

ESS-DIVE Overview

2024 ESS CI Meeting

Shreyas Cholia & Charu Varadharajan



April 15, 2024

Presentation Overview



- ESS-DIVE feature updates
- What's next: Upcoming plans
- Community funds
- Discussion about ESS-DIVE's future plans

Key Takeaways: Learn about new ESS-DIVE features and discussion on our future plans/community funds

What is ESS-DIVE?



Data Repository to

preserve

expand **access** to

and improve **usability** of

ESS data

data.ess-dive.lbl.gov

F indable A ccessible I nteroperable R eusable



A screenshot of the ESS-DIVE website interface. The top navigation bar includes links for DATA, PORTALS, PROJECTS, GET STARTED, ABOUT, and SUBMIT DATA, along with a "Sign in with Orcid" button. The main content area displays a list of datasets, with the first entry being "Feng Y ; Negron-Juarez R ; Romps D ; Chambers J ; ...". A large blue starburst graphic is overlaid on the page, containing the text "900+ public datasets". On the right side, there is a map of the world with a grid overlay, showing data points across various regions. The map has a "Hide Map" button and a legend for "Satellite" and "Terrain".

Use of ESS-DIVE is **mandatory** for ESS project data management plans

ESS-DIVE At-a-glance



Authorized users	Public datasets	Private datasets	Total file size	Number of Files	Contributing projects
383 (498%)	909 (144%)	266 (1673%)	2.82TB	16,762	90 (233%)
Public dataset updates	Temporal coverage	Public Portals	File downloads	Data views	Support Requests
102 (1600%)	1841 to 2024	15 (650%)	429.1K+ (504%)	753.8K+ (510%)	1333 (256%)

Growth statistics measured over the period from May 2020 - March 2024

ESS-DIVE Focus FY20-24



Community Adoption

Goal: Expand the range of ESS projects engaged and archiving with ESS-DIVE



Data/Metadata Standards

Goal: Make data widely findable and usable



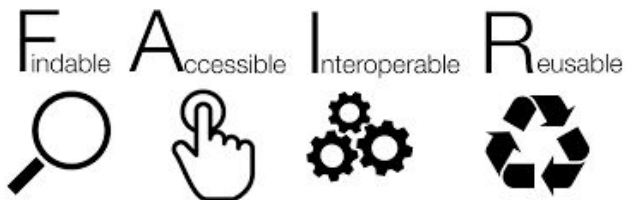
Scaling

Goal: Grow the repository capacity and resilience as demand grows



Automation

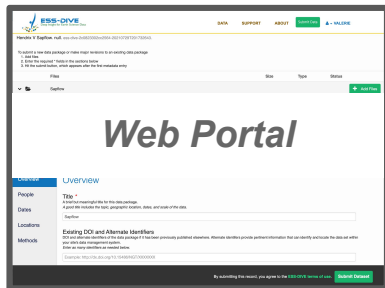
Goal: Apply automation techniques to ESS-DIVE processes to minimize human effort (human-in-the-loop).



ESS-DIVE Data Lifecycle



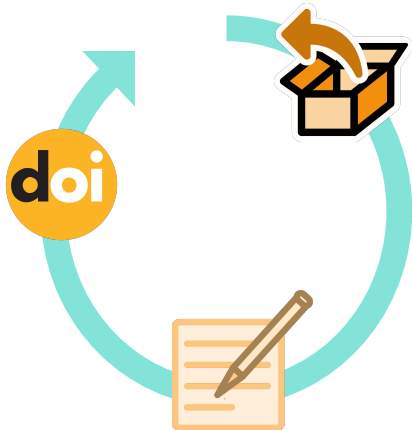
Upload



Programmatic



Publish



ESS-DIVE's *publication life cycle* allows users to iteratively curate, package, and update their data and metadata.

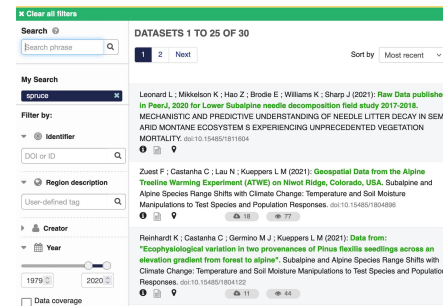
Preserve

Data Preservation Principles

- DOIs
- Quality metadata
- Change management rigor
- Redundancy
- Data auditing and reporting



Discover



OSTI.GOV



Dataset Search

Search for Datasets

Dataset Submission and Automated Review




NEW: Updated checks & reviews for reporting formats

Web UI

Overview	Overview
People	Title * A brief but meaningful title for this data package. <i>A good title includes the topic, geographic location, dates, and scale of the data.</i> <input type="text"/> Example: Sapflow and Soil Moisture Raw sensor data, Jan 2016-Apr 2016, BR-Ma2: Manaus
Dates	
Locations	
Methods	Existing DOI and Alternate Identifiers DOI and alternate identifiers of the data package if it has been previously published elsewhere. Alternate identifiers enter as many identifiers as needed below. <input type="text"/> Example: http://dx.doi.org/10.15486/NGT/XXXXXXX
	Abstract * A concise description of the purpose and content of this data package. <i>A good abstract would provide users with adequate information to determine if the data are useful for their needs.</i> <input type="text"/> Example: Raw output from the data logger connected to 9 sapflow and 5 soil moisture sensors where the sensors were installed, and other installation/maintenance details. No data processing.
Overview	Creators
People	The main researchers involved in producing the data. These include authors, owners, originators and principal investigators. <i>One or more creators is required. If none are entered, you will be set as the creator of this document. List creators in order of contribution.</i>
Dates	Creator #1 First name: <input type="text"/> Last name: <input type="text"/>
Locations	Organization name: <input type="text"/>
Methods	<i>All information entered in this section will be made public once the data package is published.</i>

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



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

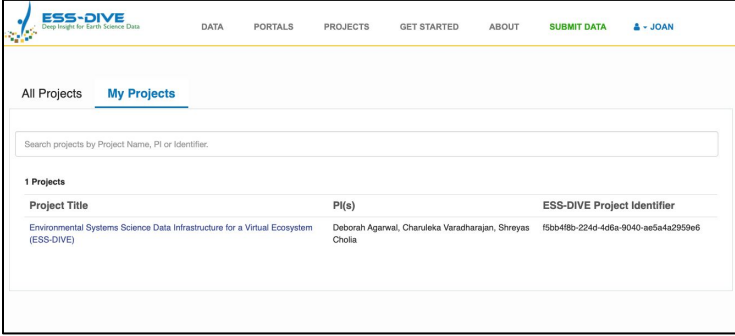
Interpretation: 100% complete

19 checks

- 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.

