

# ESS-DIVE: A DOE Earth and Environmental Data Repository



<http://ess-dive.lbl.gov>

Deb Agarwal, Charuleka Varadharajan, Shreyas Cholia, Cory Snavely, Valerie Hendrix, Joan Damerow, Fianna O'Brien, Hesham Elbashandy, Makayla Shepard, Bill Riley, Madison Burrus, Zarine Kakalia, Emily Robles, Maegen Simmonds, and Karen Whitenack



U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science





# The ESS-DIVE Team



Data Scientists and  
Software engineers



Digital Librarians



Environmental  
Scientists

Coordinator: Karen Whitenack





# ESS-DIVE: a repository serving ESS data

- Key goals

1. Broad partnership with the ESS community on repository capabilities and features
2. Robust repository that provides long-term curation of ESS data
3. Data that is Findable, Accessible, Interoperable, and Reusable

Screenshot of the ESS-DIVE website interface showing datasets and a map.

The top navigation bar includes the ESS-DIVE logo, a search bar, and links for DATA, SUPPORT, ABOUT, Submit Data, and Sign in with Orcid.

The main content area displays "DATASETS 1 TO 7 OF 7" sorted by "Most recent".

Dataset 1: Angell J; Korshover K; Planet W (1991): [Annual and Seasonal Global Variation in Total Ozone and Layer-Mean Ozone, 1958-1987 \(1991\)](#). Carbon Dioxide Information Analysis Center (CDIAC), Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States). doi:10.3334/CDIAC/ATG.NDP023

Dataset 2: Sterin A (2001): [Tropospheric and Lower Stratospheric Temperature Anomalies Based on Global Radiosonde Network Data \(1958 - 2005\)](#). Carbon Dioxide Information Analysis Center (CDIAC), Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States). doi:10.3334/CDIAC/CL1.004

Dataset 3: Lamb H (1985): [Volcanic Loading: The Dust Veil Index \(1985\) \(NDP-013\)](#). Carbon Dioxide Information Analysis Center (CDIAC), Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States). doi:10.3334/CDIAC/ATG.NDP013

Dataset 4: Andres R; Boden T; Marland G (2009): [Annual Fossil-Fuel CO<sub>2</sub> Emissions: Isomass of Emissions Gridded by One Degree Latitude by One Degree Longitude](#). Carbon Dioxide Information Analysis Center (CDIAC), Oak Ridge National Laboratory (ORNL)

A map on the right shows the Northern Hemisphere with a focus on North America, displaying data overlays for the datasets listed.

# Community Engagement Activities

- Gathered ESS project requirements through: **Site visits, monthly webinars, advisory groups**
- Data and metadata **standards** development in partnership with community
- Demos and **tutorials**
- Representing ESS data activities in conferences/meetings

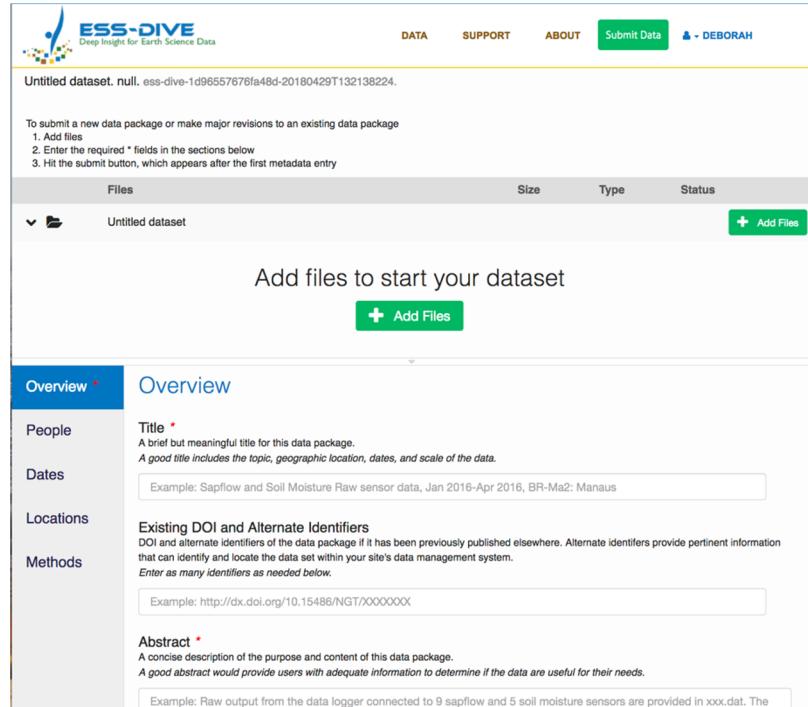
*Over 60 national and international meetings plus site visits to major ESS projects (3 years)*



# Submitting data: web portal and programmatic

- **Upload** data and metadata to create a **data package** via web page or programmatic interface (API)
- Release data packages publicly with a digital object identifier (**DOI**)
- Data usage licensing **CCby4** and **CC0**
- Preview features on <http://data-sandbox.ess-dive.lbl.gov>

**377** data packages published (3 years)



The screenshot shows the ESS-DIVE web portal interface for creating a new data package. At the top, there's a navigation bar with links for DATA, SUPPORT, ABOUT, and a green 'Submit Data' button. To the right of the button is a user profile icon for 'DEBORAH'. Below the navigation is a header for an 'Untitled dataset' with the identifier 'null. ess-dive-1d96557678fa48d-20180429T132138224'. A sub-header instructs users to submit a new data package or make revisions, listing steps: 1. Add files, 2. Enter required \* fields, 3. Hit the submit button. The main area has a table for managing files, showing one entry for 'Untitled dataset'. Below the table is a large text input field with the placeholder 'Add files to start your dataset' and a green '+ Add Files' button. On the left, there are tabs for 'Overview' (which is active), 'People', 'Dates', 'Locations', and 'Methods'. The 'Overview' tab contains sections for 'Title' (with a note about including topic, location, dates, and scale), 'Existing DOI and Alternate Identifiers' (with a note about providing alternate identifiers), and 'Abstract' (with a note about providing adequate information for usefulness). Example entries are provided for each section.

# Data Package Quality Review

## Metadata Quality Report

After running your metadata against our standard set of metadata, data, and congruency checks, we have found the following potential issues. Please assist us in improving the reusability of your research data by addressing the issues below.



► Passed 12 checks out of 14 (informational checks not included).

► Warning for 1 check. Please review these warnings.

▼ Failed 1 check. Please correct these issues.

☒ 1 of 2 method step descriptions have fewer than the minimum recommended 7 words



interpretation REQUIRED FAILURE

### ESS-DIVE Manual Package Metadata Review

\* Required

#### Abstract

Abstract should include a de  
should be understandable to  
scientific context, contain no  
needed to access attached f

Length of 100 words or mor

#### Abstract Content \*

PASS

FAIL

[Back](#)

[Next](#)

### ESS-DIVE Manual Package Metadata Review

#### FAIL: Abstract

Please copy and paste the standard response below in your reply to customer.  
If an alternative response was used, enter it in the "Other" option.

#### Response: Abstract

Revise your abstract to include a clear and concise description of the purpose and content of the data in your data package. Similar to a journal article, your data package abstract should be written for people who have not seen your related manuscripts.

Revise your abstract to meet the minimum length of 100 words.

Revise your abstract to include definitions of any acronyms used so it can be understood by anyone.

Revise your abstract so that it is written in complete sentences

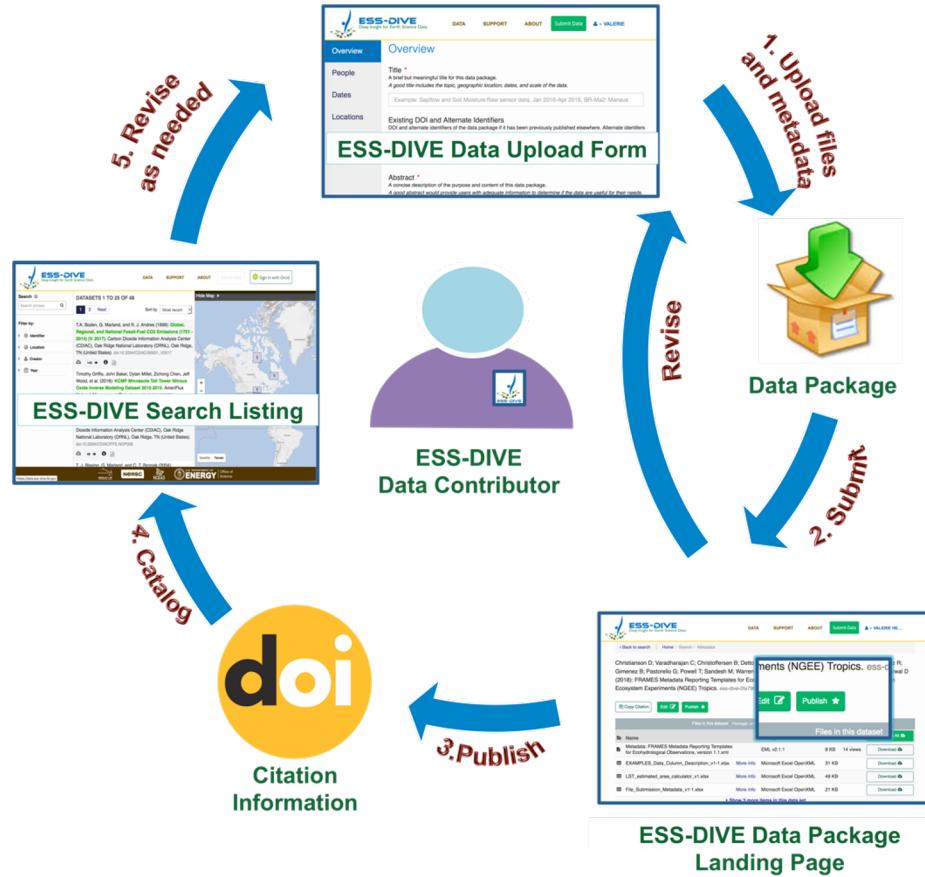
Revise your abstract to include specific details such as mention of source data for any synthesis work, software needed to view the files, clear location descriptions (avoid project-specific name with no explanation), ecosystem type involved, measurement type, etc

