



ESS-DIVE Features, Services, Plans

2025 ESS CI Meeting

Charuleka Varadharajan & Shreyas Cholia

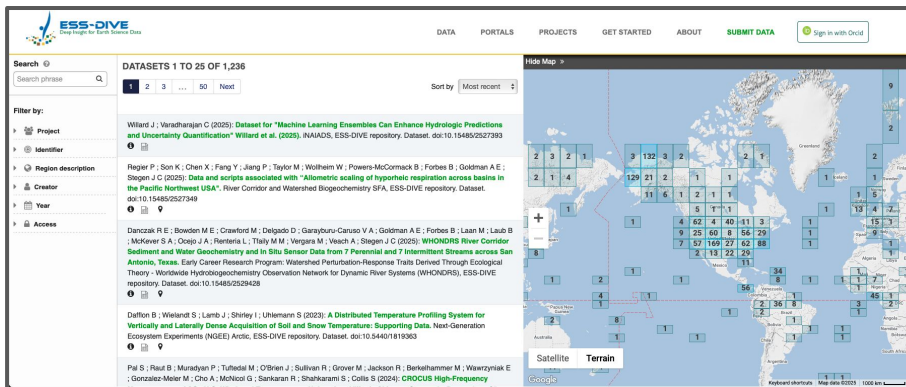


April 14, 2025

ESS-DIVE is the DOE ESS system for long-term data preservation



data.ess-dive.lbl.gov



As of March 19th, 2025



USERS

419

Registered users

100

Contributing projects

278

Events



DATA

1215

Public datasets

283

Private datasets

13.6 TB

Total file size



USAGE

594K

File downloads

1M

Data views

1200+

Citations



ESS-DIVE Roadmap



Phase 1: 2017-2020



Accessibility

Established **working repository** to store, preserve ESS data

Developed partnerships with **ESS community**

Phase 2: 2020-2024



Findability

Expanded users and storage to support ESS projects

Built community **standards** to enable **data discovery**

Phase 3: 2025-2027



Reusability

Capabilities for **data reuse, integration** with BER systems

Making it easier to contribute **well-curated data**

2027 and Beyond



Interoperability

Accelerate **knowledge discovery, predictions, AI/ML** through use of ESS-DIVE and other data

Bridging the gap between data contributors and users



Generated by DALL-E

Exciting Phase III capabilities will make it easier to contribute, find and reuse data



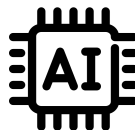
Tools & services
for **standards**
adoption



Storage &
support for
large data



Fusion
Database to
find data
within files



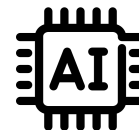
Large Language
Models for
data discovery
and curation



Tools to
integrate BER
& other data



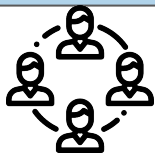
Workspaces
to **access**
data without
download



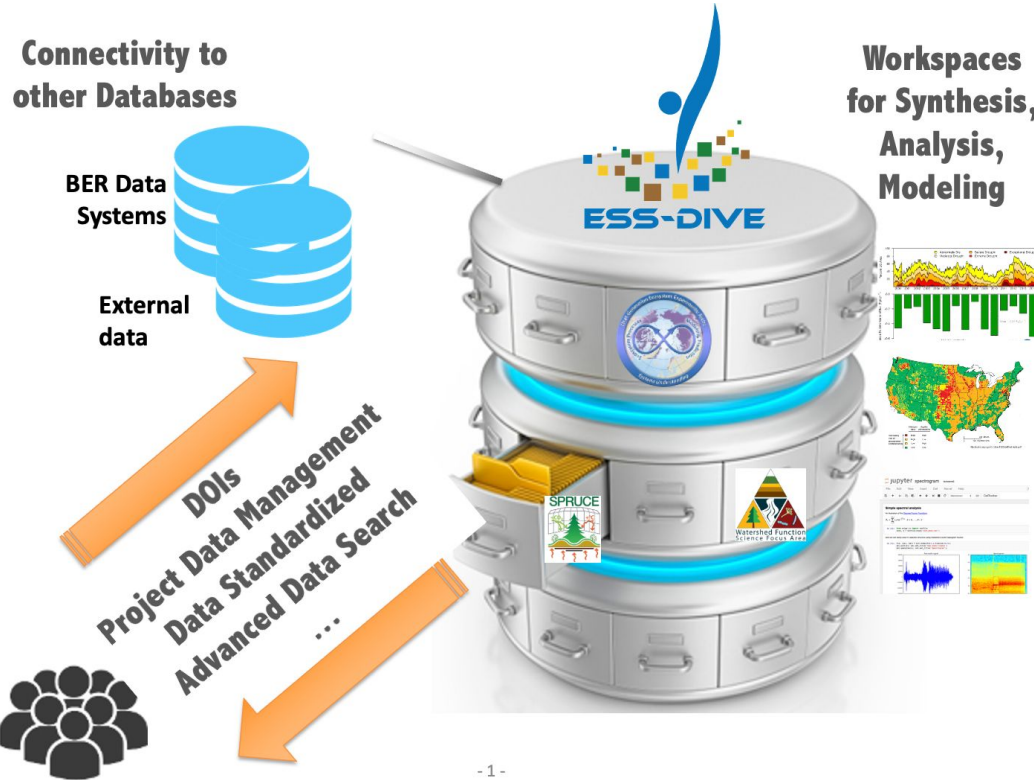
Towards
AI-ready data

Data Contributors

Data Users



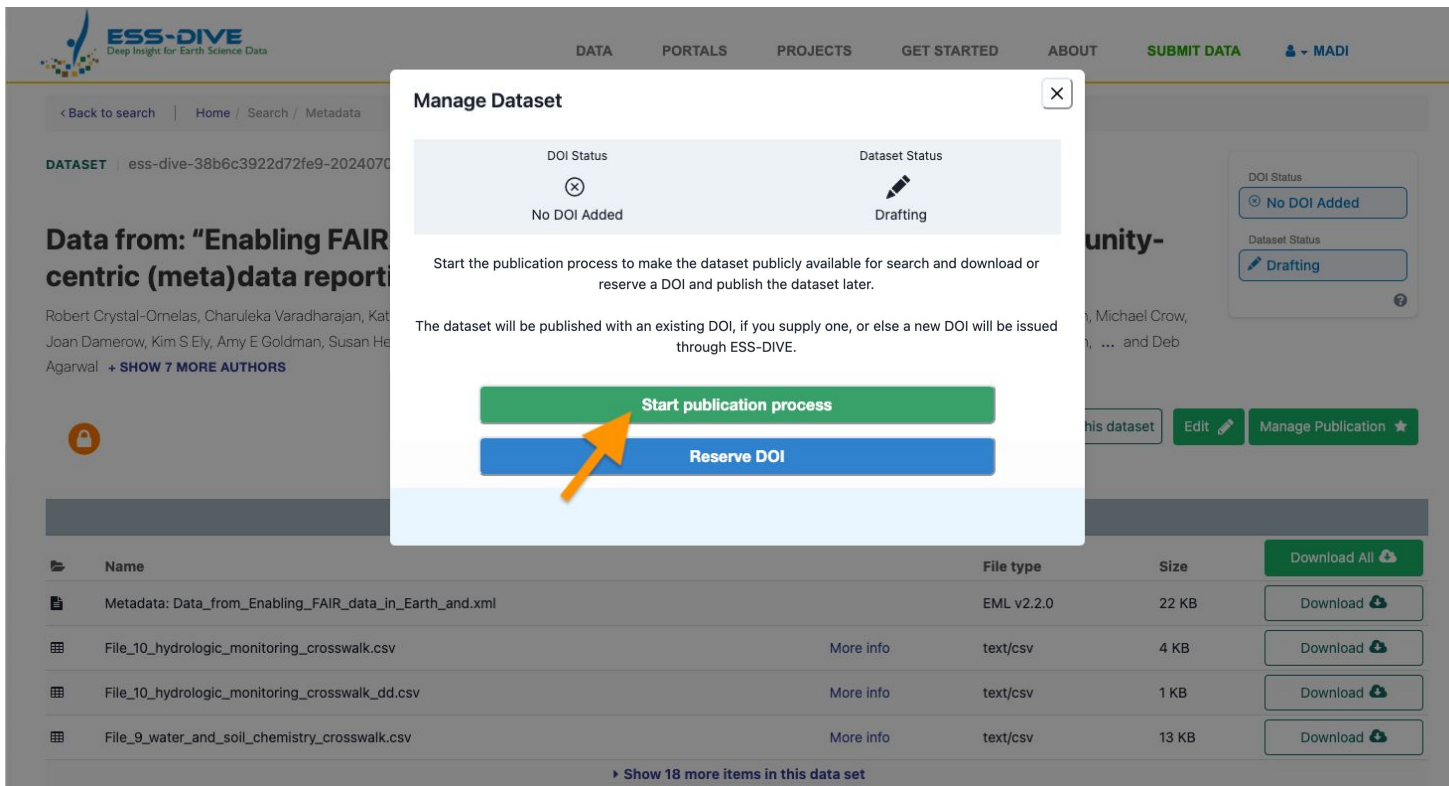
Long-term Vision: ESS Data Infrastructure for a Virtual Ecosystem (DIVE)



