

Welcome to the 2021 ESS-DIVE Community Data Workshop

Agenda

All times in PDT

**Choose a breakout room*

Complete workshop
agenda and
information:

<https://ess-dive.lbl.gov/community-data-workshop/>

Mon, May 24, 2021

09:00 AM-
11:00 AM

Opening Session
Using ESS-DIVE for your
data management

Welcome
BER - ESS Data Mgmt
Review Process
ESS-DIVE Demo
ESS-DIVE Features
Data Mgmt Plans
Q&A
Community
Presentations

11:00 AM-
12:00 PM

Break and Data Help:
Office Hour

12:00 PM-
02:00 PM

Breakout Sessions*
Data Management and
Publication Challenges
and Needs

National Lab Projects
University Projects

Tue, May 25, 2021

09:00 AM-
09:30 AM

Ready to publish data?
Decide what to include

09:30 AM-
11:00 AM

ESS-DIVE Tutorials*

ESS-DIVE for Beginners
ESS-DIVE for Advanced

11:00 AM-
12:00 PM

Break and Data Help:
Office Hour

12:00 PM-
01:00 PM

ESS Data Types*

Large & model data
Sensor data & QA/QC
Sample tracking

01:00 PM-
02:00 PM

Reporting Formats
ESS-DIVE tutorial
File-level and csv

Breakouts* Model data
Sample IDs & metadata
Water and soil quality
Leaf-gas exchange
Continuous soil
respiration
16s amplicon data
Hydrologic monitoring



ESS-DIVE

Publishing Data Packages

Which data to publish, and which repository to publish in

Zarine Kakalia
ESS-DIVE Data Management Assistant



Community Data Workshop 2021

Polls!

Takeaways from yesterday

How did you feel at the end of Day 1?



What topic from Day 1 is still unclear to you and you are hoping to understand today?

Top



What does a data package consist of?



_____ describes and gives information about data.



Presentation Overview

- Contents of a **data package**
- Differences between various **environmental data repositories**
- **Steps to create** a data package

PUBLISHING DATA PACKAGES:

What and where to publish data

Reasons to publish data

Abide by journal and funding requirements

Most journals are starting to **require** data associated with paper findings, figures, and tables to be **publicly available on a long-term data repository**

Include **DOI's**, such as those issued by ESS-DIVE, in the **Data Availability** section of a paper



