



Project Data Management on ESS-DIVE

Presenters



Charuleka Varadharajan

*ESS-DIVE Community Engagement
and Partnerships Lead*



Madison Burrus

Computer Systems Engineer

Overview



This session is designed for **ESS project PIs and data managers**

1. How to include **ESS-DIVE in your data management plan** and other project data management considerations (25 min - Charu Varadharajan)
2. **What features are available** on ESS-DIVE for project data management? (1 hour - Madison Burrus)
3. How to **get help and stay engaged** with ESS-DIVE (5 min - Charu Varadharajan)



PI ACADEMY

Fig from UC Boulder

Takeaways



Learn how ESS-DIVE can help your project with DOE ESS data management requirements

- Understand what capabilities ESS-DIVE offers that you can use in a data management plan
- Learn about current and upcoming ESS-DIVE features that you can use to organize your project data and teams
- Become aware of ESS-DIVE outreach efforts to projects

1. Including ESS-DIVE in your Data Management Plan

Charuleka Varadharajan



DOE ESS projects can preserve their data on ESS-DIVE



ESS Project Data Management Plans (DMP) must meet BER requirements for digital data management

<https://ess.science.energy.gov/ber-requirements-and-guidance-for-digital-data-management/>

- Know what DMP sections relate to ESS-DIVE
- Identify the DMP related services ESS-DIVE provides
- Steps to take before your proposal and after to use ESS-DIVE for archiving

ESS-DIVE is the DOE ESS repository for long-term data preservation.

Who is eligible to store data with ESS-DIVE?



	ESS Program Funding	Other Funding Source
Can I store data with ESS-DIVE?	Yes , you can start storing data right away	No , you need to request permission to upload data to ESS-DIVE
Rationale?	All projects funded by the ESS program should store or mirror data on ESS-DIVE	Must provide rationale for why project data is relevant to ESS program
How do I start storing data?	Fill out New Data Contributor form	Fill out New Project Request form



If your project request is approved we will follow up with instructions on how to start uploading data

Proposals and FOAs

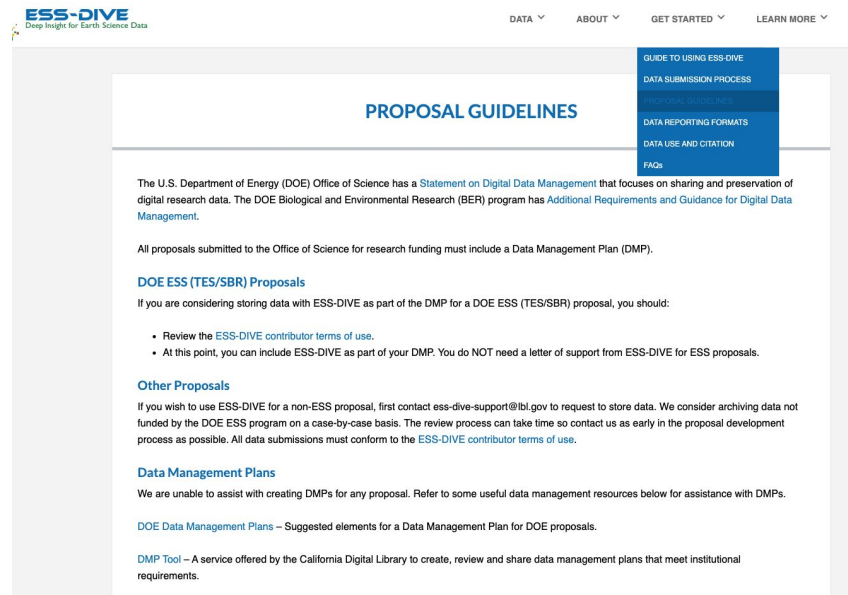


For DOE ESS FOAs:

- Review and agree to ESS-DIVE contributor terms of use
- If your data > default size limits contact ess-dive-support@lbl.gov
- Include ESS-DIVE in your DMP data sharing and preservation plan
- **No need for a letter of support from ESS-DIVE** for ESS proposals

Non-ESS proposals:

- Contact ess-dive-support@lbl.gov for permission to store



Deeper Dive: ESS-DIVE Terms of Use

Step 1: Register as a new data contributor
(<https://docs.ess-dive.lbl.gov/new-user-registration>)

Step 2: Submit data in relevant community standards and apply QA/QC

Step 3: Provide data package documentation & appropriate standardized metadata

Step 4: Agree to the data contributor license (Part B) for serving data to the public.

ESS-DIVE supports Creative Commons Attribution 4.0 and Creative Commons Public Domain policies for data usage rights

ESS-DIVE Data Contributor License Agreement (CLA)



- Allows ESS-DIVE the rights necessary to copy, store, redistribute, and share your data, metadata, and any other content of your work with the public
- The license granted is for your protection as a contributor as well as for the protection of Berkeley Lab;
- It does not change your rights to use your own contributions for any other purpose.

By accepting & agreeing to terms and conditions of the CLA, you reserve all right, title, and interest in and to Your Contributions.