Accelerate knowledge generation, predictions, AI/ML through **easy data management, integration and use of** ESS-DIVE and other data

Improving Data Management, Reuse

Streamlining Publication Process



The screenshot shows the ESS-DIVE website with a 'Manage Dataset' modal open. The modal has a title bar with a close button. Inside, it displays the current DOI Status as 'No DOI Added' (with a crossed circle icon) and the Dataset Status as 'Drafting' (with a pencil icon). Below this, it provides instructions: 'Start the publication process to make the dataset publicly available for search and download or reserve a DOI and publish the dataset later.' and 'The dataset will be published with an existing DOI, if you supply one, or else a new DOI will be issued through ESS-DIVE.' At the bottom of the modal are two buttons: 'Start publication process' (green) and 'Reserve DOI' (blue). An orange arrow points to the 'Start publication process' button. In the background, the dataset page for 'Data from: "Enabling FAIR centric (meta)data report"' is visible, showing authors, a file list, and a 'Download All' button.

Manage Dataset

DOI Status: No DOI Added

Dataset Status: Drafting

Start the publication process to make the dataset publicly available for search and download or reserve a DOI and publish the dataset later.

The dataset will be published with an existing DOI, if you supply one, or else a new DOI will be issued through ESS-DIVE.

Start publication process

Reserve DOI

Name	File type	Size	Download All
Metadata: Data_from_Enabling_FAIR_data_in_Earth_and.xml	EML v2.2.0	22 KB	Download
File_10_hydrologic_monitoring_crosswalk.csv	text/csv	4 KB	Download
File_10_hydrologic_monitoring_crosswalk_dd.csv	text/csv	1 KB	Download
File_9_water_and_soil_chemistry_crosswalk.csv	text/csv	13 KB	Download

► Show 18 more items in this data set

Tracking Dataset Status



All Available Statuses

DOI Status

Dataset Status



No DOI
Added



Drafting



Reserved



In Review



Active



Published



Non
ESS-DIVE DOI

A screenshot of the ESS-DIVE dataset page for the dataset "Enabling FAIR data in Earth and environmental science with community-centric (meta)data reporting formats". The page shows the dataset title, authors, and a table of files. A red box highlights the DOI Status and Dataset Status dropdown menus on the right side of the page. The DOI Status is set to "No DOI Added" and the Dataset Status is set to "Drafting".

Home / Search / Metadata

DATASET | ess-dive-17a238a24fed3ac-20240703T223219294

Data from: "Enabling FAIR data in Earth and environmental science with community-centric (meta)data reporting formats"

Robert Crystal-Ornelas, Charuleka Varadharajan, Kathleen Beilsmith, Ben Bond-Lamberty, Kristin Boye, Madison Burrus, Shreyas Cholia, Danielle S Christianson, Michael Crow, Joan Damerow, Kim S Ely, Amy E Goldman, Susan Heinz, Valerie C Hendrix, Zarine Kakalia, Kayla Mathes, Fianna O'Brien, Dylan O'Ryan, Stephanie C Pennington, ... and Deb Agarwal + **SHOW 7 MORE AUTHORS**

[Cite this dataset](#) [Edit](#) [Manage Publication](#)

Files in this dataset Package: ess-dive-631b4c51b450e99-20240703T223219287

Name	File type	Size	Download All
Metadata: Data_from_Enabling_FAIR_data_in_Earth_and.xml	EML v2.2.0	22 KB	Download
File 10_hydrologic_monitoring_processwalk.pdf	More info	text/pdf	Download

Upgrade to DOE OSTI Services



- ESS-DIVE manages DOIs through DOE's Office of Science and Technical Information
- OSTI cannot issue or manage DOIs during a major upgrade from ELINK 1.0 to ELINK 2.0 from **May 23rd - June 1st**
- ESS-DIVE has worked closely with OSTI to:
 - Beta test ELINK 2.0;
 - Provide feedback on functionality;
 - Enable ESS-DIVE publication system to switch between v1 and v2.
- ESS-DIVE cannot publish data or reserve DOIs during the upgrade. All other services will be available.

New Tool to Download Project Publications



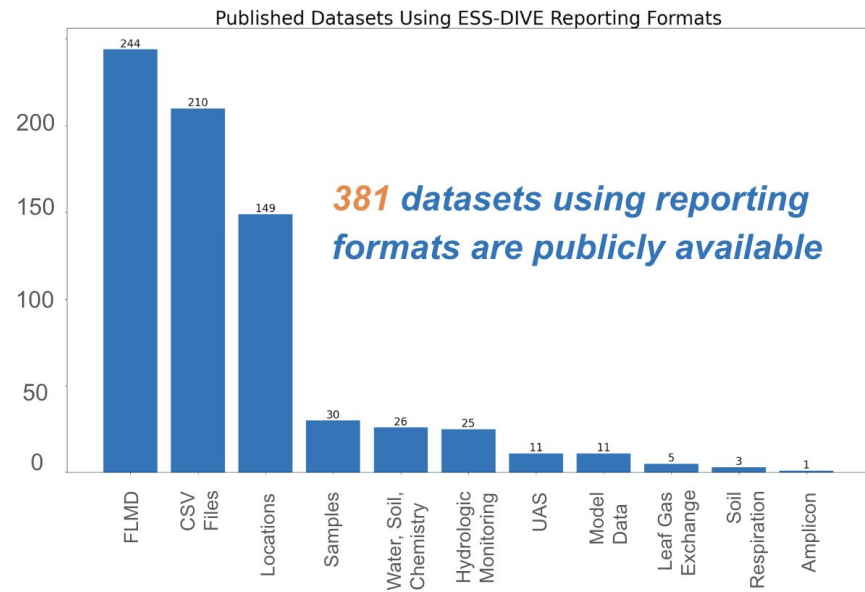
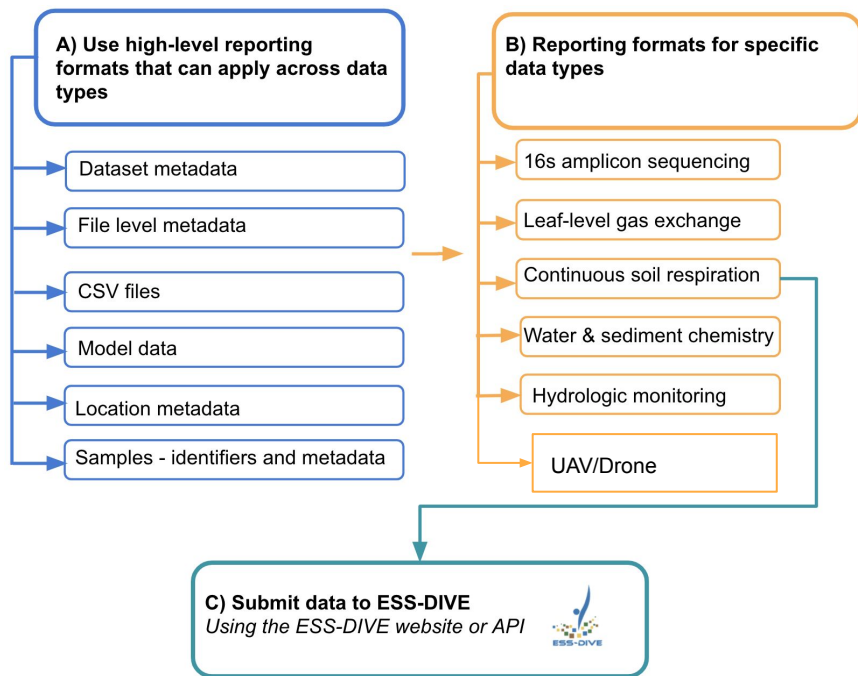
New **Jupyter Notebook tool** to obtain information about project publications

