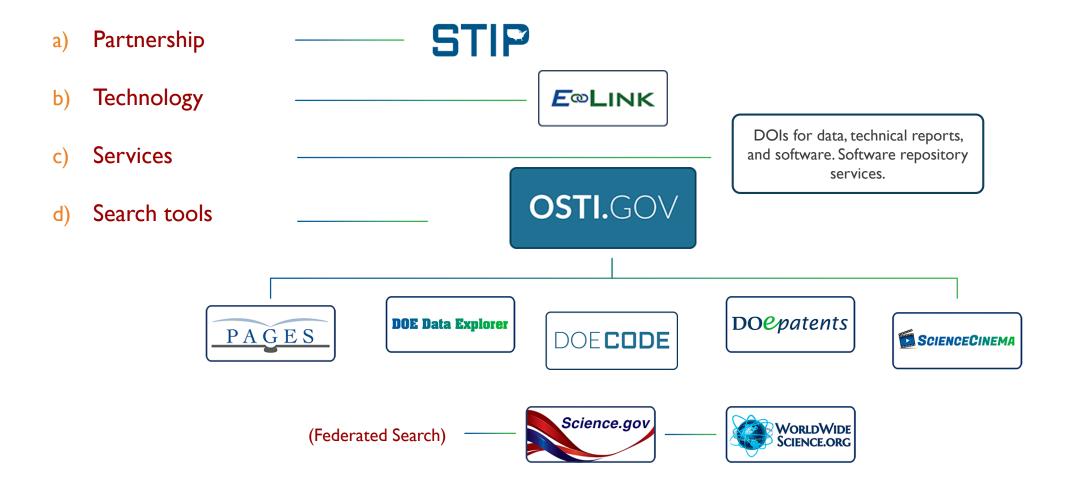
The Office of Scientific and Technical Information (OSTI) will fulfill a critical U.S. Department of Energy (DOE) mission to ensure long-term preservation of and access to the results of DOE research and development (R&D) investments. Across the full spectrum of DOE R&D programs, OSTI will provide accountability 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?



## 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, DOEsponsored data and other non-text information.
- DDE is the search tool for finding DOEfunded, 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?

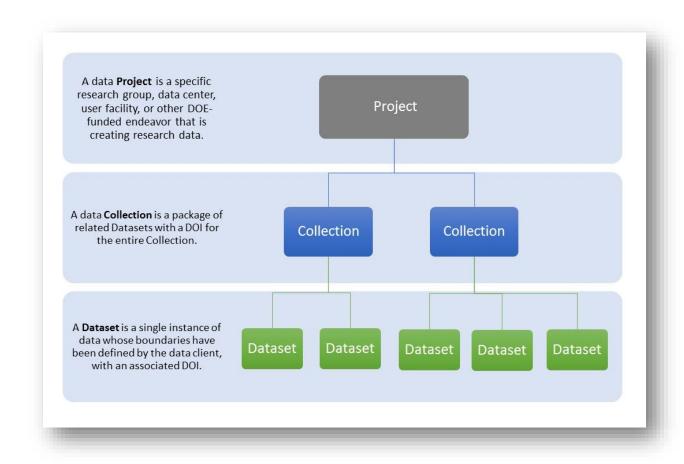
## **DOE** Data Explorer – search tool for discovering **DOE** research data

- Feedback from lab workshops and meetings with data researchers
- Provide a more intuitive way to browse/find data
  - Hierarchical
  - Interrelated
  - Customizable