Other:

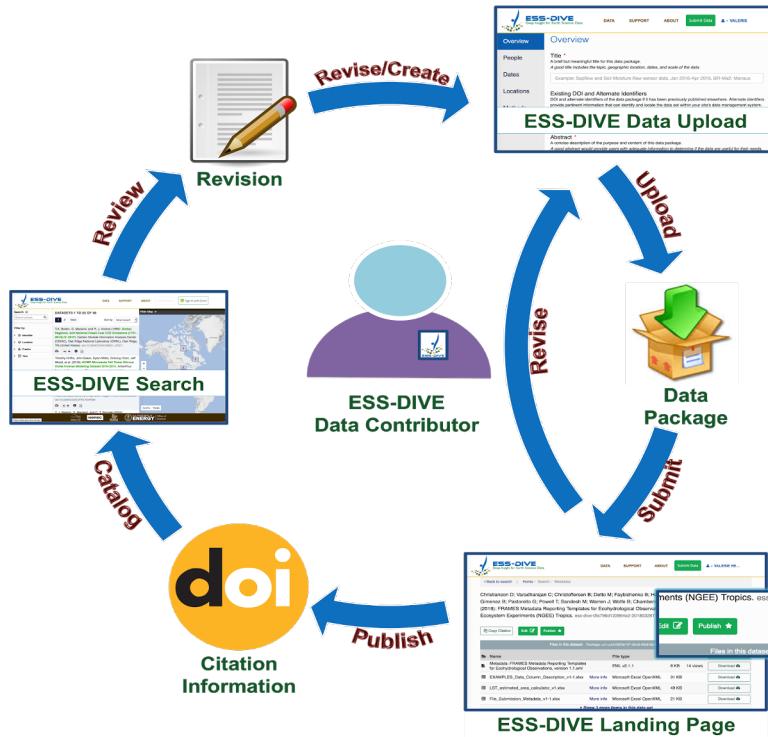
[Back](#)

[Next](#)

# Publication Cycle

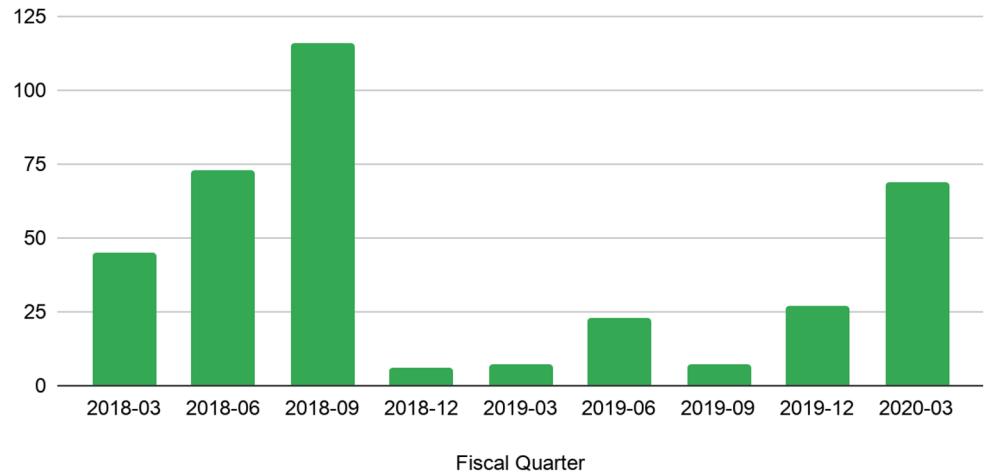


# Data Publications



## Newly Published Data Packages

Counts by Fiscal Quarter



# Prototype Data Fusion Database - Providing integrated search across data packages

Mongo Express Database: ess-dive > Collection: data\_package\_attribute

Viewing Collection: data\_package\_attribute

New Document New Index

Simple Advanced

name NO3 Regex Find

Delete all 14 documents retrieved

First Prev Next Last

data_file	data_package_id	name	count	values
WHONDRS_48hr_WS1_Geochem.csv	doi:10.15485/1509695	71851_NO3_mg_per_L_as_N	34	Below_Range,0.045,0.034,0.037
WHONDRDS_48hr_Alternahia_Geochemistry.csv	doi:10.15485/1577263	71851_NO3_mg_per_L_as_NO3	35	0.94,1.02,8.43,0.63,0.59,0.69,NaN,1
WHONDRDS_48hr_EastFork_Geochemistry.csv	doi:10.15485/1577278	71851_NO3_mg_per_L_as_NO3	34	Below_Range,16.47,16.21,18.01,18
WHONDRDS_48hr_Columbia_Geochemistry.csv	doi:10.15485/1577265	71851_NO3_mg_per_L_as_NO3	34	10.02,12.48,13.08,15.48,7.77,4.18,1
WHONDRS_48hr_WS1_Geochem.csv	doi:10.15485/1509695	71851_NO3_mg_per_L_as_NO3	34	Below_Range,0.201,0.152,0.165
WHONDRDS_48hr_Nisqually_Geochemistry.csv	doi:10.15485/1576995	71851_NO3_mg_per_L_as_NO3	34	Below_Range,0.3
WHONDRDS_48hr_Jordan_Geochemistry.csv	doi:10.15485/1577266	71851_NO3_mg_per_L_as_NO3	34	Below_Range,0.12,0.11,0.17,0.13,5
WHONDRDS_48hr_Erpe_Geochemistry.csv	doi:10.15485/1577260	71851_NO3_mg_per_L_as_NO3	51	11.62,14.09,14.23,15.95,10.17,14,3
SPRUCE_S1_porewater_chemistry_depth_2011_2013_201...	doi:10.3334/CDIAC/SPRUCE.018	NO3	170	mg N/L,-9999,0,0.03,0.09,0.05,0.08
SPRUCE_S1_groundwater_chemistry_2013_20171024.csv	doi:10.3334/CDIAC/SPRUCE.018	NO3	57	mg N/L,0.09,0.13,0.02,0.01,0.04,0

- Purpose: make standardized data within data packages searchable
- Challenge: No standards yet adopted
- Prototype: reading the non-standardized comma separated value (csv) and Excel files
- Indexing everything found

# Custom Data Portals – available in sandbox

- Enable development of a topic or project specific view of ESS-DIVE
- Context web pages describing project or topic
- Custom data search
- Metrics specific to data
- Enables long-term view of project/topic including context
- Can be private or public

 Carbon Dioxide Information Analysis Center

About Fossil-Fuel CO<sub>2</sub> Emissions Vegetation Response to CO<sub>2</sub> Trace Gas Emissions Climate Atmospheric Trace Gases Land-Use & Ecosystems FAQs Data Metrics

Search

Search these datasets

DATASETS 1 TO 25 OF 226

Sort by Most recent

[1](#) [2](#) [3](#) ... [10](#) [Next](#)

**Blake D (2005): Methane, Nonmethane Hydrocarbons, Alkyl Nitrates, and Chlorinated Carbon Compounds including 3 Chlorofluorocarbons (CFC-11, CFC-12, and CFC-113) In Whole-air Samples (April 1979 — December 2012).** doi:10.3334/CDIAC/ATG.002



Prinn R; Weiss R; Arduini J; Arnold T; DeWitt H; Fraser P; Ganeshan A; Gasore J; Harth C; Hermansen O; Kim J; Krummel P; Li S; Loh Z; Lunder C; Malone M; Manning A; Miller B; Mitrrevski B; Muhle J; O'Doherty S; Park S; Reimann S; Rigby M; Salto T; Salameh P; Schmidt R; Simmonds P; Steele L; Vollmer M; Wang H J ( ; Yao B; Yokouchi Y; Young D; Zhou L (2018): **History of chemically and radiatively important atmospheric gases from the Advanced Global Atmospheric Gases Experiment (AGAGE)**. doi:10.3334/CDIAC/ATG.DB1001



