

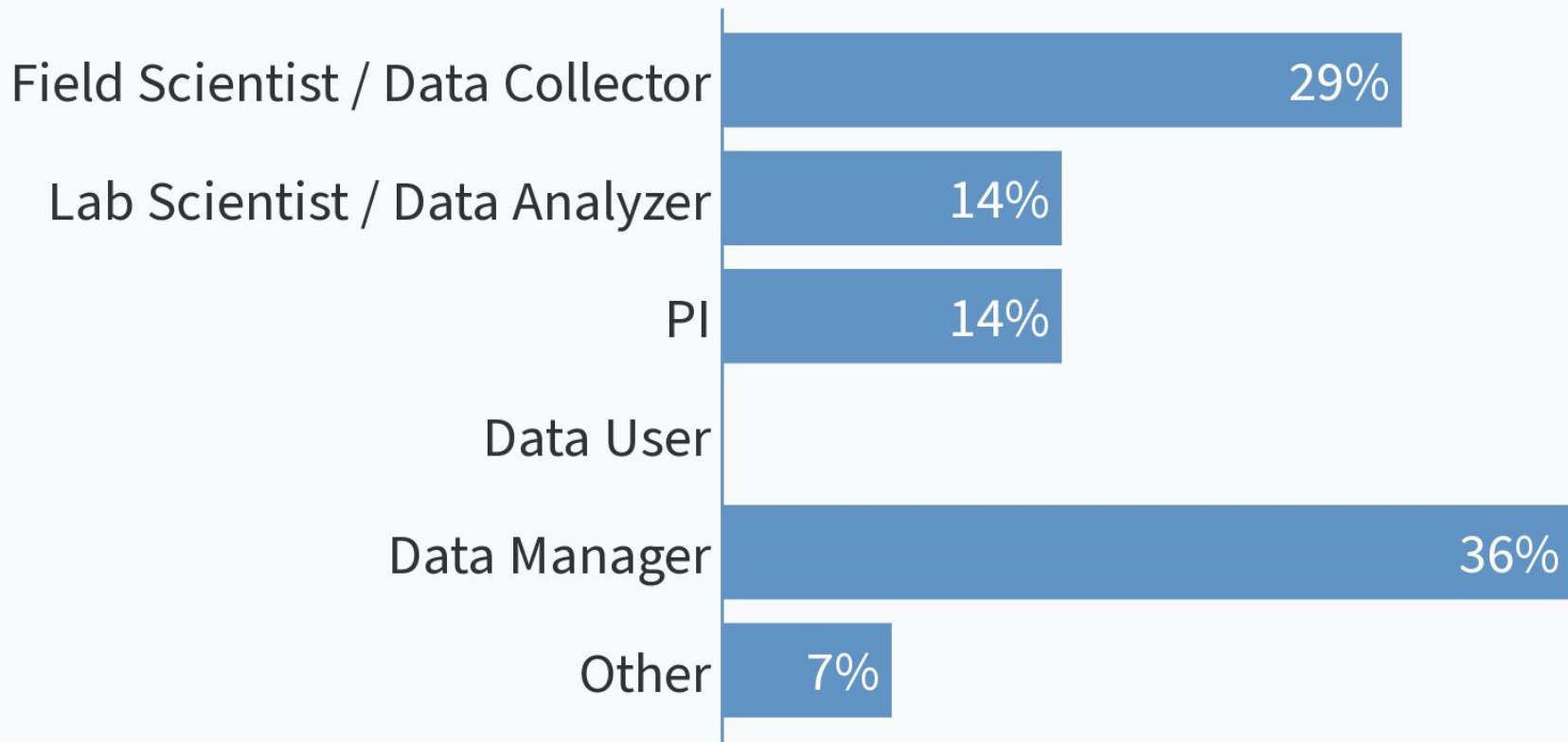


ESS-DIVE Sample Identifiers and Metadata Reporting Format Tutorial

Joan Damerow, ESS-DIVE Community Engagement Lead Scientist
Lawrence Berkeley National Laboratory



What is your primary role in your research group?



What organization are you representing today?

“ University of Wisconsin-Madison ”

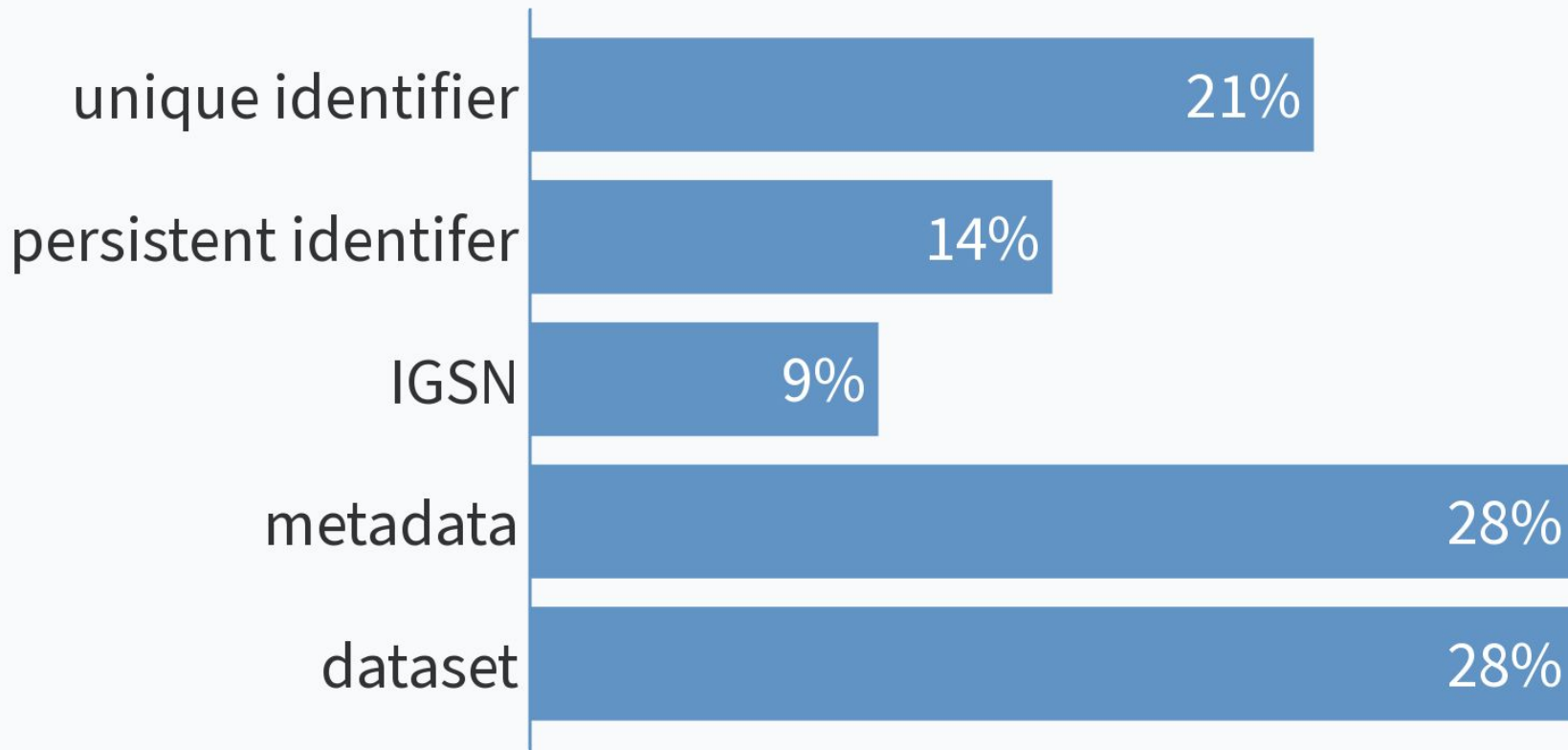
“ NGEE Arctic ”