DOI: Digital Object Identifier



Share your work with the community

Gain **publicity** from data publications, similarly to journal publications

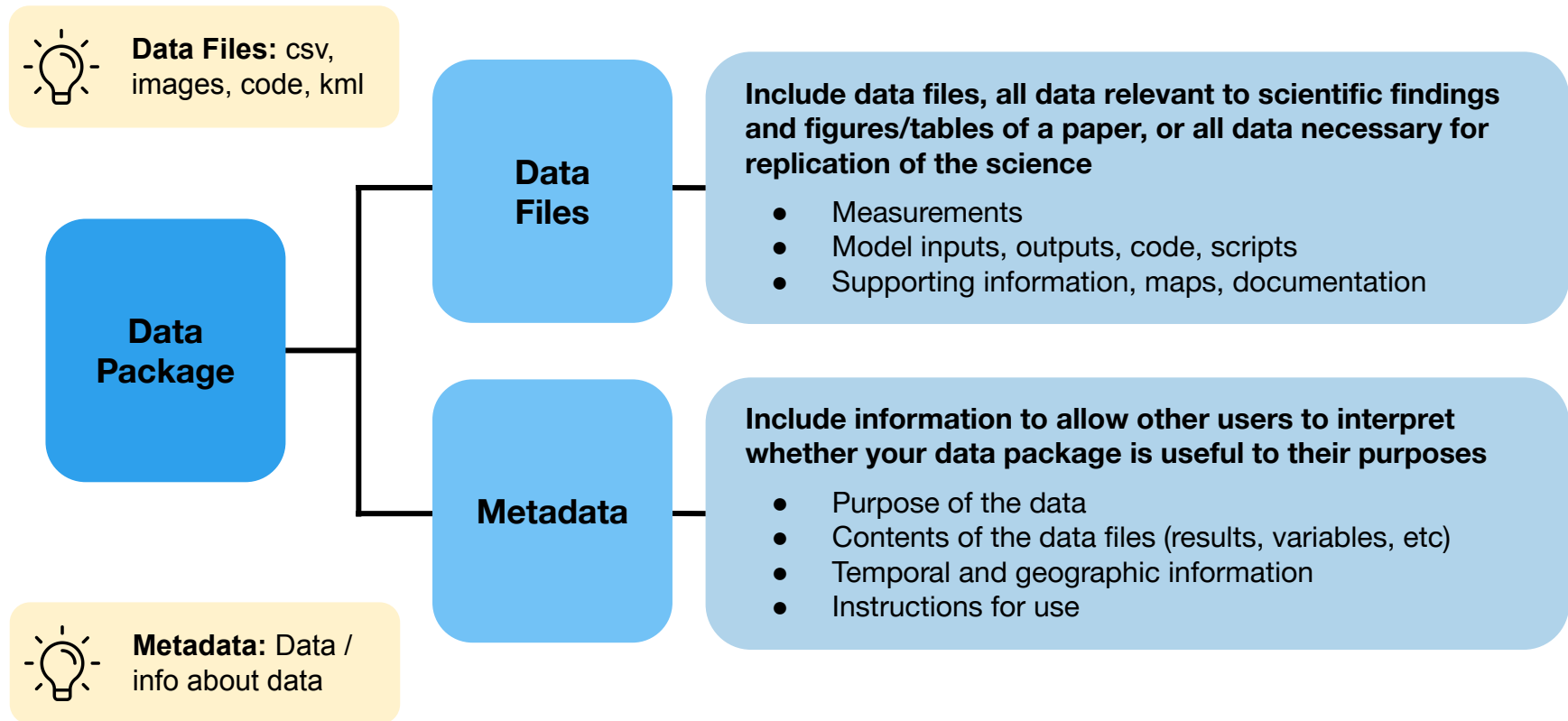
Allow others to use your work for **future studies**

Promote FAIR data practices

Findable, Accessible, Interoperable, and Reusable

Data reporting formats and metadata requirements abide by these standards

Components of a data package



Data File Types

01

ReadMe

Directory of files included with additional metadata

02

Tabular Data

Sample, analysis, observational data

03

File Level Metadata

Description of individual files within the package

04

Model Components

Model inputs, outputs, code, scripts

05

Maps

KML or KMZ files with geographic data

06

PDFs, Supporting Info

Instrument manuals, methods writeups, etc

Metadata Components

01

Title

Descriptive overview of the data package

02

Abstract

Purpose, contents, location, instructions

03

Keywords

Variables, keywords not already included in title

04

Location

Latitude and longitude, location description

05

Methods

Data collection, processing, QA/QC, error

06

Authors

In the order of contributions

Environmental data repositories



Repository: a
place to store data



Environmental Systems Science Data (ESS-DIVE)

ESS funded projects are
required to store data on
ESS-DIVE

Accepts observational,
experimental, sampling,
and modeling data



Environmental Data Initiative (EDI)



NSF Arctic Data Center (ADC)



National Center for Biotechnology Information (NCBI)



United States Geological Survey (USGS) ScienceBase



Earth System Grid Federation (ESGF)



National Microbiome Data Collaborative (NMDC)