- Dataset titles & DOIs
- No. of private and public datasets
- Year of publication
- **View and download metrics**

A screenshot of a Jupyter Notebook interface titled "Search & Download - Python". It displays a table of publications with columns for "Public", "Title", "Publication Date", and "DOI". The table lists 10 publications, each with a "True" status in the "Public" column. The interface includes a top menu bar with options like File, Edit, View, Insert, Cell, Kernel, Widgets, and Help. A toolbar below the menu contains icons for file operations, running code, and markdown editing. The table data is as follows:

	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 Seedling Experiment: Environmental Conditions, Vegetation Composition, and Seedling 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

ESS-DIVE Data Reporting Formats Adoption



- Adoption of reporting formats is resource intensive
- Improvements and tools needed

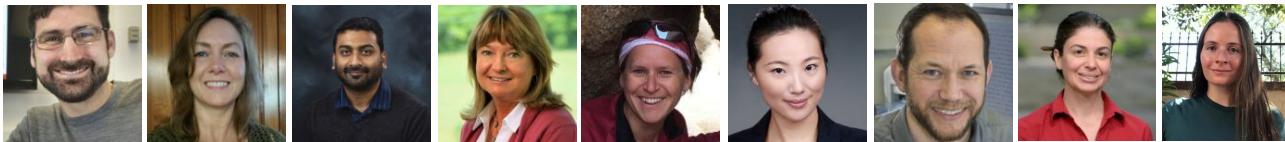
Figure source: Crystal-Ornelas, R. et al. 2022. Enabling FAIR data in Earth and environmental science with community-centric (meta)data reporting formats. *Nature Scientific Data*. <https://doi.org/10.1038/s41597-022-01606-w>

Advancing Reporting Formats and Tools







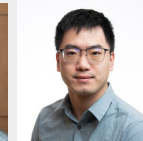


ESS-DIVE Partner Projects

2019 - 2022: Develop 12 Reporting Formats



2024: Help with Data, New and Improved RFs, Tools

						
Data Curation Support Ely (LBNL)	Hydrological Formats Revision Goldman & Forbes (PNNL)	Sensor Format & Tools Pennington & Bond-Lamberty (PNNL)	Model Data and Tools Coon & Li (ORNL, PNNL)			

Article | [Open access](#) | Published: 14 November 2022

Enabling FAIR data in Earth and environmental science with community-centric (meta)data reporting formats

Robert Crystal-Ornelas, Charuleka Varadharajan ^a, Dylan O’Ryan, Kathleen Beilsmith, Benjamin Bond-Lamberty, Kristin Boye, Madison Burrus, Shreyas Cholia, Danielle S. Christianson, Michael Crow, Joan Damerow, Kim S. Ely, Amy E. Goldman, Susan L. Heinz, Valerie C. Hendrix, Zarine Kakalia, Kayla Mathes, Fianna O’Brien, Stephanie C. Pennington, Emily Robles, Alistair Rogers, Maegen Simmonds, Terri Velliquette, Pamela Weisenborn, ... Deborah A. Agarwal ^a [+ Show authors](#)


Scientific Data **9**, Article number: 700 (2022) | [Cite this article](#)

4925 Accesses | 9 Citations | 19 Altmetric | [Metrics](#)

 Ecological Informatics
Volume 62, May 2021, 101280

A reporting format for field measurements of soil respiration

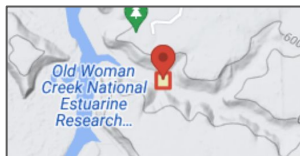
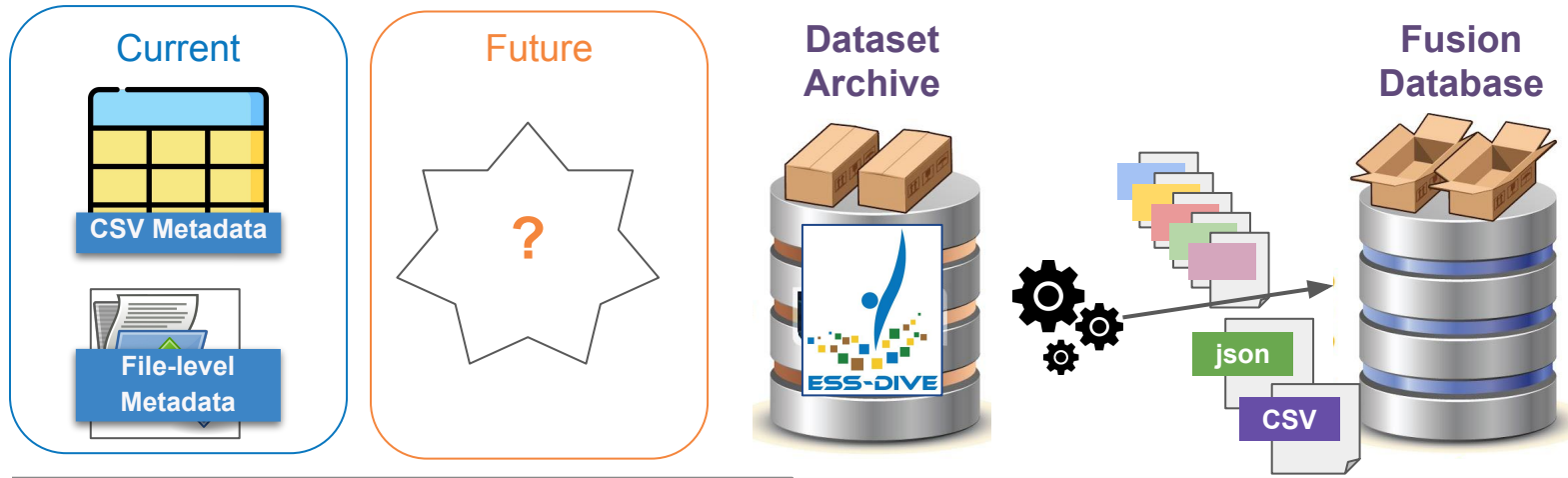
Ben Bond-Lamberty ^a, ^b, Danielle S. Christianson ^b, Robert Crystal-Ornelas ^b, Kayla Mathes ^c, Stephanie C. Pennington ^a

 Ecological Informatics
Volume 61, March 2021, 101232

A reporting format for leaf-level gas exchange data and metadata

Kim S. Ely ^a, ^b, Alistair Rogers ^a, Deborah A. Agarwal ^b, Elizabeth A. Ainsworth ^c, Loren P. Albert ^d, Ashehad Ali ^a, Jeremiah Anderson ^a, Michael J. Aspinwall ^f, Chandra Bellasio ^g, Carl Bernacchi ^e, Steve Bonnage ^h, Thomas N. Buckley ⁱ, James Bunce ^j, Angela C. Burnett ^a, Florian A. Busch ^k, Amanda Cavanagh ^l, Lucas A. Cernusak ^m, Robert Crystal-Ornelas ^a, Joan Damerow ^h, Kenneth J. Davidson ^a, Dedi Yang ^a

Reporting Formats enables Deeper Search



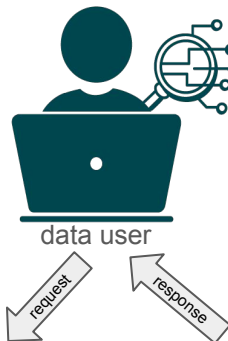
(41.376264, -82.50797)

Deep Search with the Deep Dive API

Request URL

<https://fusion.ess-dive.lbl.gov/api/v1/deepdive?rowStart=1&pageSize=25&fieldName=ch4>

<https://fusion.ess-dive.lbl.gov/>



Response body

```
{
  "url": "https://fusion.ess-dive.lbl.gov/api/v1/deepdive?rowStart=1&pageSize=25&fieldName=ch4",
  "next": "https://fusion.ess-dive.lbl.gov/api/v1/deepdive?rowStart=26&pageSize=25&fieldName=ch4",
  "previous": null,
  "pageCount": 25,
  "results": [
    {
      "field_name": "Flux_CH4_diffusive",
      "unit": "umol m2 s-1",
      "definition": "CH4 flux without ebullition (positive = to atmosphere); negative values are also acceptable",
      "data_type": "numeric",
      "total_record_count": 310,
      "missing_values_count": 12,
      "values_summary": {
        "min": -0.16142099999999998,
        "max": 3.4683778
      },
      "doi": "doi:10.15485/2229438",
      "version": "ess-dive-8a116c1b99c636d-20231111T151442562",
      "data_file": "soil_respiration_data.csv",
      "data_file_url": "https://fusion.ess-dive.lbl.gov/api/v1/deepdive/doi-10.15485/2229438/soil_respiration_data.csv"
    }
  ]
}
```

Download

Reducing Time for Publications in Reviews



Hello,

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

Dataset metadata



1. Please include email addresses for each contributor. Contributors with insufficient information may be difficult for users to locate.
2. Ensure that all formatting errors within keywords/variables are resolved (e.g., Utqia?vik, ?, Utqia?vik, Alaska).