Project Data Management Tools

- **Project teams** for internal sharing, collaboration and curation
- **Bulk data uploads, updates** using API
- Project-centric data **search** and **portals**
- **COMING SOON!!** Dataset tracking and reserving DOIs
- **NEED INPUT!!** Project data **citations** and **metrics**



Project Title	PI(s)	ESS-DIVE Project Identifier
Environmental Systems Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE)	Deborah Agarwal, Charulika Varadarajan, Shreyes Chola	f5bb4fb-224d-4d6a-904a-ae5a4a2959e6



Lunchtime Breakout: ESS-DIVE Tutorial
for PIs and Data Managers
Wed., April 17th; 12:30-1:30pm EST

PIs will walk
away with a
To-Do List



Project Teams

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 or group

8 projects have used this feature since 2023 launch



ess-dive-data-managers (10 members)


- ★ Madison Burrus (Me)
http://orcid.org/0000-0003-2296-4698
- ★ Joan Damerow
http://orcid.org/0000-0003-2601-5043
- ★ Emily Robles
http://orcid.org/0000-0003-3720-6566
- ★ Hesham Elbashandy
http://orcid.org/0000-0002-7535-8045

First 1 2 3 Last

Add Member - Search by username, email, or name OR enter a full username below.

- Dylan Projansky
- Dylan Levene
- Dylan O'Ryan
- Dylan P. O'Ryan

2. Add people to team



[DATA](#) [PORTALS](#) [PROJECTS](#) [GET STARTED](#) [ABOUT](#) [SUBMIT DATA](#) ▲ - DYLAN P.

ess-dive-data-managers

Group

Group Id: ess-dive-data-managers

●
70 contributions

↓
downloads

3 members

- ▲
Madison R. Burrus
http://orcid.org/0000-0002-0439-9173
- ▲
Dylan P. O'Ryan (Me)
http://orcid.org/0009-0006-8124-6583
- ★
Emily Ann Robles
http://orcid.org/0009-0002-3285-2198

DATASETS 1 TO 5 OF 7

1
2
Next

Sort by Most recent ▼

Damerow J.; Varadharajan C.; Boyle K.; Brodie E.; Burrus M.; Chadwick D.; Cholla S.; Crystal-Ornelas R.; Elbashandy H.; Eloy Alves R.; Ely K.; Goldman A.; Hendrix V.; Jones C.; Jones M.; Kakalia Z.; Kemner K.; Kersting A.; Maher K.; Marino N.; O'Brien F.; Parzan Z.; Robles E.; Snavely C.; Sorensen P.; Stegen J.; Weisenhorn P.; Whitenack K.; Zavarin M.; Agarwal D. (2020): [Sample Identifiers and Metadata Reporting Format for Environmental Systems Science](#). Environmental Systems Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE). ESS-DIVE repository. Dataset. doi:10.15485/1660470

Weisenhorn P.; Bellsmith K. (2022): [ESS-DIVE Reporting Format for Amplicon Abundance Table](#). Environmental Systems Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE). ESS-DIVE repository. Dataset. doi:10.15485/1865729

Crystal-Ornelas R.; Varadharajan C.; Bellsmith K.; Bond-Lamberty B.; Boyle K.; Burrus M.; Cholla S.; Christianson D S.; Crow M.; Damerow J.; Ely K S.; Goldman A E.; Heinz S.; Hendrix V C.; Kakalia Z.; Mathes K.; O'Brien F.; O'Ryan D.; Pennington S C.; Robles E.; Rogers A.; Simmonds M.; Veliquette T.; Weisenhorn P.; Welch J N.; Whitenack K.; Agarwal D. (2022): [Data from: "Enabling FAIR data in Earth and environmental science with community-centric \(metadata reporting formats\)"](#). Environmental Systems Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE). ESS-DIVE repository. Dataset. doi:10.15485/1866606

Crystal-Ornelas R.; O'Ryan D.; Christianson D.; Hendrix V C.; Agarwal D.; Varadharajan C. (2022): [ESS-DIVE Reporting Format for Location Metadata](#). Environmental Systems Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE). ESS-DIVE repository. Dataset. doi:10.15485/1865730

Boyle K.; Crystal-Ornelas R.; ...

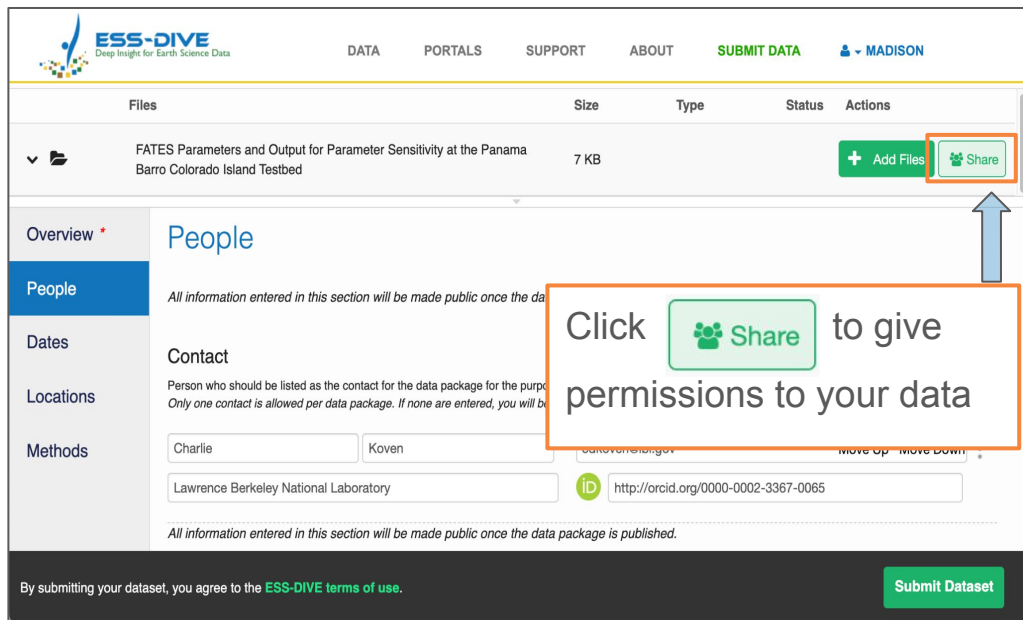
1. Share datasets with a team

New team members get immediate access to all project datasets!

9

Share Datasets

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



The screenshot shows the ESS-DIVE web interface. At the top, there is a navigation bar with links for DATA, PORTALS, SUPPORT, ABOUT, and SUBMIT DATA, along with a user profile for MADISON. Below this is a table of files. One file is listed: "FATES Parameters and Output for Parameter Sensitivity at the Panama Barro Colorado Island Testbed" with a size of 7 KB. The "Actions" column for this file contains two buttons: "Add Files" and "Share". The "Share" button is highlighted with an orange box. A blue arrow points from the "Share" button to a callout box that says "Click Share to give permissions to your data". Below the file list, there is a "People" section with a "Contact" form. The form includes fields for name (Charlie, Koven), organization (Lawrence Berkeley National Laboratory), and ORCID ID (http://orcid.org/0000-0002-3367-0065). At the bottom of the page, there is a "Submit Dataset" button.