“ LBNL ”

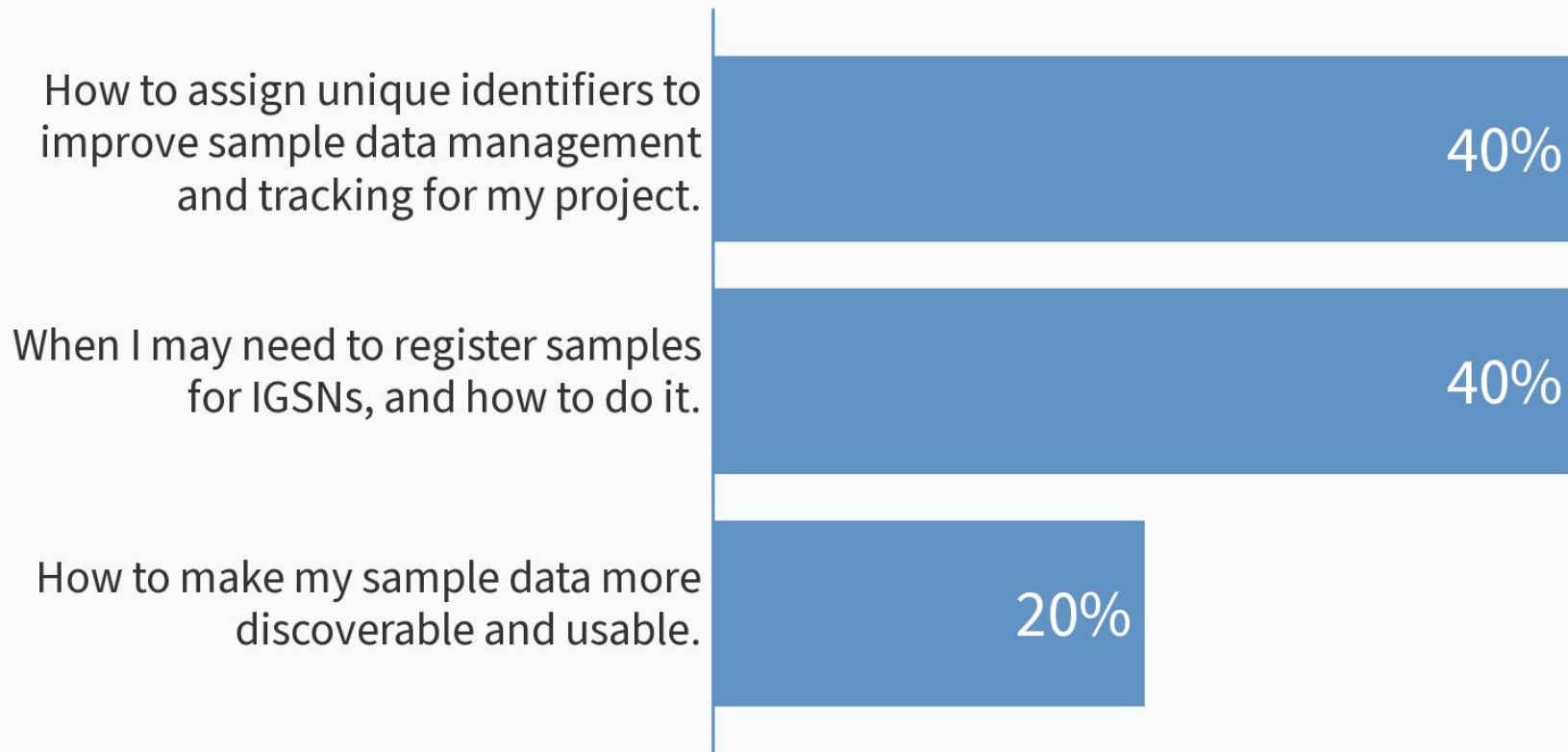
“ NGEEs/BNL ”

“ University of Alaska Fairbanks ”

Which of the following terms are you familiar with?



What are you most interested in learning about today?



