

Data Discovery

Sara Studwell

Data Team Lead & DOE Data Explorer Product Manager

studwells@osti.gov



U.S. DEPARTMENT OF
ENERGY

2019 DOE Data ID Service Workshop • September 17-18



DOE Data Explorer

DOE Data Explorer (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. The data themselves reside at national laboratories, data centers, user facilities, colleges and universities, or other websites.

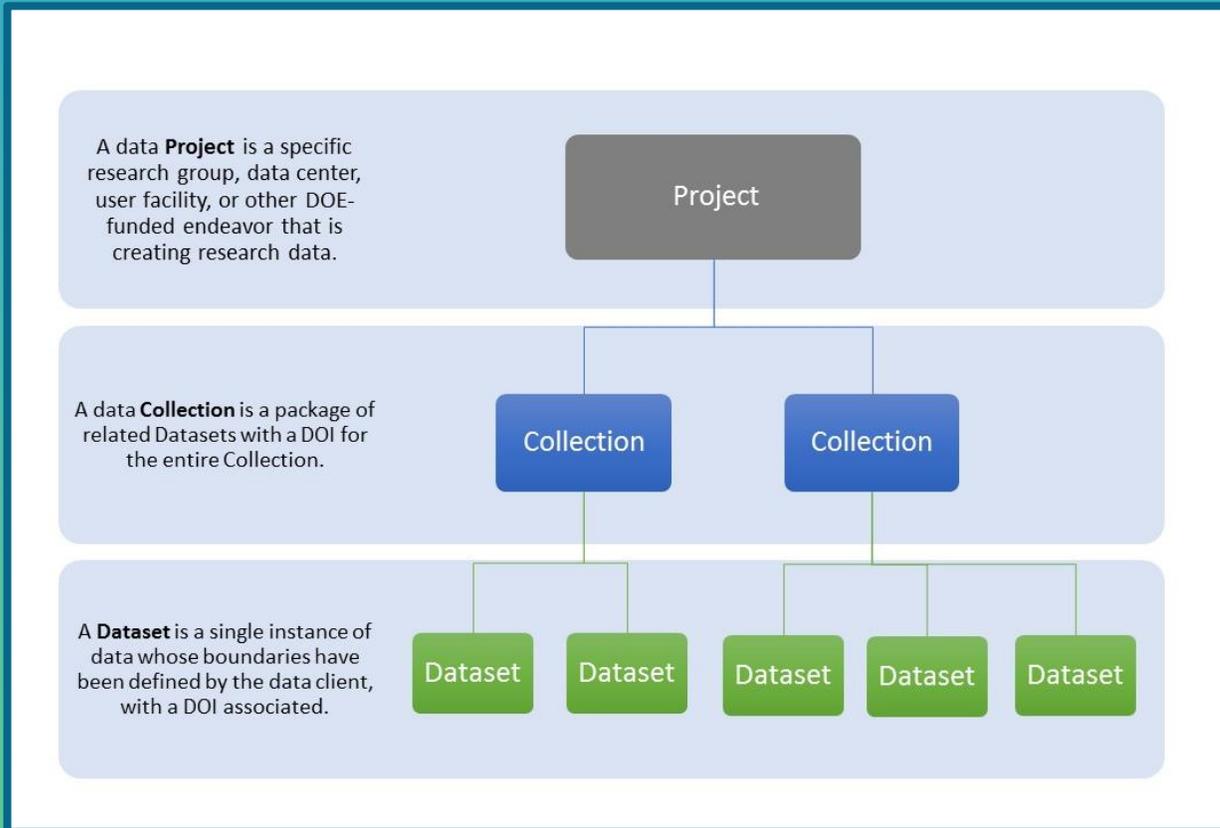


The screenshot shows the DOE Data Explorer website. At the top left is the U.S. Department of Energy seal. To its right is the title "DOE Data Explorer" and the text "U.S. Department of Energy Office of Scientific and Technical Information". Below this is a search bar with the placeholder text "Search Scientific Data" and a green search button. Underneath the search bar is a section titled "Explore DOE Data" which contains three columns: "Explore Projects", "Explore Collections", and "Explore Datasets". Each column has a tree diagram icon and a descriptive paragraph. The "Explore Projects" paragraph states: "A data Project is a specific research group, data center, user facility, or other DOE-funded endeavor that is creating research data." The "Explore Collections" paragraph states: "A data Collection is a package of related datasets with a DOI for the entire Collection." The "Explore Datasets" paragraph states: "A Dataset is a single instance of data whose boundaries have been defined by the data creator, with a DOI associated."