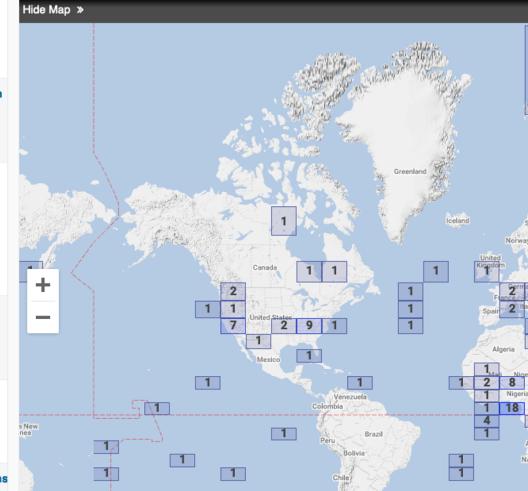
**Blasing T J; Broniak C; Marland G H (2004): Estimates of Monthly CO<sub>2</sub> Emissions and Associated 13C/12C Values from Fossil-Fuel Consumption in the U.S.A. (1981-2003).** doi:10.3334/CDIAC/FFE.001



Gibbs H K; Brown S (2007): **Geographical Distribution of Woody Biomass Carbon in Tropical Africa: An Updated Database for 2000 (NDP-055.2007, NDP-055b)).** doi:10.3334/CDIAC/ILUE.NDP055.2007



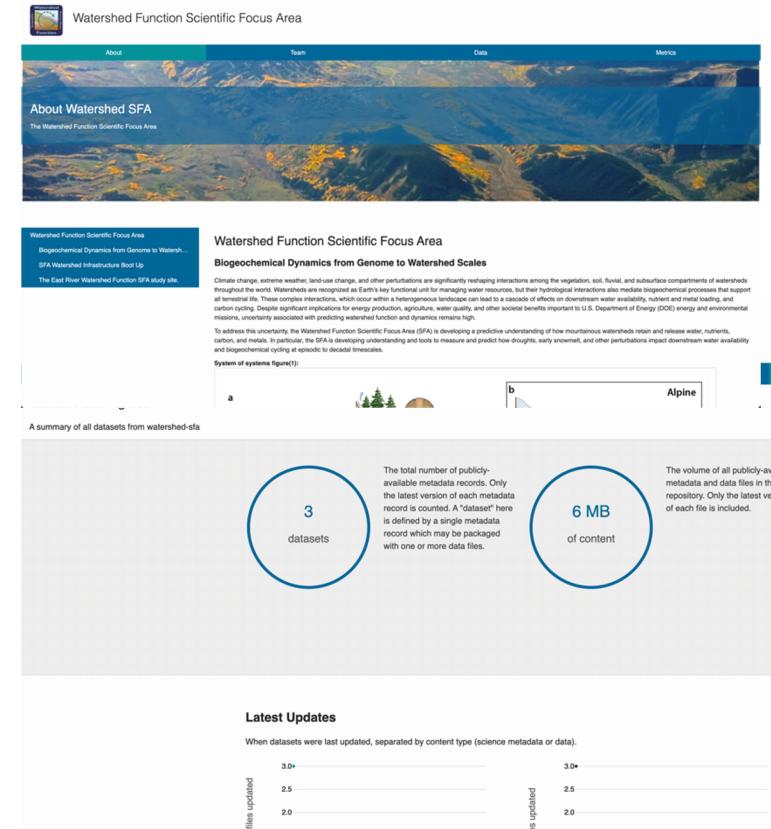
Menne M; Jr. C W; Vose R (2016): **Long-Term Daily and Monthly Climate Records from Stations Across the Continuous United States (U.S. Historical Climatology Network).** doi:10.3334/CDIAC/USHCN



# Upcoming Project Support Features:

**Project Spaces:** Project management and administration interfaces for use in managing data produced by projects.

- Enable project level curation and long-term maintenance
- Provide space for sharing data within project before it is made public
- Support data curation by project data manager
- Entry of data packages on behalf of data authors by project



The screenshot displays the Watershed Function Scientific Focus Area (SFA) interface, featuring a top navigation bar with tabs for About, Team, Data, and Metrics. Below the navigation is a large, scenic image of a mountainous landscape with autumn foliage. To the left, a sidebar provides an overview of the SFA's mission and study sites. The main content area includes a detailed description of biogeochemical dynamics, a systems diagram titled 'Systems of systems figure(1)', and a section titled 'Latest Updates' showing a line graph of file updates over time.

**Watershed Function Scientific Focus Area**

**About Watershed SFA**  
The Watershed Function Scientific Focus Area

**Watershed Function Scientific Focus Area**  
Biogeochemical Dynamics from Genome to Watershed...  
SFA Watershed Infrastructure Boot Up  
The East River Watershed SFA study site.

**Watershed Function Scientific Focus Area**  
Biogeochemical Dynamics from Genome to Watershed Scales

Climatic change, extreme weather, land-use change, and other perturbations are significantly reducing interactions among the vegetation, soil, fluvial, and subsurface compartments of watersheds throughout the world. These complex interactions, which occur within a heterogeneous landscape can lead to a cascade of effects on downstream water availability, nutrient and metal loading, and carbon cycling. This significant loss of system integrity has important implications for energy, water, and other societal benefits important to U.S. Department of Energy (DOE) energy and environmental mission, particularly associated with predicting watershed function and dynamics resulting from these changes.

To address this uncertainty, the Watershed Function Scientific Focus Area (SFA) is developing a predictive understanding of how mountainous watersheds retain and release water, nutrients, carbon, and metals. In particular, the SFA is developing understanding and tools to measure and predict how droughts, early snowmelt, and other perturbations impact downstream water availability and biogeochemical cycling at episodic to decadal timescales.

**Systems of systems figure(1)**

a Alpine

b Alpine

A summary of all datasets from watershed-sfa

3 datasets

The total number of publicly-available metadata records. Only the latest version of each metadata record is counted. A "dataset" here is defined by a single metadata record which may be packaged with one or more data files.

6 MB of content

The volume of all publicly-available metadata and data files in this repository. Only the latest version of each file is included.

**Latest Updates**

When datasets were last updated, separated by content type (science metadata or data).

3.0+ 2.5 2.0 1.5 1.0 0.5 0.0

3.0+ 2.5 2.0 1.5 1.0 0.5 0.0

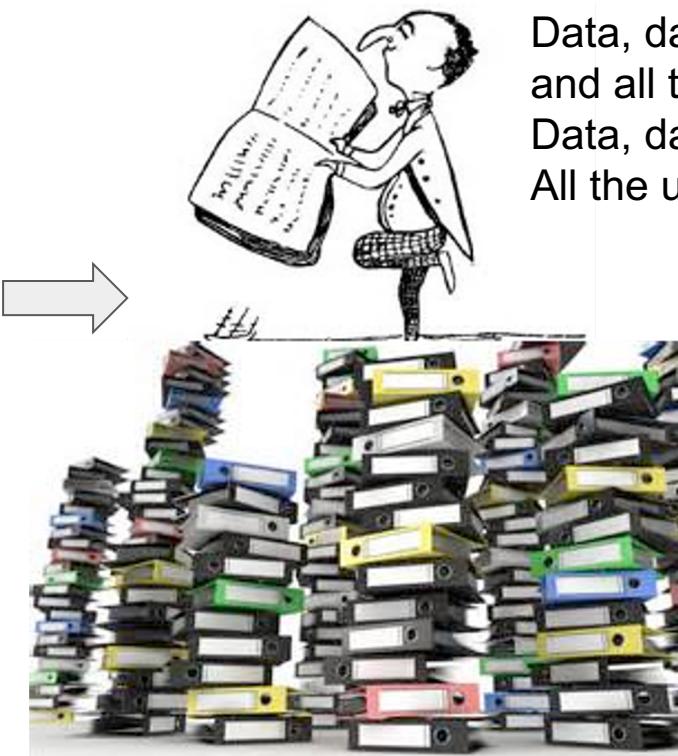
files updated files updated

# ESS-DIVE Long-Term Vision

## Connecting Users to Data



...  
Water, water, every where,  
And all the boards did shrink;  
Water, water, every where,  
Nor any drop to drink.  
...



Data, data everywhere  
and all the simulations drink  
Data, data everywhere  
All the users think

# Developing Standards

## ESS-DIVE Base Funds

- LBNL (Joan Damerow): **Sample metadata and identifiers**

## ESS-DIVE Community Funds

- ORNL (Ranjeet Devarakonda): **Comma separated value files & File-level metadata**
- PNNL (Amy Goldman): **Hydrologic monitoring** data
- SLAC (Kristin Boye): **Sample-based Water quality** data
- PNNL (Ben Bond-Lamberty): **Soil respiration** data
- BNL (Alistair Rogers): **Leaf physiology** data
- ANL (Pamela Weisenhorn): **16S Amplicon** data

# ESS-DIVE Collaboration Opportunities

- **Community funds:** What should be next?
- **Data Curation/Review:** Peer review?
- **Outreach:** Ambassadors?
- Other ideas?

