

## **ESS-DIVE New User Bootcamp**

**Emily Robles** 

**Environmental Data Management Research Associate** 









#### **Tutorial Overview**



#### **Part 1: Creating Datasets**

- Components of a dataset
- Choosing an upload method
- Hands-on: Create a dataset on the ESS-DIVE Sandbox

#### **Part 2: Data Publication Process**

- Publication cycle and review workflow
- Reviewing Assessment Reports

**Break** 

#### Part 3: Data Search and Download

- Navigating the Data Search page
- How to use filter and search capabilities
- Hands-on: Search for and download published data

# Part 4: Registering as a Data Contributor

Completing the registration form

**Remaining Questions** 





By the end of this presentation, you should be able to..

- Create and publish datasets
- Share your datasets with collaborators and team members
- Search for and download datasets from ESS-DIVE
- Utilize ESS-DIVE's help documentation, resources, and support



### **CREATING A DATASET**

### Dataset publication at a glance



#### Collect your data files



01

Collect and organize data files related to your findings, tables, and figures.

Refer to ESS-DIVE reporting formats.

#### Upload data + metadata



02

Ensure metadata complies with ESS-DIVE requirements.

Choose your upload method and submit your dataset.

Review your automated assessment report.

#### Review and publish



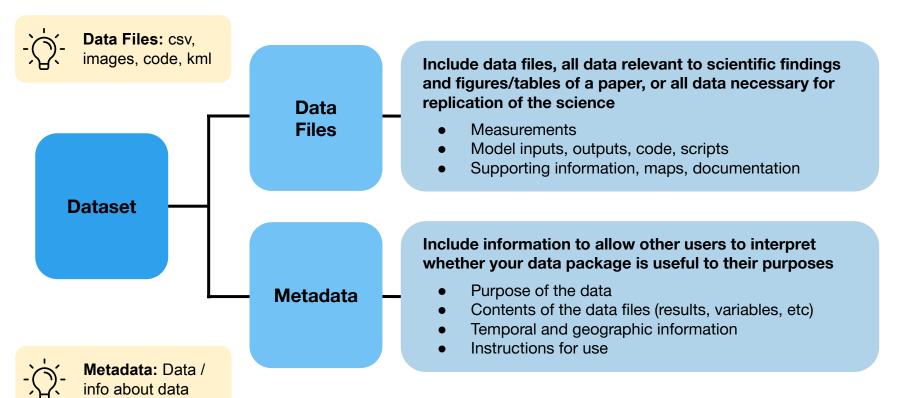
03

Request publication and address revision requests from ESS-DIVE team review.

Meet the funding/journal publishing time frame.

### Components of a dataset





### Choosing an upload method



Upload Method	File volume	No. of files	Considerations
Data Submission Web Form	< 10 GB	< 100	Immediate upload. Easy to follow guidance throughout form.
Package Service API	10 - 500 GB	< 100	Takes initial investment of time to set up, but immediate upload. Technical skills helpful but not required (see tutorials).
High Bandwidth Upload Service	> 500 GB	> 100	Longest process, and requires collaboration with Support Team and use of a separate desktop app. Can use if unfamiliar with scripting.

Contact <a href="mailto:ess-dive-support@lbl.gov">ess-dive-support@lbl.gov</a> if you need to:

(a) upload more than 10GB of data or (b) use the High Bandwidth Upload Service





#### Package Service API

- A programmatic method creating new datasets and editing existing datasets
  - Upload files and create metadata using JSON-LD
- Bulk edit/upload capabilities
- Download any public datasets

#### **Data Submission Web Form**



- Create and edit datasets one at a time
- Drag and drop or select files from your file manager to upload data
- Edit datasets through the web form

### Tools and support for dataset creation





### Help Documentation

Refer to ESS-DIVE's website and Gitbooks for detailed information on dataset requirements.



### Metadata Guide and Offline Template

Review our complete **Metadata Guidelines**.

Collaborate on metadata with co-authors using an offline template before working on ESS-DIVE.



### Sandbox Testing Server

Practice uploading data packages to our sandbox test server, which does not permanently save data.



### Support Email Service

Feel free to contact the ESS-DIVE support team with questions!

ess-dive-support@lbl.gov ess-dive.lbl.gov/contact/



# **HANDS-ON**Creating and Submitting a Dataset



### Break - Create an ORCiD

The ESS-DIVE Sandbox and Main Portal both require an ORCiD account to login and submit data.

https://orcid.org/register





## Navigate to ESS-DIVE Sandbox

Sandbox is a test system where you can practice uploading data. This data is **NOT** preserved in ESS-DIVE.

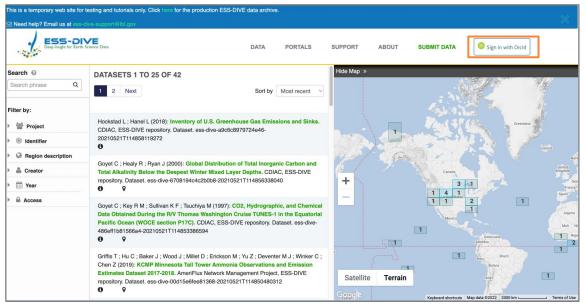
https://data-sandbox.ess-dive.lbl.gov

This is a temporary web site for testing and tutorials only. Click here for the production ESS-DIVE data archive.