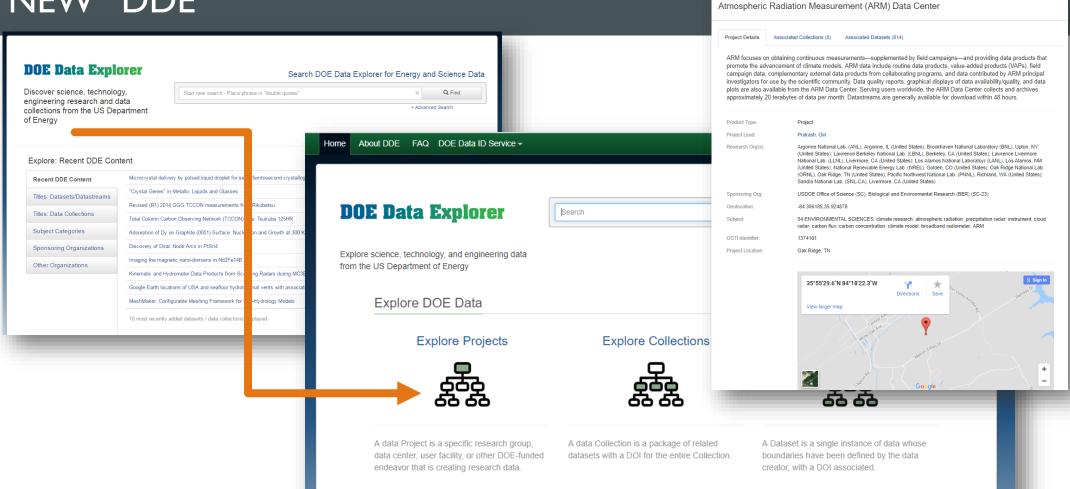
**DOE Data Explorer** 

## **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



Data Explorer / Search Results / Atmospheric Radiation Measurement (ARM) Data Center

## VISUALLY REPRESENTING THE NEW ORGANIZATIONAL STRUCTURE

**DOE** Data Explorer

Search for: All records

Kristin Persson (69.693)

Nitin Bharadwai (119)

Michael Jensen (107)

RESEARCH ORGANIZATION

Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States)

LBNL Materials Project (69,693)

Atmospheric Radiation Measuremen

(ARM) Archive, Oak Ridge National

Laboratory (ORNL), Oak Ridge, TN

Atmosphere to Electrons (A2e) Data

Archive and Portal Pacific Northwest National Laboratory (211)

PUBLICATION YEAR

**T 2017** (2,469)

**T 2016** (32,286)

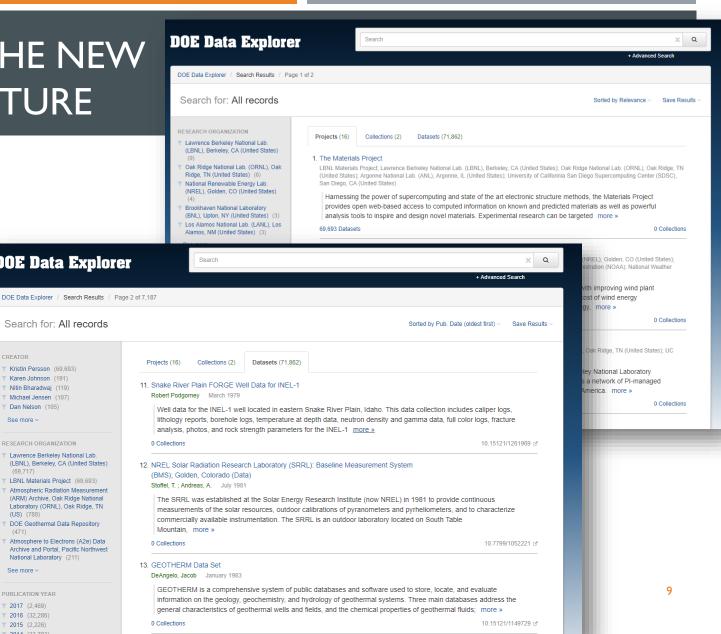
T 2015 (2,226) **T 2014** (33,783)

DOE Geothermal Data Repository

Dan Nelson (105)

See more

- 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.