Key Elements of a Data Management Plan (DMP)

- Data Collected, Generated, or Used
- **Standards and reporting formats**
- **Data Sharing and Preservation**
- Related tools, software or code
- Data protection: Security and Integrity
- DMP Oversight
- Rationale for the proposed data management plan

ESS-DIVE is DOE ESS's central location for long-term data preservation

DMP: Data Sharing and Preservation



- **ESS-DIVE provides**
 - Storage of **diverse data** types
 - Public release of data with digital object identifier (**DOI**)
 - Pre-publish DOIs may be available on request
 - **Citation** text for published data
 - Long-term data **search and download** availability
 - **Curation** of dataset metadata and basic file checks
- **ESS-DIVE does not provide**
 - Enforcement of data publication timelines
 - Guidance/obtaining permissions on what project data can be shared

DMP: Data and metadata standards

- **ESS-DIVE provides**
 - **Metadata standards** for data packages
 - Community **reporting formats** for several ESS data and metadata types
 - Basic **guidance for organizing** data packages and authorship
 - **Checks** of metadata quality (FAIR scores) and that files are readable
- **ESS-DIVE does not provide**
 - Project is expected to provide data formatting and quality checking

Other considerations for project data management with ESS-DIVE

- Incorporate **data reporting formats** early in scientific workflows
- Data associated with papers must be made publicly available **within 1 month of publication**
 - Include your data DOI in the research highlight
- Archive **software** and **code**

Archiving data in multiple data repositories

At minimum ESS projects need to store metadata on ESS-DIVE

Options if you have data in other repositories:

- Replicate the **data and metadata** at ESS-DIVE and the other repository
- Replicate the **metadata** at ESS-DIVE and **link to the data** at the other repo
- Store the **data and metadata** in ESS-DIVE and **link to the data** ESS-DIVE from the other repository
- Use the same or a different **DOIs**

2. Overview of Data Management Features

Madison Burrus



Available Data Management Features

Package Service API

Bulk data
uploads/edits

Dataset Sharing

Collaborate

External Linking

Link to other
repos and data
systems

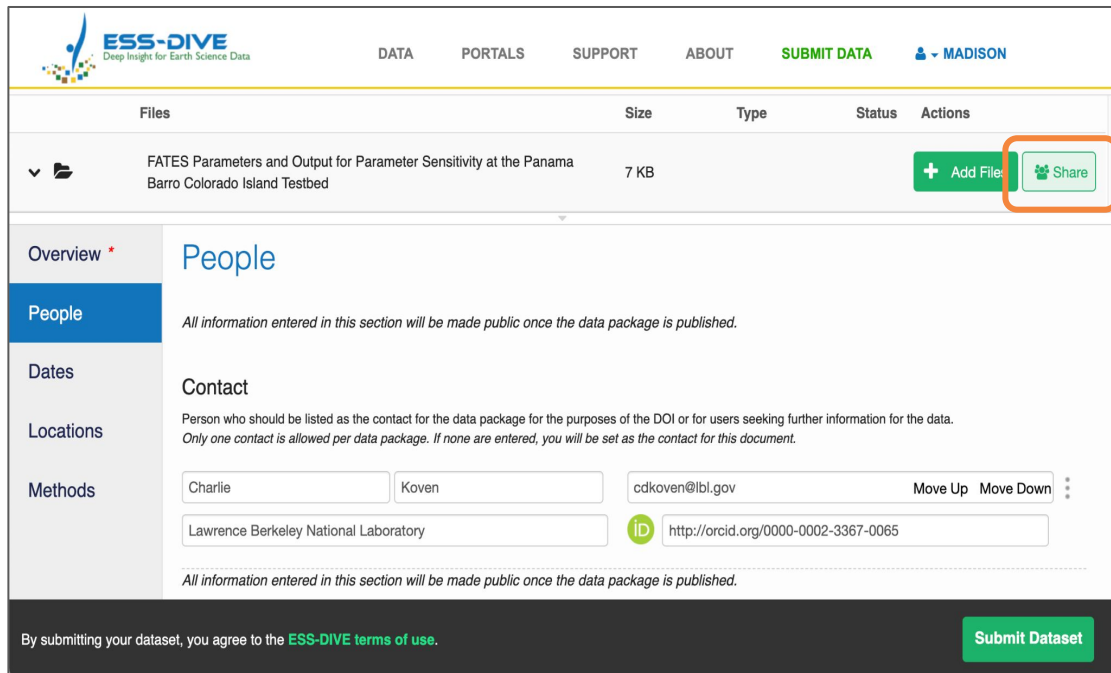
Data Portals

Group related
datasets

Collaborate: Dataset Sharing



What is dataset sharing?

A screenshot of the ESS-DIVE web interface. The top navigation bar includes links for DATA, PORTALS, SUPPORT, ABOUT, SUBMIT DATA, and a user profile for MADISON. Below this is a table of datasets. The first dataset is "FATES Parameters and Output for Parameter Sensitivity at the Panama Barro Colorado Island Testbed", which is 7 KB in size. In the "Actions" column for this dataset, there is a green "Add Files" button and a green "Share" button. An orange box highlights the "Share" button, with an orange arrow pointing from it to a larger, isolated "Share" button on the right. Below the table, there is a section for "People" where contact information can be entered. This section includes a "Contact" field with a description: "Person who should be listed as the contact for the data package for the purposes of the DOI or for users seeking further information for the data. Only one contact is allowed per data package. If none are entered, you will be set as the contact for this document." There are input fields for "Name" (containing "Charlie"), "Email" (containing "cdkoven@lbl.gov"), and "Institution" (containing "Lawrence Berkeley National Laboratory"). There is also an "ORCID iD" field with the URL "http://orcid.org/0000-0002-3367-0065". At the bottom of the page, there is a "Submit Dataset" button and a note about agreeing to the ESS-DIVE terms of use.