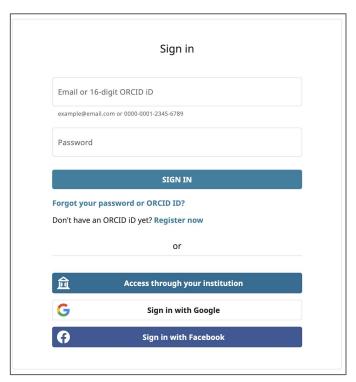
✓ Need help? Email us at ess-dive-support@lbl.gov

### Login with ORCiD account



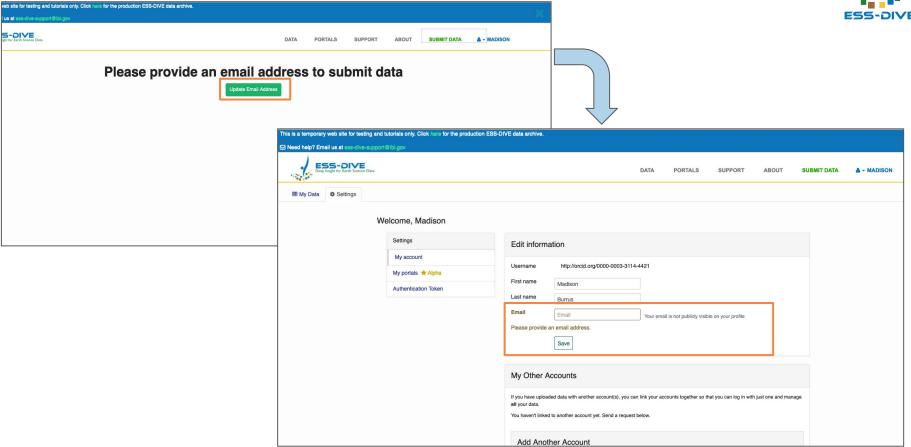






### Add email address to your profile





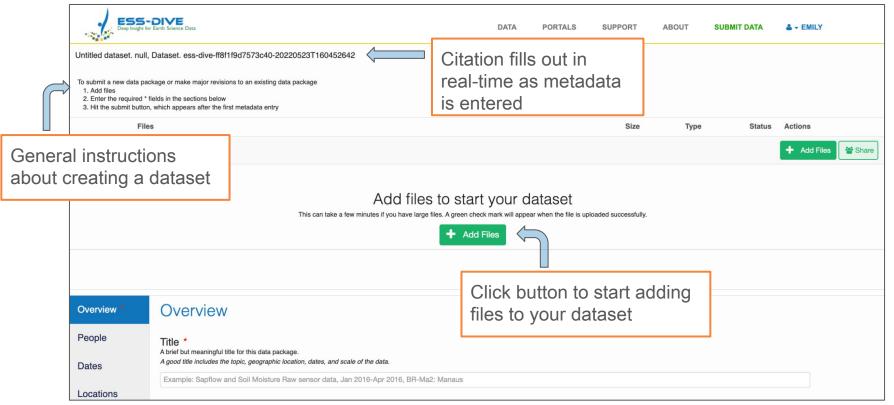
### Navigate to submission form



This is a temporary web site for te	esting and tutorials only. Click here for the production	n ESS-DIVE data archive.	
☑ Need help? Email us at ess-div	ve-support@ibl.gov		
ESS-DIV Deep Insight for Earth Scient	nce Data	DATA PORTALS SUPPORT ABOUT SUBMIT DATA A MADISON	
■ My Data 🌼 Settings			
	Welcome, Madison Settings	Edit information	
	My portals ★ Alpha  Authentication Token	Success! Your profile has been updated.	
		Username http://orcid.org/0000-0003-3114-4421  First name Madison  Last name Burrus  Email binary.burrus4@gmail.com Your email is not publicly visible on your profile	
		My Other Accounts  If you have uploaded data with another account(s), you can link your accounts together so that you can log in with just one and manage all your data.  You haven't linked to another account yet. Send a request below.  Add Another Account	

#### Dataset submission form





#### Dataset submission form



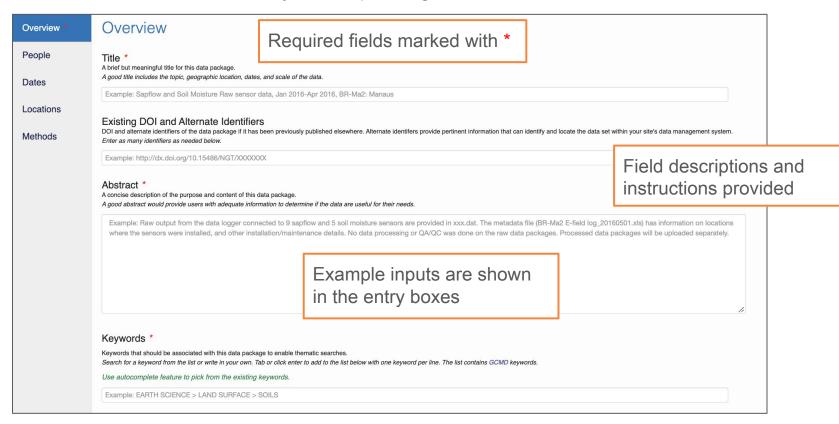
	Files	Size	Туре	Status Actions
· <b>&gt;</b>	Beginner's Tutorial Test Submission	2 KB		Remove this file
	BeginnerTutorialFile2.pdf	16 KB	Data	Remove
Ľ	BeginnerTutorialFile1.rtf	430 B	Data	Remove

Files can be removed by selecting the Remove option under Actions

#### Fill in metadata fields



#### Metadata describe the data that you are uploading



#### Fill in metadata fields



#### Keywords \*

Keywords that should be associated with this data package to enable thematic searches.

