ESS-DIVE: Advancing DOE ESS Data Management, Discovery, and Use

Shreyas Cholia[1], Joan Damerow[2], Deb Agarwal[1], Madison Burrus[2], Hesham Elbashandy[1], Valerie Hendrix[1], Matthew B. Jones[3], Emily Nagamoto[2], Rushiraj Nenuji[3], Fianna O'Brien[1], Dylan O'Ryan[2], Shalki Shrivastava[1], Jing Tao[3], Karen Whitenack[1], Charuleka Varadharajan[2]

[1]Scientific Data Division, Lawrence Berkeley National Laboratory; [2]Earth & Environmental Sciences Area, Lawrence Berkeley National Laboratory; [3]National Center for Ecological Research and Synthesis
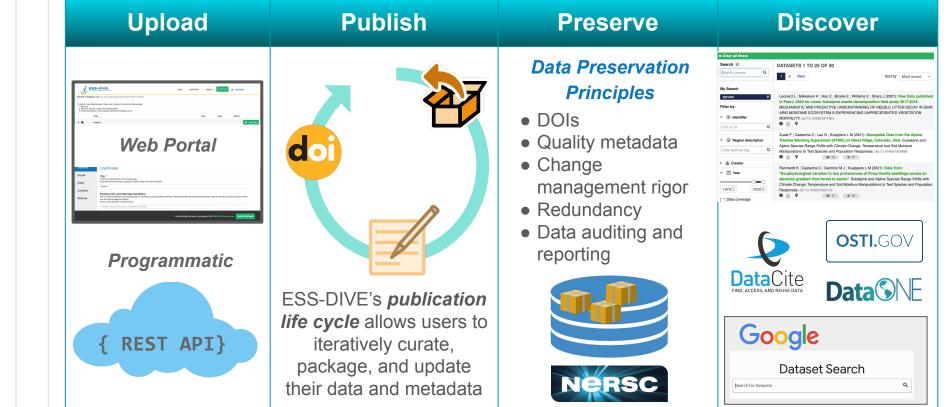
About ESS-DIVE

ESS-DIVE is a **CoreTrustSeal certified data repository** that enables data management, discovery, and use of a variety of DOE's Environmental Systems Science (ESS) data.

ESS-DIVE stores a variety of data types from ESS-sponsored research, which includes:
- Data from a wide range of natural and urban systems
- Data from field, lab experiments, modeling, remote sensing data

Website: https://ess-dive.lbl.gov
Data Access: https://data.ess-dive.lbl.gov
Features: https://ess-dive.lbl.gov/our-features-and-services

USERS	419 Registered users	100 Contributing projects	278 Events
DATA	1215 Public datasets	283 Private datasets	13.6 TB Total file size
USAGE	594K File downloads	1M Data views	16 Public portals

ESS-DIVE use as of 2025-03-19

Data Upload & Review

- **Web portal -** Web form to upload of files & metadata.
- **Large dataset support -** Secondary (Tier-2) storage layer for large datasets including direct file access and hierarchical data
- **Advanced Access -** REST API programmatic access, Globus for high performance data transfer.
- **Quality checks -** Metadata curation before publication to make data FAIR, including assessment reports
- **Standard review** of dataset metadata based on FAIR principles and reporting format review with Fusion Database

API for Data Upload/ Download
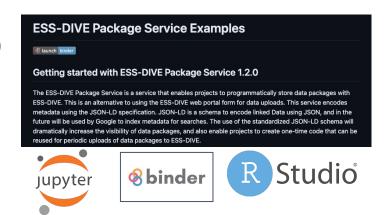
Project Data Archive

Project Data Management

ESS-DIVE supports **end-to-end data management lifecycle**, and offers project-specific features that allow researchers to **collaborate**, **share**, and **manage** data within their teams.