Review and Edit
datasets with registered
team members

Dataset Sharing Permission Types

Viewer

- **view** private datasets

Editor

- **view** private datasets
- **edit** metadata
- **add/remove** files
- **publish** datasets

Manager

- **view** private datasets
- **edit** metadata
- **add/remove** files
- **publish** datasets
- **share** datasets with others

Why contributors should share datasets

Collaborate with your team to complete datasets

- If **someone cannot complete a dataset they started**, they can share it with someone else to complete it
- Allow co-authors and team members to **review and correct** your private datasets before publishing
- **Split responsibilities**: one person can upload files while someone else curates dataset metadata

Why contributors should share datasets

Project Data Managers can report on publication progress

- **Share project datasets with PIs and data managers** to allow them to view or manage the data
- **Verify** that:
 - Datasets are being published on time
 - Report on how many active datasets are on ESS-DIVE
 - Important metadata is consistent

Dataset Sharing Demo



How to access dataset sharing and assign permissions

[Data Sharing Tutorial Webinar](#)

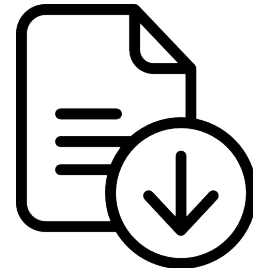
[Data Sharing Instructions Documentation](#)

Upload: Package Service API



What is the Package Service Application Programming Interface (API)?

A scripting tool used to **programmatically submit, edit, and download** datasets on ESS-DIVE.



Why use the Package Service API?

You can script code that manages multiple datasets at the same time

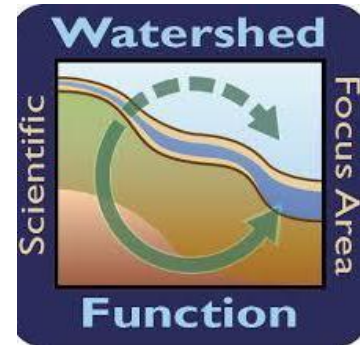
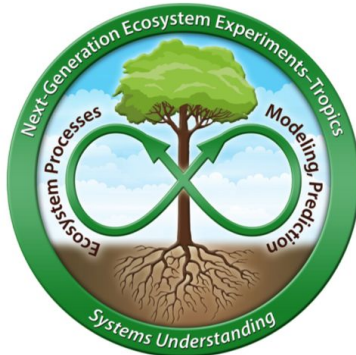
1. Curate Metadata
2. Transfer Existing Datasets
3. Report on Publication Progress
4. Easy to Use

1. Curate Metadata

- **Create multiple new datasets** with repetitive metadata
- **Copy and paste** dataset metadata into new or existing datasets (*i.e. Authors, Geographic Descriptions*)
- **Download** some or all dataset metadata and/or files
- **Share multiple datasets** with the same people/permissions at once

2. Transfer Existing Datasets

- **Mirror**, or copy, datasets from internal, team-only project archives
- **Reduce manual effort** required to duplicate existing metadata
- We have **experience supporting projects** through the process, email us to discuss if this is something your project needs



3. Report on publication progress

- **Search** through all your project's datasets for specific metadata
- **Retrieve and review** specific dataset metadata
- Easy way to create lists of data on ESS-DIVE **for annual reports!**

4. Easy to Use!

- **Easy to use** for non-programmers!
- Download or copy our tutorials scripts to **test it right away**
- Scripts can be written in **R, Python, or Java**

How to start using the API

No Demo:

this is more than we can cover in this training session

- We have a lot of documentation and webinar recordings available that detail how to use the API, please refer to these

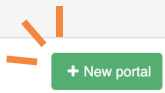
[Detailed API Tutorial Webinar](#)




[Overview API Webinar](#)

[Detailed API Instructions Documentation](#)