# Questions?

Contact: [ess-dive-support@lbl.gov](mailto:ess-dive-support@lbl.gov)



U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science



# Persistent sample identifiers and documentation standards to support efficient sample management, interoperability, and data reuse in Environmental Systems Science

*Joan Damerow, Deb Agarwal, Kristin Boye, Eoin Brodie, Madison Burrus, Shreyas Cholia, Hesham Elbashandy, Ricardo Eloy Alves, Kim Ely, Amy E Goldman, Val Hendrix, Zarine Kakalia, Ken M Kemner, Annie B Kersting, Katharine Maher, Nancy Shiao-Lynn Merino, Fianna O'Brien, Zach Perzan, Emily Robles, Cory Snavely, Patrick Sorensen, James Stegen, Pamela Weisenhorn, Karen Whitenack, Mavrik Zavarin and Charuleka Varadharajan*



U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science



# Sample Naming and Tracking



## COMMUNITY NEED

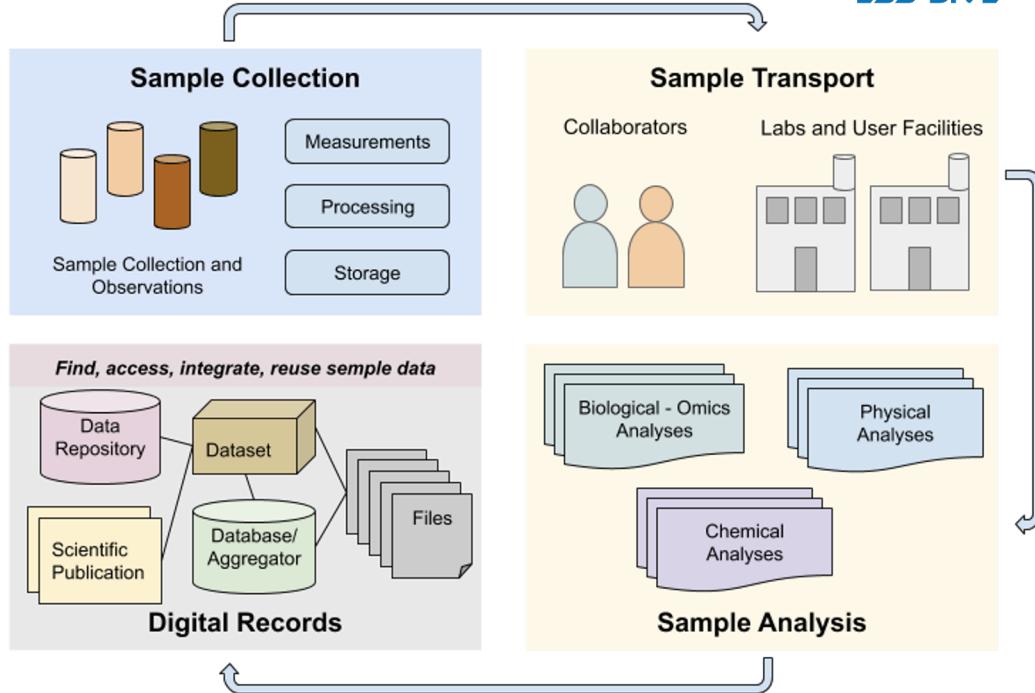
Practical, standardized sample ID/ tracking system as samples move to different physical and digital locations

## SOLUTION

Integrating and preserving sample data effectively in digital environment requires globally unique, persistent identifiers

International Geo/General Sample Numbers (IGSNs) for ESS samples

- Standardized core sample metadata, templates
- Sample landing page - metadata profile, link to related samples and publications



**IGSN: IECUR0002**

# Sample ID and Metadata Pilot Test Methods



Object Type:	Individual Sample	User Code:	SESAR					
		SYSTEM FOR EARTH SAMPLE REGISTRATION						
Sample Name	IGSN	Parent IGSN	Material	Sample description	Collection method	Latitude	Longitude	Location des

Field test of IGSN PID and metadata



Sample Metadata Research and Crosswalk

Biodiversity Information Standards TDWG

Publish data packages in ESS-DIVE



Map sample relationships and history



PID for sample tracking and publication

Incentive tools to support sample management and reuse.

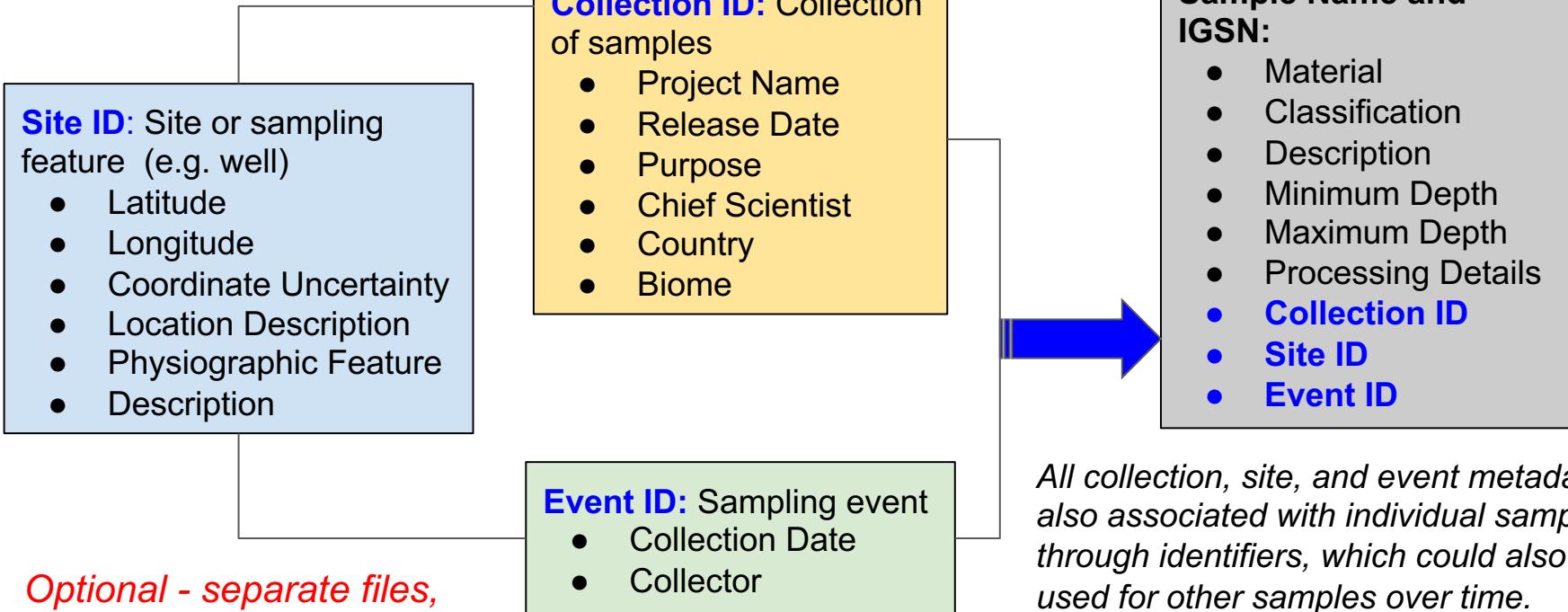


8 Projects, ~ 3,000 Registered IGSNs

Review proposed sample ID and metadata standard:

- [Crosswalk that compares existing sample metadata standards and templates](#)
- [Updated SESAR IGSN sample metadata guide](#) accounting for pilot test feedback
- Updated SESAR IGSN spreadsheet template for sample metadata

# Providing Metadata and Linking Samples at Different Levels



*Optional - separate files,  
terminology file mapping existing  
terms to standard, if needed*

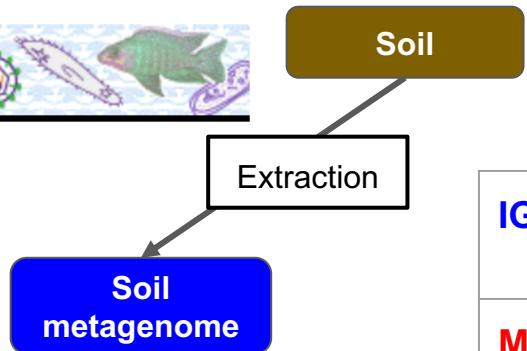
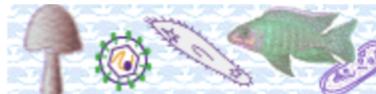
*All collection, site, and event metadata also associated with individual samples through identifiers, which could also be used for other samples over time.*

# Assigning and linking IDs to sets or chains of highly related samples and subsamples

		Relationship	1. Standard IGSNs	2. Extend primary IGSN	3. Collection or Site ID linked to IGSN
Soil Core	Collection of water samples at one place and time	FeatureOfInterest (Parent)	/GSN: IEBWE0001	/GSN: IEBWE0001	Collection ID: S19S_0006
Core Section	Varying depths	isSampleOf (Parent/child)	/GSN: IEBWE000P IEBWE000Q	extensions: IEBWE00011M IEBWE00012M	/GSN: IEBWE000P IEBWE000Q
Homogenized subsamples sent for analyses	Individual containers sent for analyses	isSampleOf (child)	/GSN: IEBWE000R IEBWE000S IEBWE000T ...	/GSN: IEBWE00012MA IEBWE00012MB IEBWE00012MC ...	extensions: IEBWE000PA IEBWE000PB IEBWE000PC ...

