Project Data Management
on ESS-DIVE

ESS PI Meeting Breakout 2022

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)

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
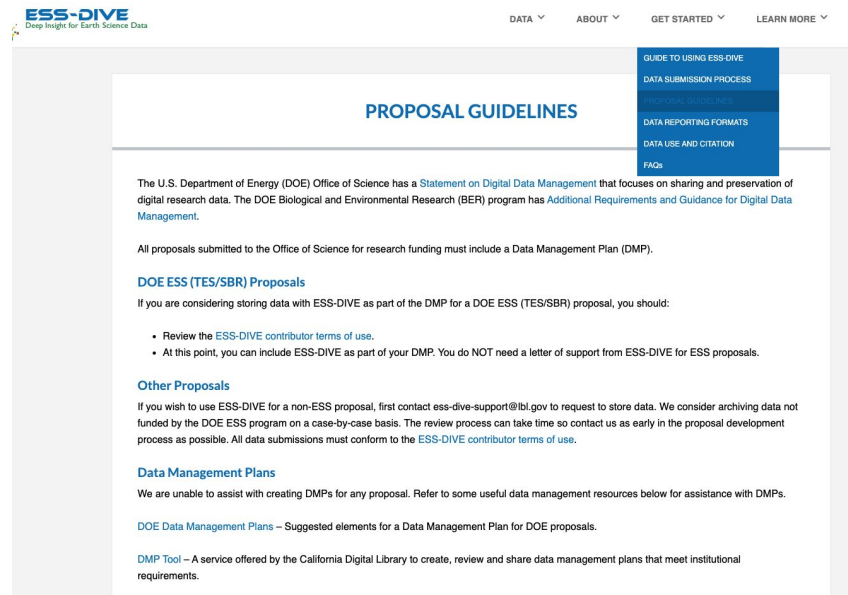
Deeper Dive: ESS-DIVE Terms of Use

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 <u>not</u> 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 <u>not</u> provide**

  - Project is expected to provide data formatting and quality checking

13

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	Dataset Sharing	External Linking	Data Portals
Bulk data uploads/edits	Collaborate	Link to other repos and data systems	Group related datasets
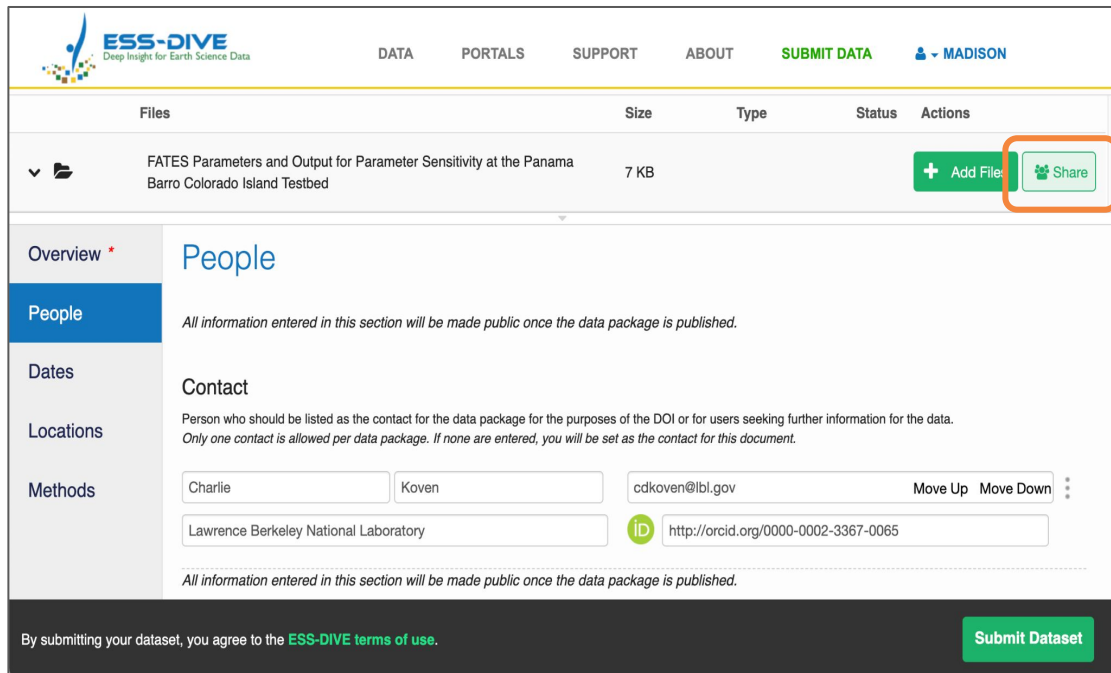
Collaborate: Dataset Sharing

What is dataset sharing?