Considerations to split up data packages

Author contributions

Based level of contributor effort for portions of data - affects author order

Data in a publication

All data (raw or processed) that went into a publication

Campaign / Time Period

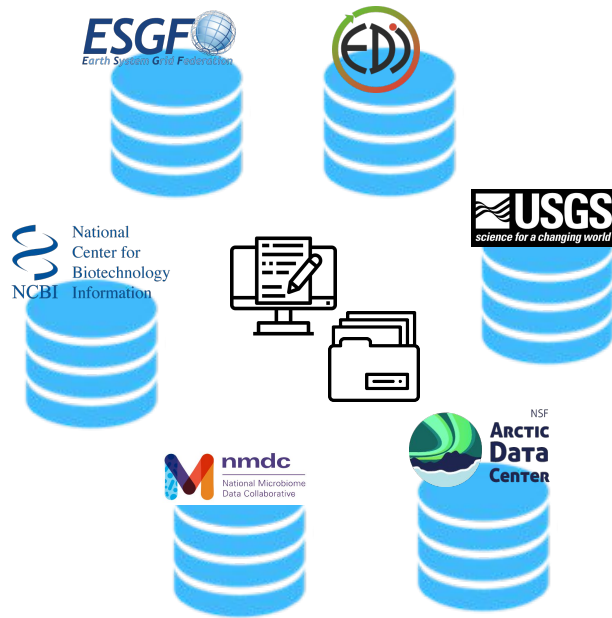
Data from a field campaign or season that need to be viewed together

Data type

Particular data type from a project - e.g. continuously generated sensor data, sample data, data synthesis product



Special Case: Mirroring data packages on multiple repositories



Copy the metadata into a new data package in ESS-DIVE.



Data upload methods can support large data



"Download All" uncompressed files	No. of uncompressed files	Volume per file	Upload Method
< 3 GB total	< 100	< 10 GB	Web Upload Form
< 3 GB total	< 100	10-100 GB	Package Service API
> 3 GB total	> 100	> 100 GB	Globus: data transfer service

PUBLISHING DATA PACKAGES:

Getting started

Steps to publish your data

Collect your data files



01

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

Refer to **ESS-DIVE reporting formats**

Choose a repository



02

Register to submit data to a data repository

Ensure the repository **fits the requirements** of your data and project

Publish with metadata



03

Publish data and metadata as a data package

Meet the funding/journal publishing time frame

BER: publish within 1 year of end of data collection, or at the time of publication

Data package creation tools ESS-DIVE offers



Help Documentation

Refer to **ESS-DIVE's website and Gitbooks** for detailed information on data package requirements

ess-dive.lbl.gov/
docs.ess-dive.lbl.gov/



Offline Metadata Guide

Collaborate on metadata with co-authors before working on ESS-DIVE

docs.ess-dive.lbl.gov/



Sandbox Testing Server

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

[data-sandbox.ess-dive.l
bl.gov/](https://data-sandbox.ess-dive.lbl.gov/)



Support Email Service

Feel free to **contact the ESS-DIVE support team** through email for any questions

[ess-dive.lbl.gov/contact/
ess-dive-support@lbl.gov](https://ess-dive.lbl.gov/contact/ess-dive-support@lbl.gov)