Tutorial Overview

- Sample identifiers and metadata
- ESS-DIVE sample ID and metadata guide/template
- Sample planning to publication workflow
- How and why to use International Geo/General Sample Numbers (IGSN)

https://bit.ly/ESSDIVE_SampleTutorial

Takeaways from this presentation

- Understand when to use **unique and persistent IDs**
- Standardize sample metadata using template
- Register samples for IGSNs
- Publish sample datasets in ESS-DIVE

*ESS-DIVE guidance for sample identifiers and metadata will enable more effective **sample planning, tracking, discovery, and reuse.***

Terminology Check: identifiers and metadata



Unique Identifier

Provides a meaningful, project-specific unique ID to organize your data

Sample Name:

RockCr001_2021-05-25



Metadata

Descriptive information about data

Sample Type: Water

Feature: Stream

Location: Rock Creek,
Crested Butte, CO



Persistent Identifiers

Globally unique IDs with permanent link/landing page, associated metadata

ORCID: People

DOI: Data, publications

IGSN: Samples

IEWFS000U

What is Sample Metadata?

Sample Collections Details

- Collector/Chief Scientist*
- Collection Date*
- Collection Time
- Collection Method Description*
- Sample Processing (MlxS)
- Field Program or Project Name*

Sample Access

- Release Date*
- Current Archive
- Current Archive Contact

Location

- Location Description
- Latitude*
- Longitude*
- Geolocation Instrument
- Elevation (start, end)
- Elevation Unit
- Country*
- Minimum/Maximum Depth in Meters (DwC)
- Minimum/Maximum Distance above Surface in Meters (DwC)

Environmental Context

- Physiographic Feature* (ENVO, MlxS)
- Biome (MlxS)

Sample Description

- *IGSN-SESAR provides*
- Sample Name*
- Object Type* (BCO)
- Material* (ENVO, PO)
- Classification
- Sample Description
- Purpose
- Size, Size Unit
- Filter Size (MlxS)
- Scientific Name (DwC)
- Sample Remarks

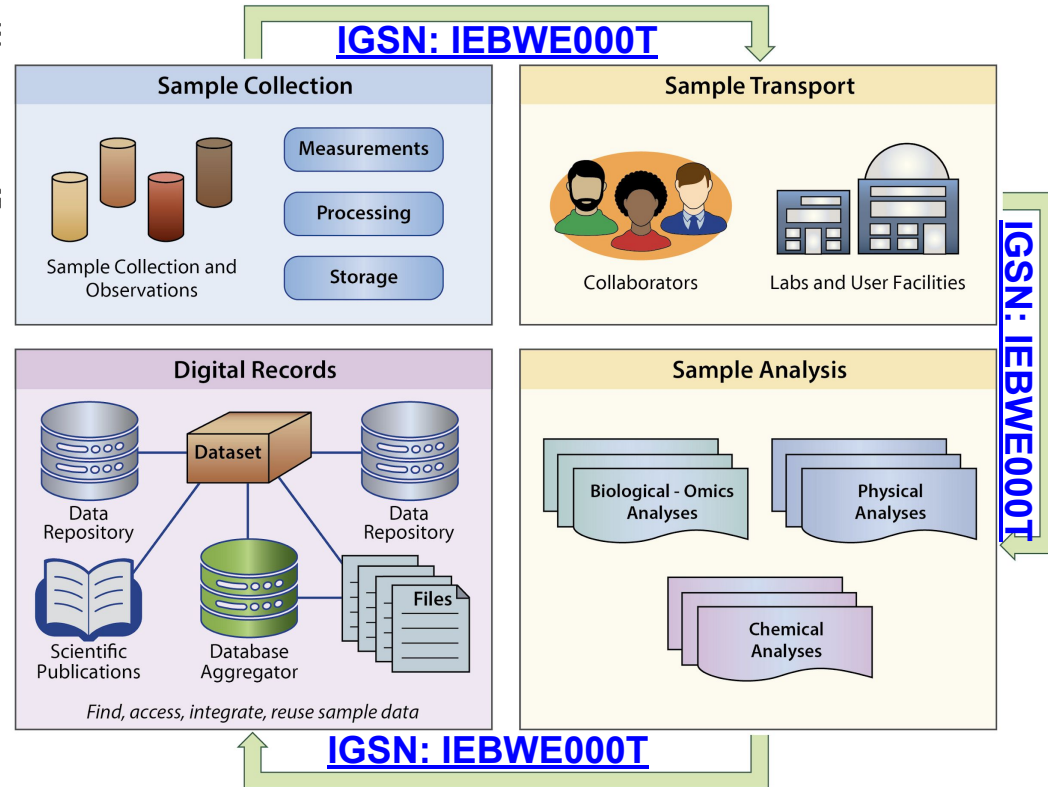
Related Identifiers

- Parent IGSN
- Collection ID (DwC)
- Event ID (DwC)
- Location ID (DwC)

Using IGSNs for Sample Tracking and Linking

When do you need persistent IDs for samples?

- 1.) Multiple datasets, journal publications
- 2.) Collaborators work on same samples
- 3.) Multiple labs for analyses
- 4.) Sample-related data in different repositories
- 5.) Archived, and used for multiple purposes over time



Persistent IDs: Landing Pages



Soil Sample
Landing Page

IGSN: IEWFS0001



IGSN: IEWFS0001
Sample Name: 115
Other Name(s):
Sample Type: Core Section
Parent IGSN: Not Provided

Description

Material: Soil
Classification: Not Provided
Field Name: Not Provided
Description: Soil cores that were collected seasonally during autumn, winter, snowmelt, and spring at a high altitude field site which is predominately montane meadow
Age (min): Not Provided
Age (max): Not Provided
Collection Method: Manual>Hammer
Collection Method Description: Soil cores were collected using soil bulk density corer attached to a slide hammer
Size: Not Provided
Geological Age: Not Provided
Geological Unit: Not Provided
Comment: Not Provided
Purpose: Not Provided

Geolocation

Latitude (WGS84): 38.917216053
Longitude (WGS84): -106.955994698

Relevant Links:

- <http://identifiers.org/gold:Gp0321263>: Soil microbial communities from the East River watershed near Crested Butte, Colorado, United States - Metgenomes (Genomes Online Database, GOLD)
- <http://identifiers.org/gold:Gp0396393>: Soil microbial communities from the East River watershed near Crested Butte, Colorado, United States - Metatranscriptomes (Genomes Online Database, GOLD)
- <https://doi.org/10.15485/1577267>: Dataset: Soil Nitrogen, Water Content, Microbial Biomass, and Archaeal, Bacterial and Fungal Communities from the East River Watershed, Colorado collected in 2016-2017.
- <https://doi.org/10.21952/WTR/1573029>: Dataset for sample collection metadata

<https://app.geosamples.org/sample/igsn/IEWFS0001>

Sorensen P ; Brodie E ; Beller H ; Wang S ; Bill M ; Bouskill N (2019): Soil Nitrogen, Water Content, Microbial Biomass, and Archaeal, Bacterial and Fungal Communities from the East River Watershed, Colorado collected in 2016-2017. Watershed Function SFA. doi:10.15485/1577267

Citations 0 Downloads 0 Views 0 Copy Citation Assessment report

Files in this dataset Package: ess-dive-4ca86f6d5ba818f-20210430T014715527688

Name	File type	Size	Login to Download
Metadata: Soil_Nitrogen_Water_Content_Microbial_Biomass_and_Archaeal_Bacterial_and_Fungal_Communities_from_the_East_River_Watershed_Colorado_collected_in_2016_2017.xml	EML v2.1.1	10 KB	Download
2017_East_River_Pumphouse_Extractable_Soil_N_Pools__1_.csv	More info Microsoft Excel	48 KB	Download
2017_East_River_Pumphouse_Archaea_and_Bacteria_Life_Strategies__1_.csv	More info Microsoft Excel	5 MB	Download
2017_East_River_Pumphouse_Fungal_Life_Strategies__2_.csv	More info Microsoft Excel	448 KB	Download

Related References

Sample Metadata:Sorensen P ; Brodie E ; Beller H ; Wang S ; Bill M ; Bouskill N (2019): Soil Nitrogen, Water Content, Microbial Biomass, and Archaeal, Bacterial and Fungal Communities from the East River Watershed, Colorado collected in 2016-2017. Watershed Function SFA. doi:10.21952/WTR/1573029

ESS-DIVE Dataset:
Soil Measurements

<https://doi.org/10.15485/1577267>

Benefits of using IGSNs Across Facilities and Data Systems



Persistent Identifier Benefits

1. **Link** and **expand** access pathways
2. **Avoid duplication** of information across platforms
3. Interpretation and **reuse**

Linking related interdisciplinary data

IGSN: IEWFS0001




IGSN: IEWFS0001
 Sample Name: 115
 Other Name(s):
 Sample Type: Core Section
 Parent IGSN: Not Provided

Description

Material: Soil
 Classification: Not Provided
 Field Name: Not Provided
 Description: Soil cores that were collected seasonally during autumn, winter, snowmelt, and spring at a high altitude field site which is predominately montane meadow
 Age (min): Not Provided
 Age (max): Not Provided
 Collection Method: Manual>Hammer
 Collection Method Description: Soil cores were collected using soil bulk density corer attached to a slide hammer
 Size: Not Provided
 Geological Age: Not Provided
 Geological Unit: Not Provided
 Comment: Not Provided
 Purpose: Not Provided

Geolocation


Latitude (WGS84): 38.917216053
 Longitude (WGS84): -106.955994698

Relevant Links:

- <http://identifiers.org/gold:Gp0321263>: Soil microbial communities from the East River watershed near Crested Butte, Colorado, United States - Metgenomes (Genomes Online Database, GOLD)
- <http://identifiers.org/gold:Gp0396393>: Soil microbial communities from the East River watershed near Crested Butte, Colorado, United States - Metatranscriptomes (Genomes Online Database, GOLD)
- <https://doi.org/10.15485/1572>: Microbial Biomass, and Archaeal the East River Watershed, Colorado
- <https://doi.org/10.21952/WTR/metadata>

Soil Sample Landing Page

Sorensen P ; Brodie E ; Beller H ; Wang S ; Bill M ; Bouskill N (2019): Soil Nitrogen Content, Microbial Biomass, and Archaeal, Bacterial and Fungal Communities from Watershed, Colorado collected in 2016-2017. Watershed Function SFA. doi:10.1101/2021.04.30.4301014715527688



Citations: 0 Downloads: 0 Views: 0 Copy Citation Assessment report

Files in this dataset Package: ess-dive-4ca8c6d5ba818f-20210430T014715527688

Name	File type	Size	Login to Download
Metadata:			
Soil_Nitrogen_Water_Content_Microbial_Biomass_and_Archaeal_Bacterial_and_Fungal_Communities_from_the_East_River_Watershed_Colorado_collected_in_2016_2017.xml	EML v2.1.1	10 KB	Download
2017_East_River_Pumphouse_Extractable_Soil_N_Pools_-_1_.csv	More info Microsoft Excel	48 KB	Download
	More info Microsoft Excel	5 MB	Download
	More info Microsoft Excel	448 KB	Download

Items in this data set

ESS-DIVE Dataset: Soil Measurements

The synchronization of microbial and plant phenology in a mountainous watershed and its importance for nutrient retention under changing hydrologic regimes.

Description The goal of the study is to observe the activation of microbial metabolic potential beneath the snowpack during winter and during the snowmelt period, as well as advanced characterization of the chemistry of carbon and nutrient transformations and assimilation by microorganisms and vegetation in response to earlier snowmelt timing.