Search for a keyword from the list or write in your own. Tab or click enter to add to the list below with one keyword per line. The list contains GCMD keywords.

Use autocomplete feature to pick from the existing keywords.

Soil

EARTH SCIENCE > AGRICULTURE > SOILS: CATEGORICAL:GCMD EARTH SCIENCE > LAND SURFACE > SOILS: CATEGORICAL:GCMD

Some fields have autocomplete entry options

#### Project\*

Enter the name of the DOE project to associate with this data package. The project is cited as the publisher of this dataset.

If multiple projects were involved, enter the project that had the largest contribution to this data package.

Search by from the list or write in your own. Tab or click enter to add to the list below.

Use autocomplete feature to pick from the existing projects. If you can not find your project name, try searching by the PI name.

#### NGEE

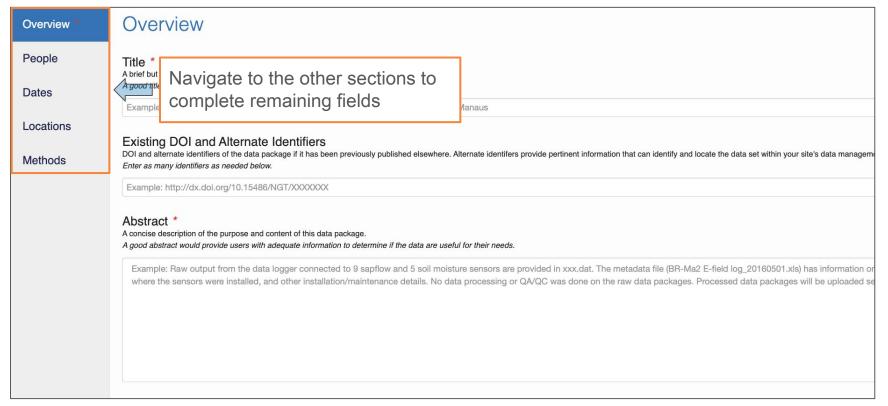
Next-Generation Ecosystem Experiments (NGEE) Tropics [PI: Jeffrey Chambers]

NASA Contribution to NGEE Tropics [PI: Douglas Morton]

Next-Generation Ecosystem Experiments (NGEE) Arctic [PI: Stan Wullschleger]

#### Fill in metadata fields





### Submitting your dataset



#### 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-Ma2: Manaus

**Dates** 

People

#### Existing DOI and Alternate Identifiers

Locations

DOI and alternate identifiers of the data package if it has been previously published elsewhere. Alternate identifiers provide pertinent information that can identify and locate the data set within your site's data management system.

Enter as many identifiers as needed below.

Methods

#### Example: http://dx.doi.org/10.15486/NGT/XXXXXXX

#### Abstract \*

A concise description of the purpose and content of this data package.

A good abstract would provide users with adequate information to determine if the data are useful for their needs.

Example: Raw output from the data logger connected file (BR-Ma2 E-field log\_20160501.xls) has information installation/maintenance details. No data processing ouploaded separately.

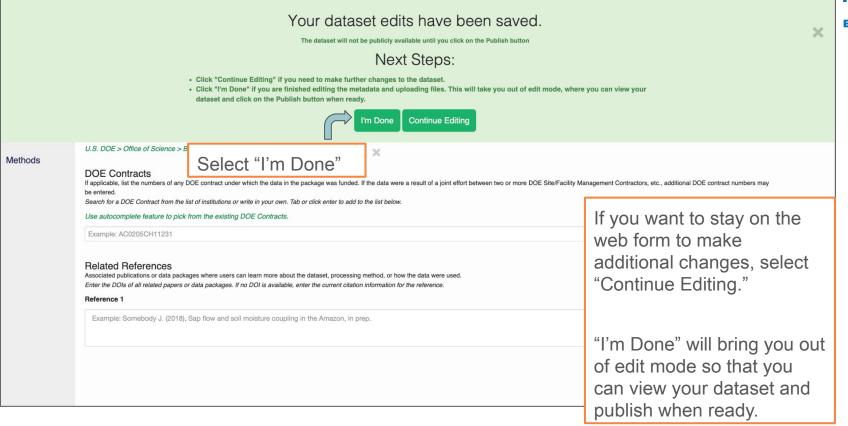
When you have finished completing at least the **required** metadata fields, you can select submit to save your dataset.

By submitting your dataset, you agree to the ESS-DIVE terms of use.