Data Preservation Principles
- DOIs
- Quality metadata
- Change management rigor
- Redundancy
- Data auditing and reporting

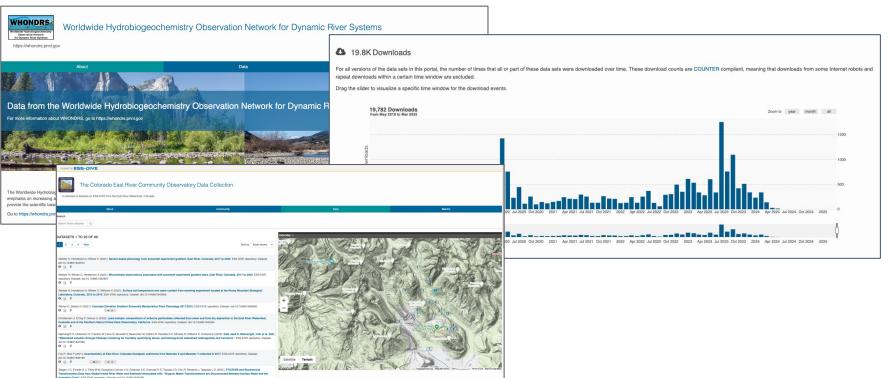ESS-DIVE's **publication life cycle** allows users to iteratively curate, package, and update their data and metadata

NEW: Revamped publication interface allow for **reservation of DOIs** before publication, clear **dataset and DOI statuses**, and intuitive **dataset publication**

Features for **collaborative project data curation**
- Dataset sharing within project teams to promote collaborative editing and curation of data packages
- Project level team management and access control
- Data search within a project
- Custom project-centric data portals - curated collections of project data or other views (e.g. field site, campaign)
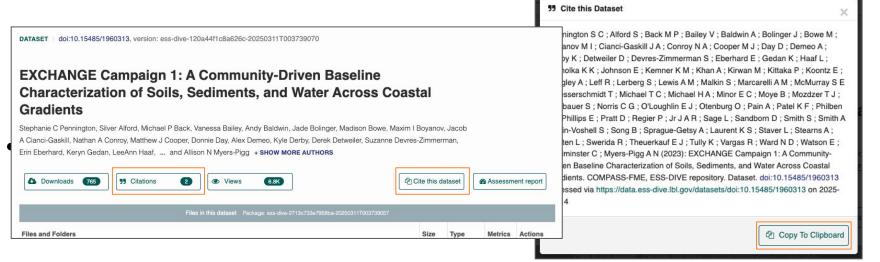- Project data citations and metrics

Link related resources on ESS-DIVE and elsewhere
- Use the external linking feature to connect data
- Provide input on linking related resources for scientific use, and complex data citations

Adopt **data and metadata reporting formats** in your project's data collection and publication workflows
- 12 data and metadata reporting formats for environmental data
- See poster: "ESS-DIVE Partner Projects" for more details

Crystal-Ornelas, R., Varadharajan, C.*, O'Ryan, D. et al. Enabling FAIR data in Earth and environmental science with community-centric (meta)data reporting formats. Sci Data 9, 700 (2022). https://doi.org/10.1038/s41597-022-01606-w

Using & Citing ESS-DIVE Data

ESS-DIVE provides many features to **access and use data**.

Dataset API (api.ess-dive.lbl.gov)
Package service API enables programmatic submission, management, search and download of datasets
- Find datasets by project, DOI, author or other search terms
- Search & download data files or metadata in bulk

FusionDB - Deep Dive API (fusion.ess-dive.lbl.gov)
ESS-DIVE Deep Dive API makes standardized data searchable
- Search within data files and find specific variables relevant to your scientific research
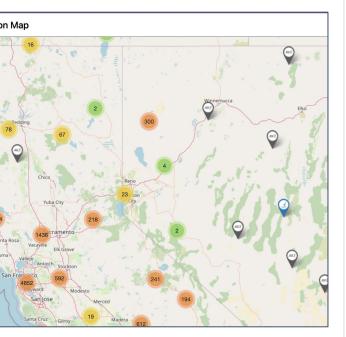
Citing ESS-DIVE Datasets
- Include ESS-DIVE datasets in the *data availability statement* and as a *reference* in your papers

BER Data Integration

BER Bio-Eco Data Integration project (BERTron) to harmonize and integrate data across ESS-DIVE, NMDC, KBase, EMSL, and JGI
- Enabling Unified Search
- Standardizing on schema terms
- Code-a-thons
- User Interviews

BER Related Data Use Cases

20+ use cases completed to connect interdisciplinary ESS-DIVE datasets to related research outputs, such as sites and samples, associated journal articles, references, derived and source data.

