How to Use ESS-DIVE Reporting Formats

Rob Crystal-Ornelas
ESS-DIVE Postdoc
Earth and Environmental Sciences Area
Berkeley Lab

Presentation Overview

- An introduction to **data reporting formats**
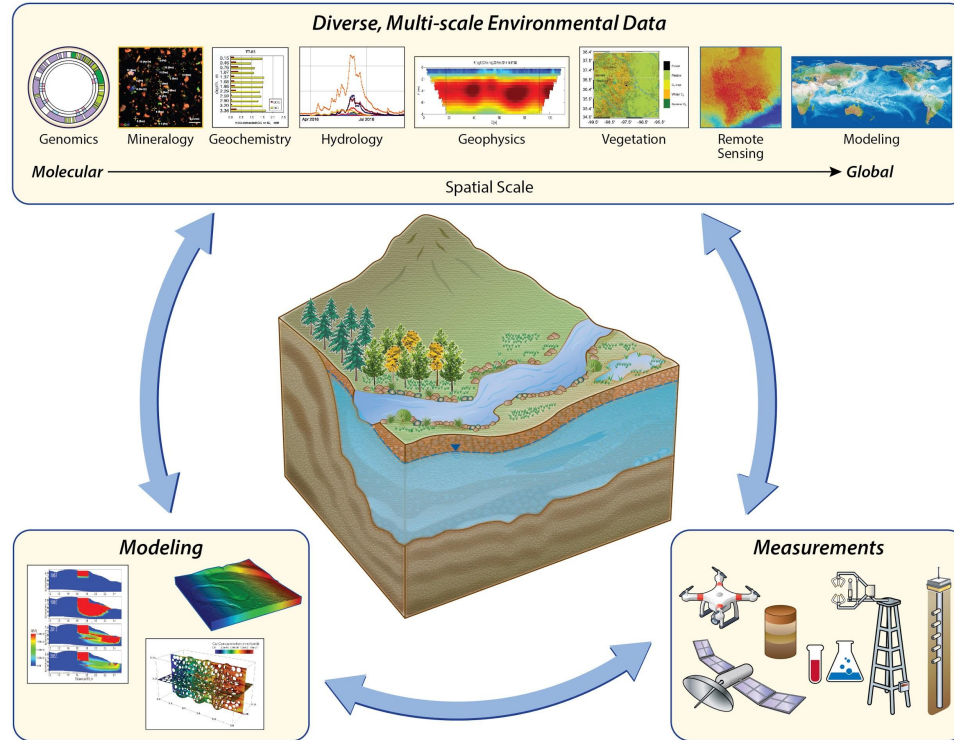
- ESS-DIVE's Community-led reporting formats

Takeaways from this presentation

- Data reporting formats can enable **finding** and **reusing** data

- Choosing the reporting format for **your data**

Reporting formats can increase data reusability & many ways to learn more

Data reporting formats can enable finding and reusing data

ESS-DIVE: A DOE cross-disciplinary **data repository** that stores highly diverse data

Varadharajan et al. (2019)

Why is it important to use a consistent format?

Year	City	Amount
1990	New York City	$1,123,456.00

(image from Dougherty & Ilyankou 2021)

Why is it important to use a consistent format?

(image from Dougherty & Ilyankou 2021)

The costs of non-reusable data

Researchers spend 80% of their time cleaning data -- not much left over for analysis (from: *Cost of not having FAIR data)*

Image: OCLC.org

The costs of non-reusable data

Researchers spend 80% of their time cleaning data -- not much left over for analysis (from: *Cost of not having FAIR data)*

But also,

"Science builds from our data" - Prof. Atul Butte

- Discovery from re-use (meta-analysis)
- Improving research methods
- Correcting/updating scientific literature & More

Image: OCLC.org

Formatted data are ready for reuse

year	city	amount
1990	New_York_City	1123456.00
1995	New_York_City	2200000.00
2005	New_York_City	-9999
2020	New_York_City	5000000.00

ESS-DIVE Community Data Reporting Formats

Clarifying terminology: Data standards and reporting formats

- Data Standards - Decades of development, accredited by governing org.