Group: Data Portals

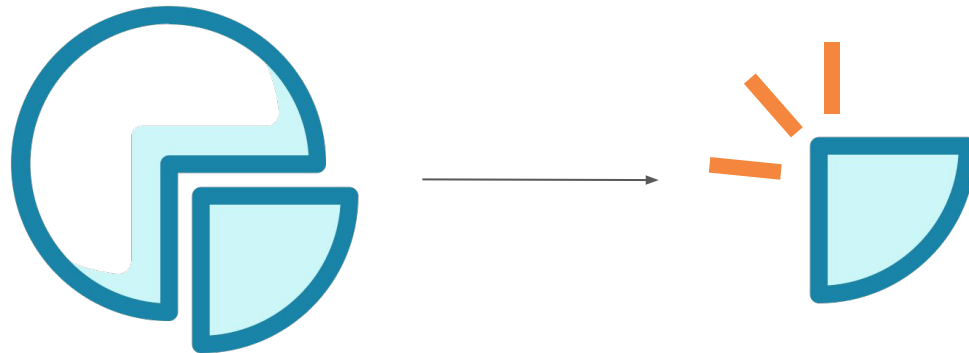
My Portals



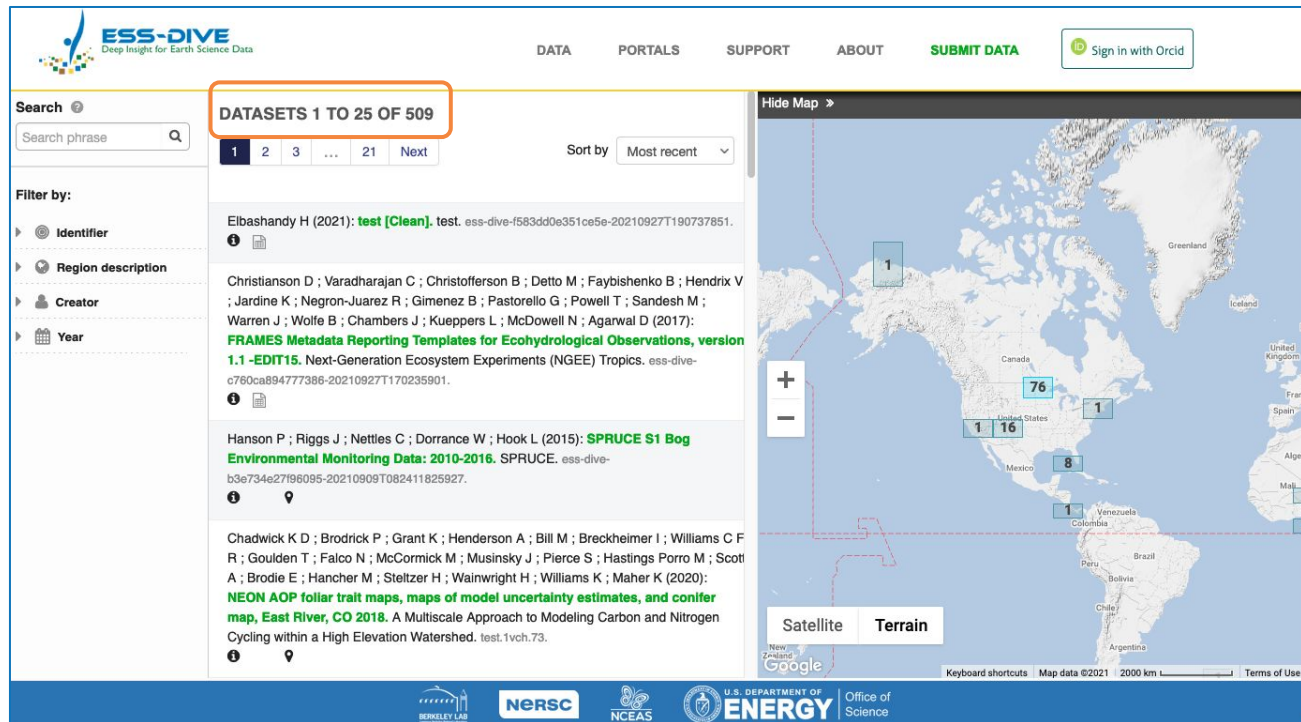
Logo	URL Name	Title	
	east-river-watershed	Data from the East River Watershed, Colorado	Edit
	PNNLRiverCorridorSFA	PNNL River Corridor SFA	Edit
	WHONDRS	Worldwide Hydrobiogeochemistry Observation Network for Dynamic River Systems	Edit

What is a data portal?

A data portal is a **collection** of any ESS-DIVE **datasets**. Data portals allow you to easily **highlight and share** related datasets.



ESS-DIVE Search Page: default view



The screenshot shows the ESS-DIVE Search Page interface. At the top, there is a navigation bar with links for DATA, PORTALS, SUPPORT, ABOUT, and SUBMIT DATA, along with a 'Sign in with Orcid' button. The main content area is divided into three sections:

- Search Bar:** A search phrase input field with a magnifying glass icon.
- Filter by:** A section with expandable filters for Identifier, Region description, Creator, and Year.
- Datasets List:** A list of datasets, with the first two highlighted:
 - Dataset 1:** Elbashandy H (2021): **test [Clean]**. test. ess-dive-f583dd0e351ce5e-20210927T190737851.
 - Dataset 2:** Christianson D ; Varadharajan C ; Christofferson B ; Detto M ; Faybishenko B ; Hendrix V ; Jardine K ; Negron-Juarez R ; Gimenez B ; Pastorello G ; Powell T ; Sandesh M ; Warren J ; Wolfe B ; Chambers J ; Kueppers L ; McDowell N ; Agarwal D (2017): **FRAMES Metadata Reporting Templates for Ecohydrological Observations, version 1.1 -EDIT15**. Next-Generation Ecosystem Experiments (NGEE) Tropics. ess-dive-c780ca894777386-20210927T170235901.


On the right side, there is a map showing the distribution of datasets across the Americas, with numbered markers (1, 76, 16, 8, 1, 1) indicating specific locations. The map includes a 'Hide Map' button and a 'Satellite/Terrain' toggle.

At the bottom, there is a footer with logos for BERKELEY LAB, NERSC, NCEAS, and the U.S. DEPARTMENT OF ENERGY Office of Science, along with keyboard shortcuts, map data, and a 2000 km scale bar.