**Submit Dataset** 

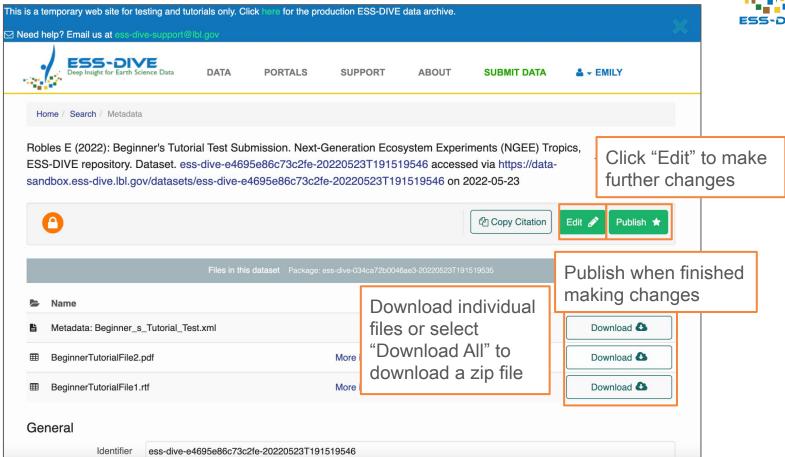
### Submitting your dataset





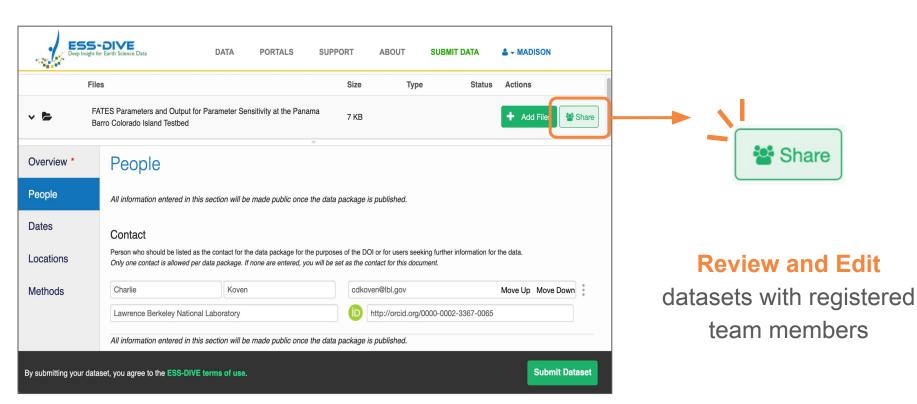
### Viewing your submitted dataset







### New feature: Sharing your dataset







#### **Viewer**

#### **Editor**

#### Manager

view private datasets

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

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



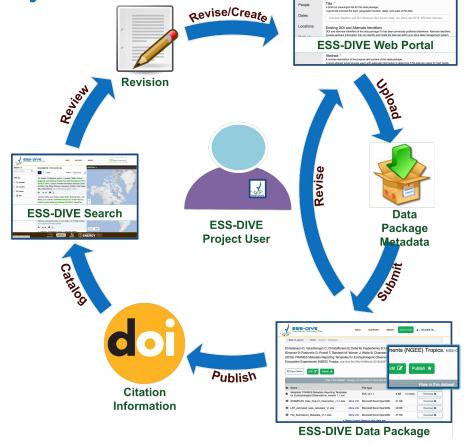
# Questions?



### DATA PUBLICATION PROCESS

### Publication cycle





Overview

### Viewing your assessment report



After running your metadata against our standard set of metadata, data, and congruency checks, we have found the following potential issues. Please assist us in improving the discoverability and reusability of your research data by addressing the issues below.

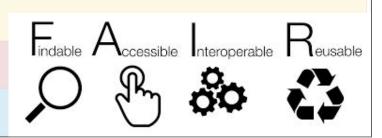


Accessible: 100% complete

Reusable: 100% complete

Interoperable: 100% complete

- Passed 19 checks out of 19 (informational checks not included).
- Warning for 0 checks.
- Failed 0 checks.
- 3 informational checks.



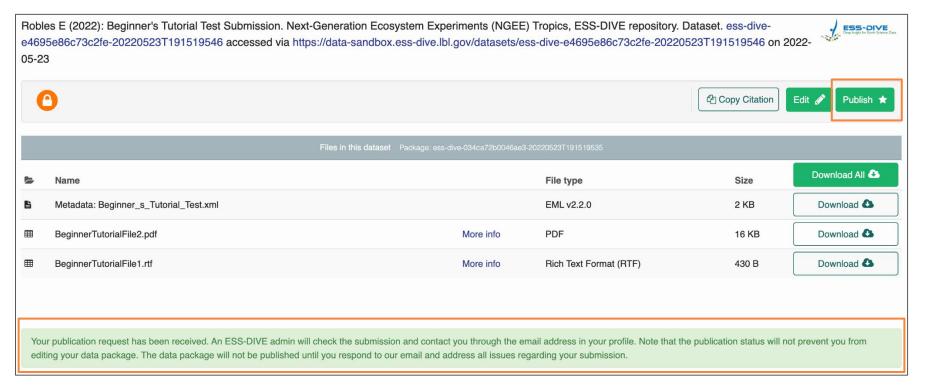
### Viewing your assessment report



<b>▼</b> F	Passed 19 checks out of 19 (informational checks not included).						
	The DOE project name was found in the controlled list of ESS projects.	0	Findable OPTIONAL SUCCESS				
	The one URL found in the metadata resolves correctly.	0	Findable REQUIRED SUCCESS				
	The funding organization "U.S. DOE > Office of Science > Biological and Environmental Research (BER)" was found in the metadata.	•	Findable OPTIONAL SUCCESS				
	The metadata identifier 'ess-dive-b5e2aede2d0f0ef-20220520T180415182358' was found and is resolvable using the DataONE resolve service, but is not publicly readable	0	Accessible OPTIONAL SUCCESS				
	No proprietary data formats found (out of 3 total formats).	0	Reusable OPTIONAL SUCCESS				
	3 creators are present.	0	Findable REQUIRED SUCCESS				
	One contact is present.	0	Findable REQUIRED SUCCESS				
	Found a contact user identifier that contains an ORCID: 'https://orcid.org/0000-0001-9262-7430'	0	Findable REQUIRED SUCCESS				