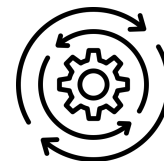
Submit Dataset

Data Contributor



Edit

Collaborator



PI / Project Manager

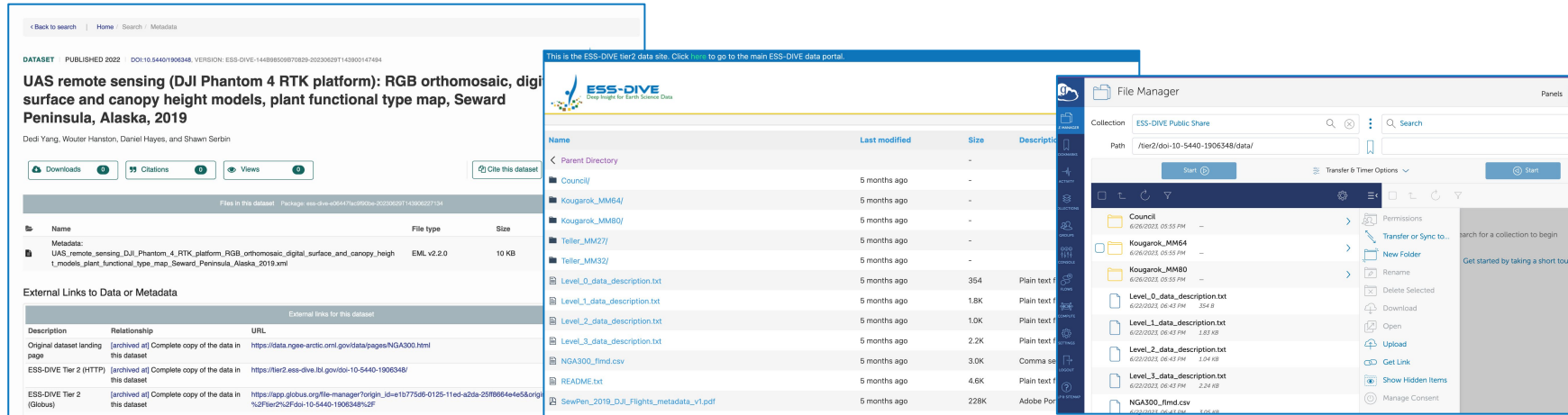
Large Data on Tier 2

Secondary, direct access storage with metadata on ESS-DIVE

- Large Data: 500GB - multiple TB range
- Browsible Folder Hierarchy
- Globus and HTTP Access
- Email ess-dive-support@lbl.gov if you need to store large data

7.3TB data published on Tier 2

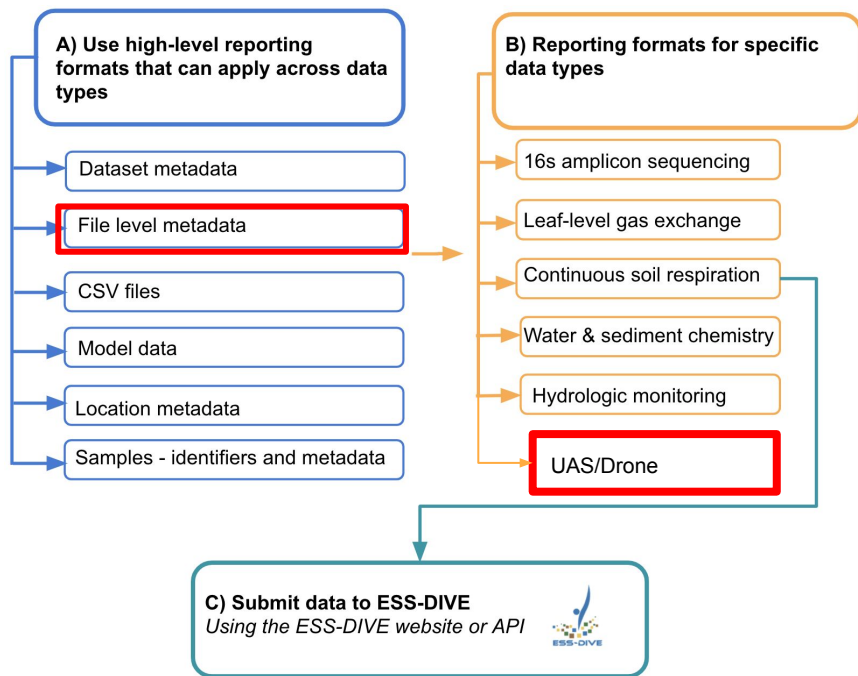
10 Tier 2 datasets in past year



The image shows a composite of three screenshots from the ESS-DIVE website and interface:

- Top Left:** A dataset page for "UAS remote sensing (DJI Phantom 4 RTK platform): RGB orthomosaic, digital surface and canopy height models, plant functional type map, Seward Peninsula, Alaska, 2019". It includes metadata, download/citation buttons, and a table of files in the dataset.
- Top Right:** A navigation banner for the ESS-DIVE Tier 2 data site.
- Bottom:** A File Manager interface showing a folder hierarchy for "ESS-DIVE Public Share" with subfolders like "Council", "Kougarok_MM64", "Kougarok_MM80", and "Teller_MM27". It lists files such as "Level_0_data_description.txt" and "Level_1_data_description.txt" with their sizes and modification dates.

ESS-DIVE Formats and Guidelines for Submitting Data Developed in Partnership with the Community



ESS-DIVE Community Space on GitHub

<https://github.com/ess-dive-community>



Hands-on Session: ESS-DIVE Data Reporting Formats
Mon., April 15th; 1:45-3:15pm EST

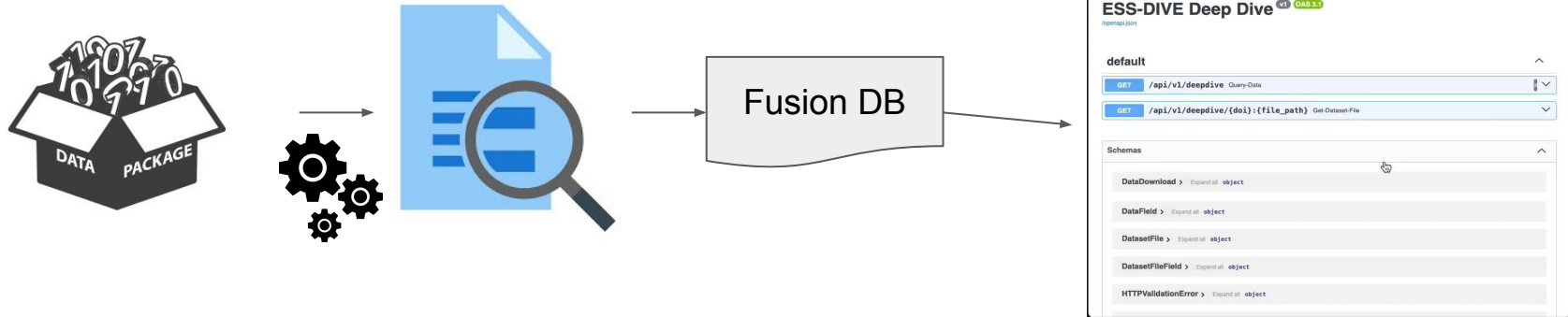
Figure source: Crystal-Ornelas, R. et al. 2022. Enabling FAIR data in Earth and environmental science with community-centric (meta)data reporting formats. *Nature Scientific Data*. <https://doi.org/10.1038/s41597-022-01606-w>