The main data Search Page is a **view of every dataset** published on ESS-DIVE

Data Portals: unique data views

Hosted by **ESS-DIVE**

 Carbon Dioxide Information Analysis Center

About Fossil-Fuel Emissions Vegetation Response to CO₂ FAQs **Data**

Search

Search these datasets

DATASETS 1 TO 25 OF 226

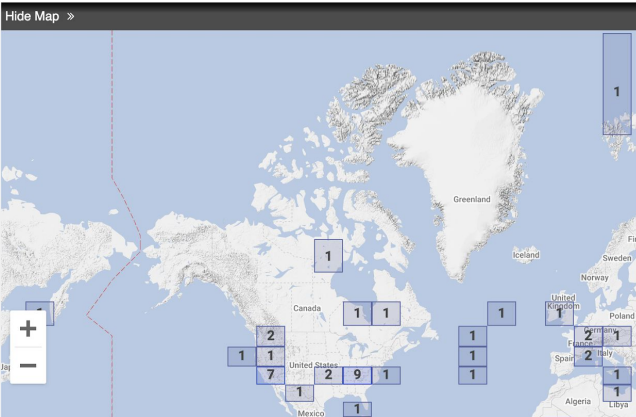
1 2 3 ... 10 Next Sort by Most recent ▾

Prinn R; Weiss R; Arduini J; Arnold T; DeWitt H; Fraser P; Ganesan A; Gasore J; Harth C; Hermansen O; Kim J; Krummel P; Li S; Loh Z; Lunder C; Maione M; Manning A; Miller B; Mitrevski B; Muhle J; O'Doherty S; Park S; Reimann S; Rigby M; Salto T; Salameh P; Schmidt R; Simmonds P; Steele L; Vollmer M; Wang H J (; Yao B; Yokouchi Y; Young D; Zhou L (2018): **History of chemically and radiatively important atmospheric gases from the Advanced Global Atmospheric Gases Experiment (AGAGE)**. doi:10.3334/CDIAC/ATG.DB1001
 ⓘ 📄 📍 99 4 📄 56 👁 904

Kovacs K; Horvath E (2000): **The United Nations Population Statistics Database**. doi:10.15485/1464266
 ⓘ 📄 👁 168


Blake D (2005): **Methane, Nonmethane Hydrocarbons, Alkyl Nitrates, and Chlorinated Carbon Compounds including 3 Chlorofluorocarbons (CFC-11, CFC-12, and CFC-113) in Whole-air Samples (April 1979 — December 2012)**. doi:10.3334/CDIAC/ATG.002
 ⓘ 📄 👁 232

Hide Map »



This data portal is a **view** of **226 datasets** related to a specific research topic

What's inside a Data Portal?




Worldwide Hydrobiogeochemistry Observation Network for Dynamic River Systems

<https://whondrs.pnnl.gov>

[About](#)[Data](#)[Metrics](#)

Data from the Worldwide Hydrobiogeochemistry Observation Network for Dynamic River Systems

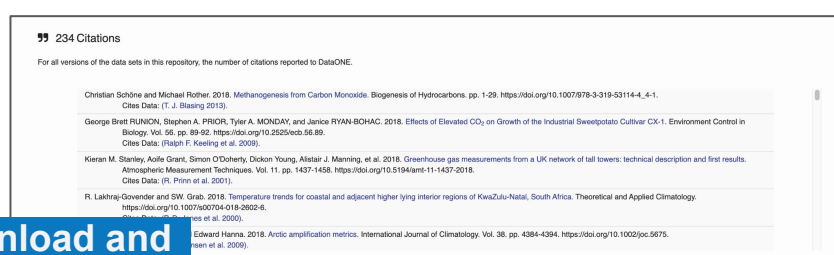
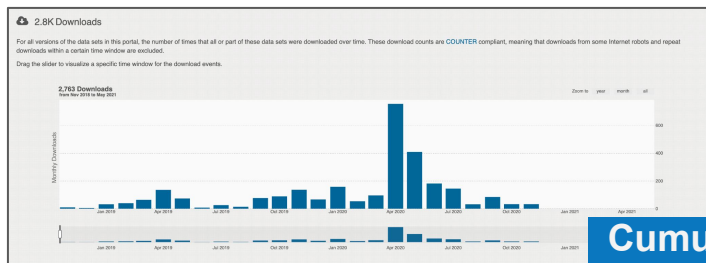
For more information about WHONDRS, go to <https://whondrs.pnnl.gov>



The Worldwide Hydrobiogeochemistry Observation Network for Dynamic River Systems (WHONDRS) is a research consortium that aims to understand coupled hydrologic, biogeochemical, and microbial function within river corridors, with an emphasis on increasing accessibility of resources and knowledge throughout the research life cycle. WHONDRS seeks to galvanize a global community around understanding these coupled systems from local to global scales and ultimately to provide the scientific basis for improved management of dynamic river corridors throughout the world.

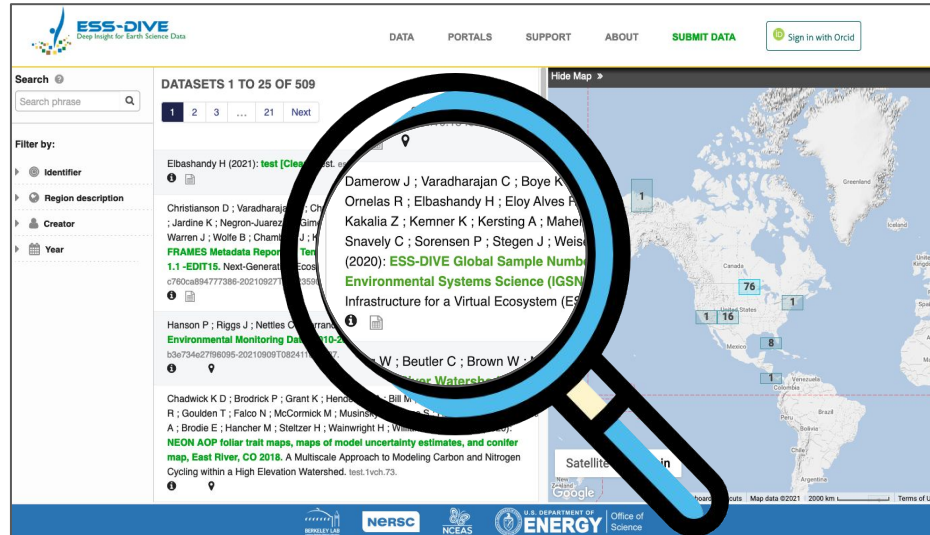
Go to <https://whondrs.pnnl.gov> to learn about WHONDRS research and ways that you can get involved.