Metabolomics: 52 Metatranscriptome: 45 Metagenome: 48 Organic Matter: 1007





Eoin Brodie

Principal investigator

<https://eesa.lbl.gov/profiles/eoin-brodie/>
<https://orcid.org/0000-0002-8453-8435>
<https://watershed.lbl.gov/>

National Microbiome Data Collaborative: Study Page

Project Information Sequencing Information Sample Information

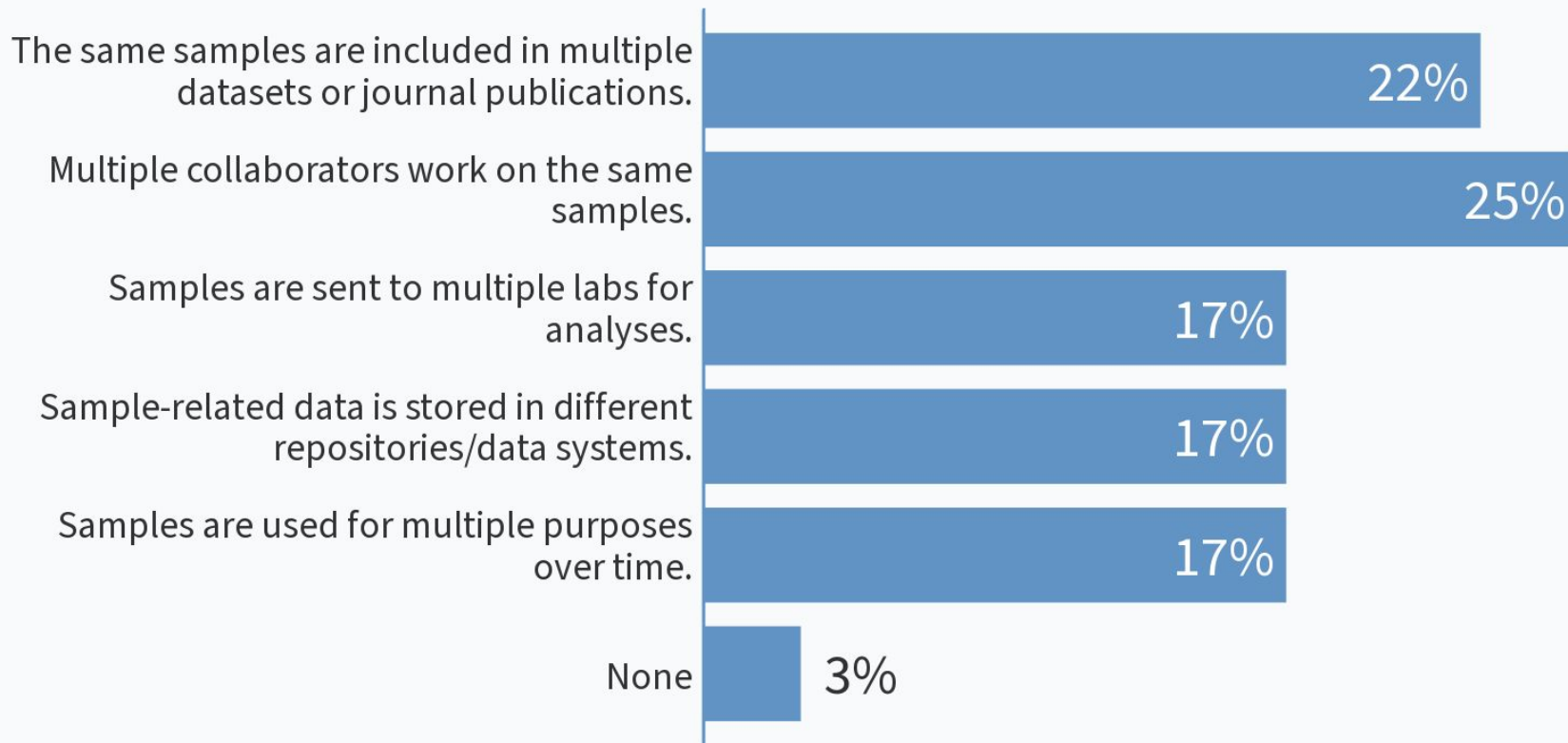



PROJECT INFORMATION

GOLD Project ID: Gp0321263
 Project Name: Soil microbial communities from the East River watershed near United States - ER_DNA_115
 Other Names:
 Legacy GOLD ID:
 NCBI BioProject Name: Soil microbial communities from the East River watershed near Crested Butte, Colorado, United States - ER_DNA_115
 NCBI BioProject ID: 5185
 NCBI BioProject Accession: PRJN
 NCBI Locus Tag: EVQ
 NCBI BioSample Accession: SAM
 PI: Eoin

JGI Sequencing Projects: Microbial Communities

Which of the following apply for your project sample data?



Sampling Planning to Publication Workflow

General Workflow - Sample IDs and Metadata Reporting Format



Step 1

Plan your sample campaign



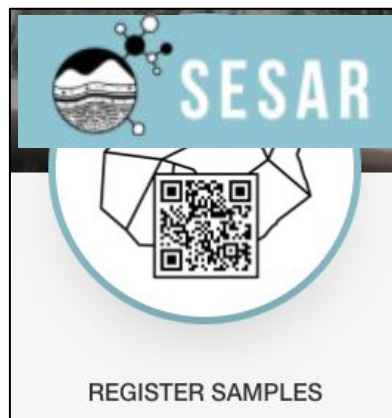
Step 2

Standard sample metadata

Object Type:	Core Section	User Code:	IEWFS	
Sample Name	IGSN	Material	Sample description	Collection method
115	IEWFS0001	Soil	Soil cores that were collected seasonally during autumn, winter, snowmelt, and spring at a high altitude field site which is predominately montane meadow Soil cores that were	Manual>Hammer