Multiple ways to create and edit data packages

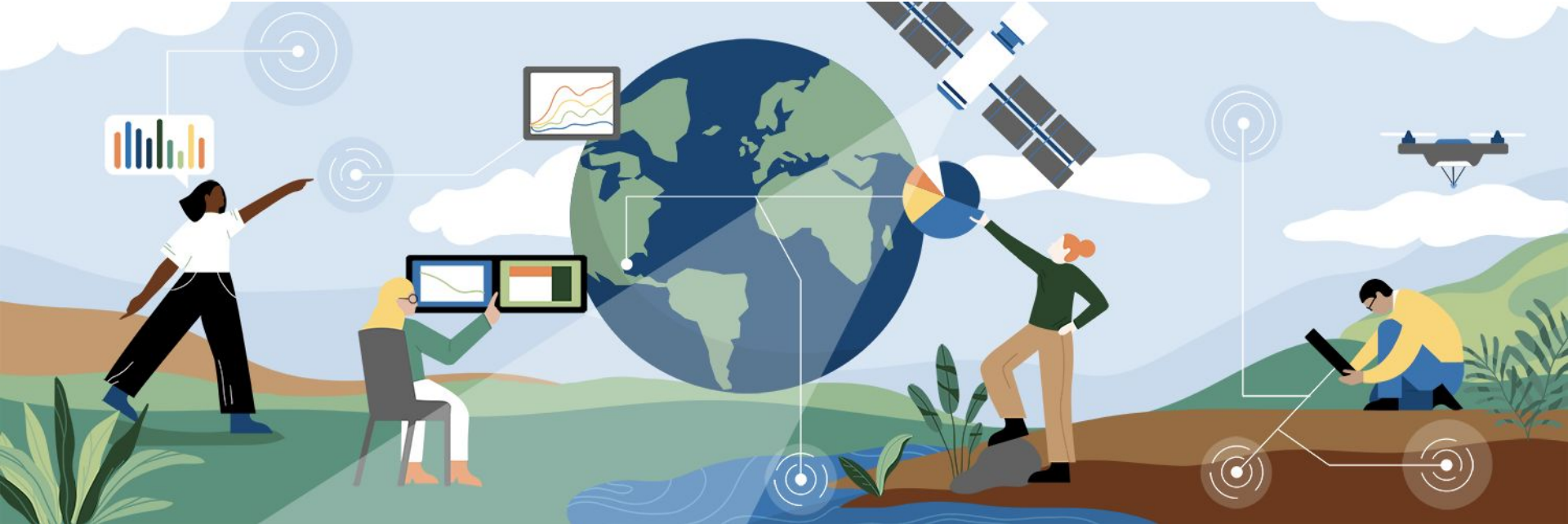


- **Web Upload Form**

- **Manually** enter data package metadata on the ESS-DIVE user interface (UI), one data package at a time. Drag and drop or select files from your file manager to upload data.
- Interacting with the web form when editing a data package on ESS-DIVE

- **Package Service API**

- A **programmatic** method for creating AND editing data package metadata on ESS-DIVE. Upload files and create metadata using JSON-LD.
- **Future Feature**: use the API to programmatically download any public data package!



Questions?

ESS-DIVE Tutorials - Join a Breakout!



Beginner Users

Join Breakout or

Add “1” before your name to join Beginner Users

“1 - Jane Smith”

1

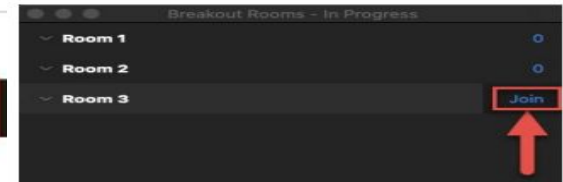
Advanced Users

Join Breakout or

Add “2” before your name to join Advanced Users

“2 - Jane Smith”

2



Data Help Office Hours - Join a Breakout!

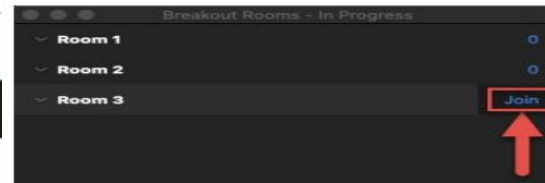


**Data Upload &
General Questions**

**Project Spaces &
Portals**

**API & Large Data
Uploads**

Reporting Formats



ESS Data Types - Join a Breakout!



Large data &
model data
archiving

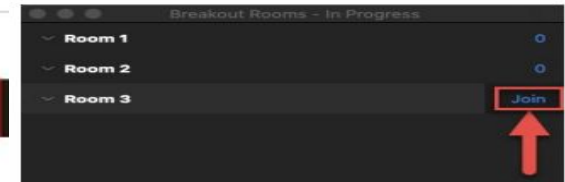
1

Sensor data
and QA/QC

2

Sample
workflows,
linking related
sample data

3



Reporting Formats - Join a Breakout!