**\*\*Subsequent IDs assigned in external labs or user facilities are separate but always associated with the sample PID**

# Sample Interoperability with NMDC, KBase, JGI, EMSL



- Links between IDs: Source material ID and subsample/analysis ID
- IGSN metadata → MIxS/MIMS
- Use of standard ontologies - mapping ontologies to terms in use
- Use cases: [WHONDRS](#), LBNL TES SFA FICUS

IGSN	Source_material_id = IGSN Link
Material	env_medium* = ENVO; isolation_source
<i>populate from material</i>	organism* ( <a href="#">e.g. soil metagenome</a> )
Physiographic feature	Env_local_scale* = ENVO
	Env_broad_scale* = ENVO
Country; Locality	Geo_loc_name* = GAZ
	samp_mat_process

# Questions?

Contact: [ess-dive-support@lbl.gov](mailto:ess-dive-support@lbl.gov)



U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science



# Developing CSV and File-level Metadata Standards

ORNL and ESS-DIVE  
Community Fund Standards

May 11, 2020

Terri Velliquette <velliquettet@ornl.gov>  
Jessica Welch  
Michael Crow  
Ranjeet Devarakonda  
Susan Heinz

ORNL is managed by UT-Battelle, LLC for the US Department of Energy

ORNL Community Fund Project is to develop two standards:

- a generic standard for the structure and content requirements of data stored in a comma-separated values file referred to as a CSV file.
- a metadata schema to describe the contents, scope, and structure of each data file within the ESS-DIVE repository referred to as file-level metadata or FLMD.

These standards will promote:

- machine-readability for the efficient extraction of file-level discovery and usable metadata.
- interoperability to facilitate the development of a future fusion database.

# An Example of File-level Metadata (FLMD)

## FLMD can provide:

- granular information that can enable data users to compare data files within a data package
- the ability to search for and locate files across an overarching data collection

**DataONE**

About News Participate Resources Education Data

DATONE SEARCH: Search Summary Jump to: DOI or ID Go

< Back to search | Search / Metadata

Claire Addis. 2019. Plant trait measurements of species at snowfences around Toolik Lake, Alaska.

<https://search.dataone.org/view/doi:10.18739/A26Q1SH2V>

**Data Table**

Entity Name: Graminoids.csv  
Download

Description: Plant trait data for graminoid species at snowfences near Toolik Lake, Alaska, 2010-2011.

Object Name: Graminoids.csv

Online Distribution Info: <https://cn.dataone.org/cn/v2/resolve/urn:uuid:5815e244-4576-4707-a196-953c5b6986b0>

Size: 18570 bytes

Authentication: 4f05e737e2ca8abb50f899326fed038 Calculated By MD5

Externally Defined Format: Format Name application/vnd.ms-excel

**Attribute Information**

Variables	Name	Label	Definition	Storage Type	Measurement Type	Measurement Domain	Code	Definition	Source
Treatment	Treatment	Treatment	Treatment applied to plant						
species									
ramet									
M_wt									
D1_wt									
D2_wt									
D3_wt									
D4_wt									
D5_wt									
D6_wt									
M_LA									
D1_LA									
D2_LA									
D3_LA									
D4_LA									
D5_LA									
D6_LA									
M_num									
D1_num									
D2_num									
D3_num									
D4_num									
D5_num									
D6_num									

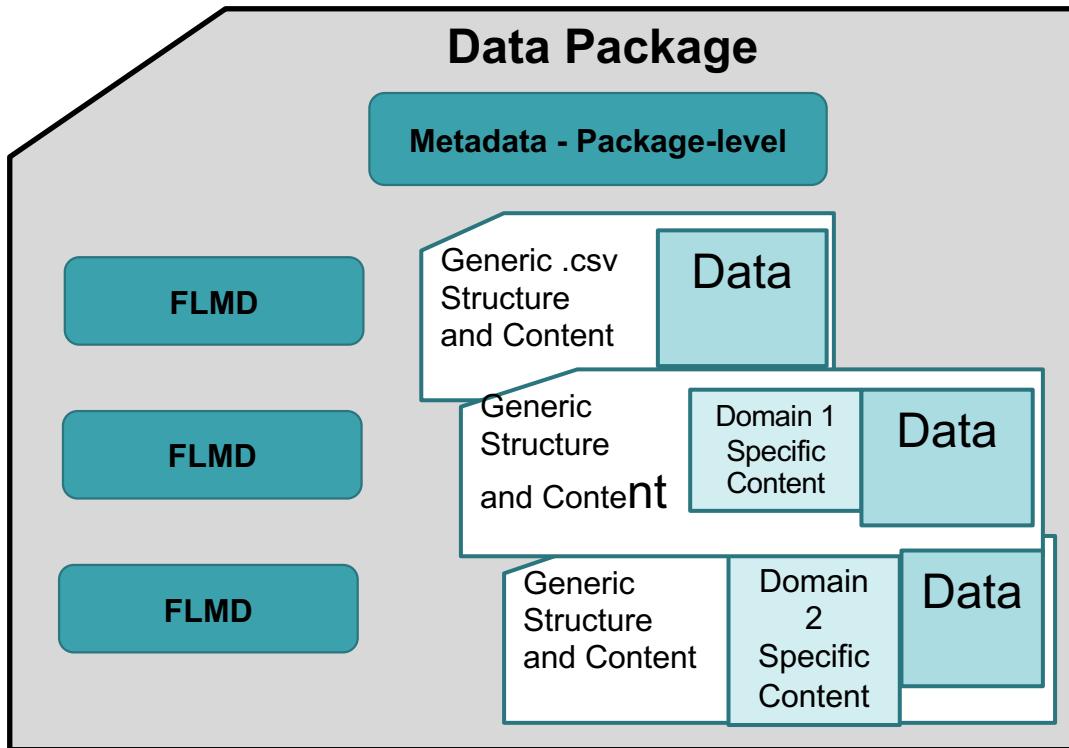
Missing Value Code: control plant collected from control side of snowfence  
snow plant collected from snow-drift side of snowfence

Accuracy Report: Coverage: Method:

Accuracy Assessment:

# CSV Standards and Format

- The generic structure and content requirements are the foundation for standardized and consistent data and metadata reporting. The goal is to have general data and domain-specific content and data fit within the larger common file structure.



## Next Steps

### Template:

- Develop example template for \*.csv file

### Format Check:

- Implement CSV file format verification script

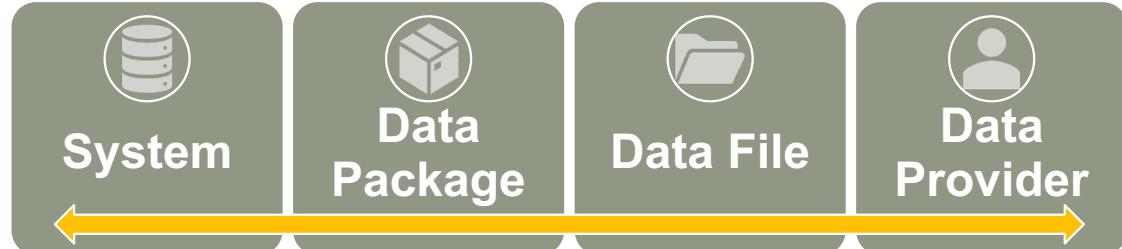
### File Level Metadata Extraction Tool:

- Implement a script that will extract FLMD from a standard \*.csv file and store the metadata for ESS-DIVE Search applications.

# Possible Sources for FLMD and Potential Uses of FLMD

## FLMD

- Can be populated from multiple sources (following standard formats)
- Can display with each data file on the Data Package Landing Page
- Can be searched
- Can travel with the data file if pulled into a fusion database





Pacific  
Northwest  
NATIONAL LABORATORY

# Hydrologic Monitoring Standards

May 11, 2020

**Amy Goldman, Huiying Ren, Josh Torgeson, Huifen Zhou**

ESS Cyberinfrastructure Working Group Annual Meeting



U.S. DEPARTMENT OF  
**ENERGY** **BATTELLE**

PNNL is operated by Battelle for the U.S. Department of Energy

# Hydrologic monitoring data often lack consistency and metadata

**Scope:** Sensors used in the field to measure parameters in water

**Parameters:** “Water level”, temperature, electrical and specific conductivity, pH, dissolved oxygen



## Challenges

- Rarely define terms
- Often not machine readable
- Rarely have information about instrumentation

Examples of ambiguous column headers	Possible questions from data users
Elevation	Surface water elevation? Riverbed elevation? Using what datum?
Level_m	Level of what? Relative to what?
Well depth	Depth to water from the well casing? Water elevation in the well?