Step 3

Register samples for IGSNs



Step 4

Submit sample datasets to ESS-DIVE



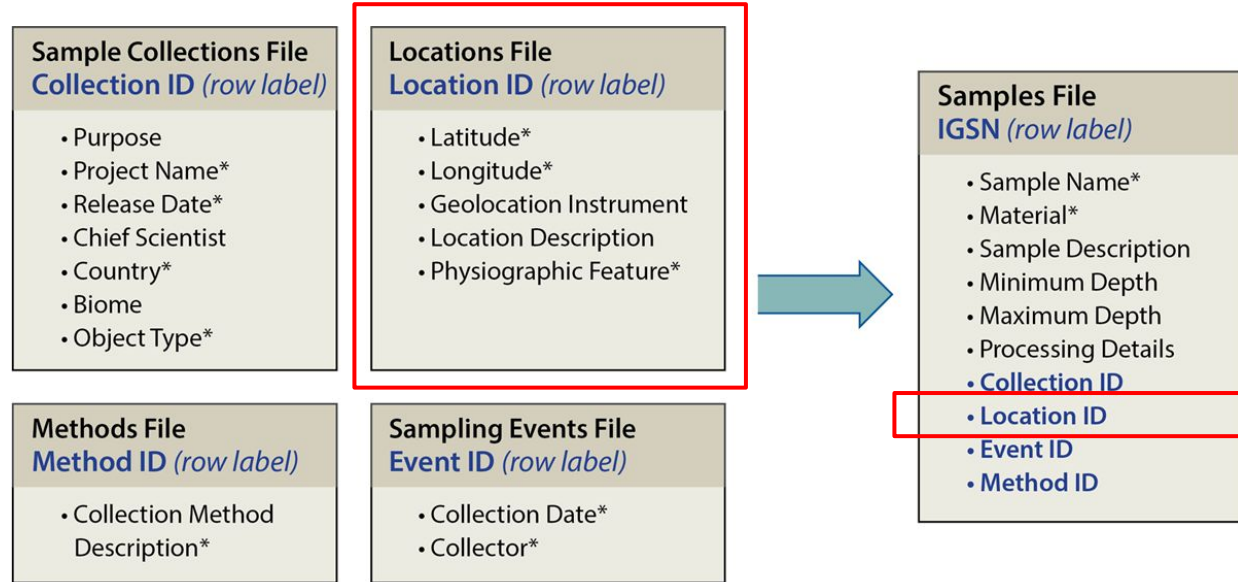
Step 1: Planning - consider file organization and IDs for your sampling campaign



Incorporate sample data management into planning your field campaign

How to do it:

Assign project-specific unique identifiers (e.g. locations, samples, subsamples)



EESA20-060

Consistently using identifiers to manage and link related metadata across files



LOCATIONS FILE

locationID	Other name(s)	Description	Latitude	Longitude
ER-GUM1	Gothic Upper Montane-1; GUM	Groundwater Monitoring Wel	38.9550494	-106.9923898
ER-RCK1	Rock Creek; Rock	Groundwater Monitoring Wel	38.9816681	-107.0055887
ER-RUS1	Rustlers Gulch; Rustlers	Groundwater Monitoring Wel	38.9889397	-107.007755
ER-SHM1	Shumway Well; Shumway	Groundwater Monitoring Wel	38.9454058	-106.9895197
ER-BRD1	Bradley Creek; Bradley	Water Quality-Discharge Sta	38.9854093	-107.00484

PARENT SAMPLE FILE

Sample Name	IGSN	locationID	Collection date	Material	Field name (info)
GUM-1_2021-03-10	IEWFS004Q	ER-GUM1	2021-03-10	Liquid>aqueous	groundwater [ENVO:01001004]
Shumway_2021-03-10	IEWFS004R	ER-SHM1	2021-03-10	Liquid>aqueous	groundwater [ENVO:01001004]
Tuttle_2021-03-10	IEWFS004S	ER-TTL1	2021-03-10	Liquid>aqueous	groundwater [ENVO:01001004]
EAQ_2021-03-03	IEWFS004T	ER-EAQ1	2021-03-03	Liquid>aqueous	surface water [ENVO:00002042]
Rustlers_2021-03-03	IEWFS004U	ER-RUS1	2021-03-03	Liquid>aqueous	surface water [ENVO:00002042]
Bradley_2021-03-03	IEWFS004V	ER-BRD1	2021-03-03	Liquid>aqueous	surface water [ENVO:00002042]

SUBSAMPLE FILE

Sample Name	Parent IGSN	IGSN	locationID	Collection date	Material
EAQ_2021-03-03_DO	IEWFS004T	IEWFS001I	ER-EAQ1	2021-03-03	Liquid>aqueous
EAQ_2021-03-03_CA	IEWFS004T	IEWFS001J	ER-EAQ1	2021-03-03	Liquid>aqueous
EAQ_2021-03-03_AM	IEWFS004T	IEWFS001K	ER-EAQ1	2021-03-03	Liquid>aqueous
EAQ_2021-03-03_AN	IEWFS004T	IEWFS001L	ER-EAQ1	2021-03-03	Liquid>aqueous
Rustlers_2021-03-03_DO	IEWFS004U	IEWFS001M	ER-RUS1	2021-03-03	Liquid>aqueous
Rustlers_2021-03-03_CA	IEWFS004U	IEWFS001N	ER-RUS1	2021-03-03	Liquid>aqueous
Rustlers_2021-03-03_AM	IEWFS004U	IEWFS001O	ER-RUS1	2021-03-03	Liquid>aqueous

Time to Practice!

