ESS-DIVE for PIs and Data Managers

data.ess-dive.lbl.gov
Welcome!

Audience Introductions
Madison Burrus

Computer Systems Engineer
Objectives

1. ESS-DIVE Feature Overview for:
   a. Data Archival *(5 mins)*
   b. Data Management Setup *(10 mins)*
   c. Reporting progress to DOE & Questionnaire *(10 mins)*

2. Preview of upcoming features *(<5 mins)*

3. To-Do List for PIs *(<5 mins)*

**Key Takeaway:** Learn what is currently possible on ESS-DIVE and become aware of features we are working on!
Minimum DMP Needs

ESS projects need to:

1. **Archive project data** on ESS-DIVE
   a. Project data previously published should be mirrored on ESS-DIVE

2. **Report data publication** progress to DOE

ESS-DIVE is here to help ESS projects meet these needs and support project science
1. Metadata & Data Submission
2. Large Data Support
3. External links for specialized and previously published data
4. Teams & Project Data Managers
5. Share datasets with individuals and teams
6. Data Portals to showcase data to the public
7. Publication and data metrics reports
8. Make ESS data FAIR with Reporting Formats & DeepDiveAPI
Data Archival
Metadata Submission & Automated Review

Overview

Title *
A brief but meaningful title for this data package. A good title includes the topic, geographic location, dates, and scale of the data.
Example: Sapflow and Soil Moisture Raw sensor data, Jan 2016-Apr 2016, BR-Mad: Manaus

Existing DOI and Alternate Identifiers
DOI and alternate identifiers of the data package if it has been previously published elsewhere. Alternate identifiers as needed below.
Example: http://dx.doi.org/10.15486/NGT/000000

Abstract *
A concise description of the purpose and content of this data package. A short abstract would provide users with adequate information to determine if the data are useful for their needs. For example: Raw output from the data logger connected to 9 sapflow and 5 soil moisture sensors in five different locations. Details on where the sensors were installed, and other installation/maintenance details. No data processing.

{ REST API }

provider = {
    "identifier": {
        "@type": "PropertyValue",
        "propertyID": "ess-dive",
        "value": "1e6d50d3-9532-43fb-a63f-bdcb4350bf0c"
    }
}

{ REST API }

Web Form

Metadata Quality Report

After running your metadata against our standard set of metadata, data, and congruency checks, we have research data by addressing the issues below.

- Identification: 100% complete
- Discovery: 67% complete
- Interpretation: 100% complete

- Passed 11 checks out of 12 (informational checks not included).
- Warning for 0 checks.
- Failed 1 check. Please correct these issues.

The abstract is only 75 word(s) long but 100 or more is required.

Receive feedback while preparing dataset metadata
Data Submission

There are three ways to upload data files to ESS-DIVE

<table>
<thead>
<tr>
<th>Upload Method</th>
<th>File volume</th>
<th>No. of files</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intuitive</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Submission Web Form</td>
<td>&lt; 10 GB</td>
<td>&lt; 100</td>
</tr>
<tr>
<td><strong>Programmatic</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dataset API</td>
<td>10 - 500 GB</td>
<td>&lt; 100</td>
</tr>
<tr>
<td><strong>Large Data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Globus Data Transfer Service</td>
<td>&gt; 500 GB</td>
<td>&gt; 100</td>
</tr>
</tbody>
</table>
Support for Large Data Volumes

Upload: Submit data using Globus’ cloud-based, intuitive upload interface

Access: Explore hierarchies and download files from ESS-DIVE’s Tier 2 landing page. Cloud-based download also supported on Globus.

Come talk to us to get started!
External Links

Store specialized data types where it makes the most sense scientifically and connect to ESS-DIVE for search and discovery.

External Links to Data or Metadata

Link your ESS-DIVE datasets to:
GitHub Code, project data archives, institutional repositories

Come talk to us to get started!
Data Management Setup
Share Datasets

Sharing enables teams of scientists to build, curate, and publish datasets together.

Click to grant permissions to your data.

Share with teams and individuals.