DDE includes data Project, data Collection, and individual Dataset records.

- A data Project is a scientific research group, data center, user facility, or other DOE-funded endeavor that is creating research data.
- A data Collection is a package of related datasets with a DOI for the entire collection.
- A dataset is a single instance of data whose boundaries have been defined by the data submitter, with a DOI associated.



Search and Navigation

- Basic and advanced search options
- Toggle between Projects, Collections, and Datasets using tabs at the top of the results list
- Refine options change based on results product type

The screenshot shows the DOE Data Explorer interface. At the top, it says 'DOE Data Explorer' and 'U.S. Department of Energy Office of Scientific and Technical Information'. There is a search bar with 'Search Scientific Data' and a magnifying glass icon. Below the search bar, there are navigation links: 'Data ID Services', 'About', 'FAQs', and 'News'. The main content area shows search results for 'wind'. It indicates '1,136 Datasets' and '1,140 Total Results'. The results are sorted by 'Relevance'. There are navigation buttons for 'Prev', 'Next', and 'Save Results'. Below the search bar, there are tabs for 'Projects (4)', 'Collections (0)', and 'Datasets (1,136)'. The first result is '1. NREL National Wind Technology Center (NWTCC): M2 Tower; Boulder, Colorado (Data)'. It lists the author 'Jager, D.; Andreas, A.' and provides a detailed description of the facility. The second result is '2. Sodar - Vaisala Triton Wind Profiler, AON3 - Raw Data' by 'Stoelinga, Mark'. It describes the measurements from eight different Vaisala Triton Wind Profiler instruments. The page also includes a 'SEARCH FOR:' section with 'KEYWORDS: wind' and a 'REFINE BY:' section with categories like 'CREATOR / AUTHOR' and 'RESEARCH ORGANIZATION'.



Additional Features

- Save/share options
- Account functionality
- News and information pages about Data ID Services
- Other Related Research

NREL National Wind Technology Center (NWTC): M2 Tower; Boulder, Colorado (Data)

Dataset Associated Project Associated Collections (0) Other Related Research

DATASET:

[View Dataset](#)

DOI: 10.7799/1052222

SAVE / SHARE:

[Export Metadata](#) ~
[Save to My Library](#)

[f](#) [t](#) [e](#) [p](#) [r](#)

Abstract

The National Wind Technology Center (NWTC), located at the foot of the Rocky Mountains near Boulder, Colorado, is a world-class research facility managed by NREL for the U.S. Department of Energy. NWTC researchers work with members of the wind energy industry to advance wind power technologies that lower the cost of wind energy through research and development of state-of-the-art wind turbine designs. NREL's Measurement and Instrument Data Center provides data from NWTC's M2 tower which are derived from instruments mounted on or near an 82 meter (270 foot) meteorological tower located at the western edge of the NWTC site and about 11 km (7 miles) west of Broomfield, and approximately 8 km (5 miles) south of Boulder, Colorado. The data represent the mean value of readings taken every two seconds and averaged over one minute. The wind speed and direction are measured at six heights on the tower and air temperature is measured at three heights. The dew point temperature, relative humidity, barometric pressure, totalized liquid precipitation, and global solar radiation are also available.

Creator(s)/Author(s): [Jaget, D.](#); [Andreas, A.](#)

09-24
/DA-5500-56489
-086028308
et
nal Renewable Energy Lab. (NREL), Golden, CO (United States)
E Office of Energy Efficiency and Renewable Energy (EERE), Solar Energy Technologies Office (EE-4S)

DOE Data Explorer

U.S. Department of Energy
Office of Scientific and Technical Information

Search Scientific Data

Data ID Services About FAQs News

Data ID Services

The Department of Energy (DOE) Office of Scientific and Technical Information (OSTI) offers two services for assigning digital object identifiers (DOIs) to datasets and data collections to help increase access to scientific research data - the DOE Data ID Service and the Interagency Data ID Service (IAD).