Data Standards Enabling Deep Search & Integration





NEW: Fusion DB and Deep Dive API

- **Deep Dive API:** <https://fusion.ess-dive.lbl.gov/>
- FusionDB looks inside data files: Indexes & validates 100+ datasets with ESS-DIVE file-level metadata
- 37% of newly published datasets (since July 23) are **searchable via Deep Dive API** with 11,563 unique data fields.
- **Automated Parsing** and **Validation** checking for FLMD reporting format



BASIN-3D: A Framework for Powering On-Demand Data Integration



v1 available:  <https://pypi.org/project/basin3d/> and  <https://github.com/BASIN-3D>

BASIN-3D provides a common query mechanism to access multiple disparate data sources. Results are transformed into a common data model.



v1 Capabilities

python package or django web framework

Synthesizes time series data from:

- **USGS Daily / Instantaneous Values**
- **EPA Water Quality eXchange (WQX)**
- **ESS-DIVE Hydrologic Monitoring RF**

Output formats: pandas dataframe, hdf5

Harmonizes variable terms into common format

Documentation with examples:

<https://basin3d.readthedocs.io>

Bringing Compute and Data Together



- Dedicated **storage and compute platform** for synthesis and analysis. (Ceph FS Storage 0.5 PB, Compute Nodes on LBL IT Lawrence Livermore)
- **NEW!!** Early Users: Community **Jupyterhub** platform for users to directly operate on ESS-DIVE data.

A screenshot of a Jupyter Notebook interface. The left pane shows a file explorer with a directory structure including folders like 'jpg_files', 'kml_files', 'lpd_files', 'rinex_files', and 'rtcm3_files', and files like 'NGA281_2018_TL4Z_UAS_L0_User_Guide.pdf' and 'nga281_fmnd.csv'. The right pane shows a code editor with JSON-LD code for creating a dataset. Below the code editor, a table is displayed with columns: File_Name, File_Description, Standard, UTC_Offset, File_Version, Contact, and Date_Start. The table contains 7 rows of data.

	File_Name	File_Description	Standard	UTC_Offset	File_Version	Contact	Date_Start
1	UAS_L0_User_Guide.pdf	rs. and flight parameters.	-9999	-9999	-9999	Adam Collins	-9999
2	jpg_files Folder	Iders by data of collection	-9999	-9999	-9999	Adam Collins	7/11/2018
3	.kml_files Folder	iring 2018 UAS campaign	-9999	-9999	-9999	Adam Collins	-9999
4	.lpd_files Folder	uide for additional details.	-9999	-9999	-9999	Adam Collins	7/11/2018
5	.rinex_files Folder	ser guide for more details.	-9999	-9999	-9999	Adam Collins	7/11/2018
6	.rtcm3_files Folder	ser guide for more details.	-9999	-9999	-9999	Adam Collins	7/11/2018
7	nga281_fmnd.csv	> files - file-level metadata	-9999	-9999	-9999	Adam Collins	-9999




Bringing Compute and Data Together



- Dedicated **storage and compute platform** for synthesis and analysis. (Ceph FS Storage 0.5 PB, Compute Nodes on LBL IT Lawrencium)
- **NEW!!** *Early Users*: Community **Jupyterhub** platform for users to directly operate on ESS-DIVE data.

Hosted by **ESS-DIVE**

Welcome to the ESS-DIVE Jupyter service
Please, login with your ORCID for access to your Jupyter notebook workspace on ESS-DIVE.

 Sign in with ORCID

If your login fails, contact ess-dive-support@lbl.gov to be put on the allow list for the ESS-DIVE Jupyter service.

Linking Related Sample Data Using Identifiers



[ESS-DIVE Sample Data Portal](#) with IGSNs: **26 datasets**

- Project Use Cases Linking BER Data: **7 datasets**

ESIP Physical Samples Curation Cluster



- Scientific Author Guide and Paper: publishing sample-based research

Exploring Data Harmonizer tool (used by NMDC) for validating ESS-DIVE sample metadata

4 STEPS TO PUBLISH OPEN EARTH SCIENCE SAMPLES



1. Describe samples with rich metadata, ideally using a standardized community template.
2. Assign or use identifiers (such as IGSNs) for samples
3. Publish and cite datasets with sample identifiers
4. Reference samples in your papers using consistent formatting

Flyer: <https://doi.org/10.6084/m9.figshare.24291148.v2>

Author Guide: <https://doi.org/10.6084/m9.figshare.24669057.v1>

Sample IDs and Related Identifier		Sample Collection Details						Location	
Sample Name	Collector/Chief Scientist	Collection date	Collection time	Collection method description	Sample processing	Field program/cruise	Latitude	Longitude	
1	EC1_K001_WATER_40ML_FILT	Donnie Day	12/14/2021				41.6228	-83.2362	
2	EC1_K001_WATER_15ML_FILT	Donnie Day	12/14/2021				41.6228	-83.2362	
3	EC1_K001_WATER_1L_UNFILT	Donnie Day	12/14/2021				41.6228	-83.2362	
4	EC1_K001_WATER_125ML_UNFILT	Donnie Day	12/14/2021				41.6228	-83.2362	
5	EC1_K001_WATER_FILTER	Donnie Day	12/14/2021				41.6228	-83.2362	
6	EC1_K001_SEDIMENT_JAR	Donnie Day	12/14/2021				41.6228	-83.2362	
7	EC1_K001_SEDIMENT_BAG	Donnie Day	12/14/2021				41.6228	-83.2362	
8	EC1_K001_WETLAND_JAR	Donnie Day	12/14/2021				41.62182	-83.23883	
9	EC1_K001_WETLAND_BAG	Donnie Day	12/14/2021				41.62182	-83.23883	
10	EC1_K001_WETLAND_RING	Donnie Day	12/14/2021				41.62182	-83.23883	
11	EC1_K001_WETLAND_JAR	Donnie Day	12/14/2021				41.61541	-83.23970	

2023 Annual Community Data Workshop



Open Data Workshop Agenda November 15-16, 2023

Wed, Nov 15, 2023

All times listed in Pacific Time zone

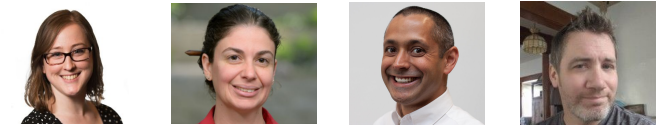
09:00AM - 10:00AM	Opening Session <ul style="list-style-type: none">Welcome and introsBER remarks on open dataESS-DIVE features updateESS-DIVE roadmap and visionQ&A
10:00AM - 11:00AM	Project Data Management and Open Science
11:00AM - 11:30AM	Break
11:30AM - 12:30PM	Community Led Discussion: <i>Project Data Publishing Decisions to Support Open Environmental System Science</i>
12:30PM - 02:00PM	Connecting Related Data for your Project

Thurs, Nov 16, 2023

**Attendees to choose their own breakout sessions in zoom*

09:00AM - 11:00AM	ESS-DIVE How-to Tutorials* <ul style="list-style-type: none">Publishing data on ESS-DIVEManaging Project Data on ESS-DIVEFinding & Accessing ESS-DIVE Data
11:00AM - 11:30AM	Break
11:30PM - 1:30PM	Advanced Search and Data Integration Enabled by ESS-DIVE Reporting Formats <ul style="list-style-type: none">Enabling Data Validation and Advanced SearchBASIN-3D Data Synthesis DemonstrationCurating High-Quality Datasets for your Project
1:30PM - 02:00PM	Workshop Wrap-Up and Future Directions

Community Presenters



ESS-DIVE Community
Workshop

Virtual, Fall 2024 – Dates TBD



What's Next?