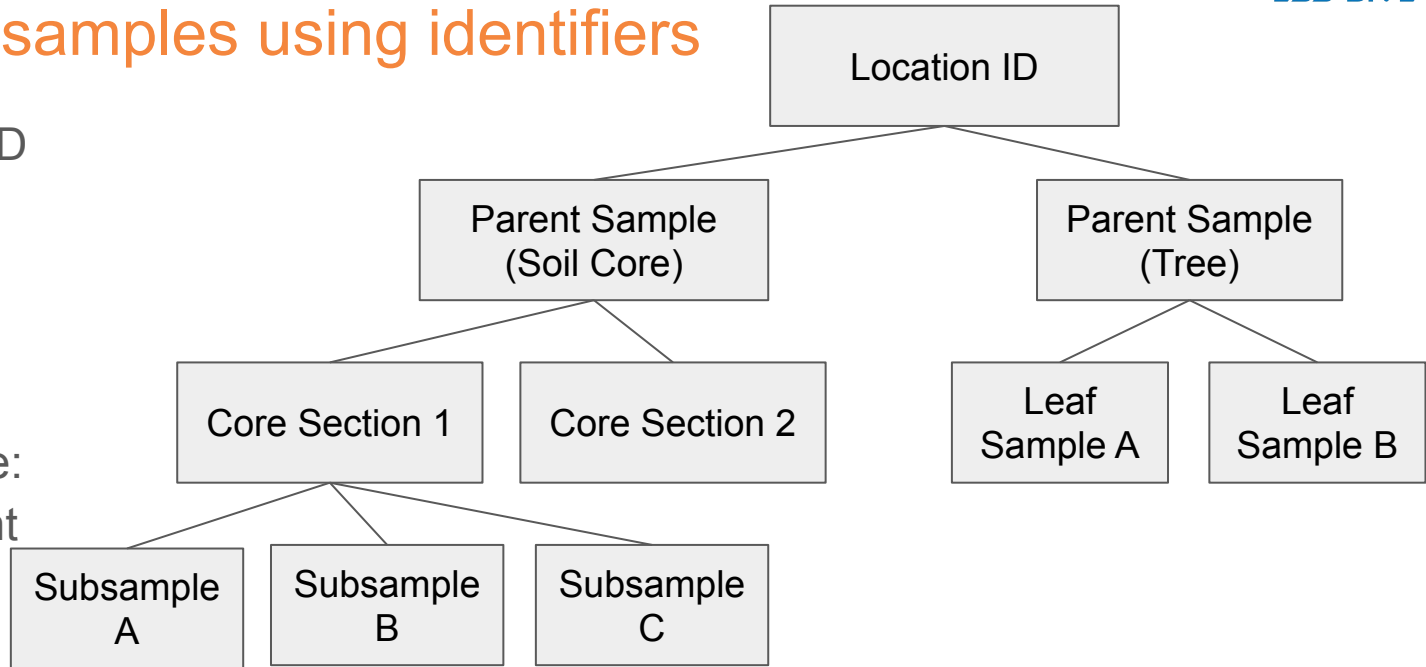
Link related samples using identifiers

Sampling Event ID

Location ID

Parent Sample

- Each sample:
record parent

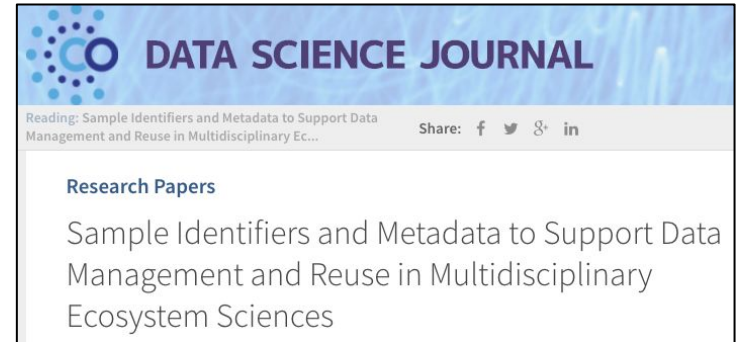


<https://bit.ly/SampleRelationshipsActivity>

Time to Practice!

Access the ESS-DIVE guide and template

- 1) *ESS-DIVE documentation for samples*
<https://ess-dive.gitbook.io/sample-id-and-metadata/>
- 2) [Instructions](#) - **download sample metadata template**
- 3) [Access metadata guide](#)
- 4) *Shared vocabularies* ([e.g. Material](#))
- 5) *Citation / References* →



Step 2: Provide standard sample metadata



Characterize samples and collection details

How to do it:

[Download](#) and complete sample metadata template

**Tip: use metadata guide for definitions and shared terms*

Object Type:	Core Section	User Code:	IEWFS						
Sample Name	IGSN	Material	Sample description	Collection method	Collection method description	Latitude	Longitude	Location description	Field program/cruise
115	IEWFS0001	Soil	Soil cores that were collected seasonally during autumn, winter, snowmelt, and spring at a high altitude field site which is predominately montane meadow	Manual>Hammer	Soil cores were collected using soil bulk density corer attached to a slide hammer	38.917216053	-106.9559947	Colorado The East River is a snow-dominated, headwater basin of the Upper Colorado River Basin located in the western United States.	Watershed Function SFA
116	IEWFS0002	Soil	Soil cores that were collected seasonally during autumn, winter, snowmelt, and spring at a high altitude field site which is predominately montane meadow	Manual>Hammer	Soil cores were collected using soil bulk density corer attached to a slide hammer	38.917216053	-106.955994698	East River Watershed, Colorado The East River is a snow-dominated, headwater basin of the Upper Colorado River Basin located in the	Watershed Function SFA
117	IEWFS0003	Soil	Soil cores that were collected seasonally during autumn, winter, snowmelt, and spring at a high altitude field site which is predominately montane meadow	Manual>Hammer	Soil cores were collected using soil bulk density corer attached to a slide hammer	38.917216053	-106.955994698	East River Watershed, Colorado The East River is a snow-dominated, headwater basin of the Upper Colorado River Basin located in the	Watershed Function SFA