### Selecting "Publish"



# ESS-DIVE

### What happens after you click *Publish*?

- QUEUE Your publication request will enter the ESS-DIVE support queue.
- CONFIRM ESS-DIVE support team will email you to confirm that your request has been received.
- REVIEW ESS-DIVE support team will review your data package and email you with revision requests if necessary.
- PUBLISH Once the data package has been reviewed it will be assigned a DOI and shared publicly.





### What happens after you click *Publish*?

#### **CONFIRM**

You will receive an email from the ESS-DIVE support team confirming your request. Hello,

We have received your request to Publish "Air temperature and relative humidity raw data from June 2017-Oct 2018 at the B34 tower in Manaus, Brazil" and will notify you when we have fully reviewed your data set. Note that the publication status will not prevent you from editing your data package.

Your data package will be published with an existing DOI, if you supplied one. Otherwise, a new DOI will be issued through the ESS-DIVE system.

If you did not want to request that your data be published, please reply to this e-mail and we will stop the publication process.

Thanks,

The ESS-DIVE Support Team



### What happens after you click *Publish*?

#### **REVIEW**

After an initial review, a revision request may be sent over email.

Revision requests must be resolved before publication.

Hello,

We have finished reviewing your data package and request the following improvements to your metadata before publication:

1. Revise your abstract to include a clear and concise description of the contents of your data package. Describe the file types in your data package and any software needed to utilize them, if applicable. Also, include definitions of any acronyms used so your abstract can be understood by anyone and add a statement about the purpose for why these data were generated and the research question it is intended to answer.

Note that once the data package is published, it will be publicly available for search and download. The publication action cannot be undone, however, you will still be able to edit or retire the data package. As the dataset contact, you are responsible for obtaining consent for publication from the data package authors and notifying them about its status. Please respond to this message once you are finished making these revisions so that we can continue the approval process.

Thank you,

The ESS-DIVE Support Team



### What happens after you click *Publish*?

#### **PUBLISH**

You will receive an email from the ESS-DIVE support team confirming published dataset. Hello,

Your data package has been published. Go to https://data.ess-dive.lbl.gov/datasets/doi:10.15485/1868939 to view.

Here is the citation with the DOI:

Newcomer M; Rosemary C; Williams K (2022): Machine Learning Assisted Gap-Filled Discharge Data for the East River Community Watershed, Colorado, for Water Years 2014-2021. Watershed Function SFA, ESS-DIVE repository. Dataset. doi:10.15485/1868939 accessed via https://data.ess-dive.lbl.gov/datasets/doi:10.15485/1868939 on 2022-05-20

You may make minor edits and corrections to the data package metadata and files. Any edits you submit now will be immediately made public on ESS-DIVE. For major revisions to a data package, you must create a new data package by clicking the "Submit Data" button.

Contact the ESS-DIVE Support Team ess-dive-support@lbl.gov if you need further information.

### Summary of publication



#### **Request Publication**



01

Review your dataset's Automated Assessment Report.

Request publication.

Receive **confirmation email** from ESS-DIVE team.

#### **Review and Revise**



02

Address revision request email from ESS-DIVE

If no revisions are needed, your dataset will be published.

Revise based on recommendations and respond once finished.

#### **Dataset Published**



03

You will receive a final confirmation message with your dataset DOI and citation.

Your dataset is now publicly available on data.ess-dive.lbl.gov

Revise at any time.



# Questions?



# **HANDS-ON:**

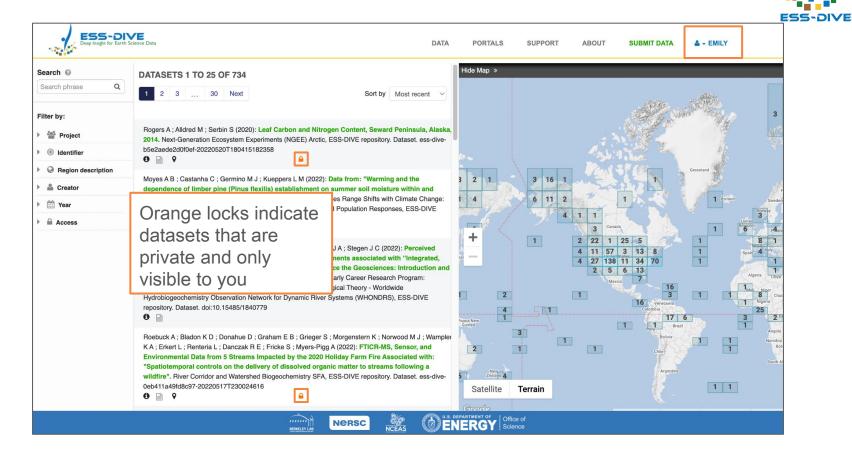
Search for and download ESS-DIVE datasets