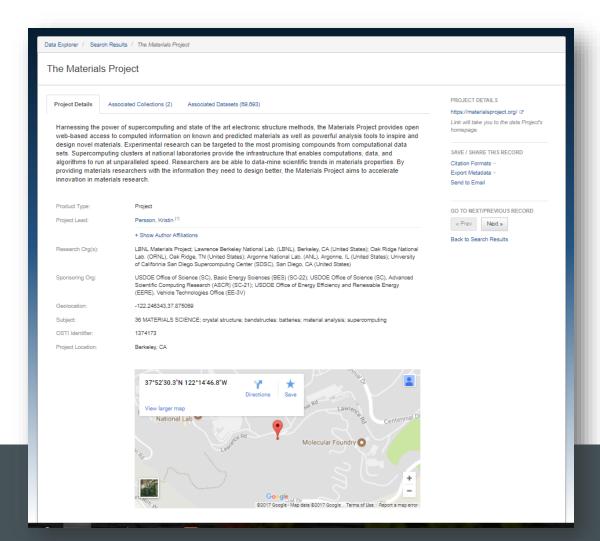


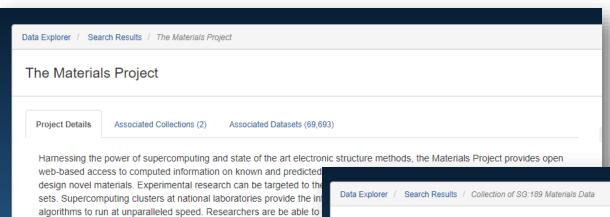
## SEARCH AND NAVIGATION USING THE NEW RELATIONS

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





providing materials researchers with the information they need to de

Persson, Kristin [1]

+ Show Author Affiliations

LBNL Materials Project; Lawrence Berkel Nat

of Califorinia San Diego Supercomputing Conter

Lab. (ORNL), Oak Ridge, TN (United State

USDOE Office of Science (SC), Basic Energ

Scientific Computing Research (ASCR) (SC

(EERE), Vehicle Technologies Office (EE-3V

Project

innovation in materials research.

Product Type:

Project Lead:

Research Org(s):

Sponsoring Org:

Collection of SG:189 Materials Data Collection Details Associated Project Associated Datasets (3) Computed materials data using density functional theory calculations bulk materials by solving approximations to the Schrodinger equation cs/calculations Creators: Persson, Kristin 2017-04-14 Publication Date: Report Number(s): mp-1024071; mp-1025404; mp-121484 DOE Contract Number: EDCBEE: AC02-05CH11231 Product Type: Collection

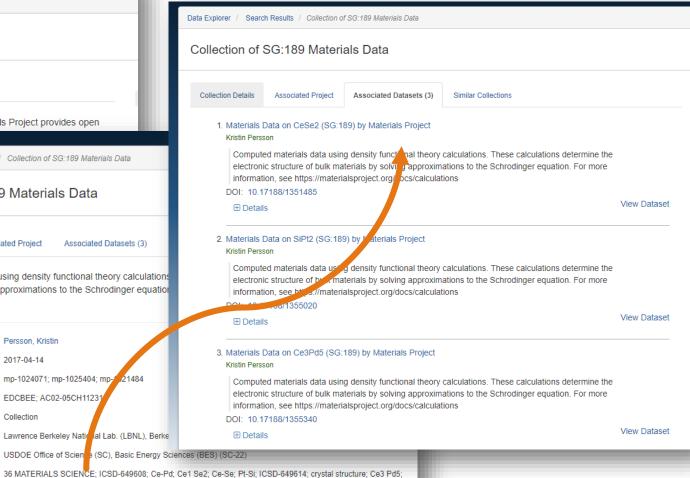
Pt2 Si1

1389158

ch Org(s):

Sponsoring Org. Subject:

OSTI Identifier:



Some links on this page may take you to non-federal websites. Their policies may differ from this site.

## DATA ID SERVICE

- 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.



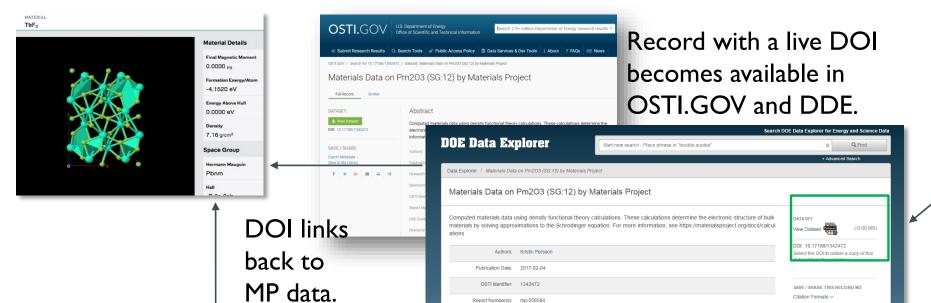
submits to



assigns a

**DOI**"10.17188/1342472"

**Data**Cite



validated and minted by



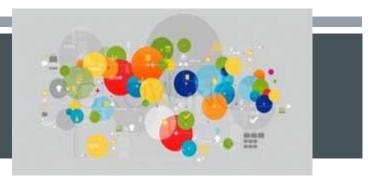
Journal article cites MP data.