Custom Descriptive Pages



Cumulative Download and Citation Metrics

Why use data portals?



- Increase **Findability** of project datasets
- Elevate **public project identity** in ESS-DIVE
- Collect **cumulative metrics** across all project datasets
- Link to project data portals in **annual reports** and **paper publications**

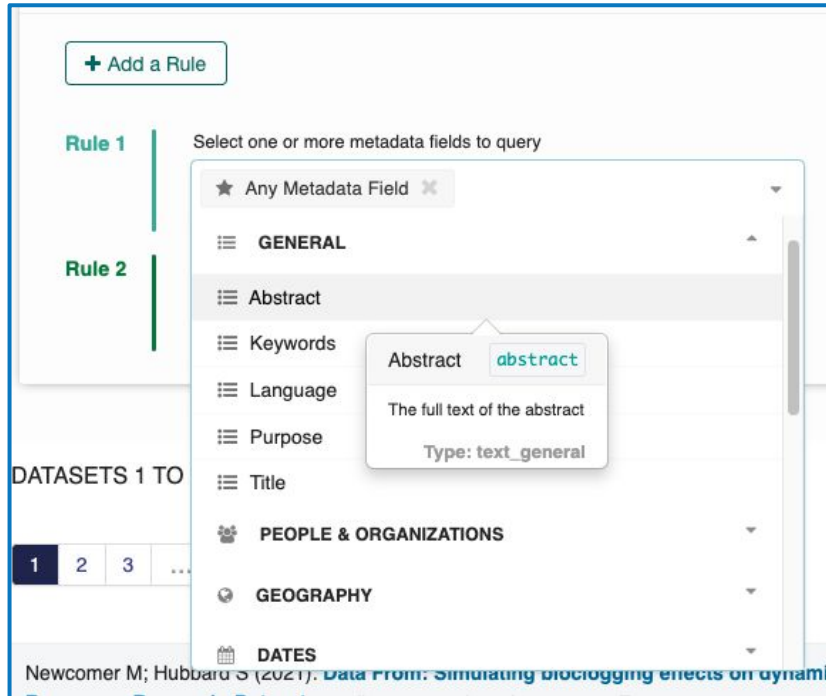
Public Listing of Data Portals

A screenshot of the ESS-DIVE website's "PORTALS" page. The header includes the ESS-DIVE logo, navigation links (DATA, PORTALS, SUPPORT, ABOUT, SUBMIT DATA), and a "Sign in with Orcid" button. The main content area is titled "All portals" and lists five data portals, each with a logo, title, and brief description.

Portal Logo	Portal Title	Description
	Alpine Treeline Warming Experiment Data Portal	The Alpine Treeline Warming Experiment (ATWE) ran from 2008-2016 on Niwot Ridge, in the Front Range of the Colorado Rocky Mountains. It combined common gardens with climate manipulations, using infrared heaters to warm soil and plant surfaces by an amount comparable to current average projections...
	Carbon Dioxide Information Analysis Center	The Carbon Dioxide Information Analysis Center (CDIAC) had the primary responsibility for providing the US government and research community with global warming data and analysis as it pertains to energy issues. The CDIAC archive data holdings have been transferred predominantly to the U.S....
	Data from the East River Watershed, Colorado	This is a data portal to collect all data collected from the East River Watershed, Colorado.
	Worldwide Hydrobiogeochemistry Observation Network for Dynamic River Systems	https://whondrs.pnnl.gov
	SPRUCE Experiment Data Portal	TBD - description

Your project data portal will be listed on **ESS-DIVE's portal page**
Anyone from the public can access this page and locate your data portal

Data Portal Demo



How to add project datasets to a portal:

Creating a data portal is easy!

[Data Portal Tutorial Webinar](#)

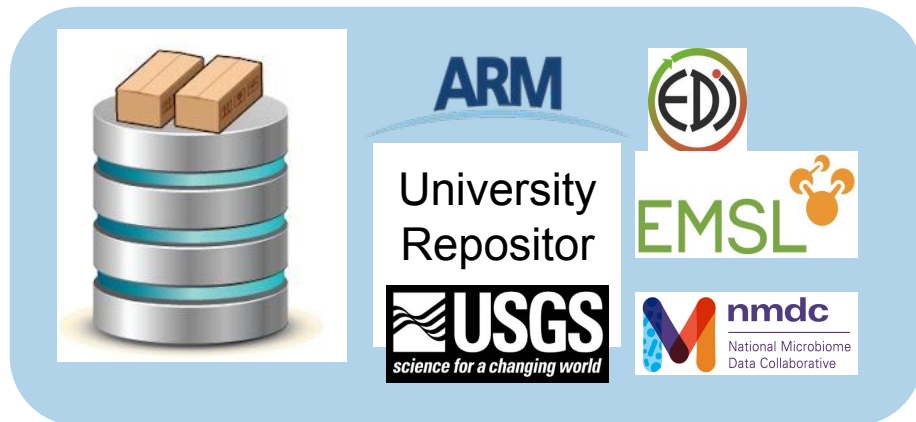
[Data Portal Instructions Documentation](#)

Integrate: External Linking




What is External Linking?

Data and complex data systems **stored on other repositories** can be connected to ESS-DIVE metadata **without needing to upload** it directly to the dataset



What does external linking look like?