DOE Data ID Service

Interagency Data ID Service

Data ID Service Workshops



Other Related Research



DOE Data Explorer | U.S. Department of Energy Office of Scientific and Technical Information | Search Scientific Data

DOE Data Explorer / Search for 1246084(filtered) / Dataset: AmeriFlux US-Ne1 Mead - irrigated continuous maize site

AmeriFlux US-Ne1 Mead - irrigated continuous maize site

Dataset | Associated Project | Associated Collections (0) | **Cited by (3)** | Other Related Research

All Cited By >

- Dataset (1)
- Journal (2)

Search

Sort by title
Sort by date

[x clear filter / sort]

Works referencing / citing this record:

FLUXNET2015 US-Ne1 Mead - irrigated continuous maize site [↗](#)

DATASET, JANUARY 2015

Suyker, Andy
FLUXNET2015
DOI: 10.18140/flx/1440084 [↗](#)

Effect of the revisit interval and temporal upscaling methods on the accuracy of remotely sensed evapotranspiration estimates [↗](#)

JOURNAL, JANUARY 2017

Alfieri, Joseph G.; Anderson, Martha C.; Kustas, William P.
Hydrology and Earth System Sciences, Vol. 21, Issue 1, p. 83-98
DOI: 10.5194/hess-21-83-2017 [↗](#)

DOE Data Explorer | U.S. Department of Energy Office of Scientific and Technical Information | Search Scientific Data

DOE Data Explorer / Search for 1438376(filtered) / Dataset: SPRUCE Ground Observations of Phenology in Experimental Plots, 2016-2017

SPRUCE Ground Observations of Phenology in Experimental Plots, 2016-2017

Dataset | Associated Project | Associated Collections (0) | References (1) | Cited by (2) | **Other Related Research**

Similar Records >

Related Works

Dataset (1)

Similar records in DOE Data Explorer and OSTI.GOV collections:

SPRUCE Ground Observations of Phenology in Experimental Plots, 2018

DATASET

Schädel, C.; Nettles, W.R.; Heiderman, R.R.; ...

This data set consists of phenological transition dates, as derived from direct observations of vegetative and reproductive phenology recorded by a human observer, from the SPRUCE experiment during 2018, the third full year of whole-ecosystem warming (Hanson et al. 2017). Both spring and autumn phenological events are included. Since April 2016, human observers have been directly tracking the phenology of both woody and herbaceous species on a weekly schedule within the SPRUCE experimental chambers. The observed date report here is the first survey date on which an event/phenophase was definitively observed. The SPRUCE experiment is located within the S1 peat [more »](#)

DOI: 10.25581/spruce.073/1557731 | [View Dataset](#)

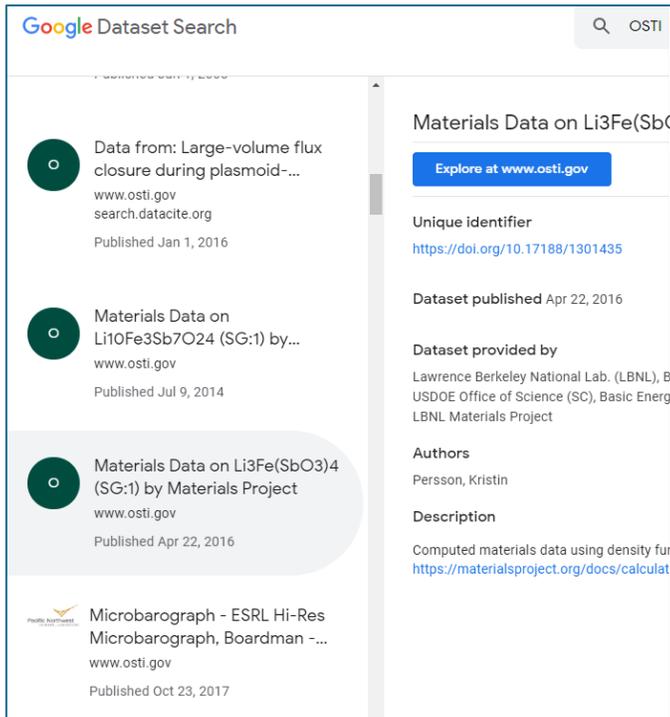
SPRUCE Vegetation Phenology in Experimental Plots from Phenocam Imagery, 2015-2017