SAVE / SHARE THIS RECORD NO

## BENEFITS TO ASSIGNING DOIS

- I. 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





- 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





Create code inside or outside DOE CODE.

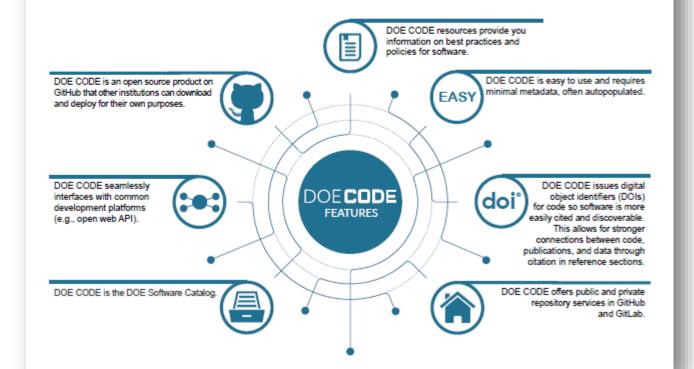


Submit your DOE-funded code.

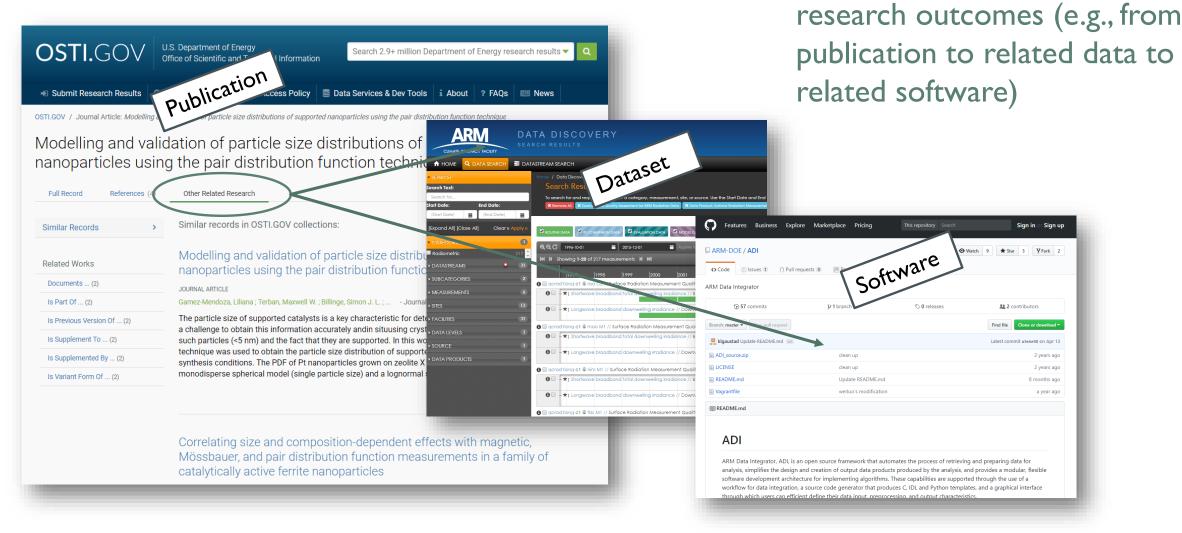


DISCOVER

Discover DOE-funded software and code.



## STRATEGIC PRIORITIES INTERLINKING RESEARCH RESULTS



End goal: interlink all related

## ADDITIONAL AREAS FOR EXPLORATION IN THE DATA ENVIRONMENT

- 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