• File Level Metadata

- We recommend that you define all files contained within your dataset. In particular, ensure that the following files are described: 2201_BGC_Soil_Acid_Base_Titration_Data_dd.csv; 2201_BGC_Soil_Characterization_dd.csv; 2201_BGC_Soil_Water_Retention_Data_dd.csv; flmd.csv

File-level metadata



Automated machine readability checks with Fusion Database Processing Pipeline

Future: Automating Reviews using LLMs



Original title: Integrated Hourly Meteorological Database of 20 Meteorological Stations (1981-2022) for Watershed Function SFA Hydrological Modeling

Model: openai/gpt-4o

Response: The 'name' field in the provided JSON is: "Integrated Hourly Meteorological Database of 20 Meteorological Stations (1981-2022) for Watershed Function SFA Hydrological Modeling".

Let's check it against the rules:

1. ****Succinctness**:** The title is 15 words long, which is within the 7 to 20 words range.
2. ****Unexplained Acronyms**:** The acronym "SFA" is not explained in the title.
3. ****Key Information**:** The title includes the type of data (Integrated Hourly Meteorological Database), the number of stations (20), the date range (1981-2022), and the purpose (Hydrological Modeling). However, it does not specify the location.

Suggested edit to improve the title:

"Integrated Hourly Meteorological Database (1981-2022) for 20 Stations in East River Watershed for Hydrological Modeling"
This revised title is 18 words long, explains the location, and avoids unexplained acronyms.

Data Integration

Bio-Eco Data Integration: 5 BER Resources Collaborating to Prototype a Unified Data Access Layer



Unifying Access Layer Components

BERtron - Global Search supported by common APIs - **find and reuse** data

Data Transfer Service - **maintain provenance, propagate credit**



Sample metadata,
standardized data
products



Biogeochemical
measurements,
sensor data

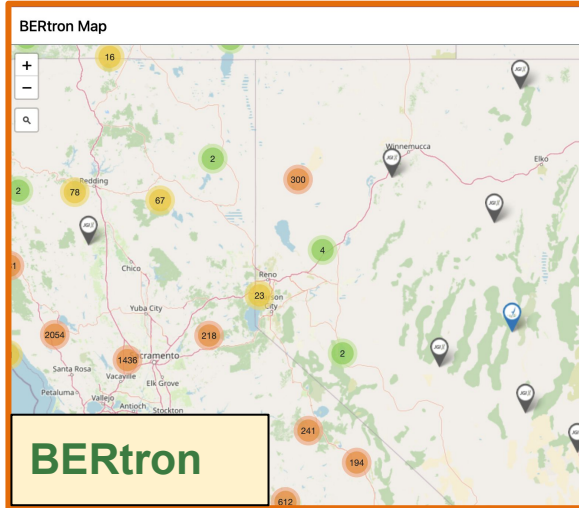


Proteomics,
metabolomics,
imaging



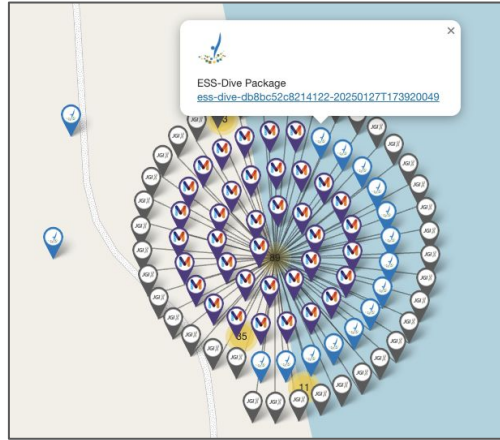
(meta)genomics,
-transcriptomics,
metadata

BERtron map to find ESS-DIVE data co-located with biological data

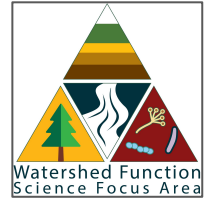
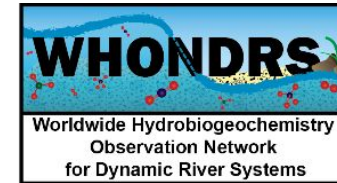
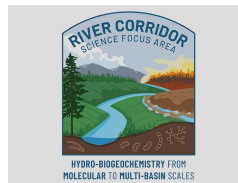


<https://ber-data.github.io/bertron/map/index.html>

* w/ LLMs @ 1000 Jam



Hanford reach,
Columbia River



... and many more projects
with bio-eco data

Upcoming: Automated Metadata Exchange with NMDC and ARM



Creating the GROW (Genome Resolved Open Watershed) Database: Leveraging Distributed Research Networks to Understand Watershed Systems

Description Streams and rivers represent key functioning units in Earth's terrestrial system that provide services essential for global energy, food and water security, as well as human recreation. Rivers are responsible for the transport and nutrient processing of terrestrial water and solutes across landscapes with discharge into reservoirs, estuaries, and oceans. Despite these important ecological roles, we know surprisingly little about the microbiomes in river and stream systems. Here, we will address this knowledge gap by sampling over 250 rivers, including the 35 larger rivers in the United States, many of which will be collected over time or across interconnected hydrological units, with a sub-set also receiving metatranscriptomes and metabolome data. To accomplish this breadth and intensity of sampling, we have formed a collaborative 15-member research consortium composed of hydrologists, geochemists, ecologists, and microbiologists. The overarching objective is to genomically resolve the taxonomy, the metabolic potential and expression, and distribution of riverine microorganisms and traits within and across watersheds.

Team
 Peter Raymond, Jacob Hosen, Mikayla Borton, Annika Mosier, Rebecca Daly, Jerome P. Pavat, Byron Cums, Stephen Good, Matthew Sullivan, Edward Hall, Christopher Henry, Kelly Wrighton, James Steyer, Robert Danczak, Michael Wilkins

Funding Sources
 The U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (BER) under contract DE-AC05-76OR21400 (PNNL) for the River Corridor Hydrobiogeochemistry Science Focus Area, BER Genomic Sciences Early Career Research Program award DE-SC0019746 to K. Wrighton.

NMDC Details
 ID: 4946187-11-9020389
 Sample count: 430

Additional Resources
 Links
 JGI GOLD Metadata
<https://gold.jgi.doe.gov/study/161-G00149396>
<https://www.pnnl.gov/projects/river-corridor>

Data Sets
 Toyoda, J. G., Goldman, A. E., Aron, S., Bar-Zeev, E., Chu, R. K., Danczak, R. E., Daly, R. A., Forbes, B., Ganayaburu-Caruso, V. A., Graham, J. B., Gu, X., Miron, J. J., Ben, H., Brenner, L., Resch, C. T., Trask, M., Toku, N., Torgerson, J. M., Wells, J. J., WHONDRS Consortium, T. (2020). WHONDRS Summer 2019 Sampling Campaign: Global River Corridor Surface Water FTICR MS, NPOC, TN, Arsons, Stable Isotopes, Bacterial Abundance, and Dissolved Inorganic Carbon (v6) [Data set]. Environmental System Science Data Infrastructure for a Virtual Ecosystem, River Corridor and Watershed Biogeochemistry SFA. <https://doi.org/10.15485/1603775>
 Provider: ESS-DIVE
 Goldman, A. E., Aron, S., Bar-Zeev, E., Chu, R. K., Danczak, R. E., Daly, R. A., Delgado, D., Fansler, S., Forbes, B., Ganayaburu-Caruso, V. A., Graham, J. B., Gu, X., Miron, J. J., Ben, H., Brenner, L., Resch, C. T., Root, K. A., WHONDRS Consortium, T. (2020). WHONDRS Summer 2019 Sampling Campaign: Global River Corridor Sediment FTICR MS, Dissolved Organic Carbon, Aerobic Respiration, Elemental Composition, Grain Size, Total Nitrogen and Organic Carbon Content, Bacterial Abundance, and Stable Isotopes (v8) [Data set]. Environmental System Science Data Infrastructure for a Virtual Ecosystem, River Corridor and Watershed Biogeochemistry SFA. <https://doi.org/10.15485/1729719>
 Provider: ESS-DIVE
 Borton, M., Wrighton, K., Sarkkin, B., Henry, C., Wood-Charlton, J. L., Edrington, J., Faria, J., & Cannon, S. (2022). GROWdb US River Systems - Samples [Data set]. Colorado State University, Fort Collins, CO. <https://doi.org/10.25982/109073.30/1895615>
 Provider: KBASE