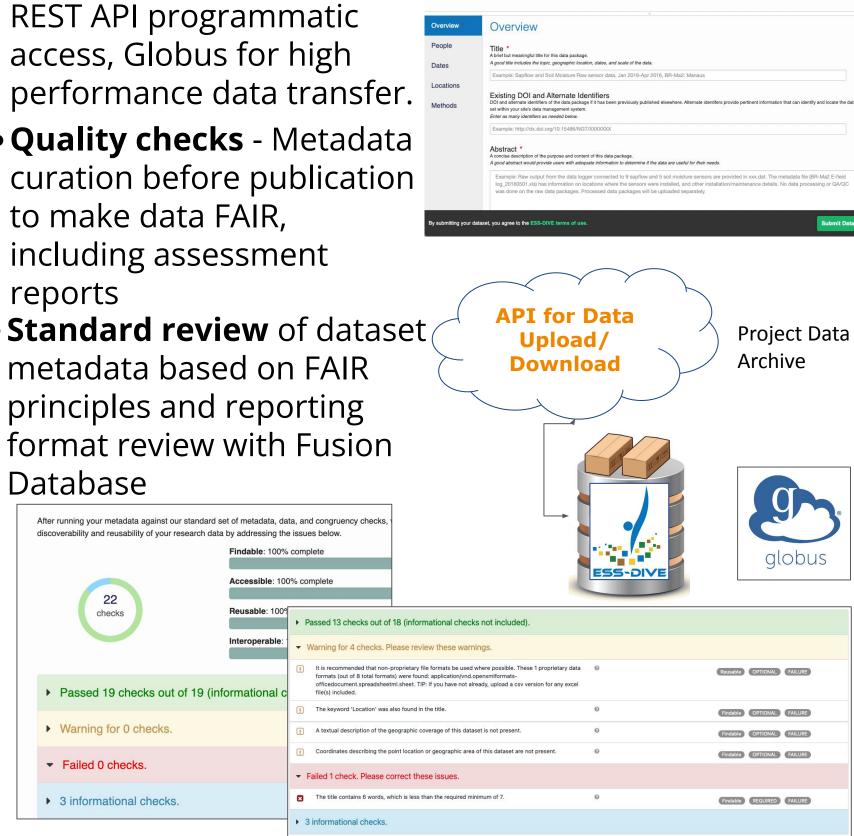**Common ways data are connected:**
- **Project, Sites, Samples**
- **Associated Journal Article** - Paper dataset created for (supplement)
- **References** - related datasets, workflows, journal articles, code
- **Data derived from this dataset** - child datasets
- **Source data** - original source data used in dataset (e.g. EMSL), datasets included in a collection

Help track provenance and citations

Interested in adding related works to your ESS-DIVE dataset? Contact us at ess-dive-support@lbl.gov

Learn More

Connect with us in various ways:
- Helpdesk, tutorials, webinars, annual workshop
- User interviews to solicit feedback on ESS-DIVE features
- PI meeting tutorials, participation in ESS Cyberinfrastructure Working Groups
- ESS-DIVE Archive Partnership Board consisting of leads of major ESS projects provides input into ESS-DIVE features and activities
- Attend the "ESS-DIVE Tutorial for Using Data" and "ESS-DIVE Tutorial for PIs and Data Managers"

Questions? ess-dive-support@lbl.gov
ESS-DIVE Documentation: https://docs.ess-dive.lbl.gov/
Webinars: https://ess-dive.lbl.gov/webinars/
Tutorials: https://ess-dive.lbl.gov/video-tutorials/

Acknowledgements

BERKELEY LAB

EARTH & ENVIRONMENTAL SCIENCES

ESS-DIVE

U.S. DEPARTMENT of ENERGY