**ESS-DIVE**
Deep Insight for Earth Science Data

DATAPORTALSUPPORTABOUTSUBMIT DATA

Sign in with Orcid

< Back to search | Home / Search / Metadata

Molotch N ; Lestak L ; Kueppers L ; Petzelka J ; Meromy L ; Schneider D ; Barnhart T ; Baker E (2022): Snow depth sensor measurement data for Upper Sub Alpine site, 2010 - 2015. Subalpine and Alpine Species Range Shifts with Climate Change: Temperature and Soil Moisture Manipulations to Test Species and Population Responses (Alpine Treeline Warming Experiment), ESS-DIVE repository. Dataset. doi:10.6073/pasta/388c70359c8d0cd79dc3875d6510071d accessed via <https://data.ess-dive.lbl.gov/datasets/doi:10.6073/pasta/388c70359c8d0cd79dc3875d6510071d> on 2022-05-17

Downloads 29Citations 0Views 24Copy CitationAssessment report

Files in this dataset Package: ess-dive-44b1b49c5d7a7ff-20220414T225815865

Name	File type	Size	Download All
Metadata: Snow_depth_sensor_measurement_data_for_Upper_Sub.xml	EML v2.2.0	15 KB	Download

External Links to Data or Metadata

External links for this dataset		
Description	Relationship	URL
doi:10.6073/pasta/388c70359c8d0cd79dc3875d6510071d	[same as] Original publication of this dataset (where the data+metadata can be found)	https://doi.org/10.6073/pasta/388c70359c8d0cd79dc3875d6510071d

What does external linking look like?

Downloads 29

Citations 0

Views 24

Copy Citation

Assessment report

Files in this dataset Package: ess-dive-44b1b49c5d7a7ff-20220414T225815865

Name	File type	Size	Download All
Metadata: Snow_depth_sensor_measurement_data_for_Upper_Sub.xml	EML v2.2.0	15 KB	Download

External Links to Data or Metadata

External links for this dataset		
Description	Relationship	URL
doi:10.6073/pasta/388c70359c8d0cd79dc3875d6510071d	[same as] Original publication of this dataset (where the data+metadata can be found)	https://doi.org/10.6073/pasta/388c70359c8d0cd79dc3875d6510071d

Why include external links?

The **ESS publication policy requires** mirroring datasets on ESS-DIVE even if the data are already published with a DOI

Why include external links?

The **ESS publication policy requires** mirroring datasets on ESS-DIVE even if the data are already published with a DOI

Consider using this feature if:

- Your data is spread across **established** repositories (e.g. NMDC, EDI, ESS-DIVE) and you **don't want to replicate** existing public datasets
- You want to establish links between data on different repositories
- Individual files in a dataset are more useful when viewed from a particular **data or analysis platform** (e.g. *google earth engine*)



External Linking Demo

Quick review of what external links look like and where to find them

[External Linking Tutorial Webinar](#)

[External Linking Instructions and Documentation](#)

Review and Next Steps

Review: Available Data Management Features



Package Service API

Create and edit
many datasets

Dataset Sharing

Collaborate on
editing datasets
with project
members

External Linking

Connect
datasets to
existing data
products and
publications

Data Portals

Collect project
datasets into a
publicly
accessible
webpage

Next Steps: Detailed How-To Instructions

A screenshot of the ESS-DIVE Docs website. The page has a light gray background. At the top left is the ESS-DIVE logo and the text "ESS-DIVE Docs". To the right of the logo are four navigation links: "Home", "About", "Data Portal", and "Contact Us". On the left side, there is a vertical menu with several items: "Data Submission Tutorials" (with a dropdown arrow), "Web Form Tutorial", "Package Service Tutorial" (with a green checkmark), "R Example", "Python Example", "Java Example", "API Updates and Changes", "Package Level Metadata Guide", "Data Reporting Formats", "Sharing Data" (with a right arrow), "Link to External Data Sources", "Register Dataset Citations", "PORTALS", "Why Use Data Portals" (with a right arrow), and "Create a Portal" (with a right arrow). The main content area on the right has a large "Welcome" heading. Below it is a paragraph: "This website serves as an in-depth guide to ESS-DIVE and is home to all help documentation. Choose one of the four sections below to get started, or navigate using the left-hand menu." Below this paragraph are two sections: "ESS-DIVE Main Data Portal: <https://data.ess-dive.lbl.gov>" and "Help Desk: ess-dive-support@lbl.gov". Below these is a section titled "Preferred Browsers" with a bulleted list: "Chrome", "Firefox", "Safari", and "May have issues with Internet Explorer". At the bottom of the main content area are two boxes. The first box has a document icon and the text "Searching & Accessing Data". The second box has a document icon and the text "Contributing Data".

Sharing Data

docs.ess-dive.lbl.gov/contributing-data/sharing-data

Package Service API

docs.ess-dive.lbl.gov/contributing-data/data-submission-guidelines/package-service-tutorial