Model data
(*Madison Burrus*)

1

Sample identifiers
and metadata
(*Joan Damerow*)

2

Water quality/soil
samples
(*Kristin Boye*)

3

Leaf-gas exchange
(*Kim Ely*)

4

Continuous soil
respiration
(*Ben Bond-Lamberty*)

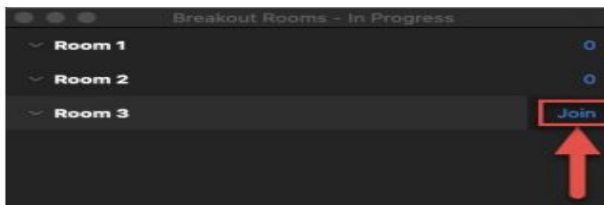
5

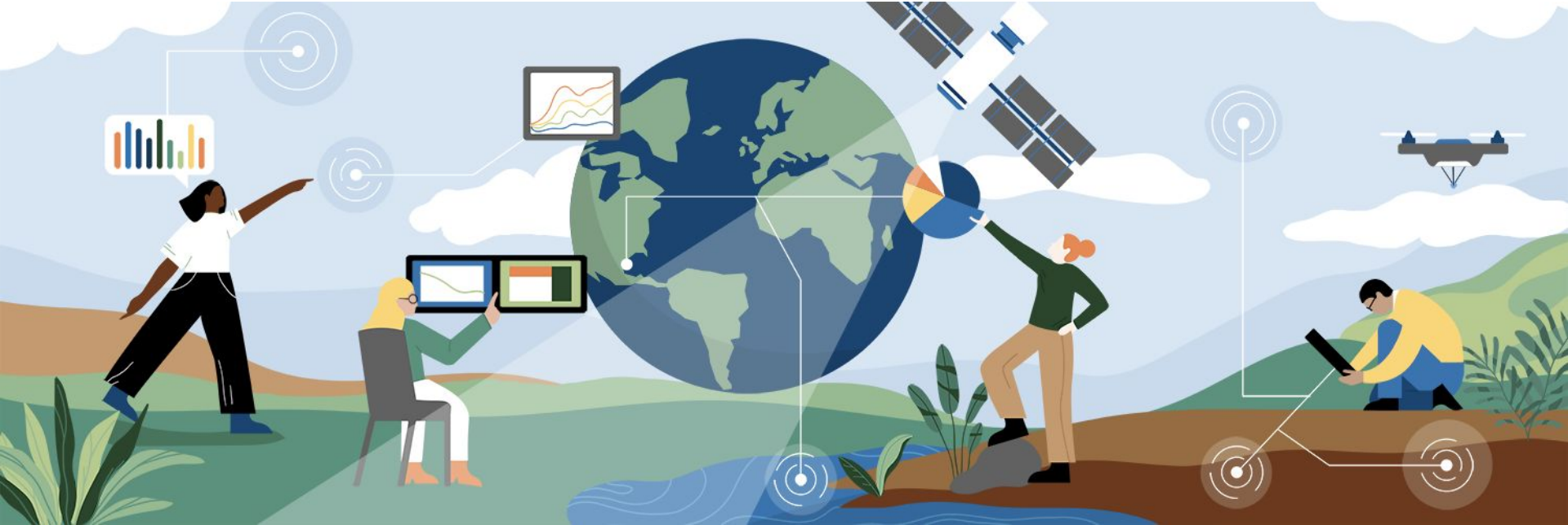
16s amplicon
sequencing
(*Pamela Weisenhorn*)

6

Hydrologic
monitoring
(*Amy Goldman*)

7





Thank you for attending!

Event Feedback:

<http://bit.ly/CommunityDataWorkshopSurvey>

ESS-DIVE RESOURCES:

Glossary

ESS-DIVE Glossary

List of common ESS-DIVE 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.

ESS-DIVE Glossary (*cont.*)

- **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-DIVE Glossary (*cont.*)

- **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.

ESS-DIVE Glossary (*cont.*)



- **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.

ESS-DIVE Glossary (*cont.*)



- **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.