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.
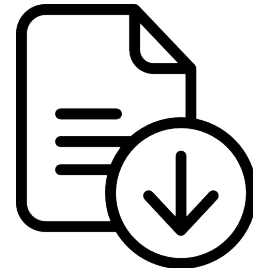
Why use the Package Service API?

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
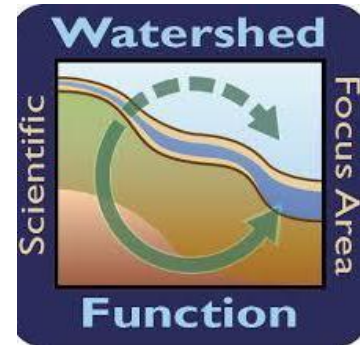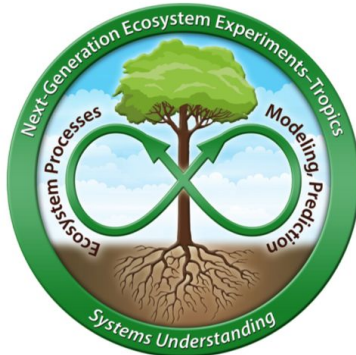
3. Report on publication progress

- **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
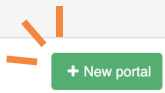
Detailed API Tutorial Webinar

Overview API Webinar

Detailed API Instructions Documentation

Group: Data Portals
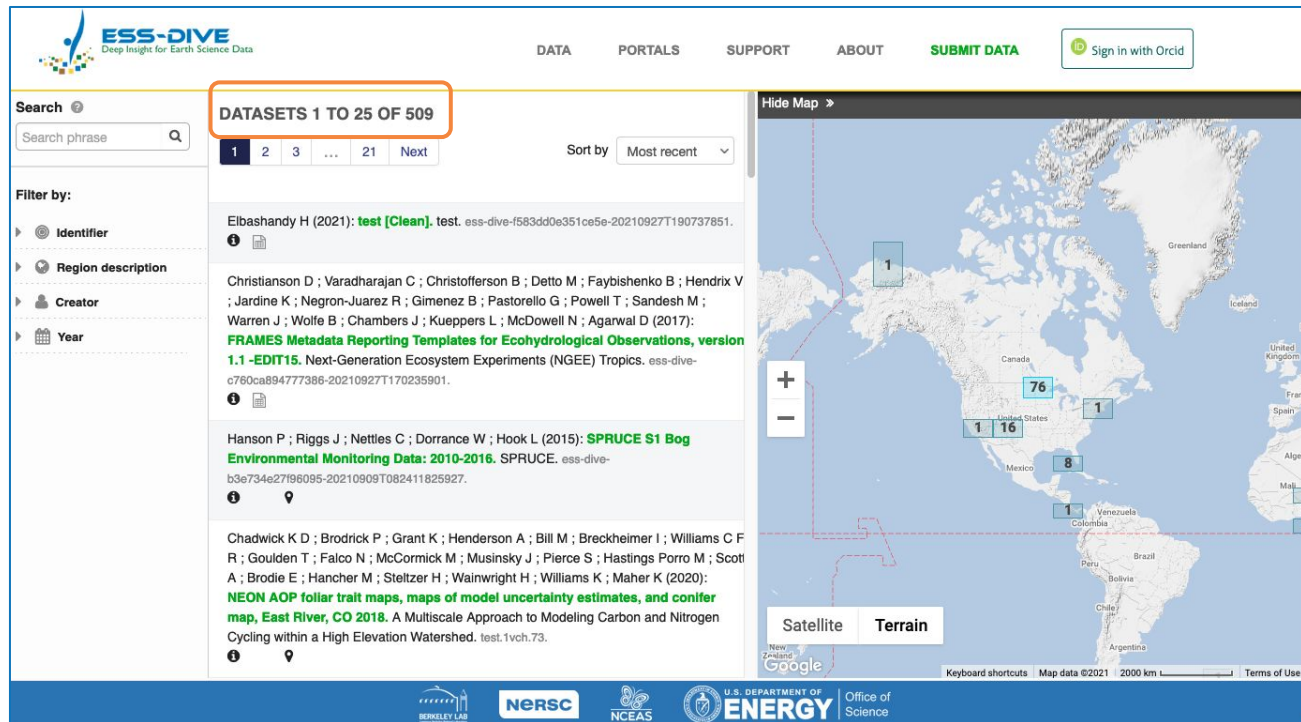
What is a data portal?