Data Contributor

Collaborator

PI / Project Manager
Teams are **groups of ESS-DIVE users** that have the same permission to access datasets.

Teams provide an easier way to share with your project.

1. Share datasets with a team

2. Add people to team

New team members get immediate access to all project datasets!
PIs and Project Data Managers

A PI or project data manager is a role in ESS-DIVE that enables someone to access your projects and teams.

⭐ PI must approve managers

Approved project managers gain access to your project page

Create teams for your project
Benefits of Sharing Datasets

- **Visibility**: Team members can share datasets with PIs or data managers to allow them to see private project datasets on ESS-DIVE and report that to DOE

- **Informed curation**: Sharing gives PIs the opportunity to help curate datasets before publication according to project needs
Data Portals

Worldwide Hydrobiogeochemistry Observation Network for Dynamic River Systems

Collection of Datasets

Custom Descriptive Pages

Metrics: Summary of Holdings, View/Download/Citation Counts
All Data Portal Metrics

- Total number of datasets
- Volume (bytes) of metadata records & data files
- Most recent date updated
- File types
- Years data collected
- Views, Downloads, Citations

https://data.ess-dive.lbl.gov/portals/ngeetropics/Metrics
Flexibility of Data Portal Branding

Data Portals can represent any theme

Create multiple portals to organize different scientific initiatives:

- Experiments
- Study Areas
- Projects
- Subcontracts or Sub-projects

Customize branding of **banner images**, **logos**, and **URLs** to center around the theme

- https://data.ess-dive.lbl.gov/portals/ngaetropics
- https://data.ess-dive.lbl.gov/portals/compass
- https://data.ess-dive.lbl.gov/portals/EXCHANGE

Data is published under one funding project
Project Data Portals

Collect all project data in one place

Mirror existing project websites onto a data portal at the end of project’s life for long term maintenance

Preserve information regarding the project's objectives, scopes, and organization for years to come

Option to manage your project datasets from this view

https://data.ess-dive.lbl.gov/portals/CDIAC

Add data portal URLs to project or institutional webpages
Custom Data Filters

**Data filters** customize how you look up data

They enable **targeted searches** within a collection

Can be used to:

- **Expand report** of project publication progress to DOE
- **Internally track** datasets by type or project-specific category
Questions?
Report Data Publication Metrics to DOE
ESS-DIVE offers a ready-made Jupyter Notebook tool that can grab information about publication progress:

- Dataset titles & DOIs
- No. of private and public datasets
- Year of publication

Run notebook in browser using Binder, no experience required!

Download a list of all project datasets

Setup search terms once, reuse notebook whenever needed
Link to Project Data Portal Metrics

Provide data portal URL directly in project reports, or copy metric numbers

Next Generation Ecosystem Experiments Tropics (NGEE-Tropics)

NGEE-Tropics is carrying out field and lab studies to generate key new data, which is integrated with diverse existing datasets to improve tropical forest representation in Earth system models.

Summary

65 datasets

77 GB of content

Latest Updates

File formats

108 Metabolite files

150 Data files
What’s Next?

Tentative FY24 plans:

Grant access to all project datasets by PI request

Augment current metrics to include:

- View and download metrics
- Datasets published over time
- Display additional dataset metadata of interest
What do you want project metrics to look like?

**We need your input** to create ESS-wide project data metrics!

What do you want to add to our current metrics?

What do you wish you could report to DOE?

Are there visualizations you’d like to have?

**Goal:** provide a consistent set of metrics to all ESS projects on ESS-DIVE that can be accessed on-demand

https://go.lbl.gov/essdive-metrics-feedback
Making ESS Data FAIR
Reporting Formats

Reporting formats are instructions, templates, and tools for consistently formatting data within a discipline.

These data standards were developed by ESS-DIVE and community partners to standardize metadata and data files of data types commonly collected by DOE ESS projects.