Tracking Dataset Status



All Available Statuses

DOI Statuses

DOI Status

No DOI Added

DOI Status

Reserved

DOI Status

Pending

DOI Status

Live

Review Statuses

Review Status

Drafting

Review Status

Review & Revise

Review Status

Published

The current CDIAC data listing is incomplete, and will be updated by May 2018. Click [here](#) for the CDIAC transition website

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

ESS-DIVE
Deep Insights for Earth Science Data

DATA PORTALS PROJECTS GET STARTED ABOUT SUBMIT DATA HESHAM

< Back to search | Home | Search / Metadata

DATASET | ess-dive-48192868f7c787f-20240322T102251168108

SPRUCE S1 Bog Peat Depth Determined by Push Probe and GPR: 2009-2010

L Slater, P J Hanson, and L A Hook

🔒 Downloads 0 🗨 Citations 0 👁 Views 0

🔗 Cite this dataset 📄 Assessment report ✎ Edit 🌟 Manage Publication

Files in this dataset | Package: ess-dive-b07fab4fdccaf-20240322T102251734872

Name	File type	Size	Download All
Metadata: SPRUCE_S1_Bog_Peat_Depth_Determined_by_Push_Probe_and_GPR_2009_2010.xml	EML v2.2.0	4 KB	Download

General

Identifier | ess-dive-48192868f7c787f-20240322T102251168108

DOI Status: No DOI
Dataset Status: Draft

Streamlining Publication Process




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

A screenshot of the ESS-DIVE web interface showing a 'Manage Dataset' modal window. The modal is centered and has a white background with a blue border. It contains the following elements:

- Dataset Status:** A light blue box with a pencil icon and the word 'Draft'.
- DOI Status:** A light blue box with a crossed-out circle icon and the text 'No DOI'.
- Text:** 'Publishing the dataset will make it publicly available for search and download.' and 'The data package will be published with an existing DOI, if you supply one, or else a new DOI will be issued through ESS-DIVE.'
- Buttons:** A green button labeled 'Start publication process' and a blue button labeled 'Reserve DOI'.

Two orange arrows point from the 'Downloads' and 'Citations' buttons in the background to the 'Start publication process' button. The background interface shows the ESS-DIVE logo, navigation links, and a dataset entry for 'SPRUCE S1 Bog Peat Dep' with a file listing table below it.

 *All designs subject to change*

Improving DOI Management Transfer



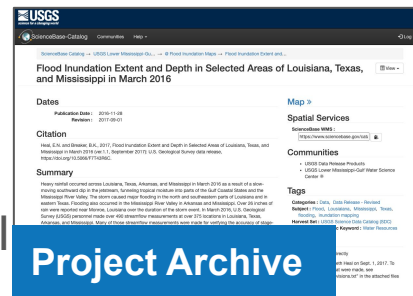
Make it easier for ESS-DIVE to manage your OSTI DOIs

- Mirror your existing datasets on ESS-DIVE
- Work with ESS-DIVE team to transfer ownership of OSTI records

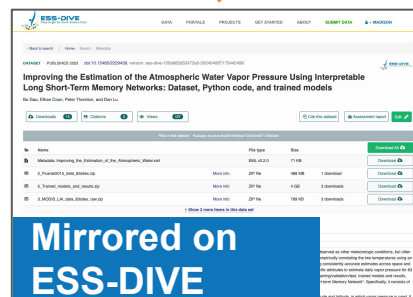
Current work: Collaborating with OSTI to streamline ownership transfers & automate record updates for transferred DOIs

Existing DOI records will be **automatically updated** with latest changes to associated ESS-DIVE datasets

Ask us about updating your DOI landing pages!



doi:10.21952/1545781



Update OSTI record on ESS-DIVE

Project Data Management Improvements



PIs and data managers default access for their project data



The screenshot shows the JupyterLab interface. At the top, there's a file browser showing a folder 'LBL Wildland Fire Dataset' containing a file 'scratch.txt' (147 B, Data type). Below this is a form for adding a new dataset. The form has sections for 'Title', 'Existing DOI and Alternate Identifiers', and 'Abstract'. The 'Title' field contains 'LBL Wildland Fire Dataset'. The 'Abstract' field contains 'WFS Data taken at test field site'.

ESS-wide project data reporting metrics
Need your input!



The screenshot shows the JupyterLab interface with a search and download page for Python data. The page title is 'Search & Download - Python Last Checkpoint: 11/14/2022 (autosaved)'. The page displays a table of search results with columns for 'Public', 'Title', 'Publication Date', and 'DOI'. The table contains 10 rows of data.

	Public	Title	Publication Date	DOI
0	True	iButton and Tinytag snow/ground interface temperature measurements at Teller 27 and Kougarak 64 from 2022-2023, Seward Peninsula, Alaska	2024	doi:10.15485/2319246
1	True	Subsurface electrical conductivity across the BEO site inferred using a capacitively coupled resistivity survey in May 2013, Utqiagvik, Alaska	2024	doi:10.15485/2335800
2	True	Representativeness-based Sampling Network Design for the State of Alaska	2013	doi:10.5440/1108686
3	True	Time-lapse photography at BEO, Utqiagvik (Barrow), Alaska, 2014	2016	doi:10.5440/1999388
4	True	Surface and Active Layer Pore Water Chemistry from Ice Wedge Polygons, Utqiagvik (Barrow), Alaska, 2013-2014	2017	doi:10.5440/1226245
5	True	Soil Water Characteristics of Cores from Low- and High-Centered Polygons, Utqiagvik (Barrow), Alaska, 2012	2017	doi:10.5440/1299259
6	True	Shrub Seeding Experiment: Environmental Conditions, Vegetation Composition, and Seeding Recruitment, Seward Peninsula, Alaska, 2018-2019	2024	doi:10.5440/1785119
7	True	Drone flight visible images, digital elevation maps, and geotiffs, Council, Seward Peninsula, Alaska, USA, July 2017	2023	doi:10.15485/2212790
8	True	Drone flight visible images, digital elevation maps, and geotiffs, Kougarak, Seward Peninsula, Alaska, USA, July 2017	2023	doi:10.15485/2228331
9	True	Drone flight visible images, digital elevation maps, and geotiffs, Teller, Seward Peninsula, Alaska, USA, July 2017	2023	doi:10.15485/2228330
10	True	Continuous soil temperature and soil deformation measurements, Teller road Mile 47, Nome, Alaska.	2023	doi:10.15485/2251663

<https://go.lbl.gov/essdive-metrics-feedback>

Hierarchical Data Views



Feature to be released late 2024

Downloads **0** Citations **0** Views **0** [Cite this dataset](#) [Assessment report](#)

Files in this dataset Package: resource_map_um:uuid:edb8385f-54df-49f9-982d-b23e7df27e74