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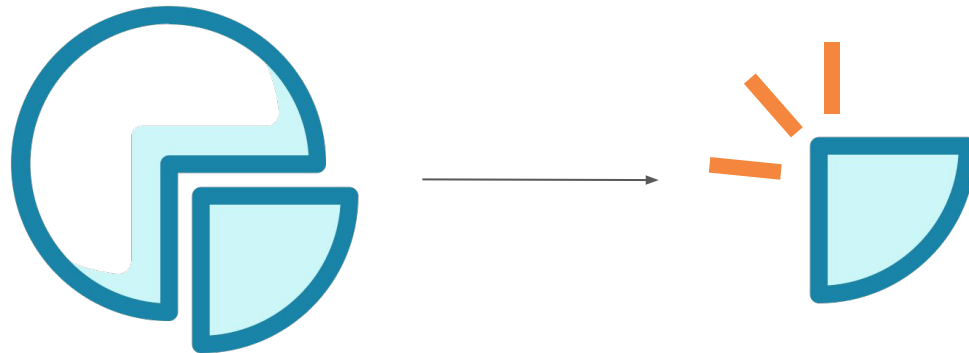
ESS-DIVE Search Page: default view

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

Data Portals: unique data views

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

What's inside a Data Portal?

Custom Descriptive Pages
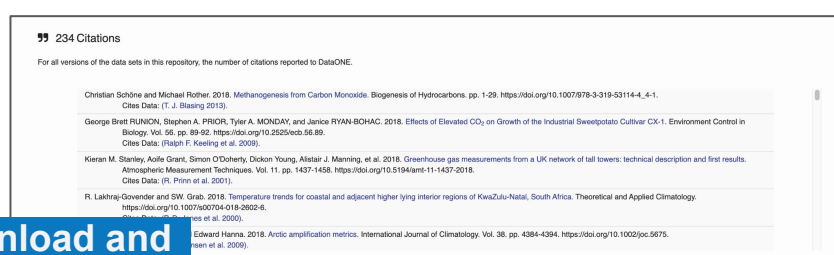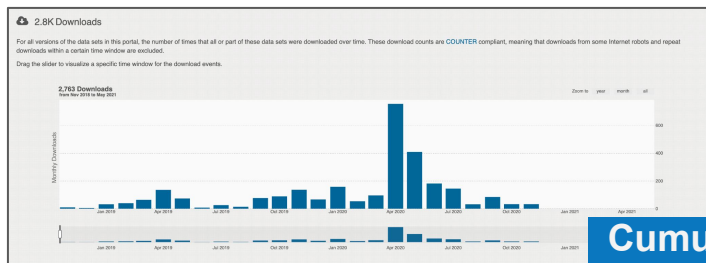
Cumulative Download and Citation Metrics