# Switch to the main ESS-DIVE search page

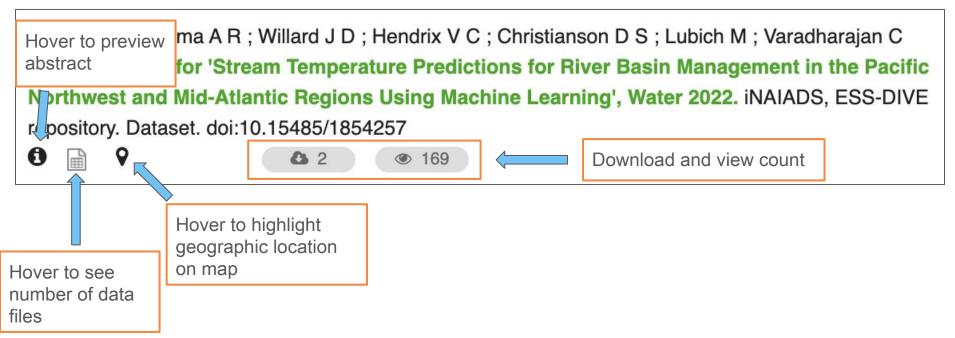
https://data.ess-dive.lbl.gov

# Main search page

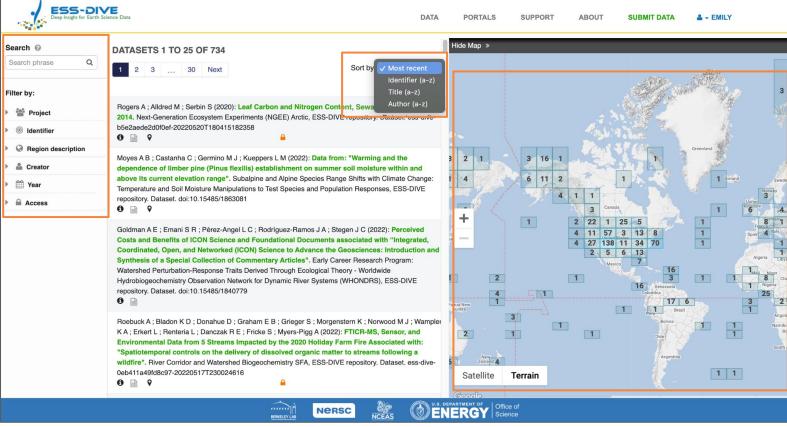


# Main search page

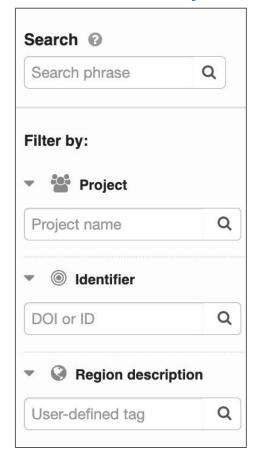






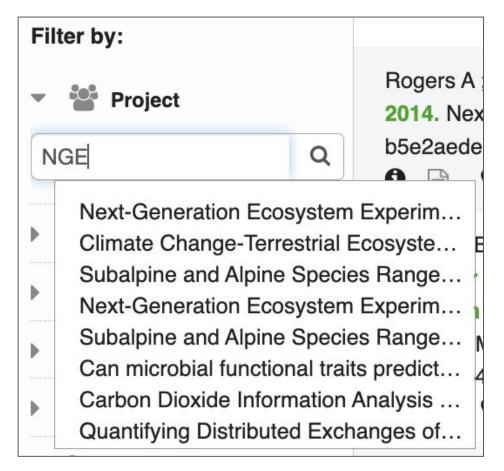




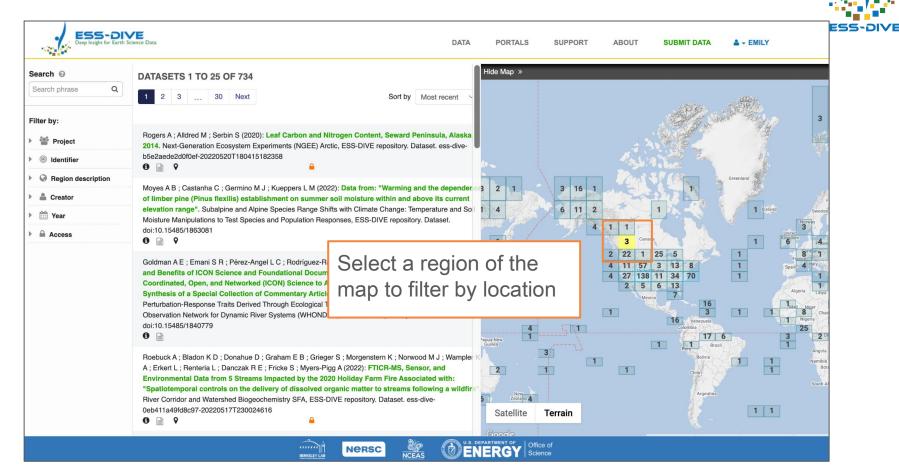




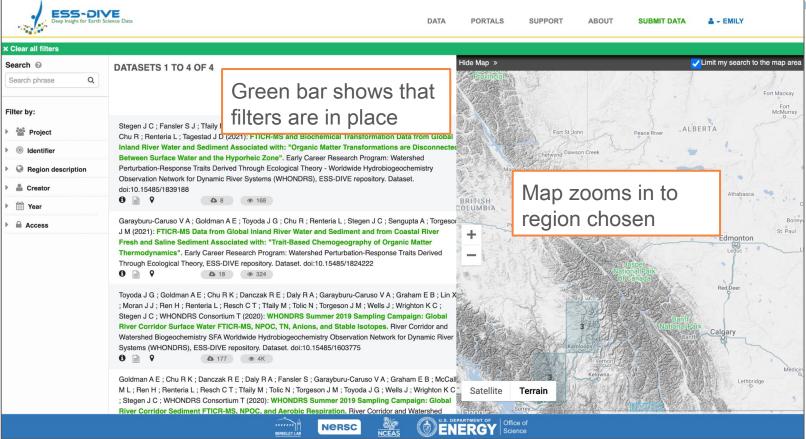




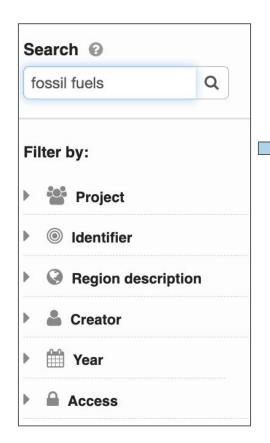
Project filter gives options from controlled list as you type

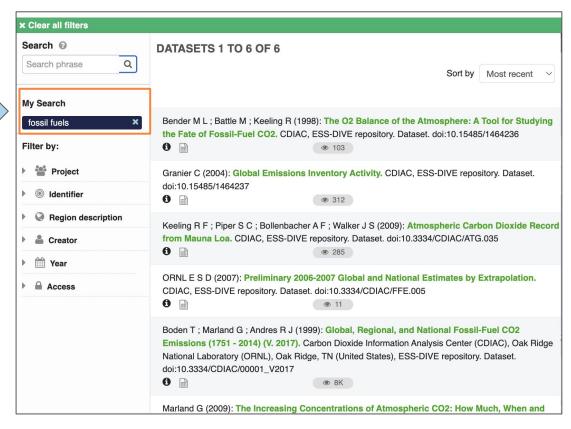




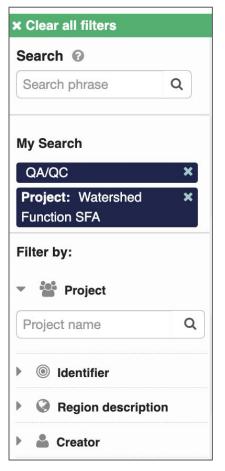


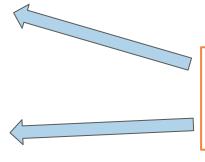








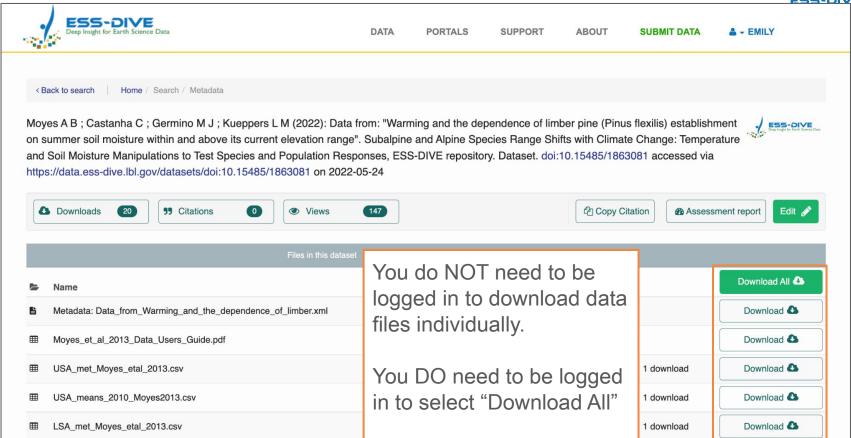




Clear filters individually or by selecting the green "Clear all filters" banner