Files and Folders	Size	Type	Metrics	Actions
▼ test 7				
level1.png	1.20 MiB	image/png		
Metadata: test.xml	2 KiB	EML v2.2.0		
▼ photos				
level2_1.png	1.66 MiB	image/png		
level2_2.png	1.93 MiB	image/png		

BER Data Systems and Project Use Cases



Collaborative paper

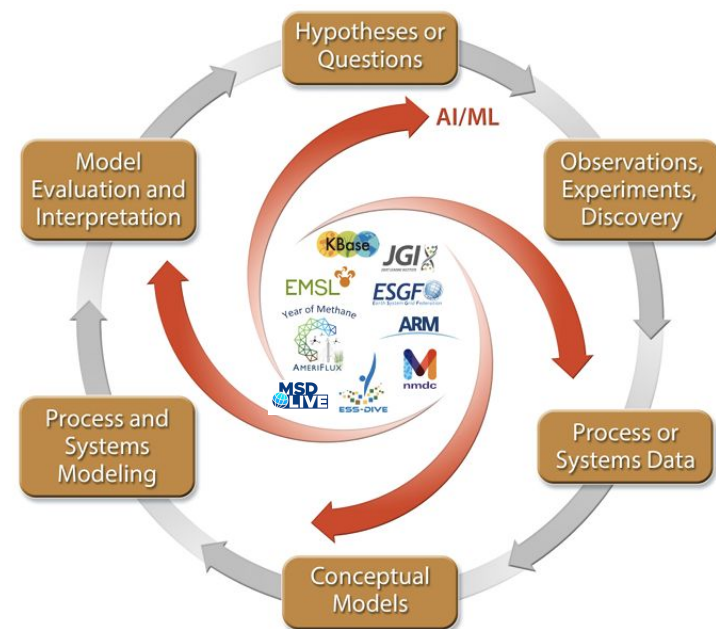
1. Scientific use cases to link and exchange (meta)data
2. Identify community practices and infrastructure needs

Approach:

- 8 BER Data Systems, 10+ project use cases
- Connect related field to lab to model (meta)data
- Samples, 'omics, sensors, remote sensing, model outputs, processed data, papers



Breakout: Strategic Data Management to Advance Environmental Systems Science
Tues., April 16th; 11:15-12:45pm EST



Connecting BER Data for ModEx

Figure from AI4CH4 workshop report (US DOE 2023)
<https://ess.science.energy.gov/ai4ch4/>

Expanding Cross-Repository Metadata Integration



External Repositories



Project Data
Archive



University
Repository



EMSL



Exploring:

- Cross-linked data
- Enhancing search across repositories
- Indexing metadata and connecting to other repos

External Links to Data or Metadata

External links for this dataset

Description	Relationship	URL
Soil thickness estimation v1.0 (archived at Zenodo)	[archived at] Complete copy of the data in this dataset	http://doi.org/10.5281/zenodo.4445383

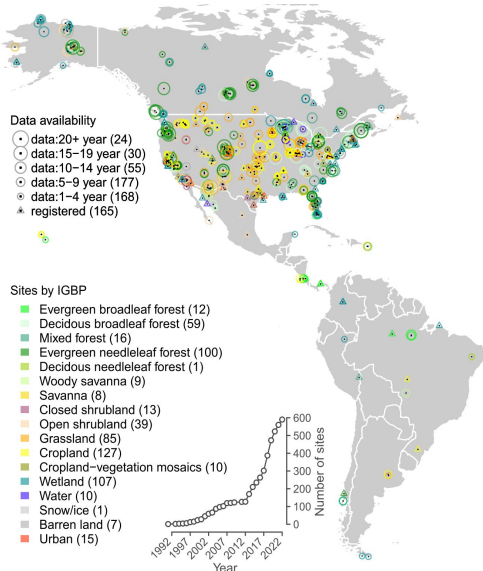
BASIN-3D Enhancements



AmeriFlux **FLUXNET data product** with gap-filled, QA/QC'd data for 196 sites across the Americas (subset of figure)

- C, H₂O, energy fluxes
- Meteorological data
- NPP, RECO

ARM Atmospheric & meteorological measurements co-located with ESS sites/data (e.g. SAIL at East River Watershed, Colorado, Ameriflux/ARM sites)



Version 2 of BASIN-3D will:

- Synthesize time series data from: Ameriflux FLUXNET, ARM, USGS Daily & Instantaneous Values, EPA Water Quality eXchange (WQX), ESS-DIVE Hydrologic Monitoring RF
- Authenticate users for data source
- Centrally handle data provided in files
- Customize missing value support
- Have citations added to metadata



Exploring Use of Large Language Models



Advanced search:

- Dataset recommendations based on plain language queries
- Summarization, extraction of key themes from published datasets
- Question/answering via chat

Data Integration:

- Variable harmonization, semantic matching
- Automated review of reporting format terms to augment and clean metadata

A screenshot of the CurateGPT! alpha search interface. It shows a sidebar with various settings: "Choose operation" (Chat), "Choose collection" (ess_deepdive), "Choose model" (gpt-3.5-turbo), "Extraction strategy" (Basic), and "Background knowledge" (ont_envo). Below these settings, it says "Cart: 0 items" and "Developed by the Monarch Initiative".

CurateGPT! *alpha*

Chat with a knowledge base

Ask me anything (within the scope of ess_deepdive!)

what datasets are relevant for microbial respiration?

Detail



Examples:

► Details

Chat

The datasets relevant for microbial respiration are ER_daily_1, Total_Ecosystem_Respiration_Square_Z, and ER_97point5pct_8.

Add to your cart

References

Reference 1

```
field_name: ER_daily
unit: grams_oxygen_per_meter_squared_per_day
definition: Ecosystem respiration estimate from streamMetabolizer. Dates with insu
data removed.
data_type: numeric
total_record_count: 770
values_summary:
  min: -30.55215904
  max: 1.225809733
doi: doi:10.15485/1985922
version: ess-dive-8aec26685468a26-20230912T185518709
data_file: Roley_CR_Metabolism_Data_Package.zip/metabolism_data.csv
data_file_url: https://fusion.ess-dive.lbl.gov/deepdive/doi:10.15485/1985922:Roley_CR_Metabol
id: https://fusion.ess-dive.lbl.gov/deepdive/doi:10.15485/1985922:Roley_CR_Metabol
```

Courtesy Chris Mungall (LBNL)