Why use data portals?
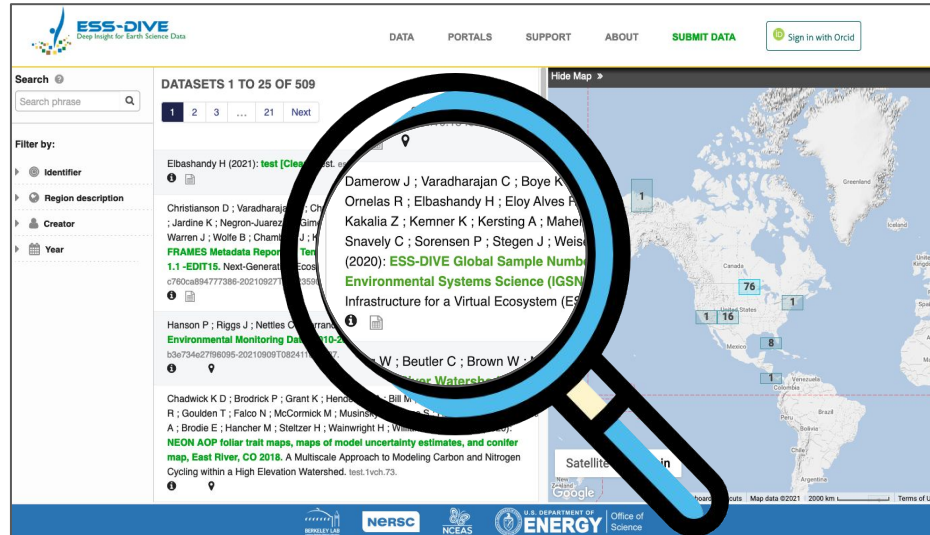
- **Increase Findability** of project datasets

- Elevate **public project identity** in ESS-DIVE

- Collect **cumulative metrics** across all project datasets
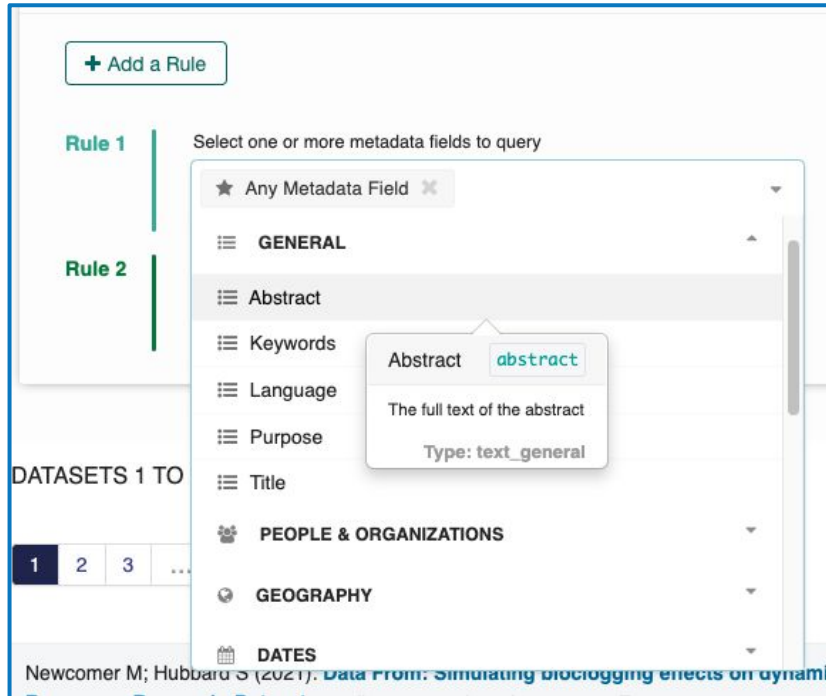
- Link to project data portals in **annual reports** and **paper publications**

Public Listing of Data Portals

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:

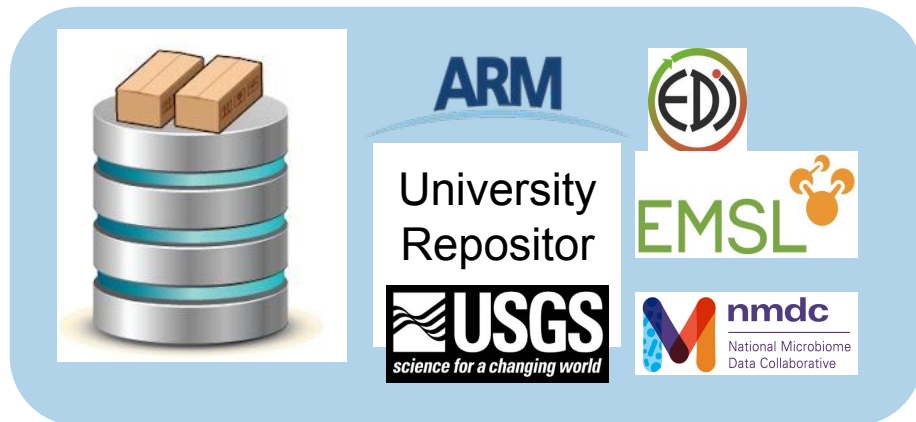**Creating a data portal is easy!**

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?

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

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
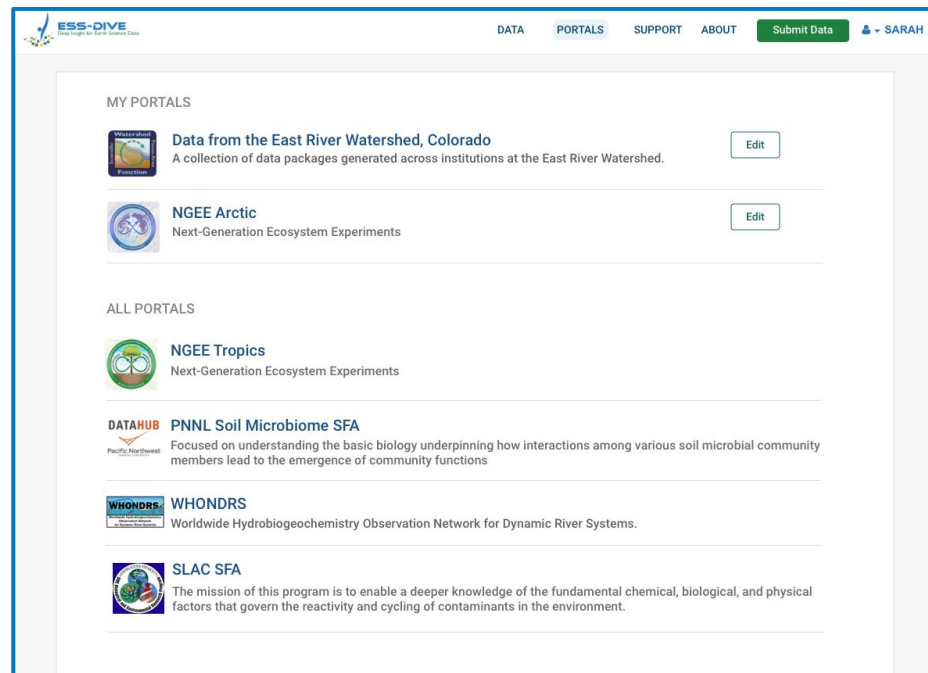*Charuleka Varadharajan*

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

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 <u>is when the funding was awarded.</u>

- 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** <u>ess-dive-support@lbl.gov</u> **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

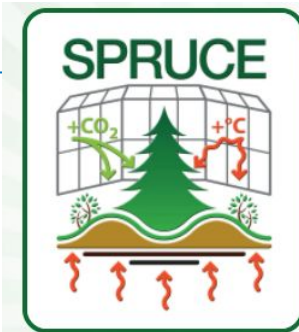# Customized project virtual "site visits"

Personalized email outreach campaign

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

Dylan O'Ryan <jira@ess-dive.atlassian.net>   4:21 PM (0 minutes ago)

to me

_____

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.
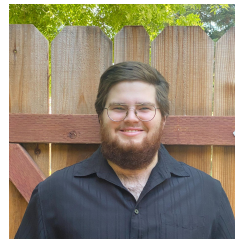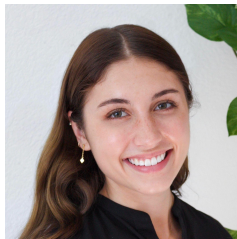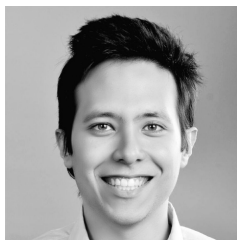
Join our community mailing list

- 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

Questions?