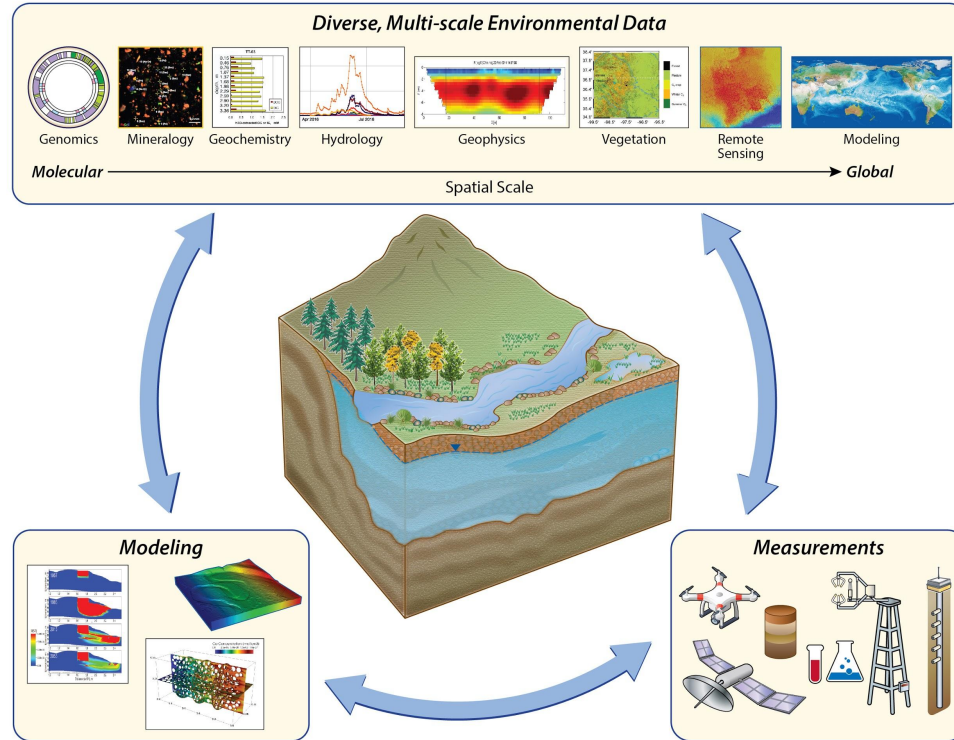# How to Use ESS-DIVE Reporting Formats

Rob Crystal-Ornelas
ESS-DIVE Postdoc

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?

Year	City	Amount
1990	New York City	$1,123,456.00
1995-96		2.2 mil
2000s	NYC	No data
2020	New_York	5000000+

(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)*

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

How can data reporting formats enable science?

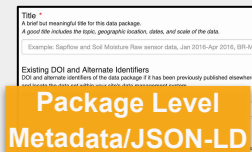# Clarifying terminology: Data standards and reporting formats

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

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

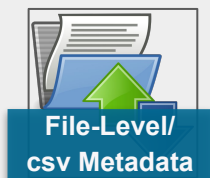# ESS-DIVE Community-led reporting formats for many types of data

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

Package Metadata Quality
Damerow (LBNL)

Sample IDs/Metadata
Damerow (LBNL)

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

Soil Respiration
Bond-Lamberty, Pennington (PNNL)
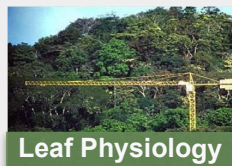
Leaf Physiology
Rogers, Ely (BNL)

Hydrologic Monitoring
Goldman (PNNL)

Water/Soil Chemistry
Boye (SLAC)

Amplicon Sequencing
Weisenhorn (ANL)

- Review of **existing formats**

- **Rounds of feedback** with researchers and projects

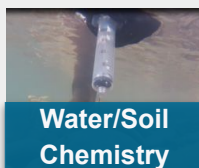- All documentation on our **GitHub Community Space**
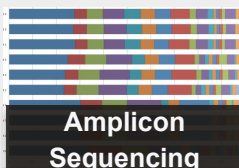
What's consistent across the reporting formats

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

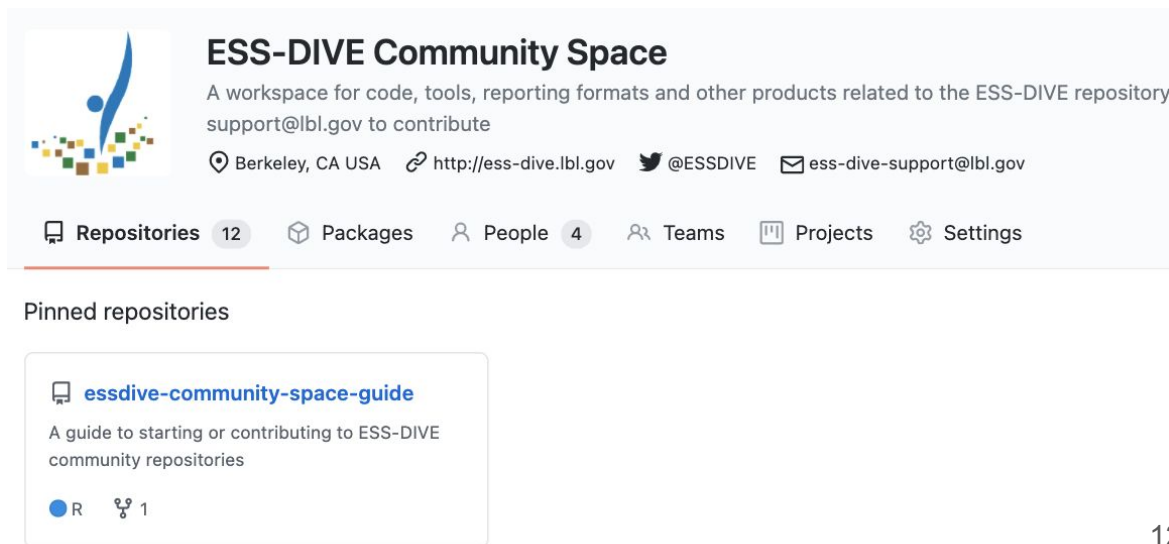*Data harmonization*

- 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

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

ESS-DIVE Glossary

- **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** - 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.

References

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