Interested in taking a closer look and providing feedback? <https://tinyurl.com/HydroStandard>

# Three-file system captures sensor metadata and allows flexible terminology

Data File	Sensor Metadata File	Terminology File
<ul style="list-style-type: none"><li>*Time stamps</li><li>Measurements</li><li>Alphanumeric code linking each parameter at a specific time point to sensor metadata file</li><li>Alphanumeric data flag codes</li></ul>	<ul style="list-style-type: none"><li>Alphanumeric sensor codes from data file</li><li>Sensor construction (make, model, serial number, accuracy, range, calibration, QAQC, notes)</li><li>Deployment detail (lat/long, *water body type, context of sensor placement, description of deployment)</li></ul>	<ul style="list-style-type: none"><li>*Terms used in headers, data flags, missing value cells</li><li>*Definition of terms</li><li>*Mapping the used terms to the terms suggested by the standard</li></ul>

Required fields are marked with \*

Columns below are a subset of the standard's suggestions. Most are optional. Required columns are marked with \*

Data File					
* DateTime_UTC	cond	WaterTemp_deg.C	DataFlag	Sensor_cond	Sensor_WaterTemp
2020-05-14 13:30:00	252	15.2	DataFlag_01	S1_SpC	S3_T

Sensor Metadata File								
SensorID	SensorMake	SensorModel	SensorSN	SensorAccuracy	* Latitude _dec.deg	* Longitude _dec.deg	* WaterBodyType	SensorContext
S1_SpC	Campbell Scientific	CS547A	41593	+/- 5% of reading	43.3195	-119.2593	groundwater [ENVO:01001004]	Well

Terminology File		
* Term	* Definition	* MapToSuggested
cond	Specific conductivity in microSiemens per centimeter	SpC_uS.per.cm

Interested in taking a closer look and providing feedback? <https://tinyurl.com/HydroStandard>

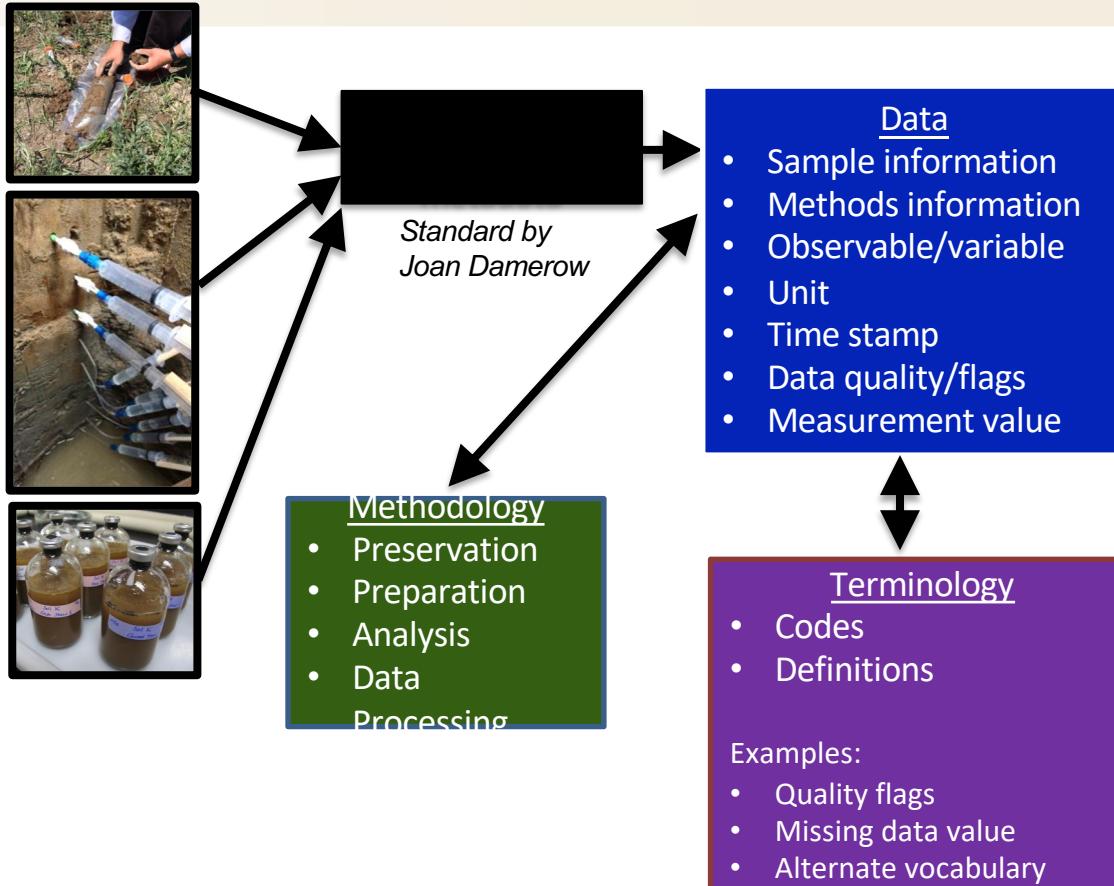


# Chemical concentrations in water, soil, and sediment samples

A community funded project for developing data and metadata standards and templates for ESS-DIVE archived data

*Kristin Boye, Associate Staff Scientist, SLAC*

# Relationships and structure overview



# Current Data File Template



Data is reported in columns with key metadata included as header rows

	A	B	C	D	E	F
1	columnHeaderRows_11	sampleID	sampleMaterial	notes	nitriteAndNitrateN	dataFlags_nitriteAndNitrateN
2	unit				milligramPerLiter	
3	unitBasis				asN	
4	analysisTimeStamp				2019-09-21 11:20	
5	preservationMethodID				store2C	
6	preparationMethodID				dilutionx2	
7	analysisMethodID				DA	
8	analysisDetectionLimit				0.04-5.0	
9	analysisPrecision				0.01	
10	dataProcessingMethodID				avg3	
11	dataSpecifier				undilutedFinal	
12	dataRows	IERVTL10Z	liquid>aqueous		2.65	
13		IERVTL18T	liquid>aqueous	only a single measurement was made for flagged measurements		
14		IERVTL19U	liquid>aqueous		1.34	dP
15						4.21
16						
17						
18						
19						
20						
21						
22						
23						
24						
25						
26						

# Current Methods and Terminology File Templates



Methods are linked through methodIDs  
(defined by user or registered IGSNs or other permanent identifier)

	methodID	methodType	methodTemp	methodLight	methodAtmosphere	methodMoisture	methodMedium	methodTime	methodDescription	methodReference	methodInstrument	methodLab
2	DA	analysisMethod	heated	ambient	air	NA			Nitrate was quantitatively reduced to nitrite by a cadmium column. Newly formed nitrite and any nitrite originally present in the sample was diazotized with sulfanilamide and subsequently coupled with N-(1-naphtyl)ethylenediamine dihydrochloride. The resulting colored azo dye was measured colorimetrically at 550 nm. The method is consistent with EPA 353.2, Rev. 2, average of 3 replicate measurements refrigerated at 2°C in the dark dilution 1:1 with milliQ water prior to analysis	<a href="https://www.epa.gov/sites/production/files/2015-08/documents/method_353-2_1993.pdf">https://www.epa.gov/sites/production/files/2015-08/documents/method_353-2_1993.pdf</a>	WestCo SmartChem 200 Discrete Analyzer	Environmental Measurements Facility (EM-1), Stanford University
3	avg3	dataProcessingMethod	NA	NA	NA	NA						
4	store2C	preservationMethod	2G	dark	air	NA						
5	dilutionx2	preparationMethod	ambient	ambient	air	NA	milliQ					
6												

Flags, codes, alternative vocabularies etc. are reported in a separate terminology file

term	definition
-9999	missing data value
dP	data flag indicating a deviating data processing method, see notes for specifications
presQ	data flag indicating a deviating preservation method resulting in questionable sample integrity, see notes for specifications

**PLEASE provide input!!!**



This is a living document and a standard in progress - it will be most **applicable**, **appropriate**, **efficient**, and **functional** if it is adopted and accepted, which means *adapted* to **YOUR** needs and practices!!!