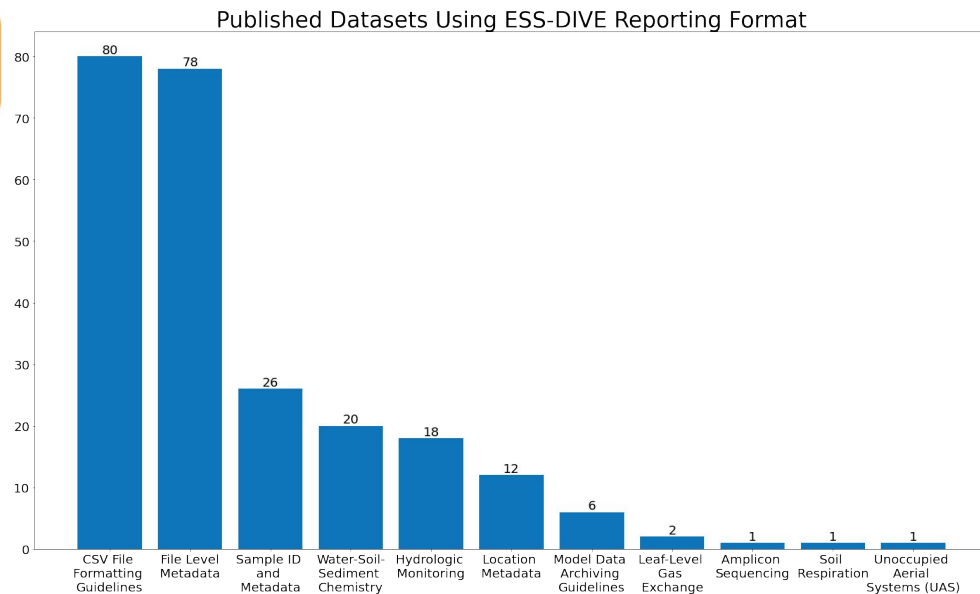
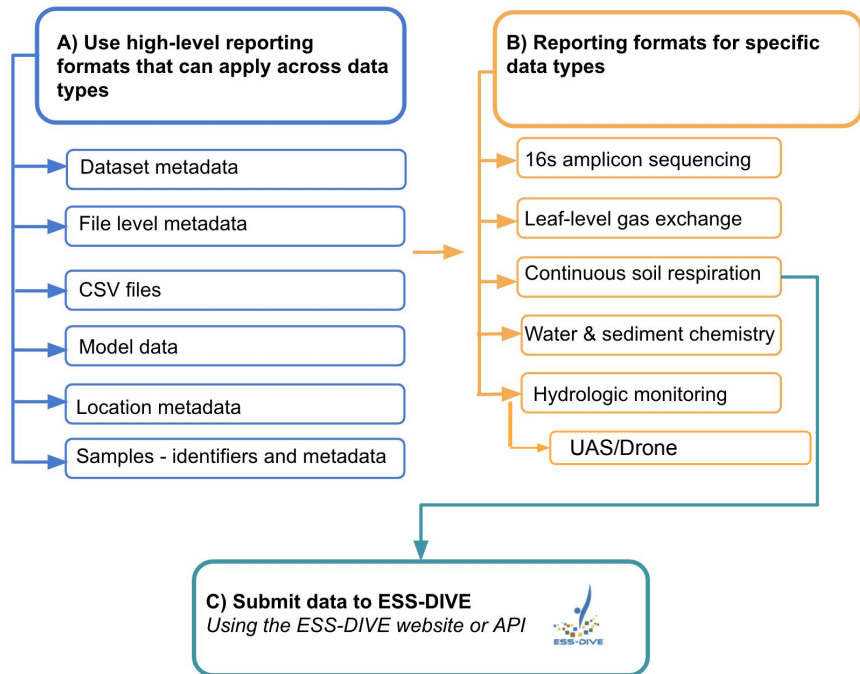
Community Funds Discussion