- Reporting Formats - Community-driven still enable data harmonization and synthesis

ESS-DIVE Community-led reporting formats for many types of data

Package Level Metadata/JSON-LD
Agarwal, Hendrix (LBNL)

Package Metadata Quality
Damerow (LBNL)

Sample IDs/Metadata
Damerow (LBNL)

Model Data Archiving
Simmonds (LBNL)

File-Level/csv Metadata
Velliquette, Heinz, Devarakonda (ORNL)

Soil Respiration
Bond-Lamberty, Pennington (PNNL)

Leaf Physiology
Rogers, Ely (BNL)

Hydrologic Monitoring
Goldman (PNNL)

Water/Soil Chemistry
Boye (SLAC)

Amplicon Sequencing
Weisenhorn (ANL)

Community Consensus Process

1. Background Research: Identify existing standards and best practices
2. Determine how standards fit into community needs
3. Propose draft of standard
4. Gather feedback
5. Revise, Test, Revise
6. Publish first version of standard
7. Implement standard in the repository, data submission and review process
8. Create tools, documentation, and training to help community adopt the standard

13

How can data reporting formats enable science?

What's consistent across the reporting formats

Data harmonization

- ISO 8601 Format for date/time

dateTime
YYYY-MM-DD HH:MM:SS
2021-05-04 17:10:02
2021-05-04

What's consistent across the reporting formats

Data harmonization

- ISO Format for date/time
- Latitude and Longitude

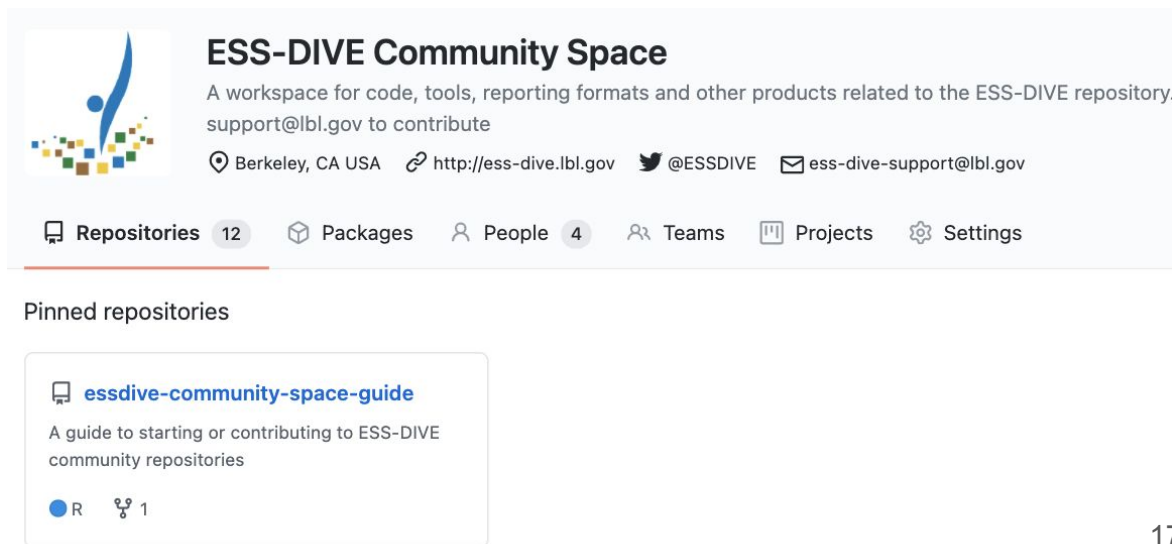| latitude	longitude
WGS84 decimal format	WGS84 decimal format