Part of:
 WHONDRS
 River Corridors Science Focus Area

Expanded ESS-DIVE (EML) to NMDC Schema Crosswalk

Metadata Field	ESS-DIVE EML	NMDC Schema	Notes / Mapping
Title	<dataset>/<title> – Dataset title	Study.title – Study title	Common field: I mapping.
Alternative Identifier(s)	<alternateIdentifier> – Alternate IDs	Study.alternative_identifiers – List of alternate IDs	Common field: I support multiple identifiers.
Abstract / Description	<dataset>/<abstract> – Dataset abstract	Study.description – Brief description of the study	Common field: I provide a textua
Keywords	<dataset>/<keywordSet> – Keywords	(No direct equivalent)	Difference: NMI not have a gene "keywords" field is captured through specific fields or controlled vocab
Publication Date	<dataset>/<pubDate> – Dataset publication date	(No explicit field)	Difference: NMI not have a dedic publication date the Study level.
Data Usage Rights (License)	<dataset>/<intellectualRights> – Usage rights/license	(No direct equivalent)	Difference: Lice information is ty managed by the

* w/ LLMs @ 1000 Jam

ESS-DIVE Datasets

BASIN-3D: An environmental data integration tool



Time Series Data Integration:

USGS NWIS Daily & Instantaneous Values

EPA Water Quality eXchange (WQX)

ESS-DIVE Hydrologic Monitoring Reporting Format

<https://basin3d.readthedocs.io>

Timeseries synthesized output formats: pandas dataframe, hdf5
Integrates with Python-based tools, e.g. Jupyter Notebooks

Planned improvements by Jan 2026



New Time Series Data Sources



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 met data product
co-located ESS sites/data
(e.g. SAIL at East River
Watershed, Colorado,
Ameriflux/ARM sites)

Extension to Spatial Data



Watershed Workflow



DayMet Daily Climate Data

1 km x 1km gridded data

- Min, Max Temperature
- Precipitation, Snow Water Equivalent
- Water Vapor Pressure
- Radiation, Day length

Related new features

- Authenticate users for data source
- Centrally handle data provided in files
- Customize missing value support
- Query by bounding box (lat / long)
- Prototype spatial data formats
- Citations added to metadata

ESS-DIVE Partner Projects

2024 ESS-DIVE Partner Projects to Support Data Standardization and Curation



ESS-DIVE Data Curation Support



Kim Ely
LBNL

Harmonization and Usability of Hydrologic Monitoring and Soil, Sediment, and Water Chemistry Reporting Formats



Amy Goldman
PNNL



Brienne Forbes
PNNL

Improving Advanced Terrestrial Simulator (ATS) Model Data Managing and Archiving Standards



Ethan Coon
ORNL



Zhi Li
PNNL

A Workflow and Reporting Format for Processing Environmental Sensor Data and Automated Generation of ESS-DIVE Compliant Metadata; Revision of Soil Respiration Reporting Format



Stephanie Pennington
PNNL



Ben Bond-Lamberty
PNNL

ESS-DIVE Data Curation Support

Kim Ely, LBNL

Supporting data contributors



WHY?

Response to requests for data curation help

WHO?

Data contributors lacking data management resources

WHAT?

Guidance on data best practices and reporting formats

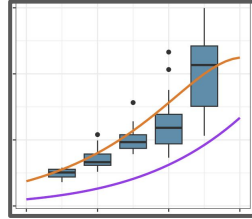
HOW?

Online meetings and email follow up, with individuals or teams

What I bring to the ESS-DIVE data curator role



Expertise across the full data life cycle

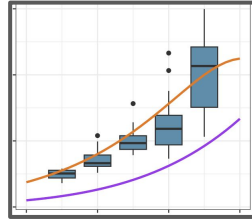


$n = > 100$

What I bring to the ESS-DIVE data curator role



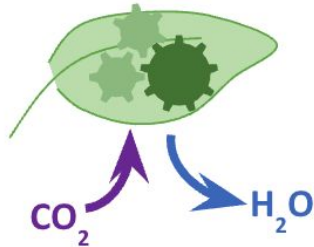
Expertise across the full data life cycle



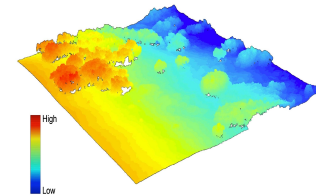
n = > 100

Delivered two domain specific reporting formats

Leaf-level gas exchange



Unoccupied Aerial Systems (UAS)



Guidance across all phases of data preparation



Assemble data



Apply reporting formats

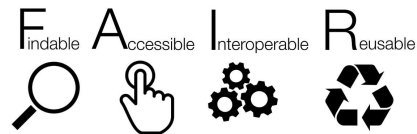
High level RF

Domain specific RF

File level metadata

Dataset metadata

Submit to ESS-DIVE



Use the ESS-DIVE Contact form to request assistance

You are contacting us about... *

Select all that are applicable

☒ ESS-DIVE Data Curation Support

☐ Submitting new data

☐ Searching for or accessing data

☐ Incorrect or missing data

☐ A question or clarification

☐ A bug

☐ A feature request

☐ Other: _____

Describe your issue *

Provide as much detail as possible

Your answer

*some details about what you are seeking
help with*

Reporting Format for Environmental Sensor Data and Automated Generation of ESS-DIVE Compliant Metadata

Stephanie Pennington & Ben Bond-Lamberty, PNNL

Processing Environmental Sensor Data

1. **Produce** an easy-to-customize R tool for processing environmental sensor data

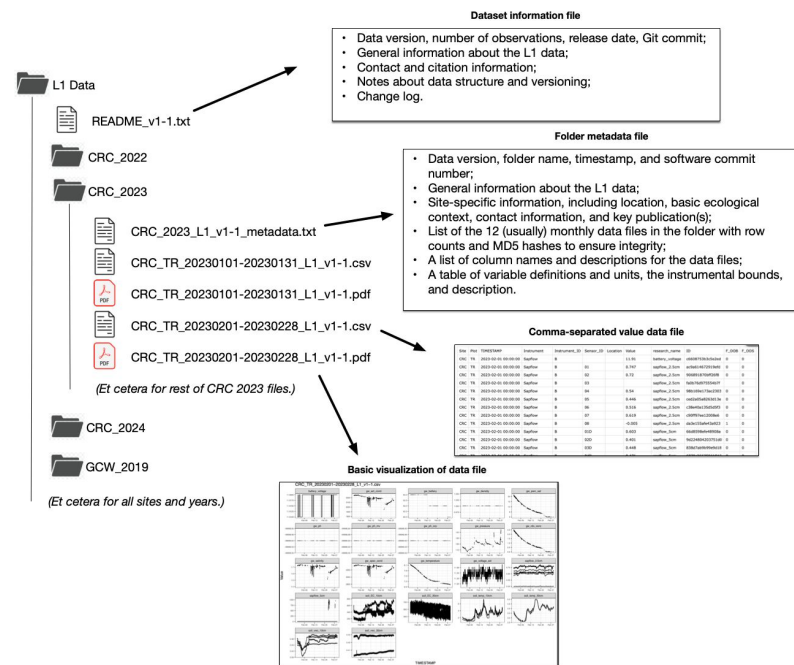
We focus on T05 data from Campbell dataloggers. These are widely used and have a standardized output format



Processing Environmental Sensor Data

1. **Produce** an easy-to-customize R tool for processing environmental sensor data