# Downloading data files







# Summary

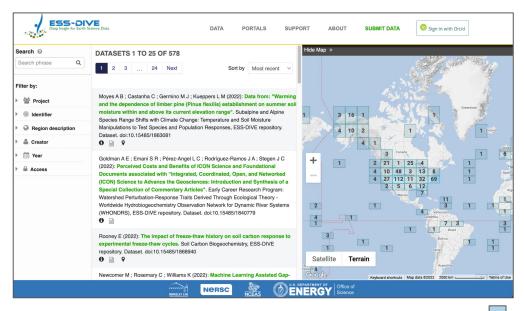
- Filter capabilities allow you to easily find datasets related to a specific project, time period, geographical location, and more
- Multiple filters can be used at the same time
- Abstract previews, data file counts, geographical location, and view and download statistics are available in the dataset list
- Data files can be downloaded from the dataset landing page you do NOT need to be logged in to download individual files



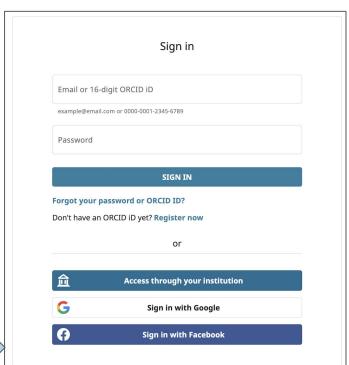
# REGISTER AS A DATA CONTRIBUTOR





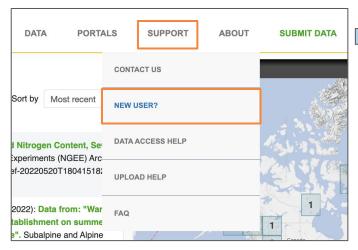


Sign in to the **main portal** using your ORCiD.