PLEASE reach out to me with comments or to be part of the review team for this standard

**kboye@slac.stanford.edu**



# Continuous soil respiration and other chamber fluxes

May 11, 2020

Ben Bond-Lamberty and Stephanie Pennington



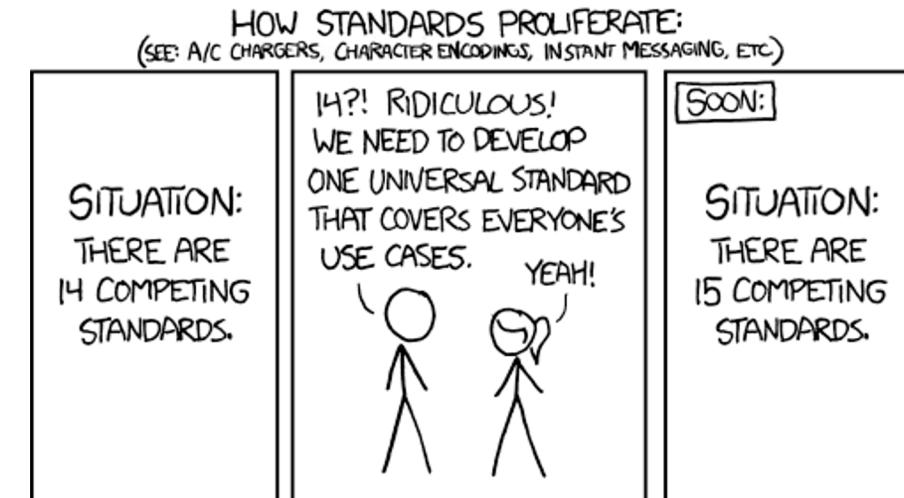
# Orphaned datasets limit science

- Large fluxes of greenhouse gases (GHGs) between soils and the atmosphere
- Continuous (automated) measurements increasingly common
- No standard or repository for them
- Potentially *huge* amount of data bearing on dynamics and response of earth system



# Philosophy, approach, standards

- “As simple as possible, but not simpler”
- Can’t be all things to all people
- Leverage historical experience
- Learn from recent ISRaD and SIDb databases
- Consultation with ICOS, FLUXNET, Ameriflux, manufacturers
- Community engagement and surveys



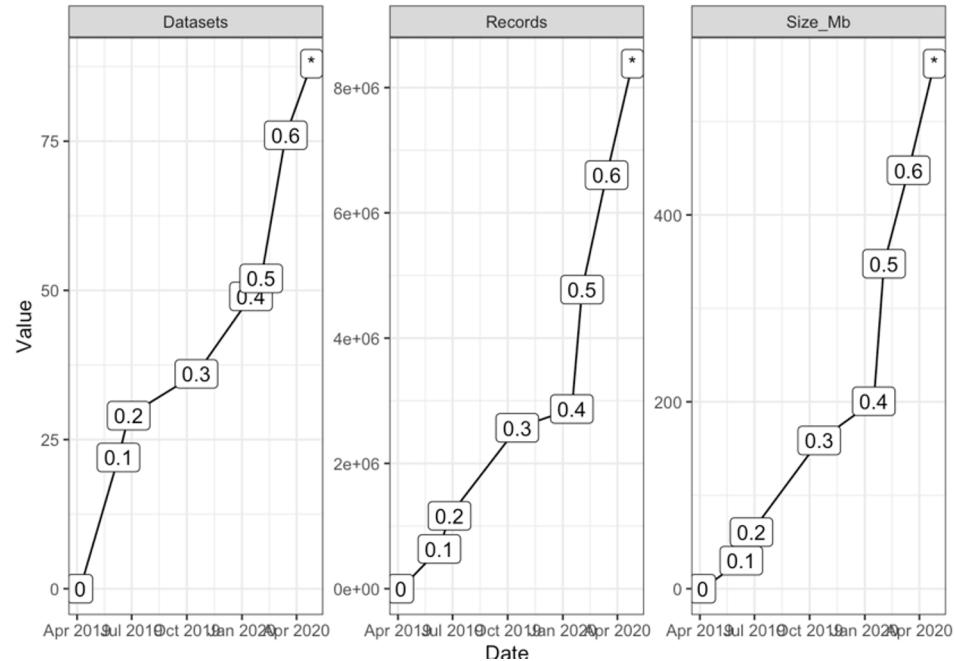
# Philosophy, approach, standards

- Location in *space* and *time*
- Measured flux(es)
- Provenance
- Contributors and reference
  
- Instrument, errors, ecosystem, errors, ancillary data...
  
- Each field added has a **cost** that has to be paid in perpetuity!

Field name	Description	Class	Units	Required
CSR_DATASET	Dataset name	character		*
CSR_SITE_NAME	Site name	character		*
CSR_LONGITUDE	Decimal longitude of site, positive = north	numeric	degrees	*
CSR_LATITUDE	Decimal latitude of site, positive = east	numeric	degrees	*
CSR_ELEVATION	Elevation of site	numeric	m	*
CSR_TIMEZONE	Timezone code, from <a href="https://en.wikipedia.org/wiki/List_of_tz_database_time_zones">https://en.wikipedia.org/wiki/List_of_tz_database_time_zones</a>	character		*
CSR_IBGP	IGBP class, from <a href="http://www.eomf.ou.edu/static/IGBP.pdf">http://www.eomf.ou.edu/static/IGBP.pdf</a>	character		*
CSR_NETWORK	Site network name	character		
CSR_SITE_ID	Site ID in network	character		
CSR_INSTRUMENT	Measurement instrument (i.e. model)	character		*

# Concurrent evolution

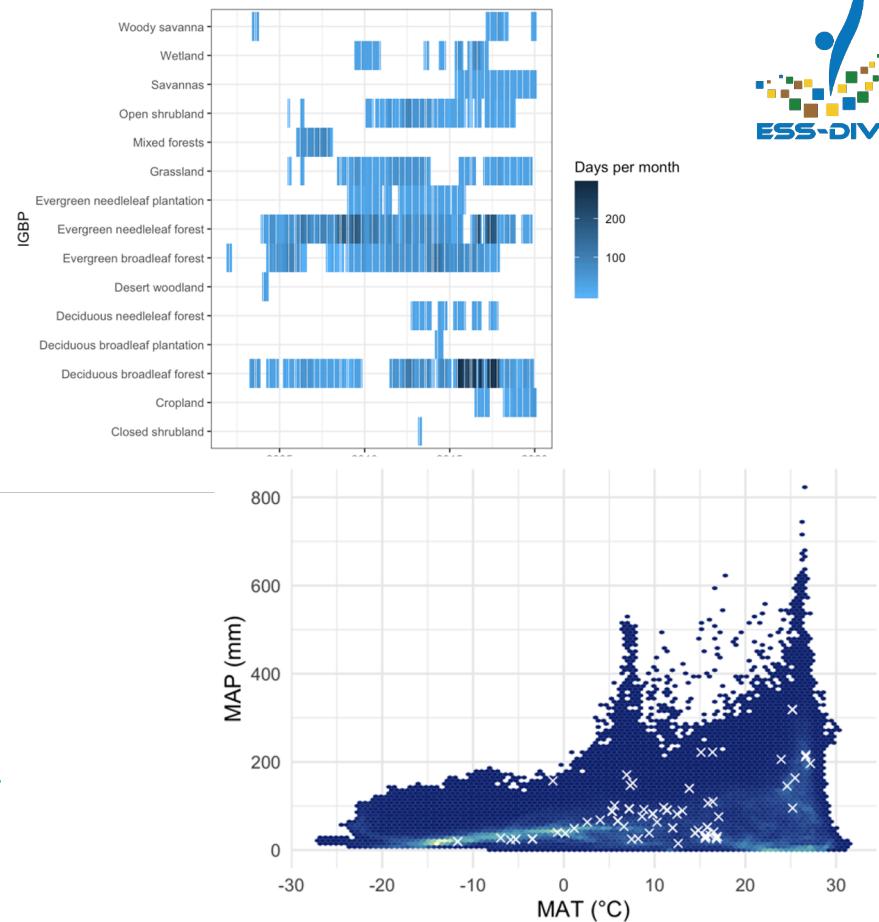
- Building *slowly* and *concurrently* :  
data and standard
- This has given real-world  
experience with kinds and  
formats of data
- ...and therefore the costs and  
benefits of different standards  
and formats
- With well-designed tools and  
software this process is worth it



# An increasingly mature data standard and database

- Data standard increasingly refined
- Concrete mapping to Ameriflux  
BADM
- *ESSD* manuscript in prep
- Data available now!
- <https://github.com/bpbond/cosore>

IGBP coverage



# ESS-DIVE data and metadata standards Leaf physiology

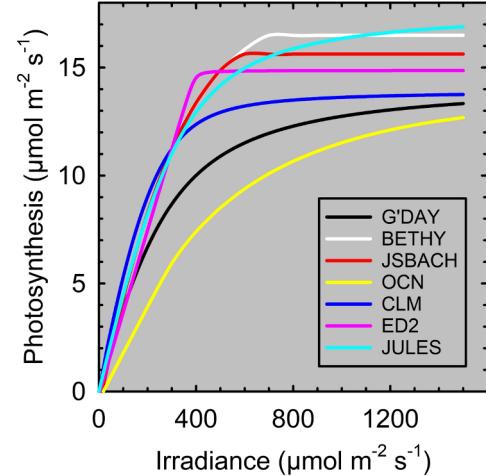
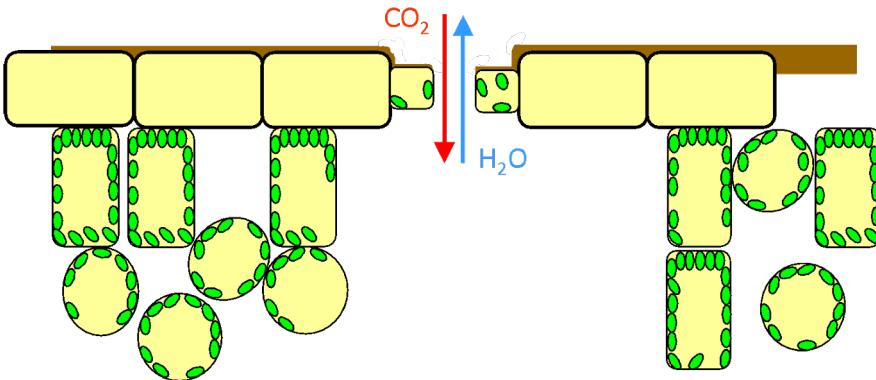
Kim Ely & Alistair Rogers