We focus on T05 data from Campbell dataloggers. These are widely used and have a standardized output format





Sensor Data Reporting Format

1. **Produce** an easy-to-customize R tool for processing environmental sensor data

We focus on T05 data from Campbell dataloggers. These are widely used and have a standardized output format

2. **Develop** a general sensor data reporting format to generate ESS-DIVE compliant formats

Focusing on sensors through *dataloggers*

Interoperability with other reporting formats

The TOA5 format looks like this:

```
"TOA5","Compass_CRC_W_301","CR1000X","28724","CR1000X.Std.05.01","CPU:COMPASS_v3  
"TIMESTAMP","RECORD","BattV_Min","Rain_mm_Tot","SlrFD_W_Avg","SlrTF_MJ_Tot","WS  
"TS","RN","Volts","mm","W/m^2","MJ/m^2","meters/second","meters/second","","met  
","","Min","Tot","Avg","Tot","WVc","WVc","Tot","Max","TMx","Avg","Max","TMx","M  
"2023-04-08 00:00:00",268,12.67,0,234.3,1.68681,1.601,77.77,0,6.12,"2023-04-07 0  
"2023-04-09 00:00:00",269,12.66,0,237.5,1.70967,1.52,104.4,0,5.46,"2023-04-08 09  
"2023-04-10 00:00:00",270,12.66,0,246.3,1.773395,1.045,95.4,0,3.9,"2023-04-09 22  
"2023-04-11 00:00:00",271,12.67,0,237.4,1.70935,0.875,143,0,4.37,"2023-04-10 10  
"2023-04-12 00:00:00",272,12.68,0,231.1,1.66359,2.62,240.3,0,10.91,"2023-04-11 7
```

Data Logger-level metadata file

Format	TOA5
Logger_name	Compass_CRC_W_301
Logger_model	CR1000X
Table_name	ClimaVue50_24hr

SCAN FOR
SURVEY



Soil Respiration Reporting Format 2.0

1. **Produce** an easy-to-customize R tool for processing environmental sensor data

We focus on T05 data from Campbell dataloggers. These are widely used and have a standardized output format

2. **Develop** a general sensor data reporting format to generate ESS-DIVE compliant formats

Focusing on sensors through *dataloggers*

Interoperability with other reporting formats

3. **Update** and expand the existing ESS-DIVE soil respiration reporting format

SCAN FOR
SURVEY



Harmonization of Water-Soil-Sediment Chemistry and Hydrologic Monitoring Reporting Formats

Amy Goldman & Brieanne Forbes, PNNL

Update on Harmonization and Increased Usability of ESS-DIVE's Hydrologic Monitoring and Water, Soil, and Sediment Chemistry Reporting Formats

Amy Goldman, Brieanne Forbes, and
Bibi Powers-McCormack

CIWG Annual Meeting 2025



Amy.Goldman@pnnl.gov
Brieanne.Forbes@pnnl.gov

This research is supported by
an ESS-DIVE Partner Project

Existing Hydrologic Monitoring RF and Soil, Sediment, and Water Chemistry RF have opportunities for harmonization



Hydrologic Monitoring RF

Sensor data

Developed by Goldman et al.



Water, Soil, and Sediment, Chemistry RF

Sample data

Developed by Boye et al.

Opportunities for harmonization

Data variables

Alphanumeric metadata codes

Metadata header rows

Header rows in the Hydrologic Monitoring RF embed metadata

# HeaderRows_6		
# HeaderRows_Format: Column_Header; Unit; InstallationMethod_ID; Instrument_Summary		
# DateTime; YYYY-MM-DD hh:mm:ss; Minidot_01; PME miniDOT Logger real time clock.		
# Temperature; degree_celsius; Minidot_01; PME miniDOT Logger with temperature sensor.		
# Dissolved_Oxygen; milligrams_per_liter; Minidot_01; PME miniDOT Logger with optical dissolved oxygen sensor (fluorescence quenching).		
DateTime	Temperature	Dissolved_Oxygen
2022-05-02 19:05:00	15.18	9.667
2022-05-02 19:06:00	15.18	9.659
2022-05-02 19:07:00	15.16	9.665

Hydrologic Monitoring RF installation method IDs are defined in a separate file

# HeaderRows_6		
# HeaderRows_Format: Column_Header; Unit; InstallationMethod_ID; Instrument_Summary		
# DateTime; YYYY-MM-DD hh:mm:ss; Minidot_01; PME miniDOT Logger real time clock.		
# Temperature; degree_celsius; Minidot_01; PME miniDOT Logger with temperature sensor.		
# Dissolved_Oxygen; milligrams_per_liter; Minidot_01; PME miniDOT Logger with optical dissolved oxygen sensor (fluorescence quenching).		
DateTime	Temperature	Dissolved_Oxygen
2022-05-02 19:05:00	15.18	9.667
2022-05-02 19:06:00	15.18	9.659
2022-05-02 19:07:00	15.16	9.665

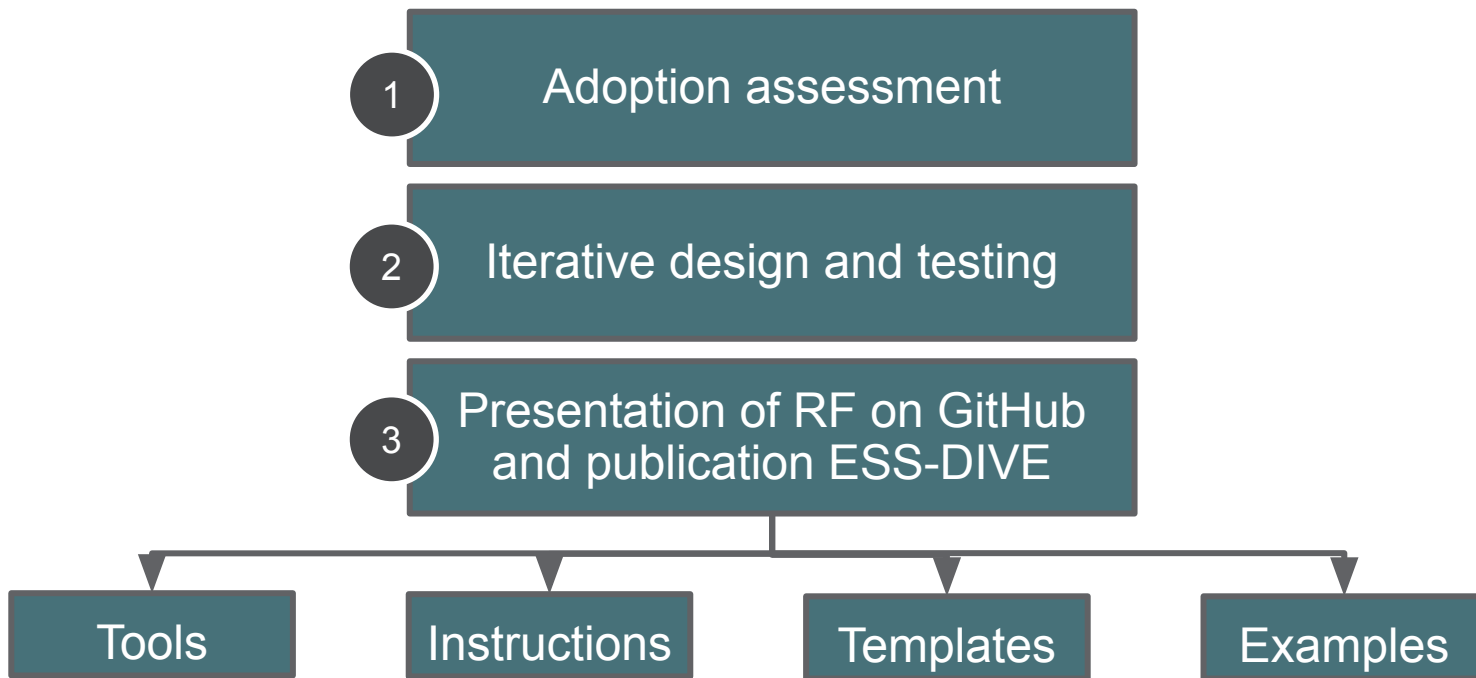
Header rows in the Water, Soil, and Sediment Chemistry RF embed metadata