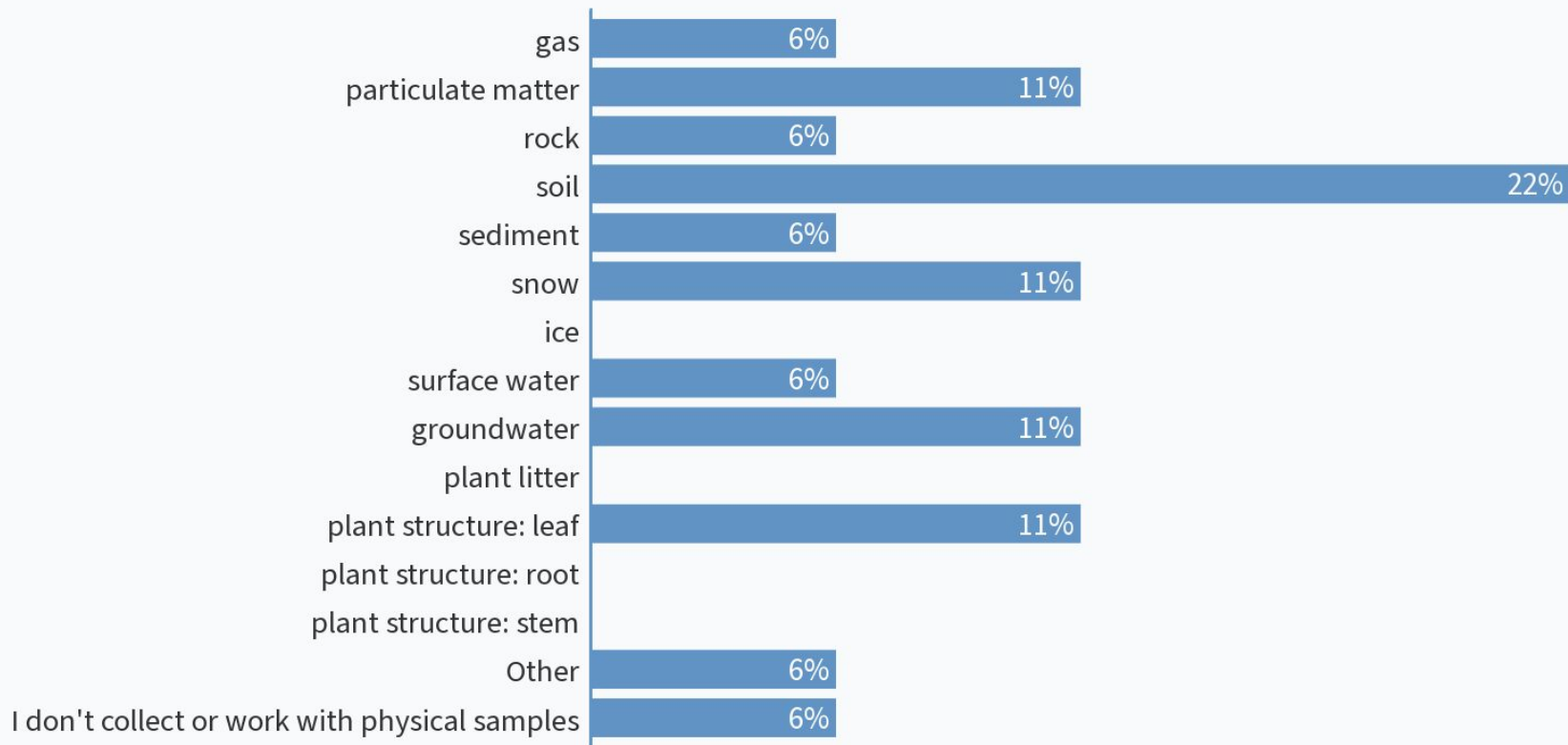
Time to Practice!

Characterize your sample material

Think of 1-2 of your samples types, and use the material vocabularies to characterize them. Links to shared vocabularies are listed in the guide.

[Link to material terms](#)

What sample type material do you most often work with?



Step 3: Register samples for IGSNs






How to do it:

- 1.) Choose a three character user code
 - Individual or project-specific (e.g. BWE)
 - Example IGSN: IEBWE0094
- 2.) Batch upload sample metadata file
- 3.) Update metadata as needed
 - Add links to related data
- 4.) Option to print labels

**Demo* Link to SESAR:

<https://www.geosamples.org/>





IGSN: IEBWE0094  

 IGSN: IEBWE0094
Sample Name: BWE201806041C6080
Other Name(s):
Sample Type: Core Section
Parent IGSN: IEBWE008U

Description

Material: Soil
Classification: Not Provided
Field Name: Not Provided
Description: Day 1685 core section from unheated control plot 1C of a deep soil warming experiment
Age (min): Not Provided
Age (max): Not Provided
Collection Method: Coring>HandHeldCorer
Collection Method Description: Collected with a multi-stage corer 5 cm in diameter using a 10 kg hand-held slide-hammer

Related Samples

Parents: IEBWE008U BWE201806041C
Siblings:
 [IEBWE0090 BWE201806041C0010](#)
 [IEBWE0091 BWE201806041C1030](#)
 [IEBWE0092 BWE201806041C3045](#)
 [IEBWE0093 BWE201806041C4560](#)

Children: No Children

Step 4: Publish your sample data on ESS-DIVE



How to do it:

- 1.) Include sample-related files: Sample metadata, location metadata, sample data
a.) Files have sample names and IGSNs
- 2.) Include IGSNs as related identifiers
- 3.) After publication: update your SESAR IGSN sample metadata with related urls (dataset DOI, url type, url description)

Sorensen P ; Brodie E ; Beller H ; Wang S ; Bill M ; Bouskill N (2019): Soil Nitrogen, Water Content, Microbial Biomass, and Archaeal, Bacterial and Fungal Communities from the East River Watershed, Colorado collected in 2016-2017. Watershed Function SFA. doi:10.15485/1577267

Citations 0 Downloads 0 Views 0 Copy Citation Assessment report

Files in this dataset Package: ess-dive-4ca86f6d5ba818f-20210430T014715527688

Name	File type	Size	Login to Download
Metadata: Soil_Nitrogen_Water_Content_Microbial_Biomass_and_Archaeal_Bacterial_and_Fungal_Communities_from_the_East_River_Watershed_Colorado_collected_in_2016_2017.xml	EML v2.1.1	10 KB	Download
2017_East_River_Pumphouse_Extractable_Soil_N_Pools__1_.csv	More info Microsoft Excel	48 KB	Download
2017_East_River_Pumphouse_Archaea_and_Bacteria_Life_Strategies__1_.csv	More info Microsoft Excel	5 MB	Download
2017_East_River_Pumphouse_Fungal_Life_Strategies__2_.csv	More info Microsoft Excel	448 KB	Download

► Show 4 more items in this data set

Sample Tracking Feedback - Poll
Questions <https://pollev.com/essdive>

**Where is your sample data stored or published? Examples:
paper only; personal files; personal or public databases;
published dataset(s)**

This poll question was skipped

What would other future researchers need to do to be able to compile and link the related sample data from your project?

This poll question was skipped

Conclusions

Sample Data Management

- Plan file organization and unique identifiers
- Standardize metadata (template)
- Register, manage, label samples - IGSNs

Use IGSNs to enable tracking samples and exchanging related information:

- Over time
- Across data systems, publications, collaborators



Questions?

Follow ESS-DIVE on Twitter! [@ESSDIVE](https://twitter.com/ESSDIVE)

Join ESS-DIVE's Community Mailing List! <http://bit.ly/essdiveMailingList>

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