DATASET



Google Dataset Search Beta

- Much like Google and Google Scholar, Google Dataset Search allows users to search and access datasets hosted in thousands of repositories.
- Currently indexes repositories that use the schema.org structured data guidelines.
- OSTI implemented Google's structured data guidelines (largely schema.org) for datasets, using both JSON-LD and microdata representations.
- Structured data changes were implemented well in advance of Google Dataset search release, ensuring OSTI STI datasets could be made available immediately.
- Structured data was tested and validated using Google structured data tools.
- Source and provenance best practices were applied, ensuring that sitemaps and canonical references for dataset STI were in place.



The FAIR Guiding Principles

To be Findable:

- (meta)data are assigned a globally unique and persistent identifier
- data are described with rich (meta)data
- (meta)data clearly and explicitly include the identifier of the data it describes
- (meta)data are registered or indexed in a searchable resource

To be Accessible:

- (meta)data are retrievable by their identifier using a standardized communications protocol
 - the protocol is open, free, and universally implementable
 - the protocol allows for an authentication and authorization procedure, where necessary
- (meta)data are accessible, even when the data are no longer available

To be Interoperable:

- (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
- (meta)data use vocabularies that follow FAIR principles
- (meta)data include qualified references to other (meta)data

To be Reusable:

- meta(data) are richly described with a plurality of accurate and relevant attributes
 - (meta)data are robust, accurate, and descriptive, adequately describing the associated data.
 - (meta)data are released with a clear and accessible data usage license
 - (meta)data are associated with detailed provenance
 - (meta)data meet domain-relevant community standards

Wilkinson, Mark D., et al. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific data* 3 (2016). DOI: 10.1038/sdata.2016.18



FAIR Guiding Principles at OSTI

Findable

- OSTI assigns a Digital Object Identifier (DOI) to every dataset metadata submission.
- OSTI offers 13 required metadata fields.
- DOI is included in the metadata record found in OSTI.GOV and DDE.
- All metadata records are registered with DataCite. Records are searchable in OSTI.GOV, DDE, DataCite, Google Dataset Search, DX.DOI.ORG.

Accessible

- Publicly available metadata is found in OSTI.GOV and DDE, as well as their respective APIs.
- OSTI provides a REST API for metadata discovery.
- Metadata records are always available at OSTI.GOV or DDE. If the URL is broken or the data is no longer available, the URL should indicate the location/availability of the data.

Interoperable

- OSTI offers dataset metadata submission via E-Link API (XML) or web submission interface.
- OSTI dataset metadata follow the DataCite schema required for obtaining a DOI, and necessary DOE-relevant metadata.
- OSTI allows for linkages to be created between research through the use of the Related Identifiers metadata field.

Reusable

- OSTI offers 13 required metadata fields.
- OSTI encourages data metadata submitters to provide accessibility information.
- OSTI encourages data clients to provide provenance information on dataset landing pages
- OSTI encourages metadata submitters to adhere to respective domain-relevant community standards.



Data Metrics

- DataCite
 - <https://stats.datacite.org/>
 - Registrations by members, clients, prefixes
 - By year
 - DOI metadata findable, registered
 - Resolutions by month
 - Total attempted, successful, failed, top 10 successes/failures
- OSTI
 - Submitted Datasets
 - By site, date range
 - Submitted Collections
 - By site, date range
- Repository
 - Can be accomplished with Google Analytics or analog counts
 - Pageviews, downloads, referrals, etc.



Questions?

ddecomments@osti.gov

studwells@osti.gov



U.S. DEPARTMENT OF
ENERGY

2019 DOE Data ID Service Workshop • September 17-18