#Columns	4			
#Header_Rows	6			
Field_Name	Sample_Name	IGSN	Material	Percent_Fine_Sand
Unit	N/A	N/A	N/A	percent
Unit_Basis	N/A	N/A	N/A	as_percent_of_dry_weight
MethodID_Preparation	N/A	N/A	N/A	GRN_T_PREP_001
MethodID_Analysis	N/A	N/A	N/A	GRN_T_AN_001
MethodID_DataProcessing	N/A	N/A	N/A	GRN_T_DP_001
#Start_Data	CM_001_GRN	10.58052/IEWDR01Q5	Sediment	3.3
N/A	CM_002_GRN	10.58052/IEWDR01Q6	Sediment	5.5
N/A	CM_003_GRN	10.58052/IEWDR01Q7	Sediment	4.2

Water, Soil, and Sediment Chemistry RF method IDs are defined in a separate file

#Columns	4			
#Header_Rows	6			
Field_Name	Sample_Name	IGSN	Material	Percent_Fine_Sand
Unit	N/A	N/A	N/A	percent
Unit_Basis	N/A	N/A	N/A	as_percent_of_dry_weight
MethodID_Preparation	N/A	N/A	N/A	GRN_T_PREP_001
MethodID_Analysis	N/A	N/A	N/A	GRN_T_AN_001
MethodID_DataProcessing	N/A	N/A	N/A	GRN_T_DP_001
#Start_Data	CM_001_GRN	10.58052/IEWDR01Q5	Sediment	3.3
N/A	CM_002_GRN	10.58052/IEWDR01Q6	Sediment	5.5
N/A	CM_003_GRN	10.58052/IEWDR01Q7	Sediment	4.2

Partner project includes phases of feedback and testing before publication



Current use of Hydrologic Monitoring RF and Water, Soil, and Sediment Chemistry RF is low

1.6%

datasets on ESS-DIVE use the
Hydrologic Monitoring RF

18 datasets out of 1,128

10 of these datasets (56%) originate
from our team

2.0%

datasets on ESS-DIVE use the
Water, Soil, and Sediment
Chemistry RF

23 datasets out of 1,128

14 of these datasets (61%) originate
from our team

Currently collecting adoption assessment information to identify barriers to RF use

Preliminary Survey Results

Top 3 reasons to choose not to adopt RFs (n = 13)

1

Did not know RF existed

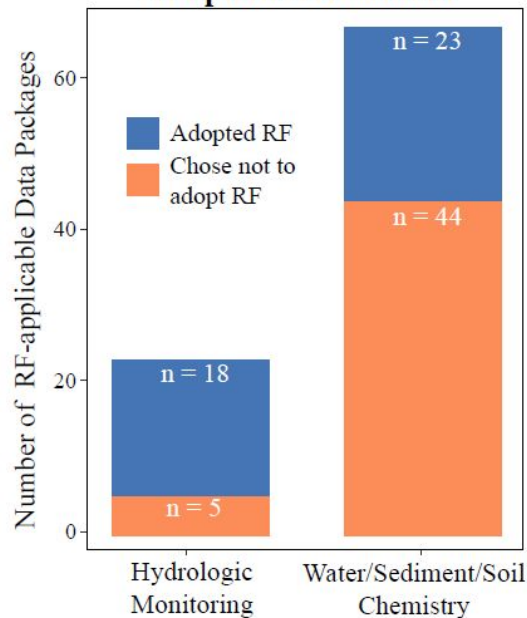
2

RF was not required

3

Did not have time/resources

Preliminary Reporting Format (RF) Adoption Assessment



29 ESS Projects are represented in this assessment

If you are interested in providing feedback as we generate and test changes, please sign up

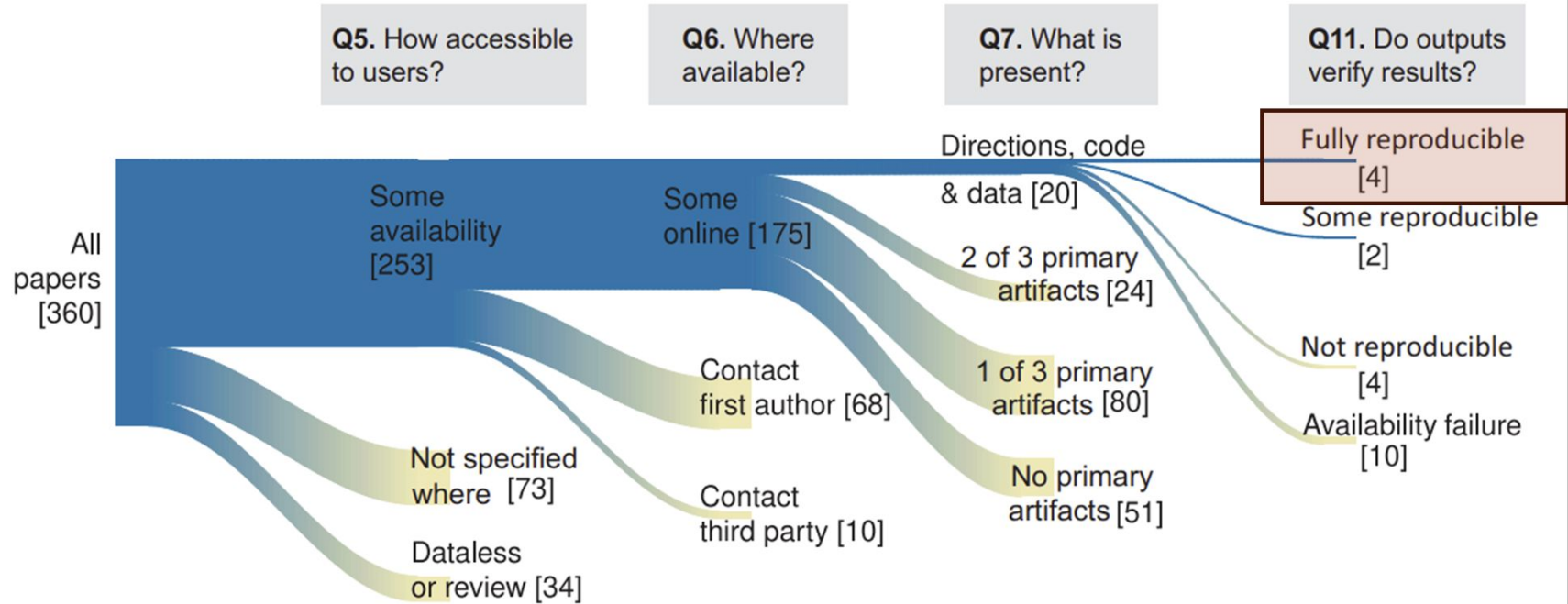
<https://tinyurl.com/RF24-signup>



Advanced Terrestrial Simulator (ATS) model data managing and archiving standards

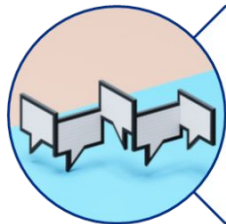
Ethan Coon, ORNL & Zhi Li, PNNL

The challenge of reproducibility: Only 1% data packages are fully reproducible

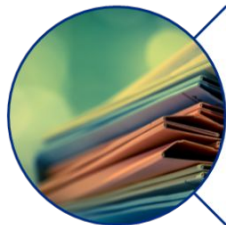


Project goals

ATS

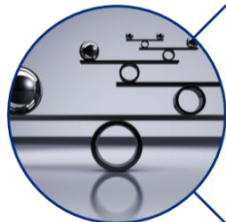


Working with both the ATS modeling community and ESS-DIVE, we aim to enhance, optimize, and standardize the existing model data archiving processes.



Develop documentation, demonstrations, and a short tutorial to train the ATS user base in improving their submissions to ESS-DIVE.