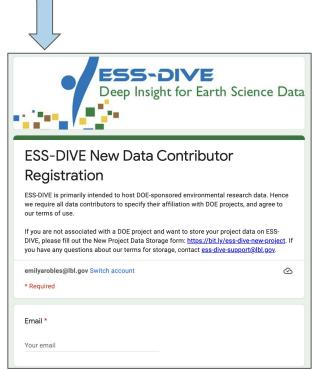


# Register as a data contributor





New User will bring you to the New Contributor
Registration documentation with the <u>registration form</u>.





Complete the form and wait for an email response with instructions to proceed.

# Tools and support for dataset creation





# Help Documentation

Refer to ESS-DIVE's website and Gitbooks for detailed information on dataset requirements.



# Metadata Guide and Offline Template

Review our complete **Metadata Guidelines**.

Collaborate on metadata with co-authors using an offline template before working on ESS-DIVE.



# Sandbox Testing Server

Practice uploading data packages to our sandbox test server, which does not permanently save data.



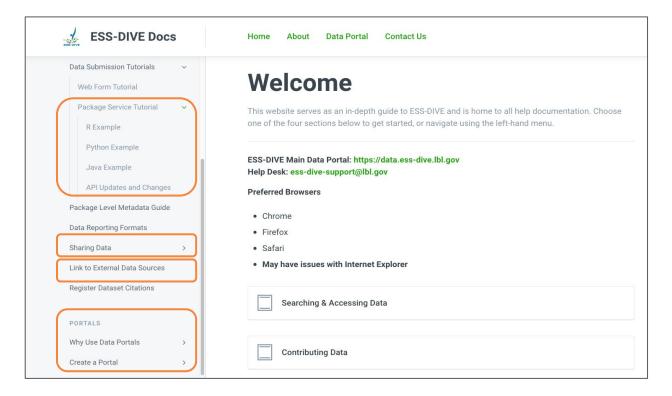
# Support Email Service

Feel free to contact the ESS-DIVE support team with questions!

ess-dive-support@lbl.gov ess-dive.lbl.gov/contact/



# Complete instructions are in our documentation



### **Sharing Data**

docs.ess-dive.lbl.gov/contributing-data/shar ing-data

### Package Service API

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

## **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



# Questions?

## **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

## **Acknowledgements**

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

Funding: EESSD Data Management





# **ESS-DIVE Glossary**

## List of common ESS-DIVE terms (only include relevant terms):

- DataONE The Data Observation Network for Earth (DataONE) is a
  distributed framework and sustainable cyberinfrastructure that provides open
  and secure access to Earth observational data. ESS-DIVE is a DataONE
  member.
- DOE The U.S. Department of Energy (DOE) is a Cabinet-level department of the United States whose mission is to ensure America's security and prosperity by addressing its energy, environmental and nuclear challenges through transformative science and technology solutions.





- DOI A Digital Object Identifier (DOI) is a unique alphanumeric string
  assigned by a registration agency (e.g., The Office of Scientific and Technical
  Information (OSTI)) to identify content and provide a persistent link to its
  location on the internet. ESS-DIVE assigns a DOI when your data package is
  published and made available electronically.
- ESGF The Earth System Grid Federation (ESGF) is a collaboration that develops, deploys and maintains software infrastructure for the management, dissemination, and analysis of model output and observational data. It is an interagency and international effort led by the U.S. Department of Energy with various co-funding agencies.





- ESS Environmental Systems Science (ESS) is a U.S. Department of Energy
  Office of Science program under the Biological and Environmental Research
  Program seeking to advance a robust predictive understanding of terrestrial
  surface and subsurface ecosystems.
- ESS-DIVE Environmental System Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE) is a U.S. Department of Energy archive for earth and environmental science data, models and software generated from research on terrestrial and subsurface environments.





- NCEAS The National Center for Ecological Analysis and Synthesis (NCEAS) at UC Santa Barbara partners with ESS-DIVE on data preservation. NCEAS is a DataONE member and a recognized expert in ecological data, digital libraries, and standards for data format.
- NERSC The National Energy Research Supercomputing Center (NERSC) is the primary scientific computing facility for the Office of Science in the Department of Energy, and is the primary data storage center for ESS-DIVE.
- ORCiD The Open Researcher and Contributor IDentifier provides anyone a
  persistent digital identifier (an ORCID iD) that distinguishes researchers from
  one another and provides a record that supports automatic links among all
  professional activities.





- OSTI The Office of Scientific and Technical Information (OSTI), a unit of the U.S. Department of Energy Office of Science, partners with ESS-DIVE on providing Digital Object Identifiers and metadata enhancement for data packages submitted to ESS-DIVE.
- Additional terms to consider defining: KBase, EMSL, JGI GOLD, NMDC,
  Dynamic data citation, Data package, Metadata, API, Text file, Comma
  separated file, NetCDF file, FAIR, ID, ISGN, Data Repository, Digital Library,
  Project Space, Data Visibility, Data Usability, Data Accessibility, JIRA, GitHub,
  Community Funds, Large/Big Data, File-level metadata, Sample-level
  metadata, 'Scalability, Resilience, and Integration', User & Preservation
  Services, Community Engagement and Partnerships