External Linking

docs.ess-dive.lbl.gov/contributing-data/link-to-external-data-sources

Portals

docs.ess-dive.lbl.gov/portals/why-use-data-portals

Future Feature: Project Spaces

Charuleka Varadharajan

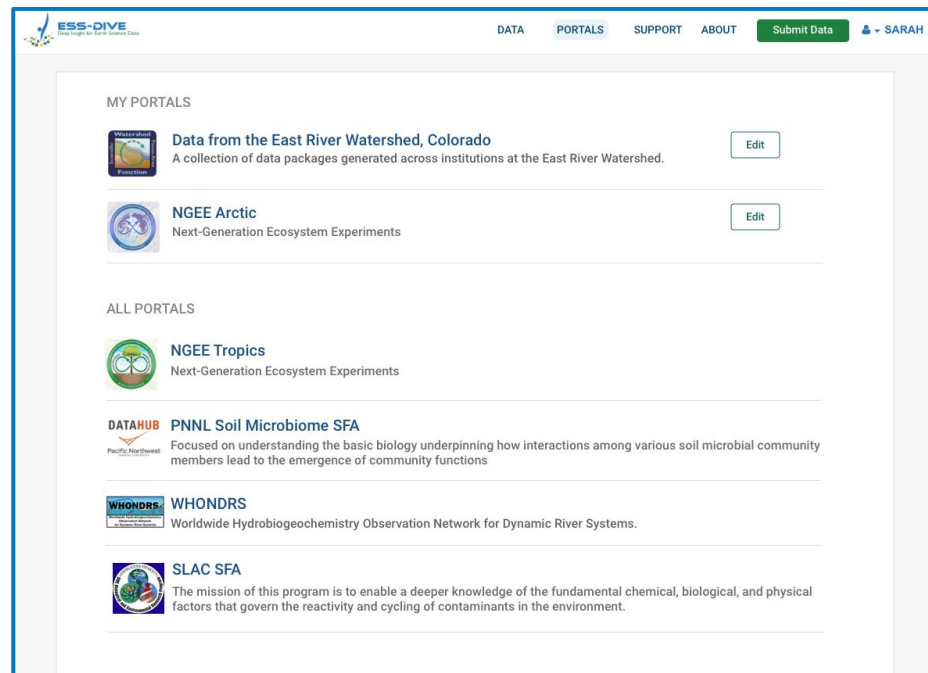
Upcoming: Project Spaces to Enable Project Data and User Management, Team Collaboration



Desired Features from User Interviews

- Creating project data **collections**
- Project **administration** of users and **curation** of data packages
- **Internal sharing within project teams** to promote collaboration

Coming Fall 2022!



Add project search filter and private tab features



Search ?

Filter by:

Project

Identifier

Region description

Creator

Year

☐ Data coverage ?

☐ Publish year ?

Access

☐ Private datasets only

data.ess-dive.lbl.gov

Filters allow you to quickly find datasets associated with your project.

Select “**Private datasets only**” to view datasets that have not been published yet.

It's important to keep project information updated on ESS-DIVE



- **Project information may change over time.** The default project information provided to ESS-DIVE is when the funding was awarded.
- Don't let this affect your data! Inconsistent **project titles** can lead to incorrect data citations and make it difficult to manage project data.
- **We can help!** PIs or project managers can **send a request to ess-dive-support@lbl.gov to update project information.** Look out for further announcements from ESS-DIVE asking projects to update information.

3. How to get help and stay engaged with the ESS-DIVE team

Charuleka Varadharajan

Customized project virtual “site visits”



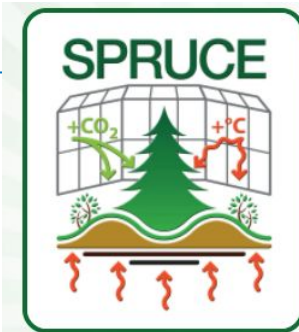
Using ESS-DIVE for Project Data Management and Publication

Presented by Joan Damerow and Rob Crystal-Ornelas
Earth and Environmental Sciences Area
Berkeley Lab



Using ESS-DIVE's CSV and File-Level Metadata reporting formats

Presented by Rob Crystal-Ornelas
Earth and Environmental Sciences Area
Berkeley Lab



Personalized email outreach campaign



EDO-32 Checking in + How can we help you submit your datasets? > Inbox x



Dylan O'Ryan <jira@ess-dive.atlassian.net>
to me ▾

4:21 PM (0 minutes ago)



Hello,

My name is Dylan O'Ryan, and I work in data curation and community support for [ESS-DIVE](#) (DOE's Environmental Systems Science data repository).

I'm reaching out on behalf of the ESS-DIVE team regarding your recent publication. Is there any DOE ESS data associated with this publication? If there is, do you need any help with submission and publication of the data on ESS-DIVE?

Warm regards and with thanks,
Dylan *on behalf of the ESS-DIVE team*

To reply to this email, please write above the line at the top and keep the same subject line.

[View request](#) · [Turn off this request's notifications](#)

ESS-DIVE sent you this message using JIRA Service Desk.

Join our community mailing list

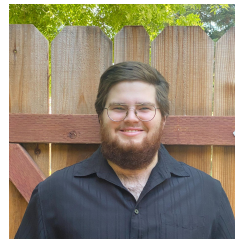
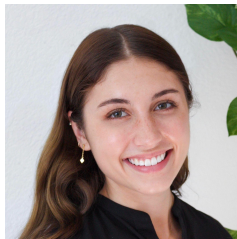
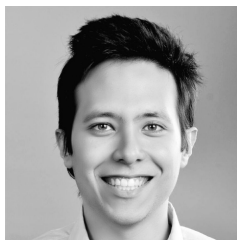


- Learn about the **latest feature upgrades** in the quarterly newsletter
- Get notified of service **outages**
- Hear about **upcoming webinars**

ESS-DIVE Team



ESS-DIVE Leadership



Contact us at any time at ess-dive-support@lbl.gov

Connect With Our Team!

To get help:

ess-dive.lbl.gov

ess-dive-support@lbl.gov

docs.ess-dive.lbl.gov



To stay updated:

ess-dive-community@lbl.gov

[@essdive](https://twitter.com/essdive)



Acknowledgements

Advisory Groups: ESS-DIVE Archive Partnership Board, ESS Cyberinfrastructure Working Groups

Funding: EESSD Data Management



Questions?