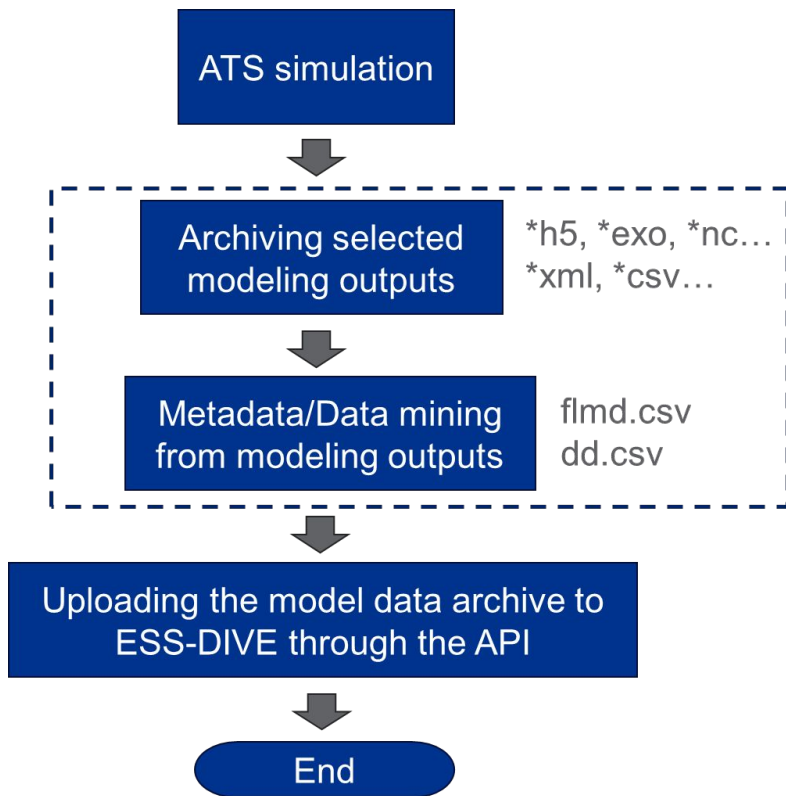
**and
other
models**



Provide a standard example for MDAs and their automation adaptable for use in other ESS codebases.

Simmonds et al (2022). Guidelines for publicly archiving terrestrial model data to enhance usability, intercomparison, and synthesis.

Prototype: Information mining from ATS outputs to meet the Reporting Format



ATS to ESS-DIVE workflow

This workflow is a step-by-step guide to archive ATS data associated with a manuscript on ESS-DIVE (<https://docs.ess-dive.lbl.gov/>). Purpose of this guide is to standardize and automate the data archiving process.

A few concepts:

- **Simulation Directory**: where ATS runs, typically on HPC
- **Data Package Directory**: where the data package is located, can be anywhere (HPC or local machine)
- **rsync**: a file transfer command line tool

	where to archive	format examples
Files not related to the manuscript	Gitlab	some files
Files related to the manuscript	ESS-DIVE	exo, xml, h5, csv

This command is to sync all h5 files in the simulation directory to the data package directory.

```
rsync -avzP --include=* --include='*.exo' --include='*.xml' --include='*.h5' --include='*.csv' --exclude=* --prune-empty-dirs ./<Simulation Directory>/ ./<Data Package Directory>/
```

This workflow creates the following files:

```
[57]: import subprocess
```

```
[58]: N = 20
write_new_csv = False
inplace = False
```

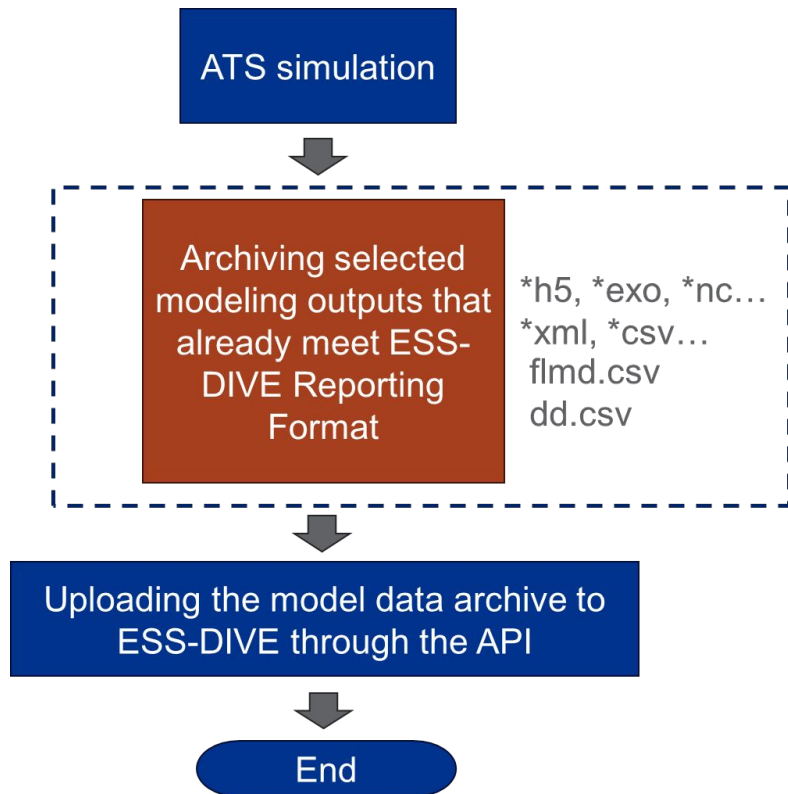
Find all files in this data package

```
[59]: paths_and_files = subprocess.check_output(['find -name "*"'], shell=True, encoding='utf-8').split('\n')
print(paths_and_files[1:N], f'...{len(paths_and_files)-N} more')

['./figures/swr_all.pdf', './figures/q_eval.pdf', './figures/naches_swr.png', './figures/setup.pdf',
'./figures/manning.pdf', './figures/bns.png', './figures/swr_fr_yr.pdf', './figures/flux.pdf', './figures/climate.pdf',
'./figures/et_eval.pdf', './figures/lai_et.pdf', './figures/naches_rain.pdf', './f
```



Proposed method: Generating Reporting Format-aware ATS outputs



ATS to ESS-DIVE workflow

This workflow is a step-by-step guide to archive ATS data associated with a manuscript on ESS-DIVE. It aims to improve ATS's ability to generate standardized MDAs, including,

- (1) Outputting ATS observation and visualization files in CF Standard Names, where to archive format examples

- (2) Implementing CSV-standard compliant ATS observation files, Files related to the manuscript ESS-DIVE exo, xml, h5, csv

- (3) Developing scripts for converting ATS's native output format to netCDF for better archival with ESS-DIVE standards, and,

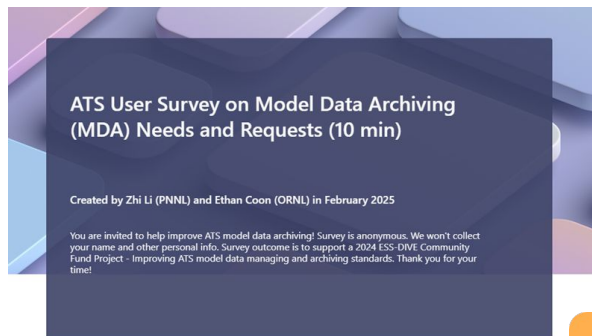
- (4) Developing scripts for setting up, automating, and submitting MDAs to ESS-DIVE, including the semi-automated generation of File Level Metadata, Data Dictionaries, etc.

```
[57] paths_and_files = subprocess.check_output(['find -name "*"'], shell=True, encoding='utf-8').split("\n")
[58] print(paths_and_files)
[59] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[60] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[61] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[62] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[63] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[64] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[65] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[66] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[67] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[68] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[69] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[70] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[71] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[72] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[73] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[74] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[75] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[76] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[77] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[78] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[79] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[80] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[81] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[82] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[83] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[84] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[85] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[86] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[87] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[88] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[89] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[90] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[91] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[92] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[93] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[94] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[95] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[96] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[97] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[98] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
[99] paths_and_files = [paths_and_files[i] for i in range(1, len(paths_and_files) if paths_and_files[i] != '')]
```


Deliverables

- Survey the ATS user group for needs in MDA and discuss with ESS-DIVE staff to identify best practices
- Improve ATS's ability to generate standardized MDAs
- Create Tutorial/Best Practices/How-To documents and videos

ATS User Survey on MDA Needs and Requests



SCAN FOR SURVEY

2. Do you regard yourself as a model data producer or consumer? *

For example, half producer half consumer=5

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

pure producer

pure consumer

3. How many years of experience do you have on ATS? *

If 10+, select 10

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

beginner

guru

13. What specific difficulties do you typically have to archive ATS data on ESS-DIVE? (Select all that apply, or input your own) *

File format (e.g., I'm not clear about

☐ how to effectively archive human-unreadable files)

File size (e.g., I don't know how to

☐ narrow down the files to archive out of the 3 TB ATS outputs I have)