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

Adoption of ESS-DIVE Reporting Formats

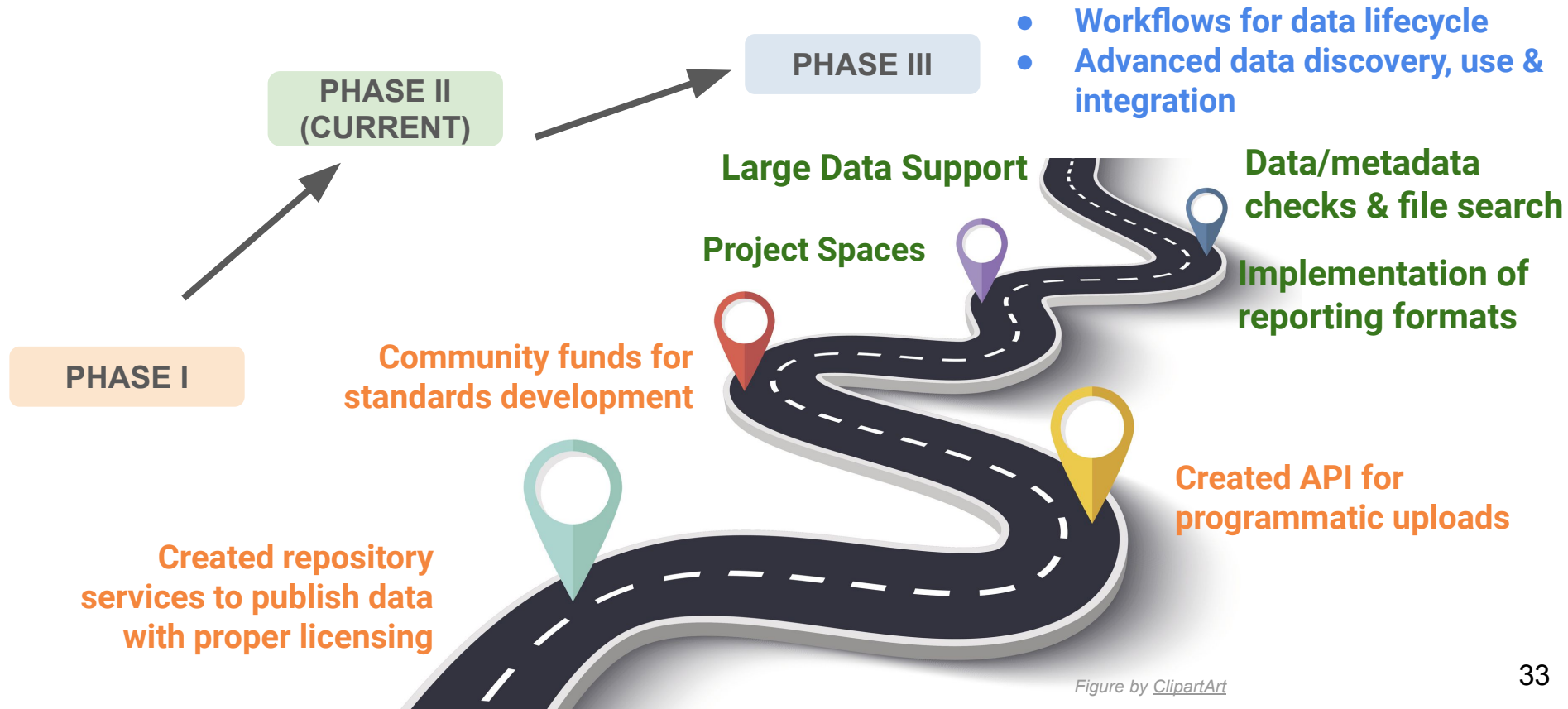


96 published datasets use one or more reporting formats

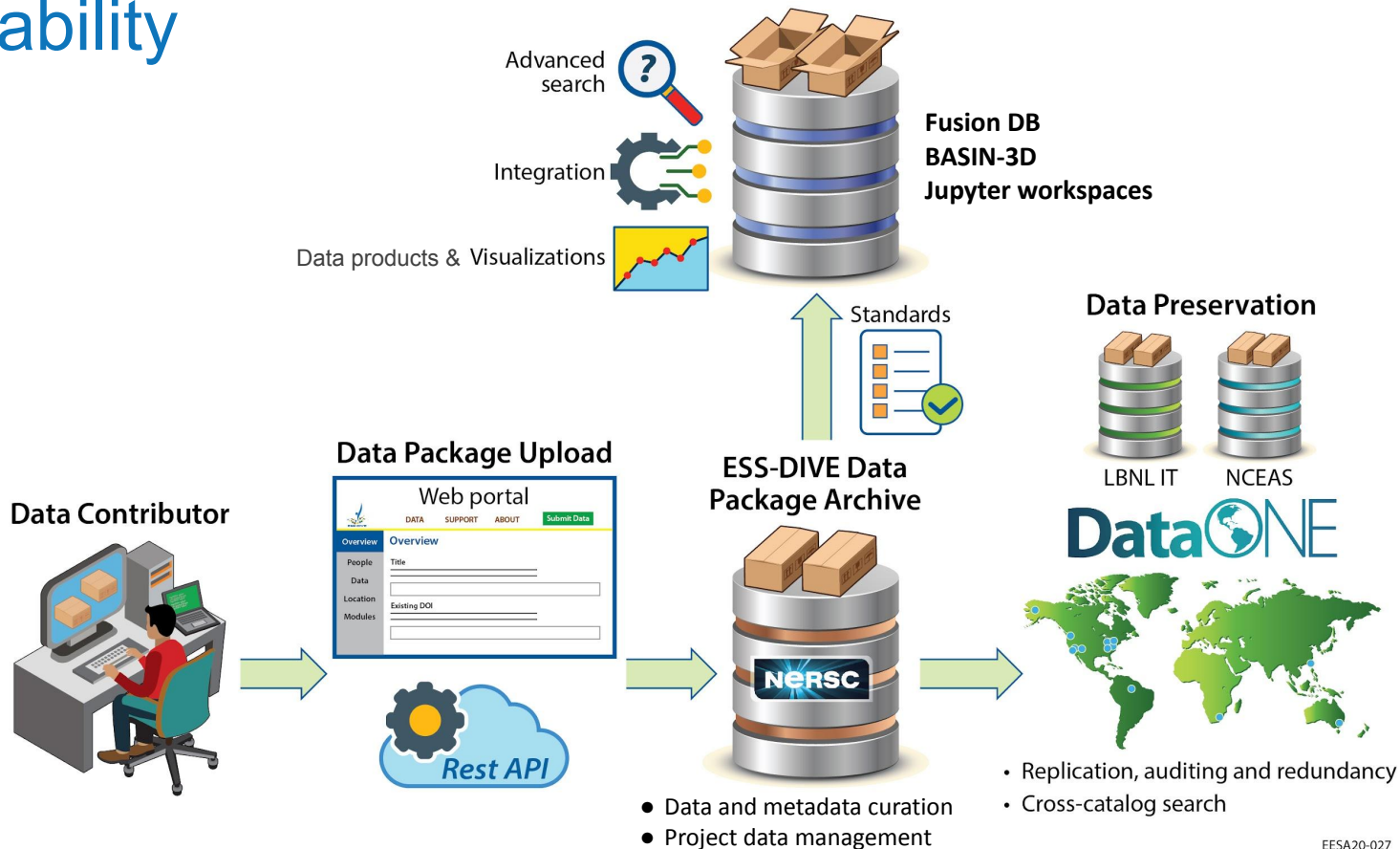
Figure source: Crystal-Ornelas, R. et al. 2022. Enabling FAIR data in Earth and environmental science with community-centric (meta)data reporting formats. *Nature Scientific Data*. <https://doi.org/10.1038/s41597-022-01606-w>

Discussion on ESS-DIVE's Future Plans

Community Engagement Drives ESS-DIVE Services and Feature Development



Vision: Easy Data Submission & Broad Data Availability



We need your input!



As a community resource how can ESS-DIVE serve your project data management needs?

- Enhancements to current features or new features (including data reporting formats)
- Pain points with current data management workflows or capabilities
- New areas: data integration, data products, Jupyter workspaces
- What kind of support & engagement would you need from ESS-DIVE staff?
- Has our current community engagement worked well and should we make changes?
- Anything else you would like to tell us about?

★ **Informal Office Hours / Help** – Mon., April 15th; 5:30-6:00pm EST

Email us at ess-dive-support@lbl.gov to submit your feedback!

ESS-DIVE at the CI/PI meetings



ESS-DIVE Updates / Planning Discussion
Mon., April 15th; 9:30-10:30am EST



Hands-on Session: ESS-DIVE Data Reporting Formats
Mon., April 15th; 1:45-3:15pm EST



Hands-on Session: Accessing HPC resources with Jupyter notebooks
Mon., April 15th; 1:45-3:15pm EST



Round Robin: Model data archiving and code management
Mon., April 15th; 3:30-5:00pm EST

★ **Informal Office Hours / Help** – Tue., April 16th; 6:30pm EST



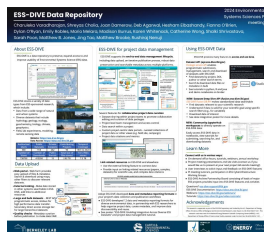
Strategic Data Management to Advance Environmental Systems Science
Tues., April 16th; 11:15-12:45pm EST



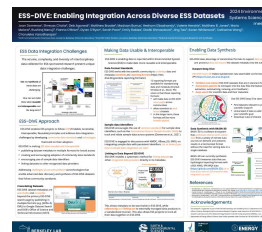
Inclusive and Equitable Research: Best Practices and Lessons Learned
Tues., April 16th; 5:30-6:30pm EST



Lunchtime Breakout: ESS-DIVE Tutorial for PIs and Data Managers
Wed., April 17th; 12:30-1:30pm EST



ESS-DIVE Data Repository Overview
Poster Session II: Tues., April 16th, 4:00-5:30pm EST; Poster #12



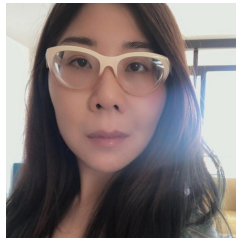
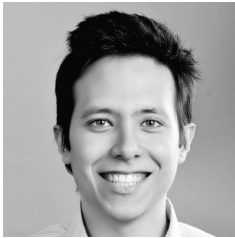
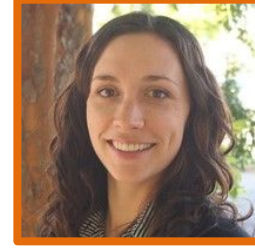
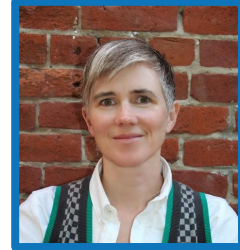
ESS-DIVE: Enabling Integration Across Diverse ESS Datasets
Poster Session II: Tues., April 16th, 4:00-5:30pm EST; Poster #18



Questions?

Connect with the ESS-DIVE Team!

ESS-DIVE Leadership



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