Barrow Environmental Observatory,  
AK

# About leaf physiology data

- Provides mechanistic understanding of plant and ecosystem fluxes of carbon and water
- Provides important parameterization for terrestrial biosphere models
- Necessary to understand the response of plants to global change



# Data challenges

Specialist and time consuming data collection, generally focused on limited species and geographic areas

Standards improve data accessibility

Re-use enhances data value

Model

Parameterization

Process

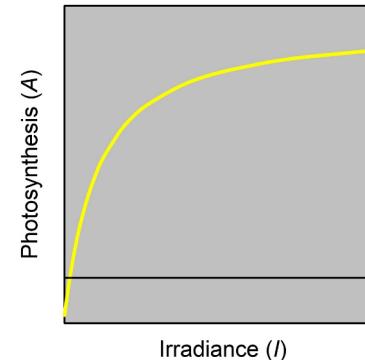
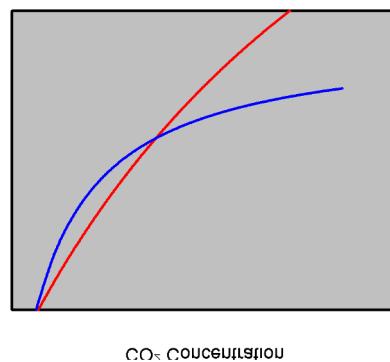
Knowledge

Global

Synthesis

# Data standard scope

- Measurements made on leaves with portable gas exchange instruments
- Photosynthesis, respiration, and stomatal conductance
- Response to changing temperature, CO<sub>2</sub>, and light levels
- Derived parameters used in models, e.g., V<sub>c,max</sub>, J<sub>max</sub>



# Data standard components

- General metadata, as prescribed by file-level and sample metadata standards
- Detailed metadata for methods and experimental parameters
- Standardization of required data variables: names, definitions & units
- Required variables for different measurement types
- Inclusion of raw data to future-proof data collection effort

Metadata ▾

Methods ▾

Variable definitions ▾

Required variables ▾

Instrument output translation ▾

# Current status and plan

- Research of existing standards complete; draft standard prepared
- Next, review with BNL and ESS-DIVE CF
- Then, invite broad community review of v2.0
- Implement resulting standards
- Publish description of standard; manuscript in preparation
- Contributors will be invited to co-author

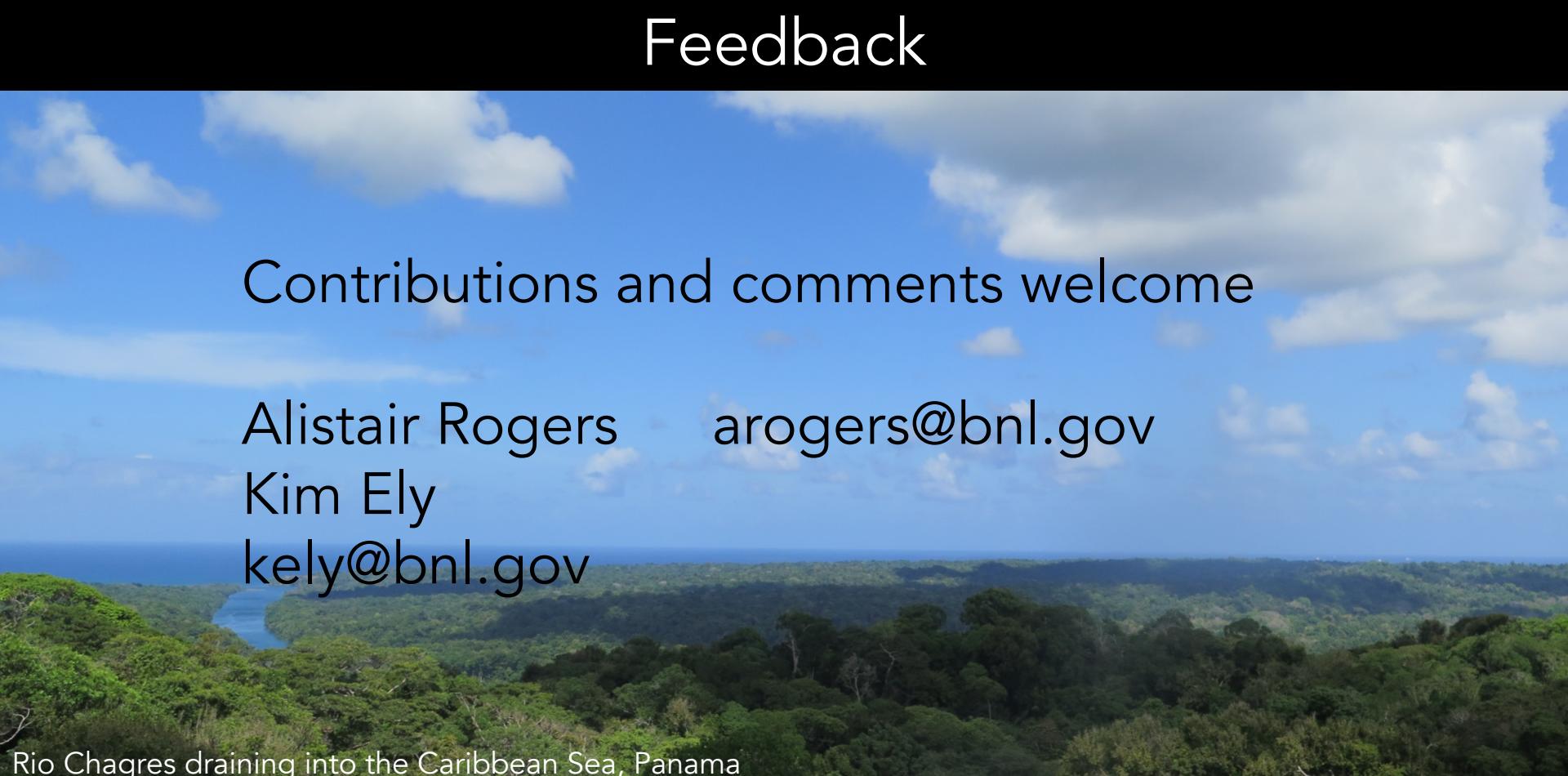


# Feedback

Contributions and comments welcome

Alistair Rogers      [arogers@bnl.gov](mailto:arogers@bnl.gov)

Kim Ely  
[kely@bnl.gov](mailto:kely@bnl.gov)



Rio Chagres draining into the Caribbean Sea, Panama

# 16S Amplicon Data and Workflows

May 11, 2020

Pamela Weisenhorn



**ESS-DIVE**  
Deep Insight for Earth Science Data

## 16S Amplicon data

16S data is widely used to examine microbial community composition (and sometimes to predict function)

Although valuable, there are many difficulties in comparing data across studies or repurposing data for other uses

Data standard needs to capture both the sample extraction and preparation as well as bioinformatic processing to ensure reproducibility and support data re-use

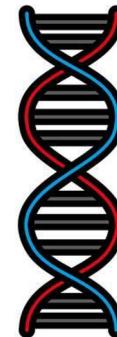
# Data standard

Two common formats

.biom and .csv

Improvements to existing approaches:

- Capture of consensus sequence information
- Key metadata related to extraction and sample preparation
- Key metadata related to bioinformatic processing





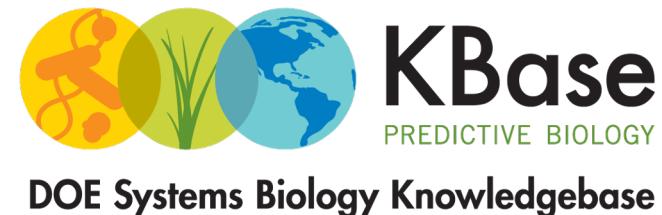
# Coordination with KBase

KBase is actively developing analysis and visualization tools for amplicon data

KBase importers will require amplicon data that adheres to the ESS-DIVE data standard in content and structure

KBase will allow export of data in ESS-DIVE compatible format

KBase is also developing templates to support the import of related sample metadata for analysis



# Questions?

Contact: [ess-dive-support@lbl.gov](mailto:ess-dive-support@lbl.gov)



U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science