No. of Reporting Formats available:

- **High-level, wide-range of data types**: 7
- **Data type specific**: 6
Benefits of Reporting Formats

The standardized data and metadata templates improve both human- and machine-readability

- Makes data more Findable, Accessible, Interoperable, Reusable
- Set consistent methods of reporting data within a project
- Allows scientists to easily work with many datasets
- Parsable standards enable advanced search
- Planned ESS-DIVE tools for advanced data search, integration, and visualization will be based around these reporting formats
DeepDive: Enhanced Data Search

The Deep Dive API is a new search capability that allows you to look for data **within dataset files** on ESS-DIVE.

Deep Dive locates and extracts data from published datasets that have **completed File Level Metadata**.

ESS-DIVE’s publication workflow has been expanded to include **file reviews** for reporting format datasets.
Suggestions for Getting Started

ESS-DIVE’s Suggestions

● **Start with one** reporting format: File Level Metadata

● **Get a head start**: Adopting reporting formats is easier before or during publication, than after

Publication Requirements

There are none! 🚫

It is **not** a requirement to adopt any reporting formats to publish data on ESS-DIVE

More capabilities to come!

Adopt reporting formats to take advantage of future programmatic tools

File Level Metadata help documentation:

https://ess-dive.gitbook.io/file-level-metadata-reporting-format
Coming soon!
New features to support data management
Community Funds

- **$1M funds** available in FY24
- Submit **white paper** proposals by May 15 (max 2 pages)
  - Email with details and template will be sent later this week
- **Preference given to Priority Topics**
  - New reporting formats
  - **New versions of existing reporting formats**: Improve machine readability and compatibility with FusionDB, BASIN-3D etc., model data archiving for large outputs/ML
  - **Community data curators**: Provide guidance to ESS community on best practices for submitting data to ESS-DIVE & help with adoption of reporting formats
  - Data integration: Tools to integrate ESS-DIVE and other BER data
  - **Community data products**: Products using ESS-DIVE & BER data for broad scientific use
- Can propose other ideas that have clear value for the ESS community
Tracking Dataset Status

All Available Statuses

**DOI Statuses**
- No DOI Added
- Reserved
- Pending
- Live

**Review Statuses**
- Drafting
- Review & Revise
- Published

---

All designs subject to change
Streamlining Publication Process

We will host a webinar with more information when these features are released

Summary: PI To-Do List
PI To-Do List

PIs can work with their data managers to:

● Designate a project data manager(s) who will be responsible for overseeing project data on ESS-DIVE

● Decide which capabilities are relevant for your research and data management plans and enforce them

● ⭐ Register as a data contributor to access all features

● ⭐ Fill out project data manager approval form

● If needed: Contact ESS-DIVE for help accessing your project datasets
PI To-Do List: Communicate with your Project

PIs and data managers can communicate decisions and expectations to your project members:

- Register as a data contributor to submit data
- Link to data published outside of ESS-DIVE
- Upload large volumes of data
- Adopt File Level Metadata reporting format
- Proactively share datasets with project teams

Depending on which features you’d like your project to use
Data Manager To-Do List

PIs or Data Managers can:

● Register as a data contributor to access all features
● Create data portal(s)
● Establish project team(s)
● *Af needed*: Download project publication report and copy data portal metrics

Depending on which features you’d like your project to use
Thank you! Let’s Stay Connected.

Contact us if you are interested in any hands on activities: ess-dive-support@lbl.gov

ESS-DIVE Documentation
docs.ess-dive.lbl.gov

Join our community mailing list!
References

Icons

- Speak up icons created by Kalashnyk - Flaticon, https://www.flaticon.com/free-icons/speak-up
- Link icons created by Smashicons - Flaticon, https://www.flaticon.com/free-icons/link
- Iteration icons created by Uniconlabs - Flaticon, https://www.flaticon.com/free-icons/iteration
- Form icons created by Freepik - Flaticon, https://www.flaticon.com/free-icons/form
- Lock icons created by Freepik - Flaticon, https://www.flaticon.com/free-icons/lock
- Download icons created by Debi Alpa Nugraha - Flaticon, https://www.flaticon.com/free-icons/download
- Paste icons created by Pixel perfect - Flaticon, https://www.flaticon.com/free-icons/paste
- Idea icons created by Pixel perfect - Flaticon, https://www.flaticon.com/free-icons/idea