What's consistent across the reporting formats

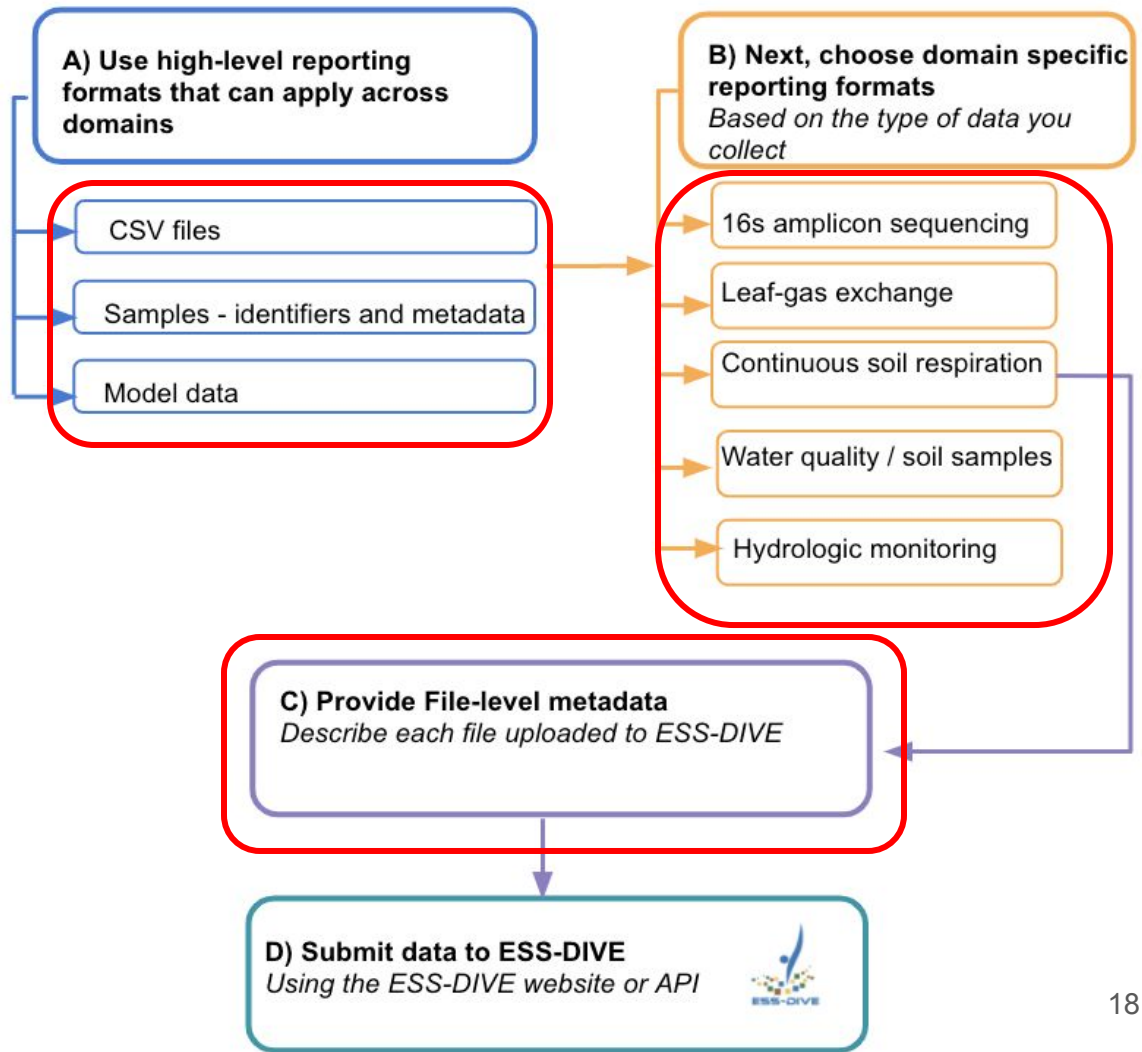*Where you find information about Reporting Formats*

- GitHub Community Space (github.com/ess-dive-community)
- GitBook Websites

ESS-DIVE Community Space

A workspace for code, tools, reporting formats and other products related to the ESS-DIVE repository. support@lbl.gov to contribute

⊙ Berkeley, CA USA   🔗 http://ess-dive.lbl.gov   🐦 @ESSDIVE   ✉ ess-dive-support@lbl.gov

📖 Repositories 12   📦 Packages   👤 People 4   👥 Teams   🗂 Projects   ⚙ Settings

Pinned repositories

📖 **essdive-community-space-guide**

A guide to starting or contributing to ESS-DIVE community repositories

🔵 R   ⑂ 1

Choose the reporting formats the fit your data

A) Use high-level reporting formats that can apply across domains

- CSV files
- Samples - identifiers and metadata
- Model data

B) Next, choose domain specific reporting formats
Based on the type of data you collect

- 16s amplicon sequencing
- Leaf-gas exchange
- Continuous soil respiration
- Water quality / soil samples
- Hydrologic monitoring

C) Provide File-level metadata
Describe each file uploaded to ESS-DIVE

D) Submit data to ESS-DIVE
Using the ESS-DIVE website or API

Summary

- **Community-led** reporting formats for your data

- Can enable data **findability** and **reuse**

- **Consistency** across reporting formats

- Support: ess-dive-support@lbl.gov

References

Dougherty, J and Ilyankou, I. (2021). Hands on Data Visualization. https://handsondataviz.org/