ESS-DIVE Sample Identifiers and Metadata Tutorial

Presented By: Joan Damerow
Community Engagement Lead Scientist for ESS-DIVE
Lawrence Berkeley National Lab
Earth & Environmental Sciences Area

Community Data Workshop 2021

Tutorial Overview

- Sample identifiers and metadata

- ESS-DIVE sample ID and metadata guide/template

- Sample planning to publication workflows

- How and why to use International Geo/General Sample Numbers (IGSN)

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

Takeaways from this presentation

- Understand when you need to use **unique and persistent IDs**

  - Unique identifiers = your name

  - Persistent identifiers = social security number

- Feel confident to start assigning unique IDs and standardizing sample metadata

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

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

IGSN: IEBWE000T

IGSN: IEBWE000T

IGSN: IEBWE000T

EESA20-037
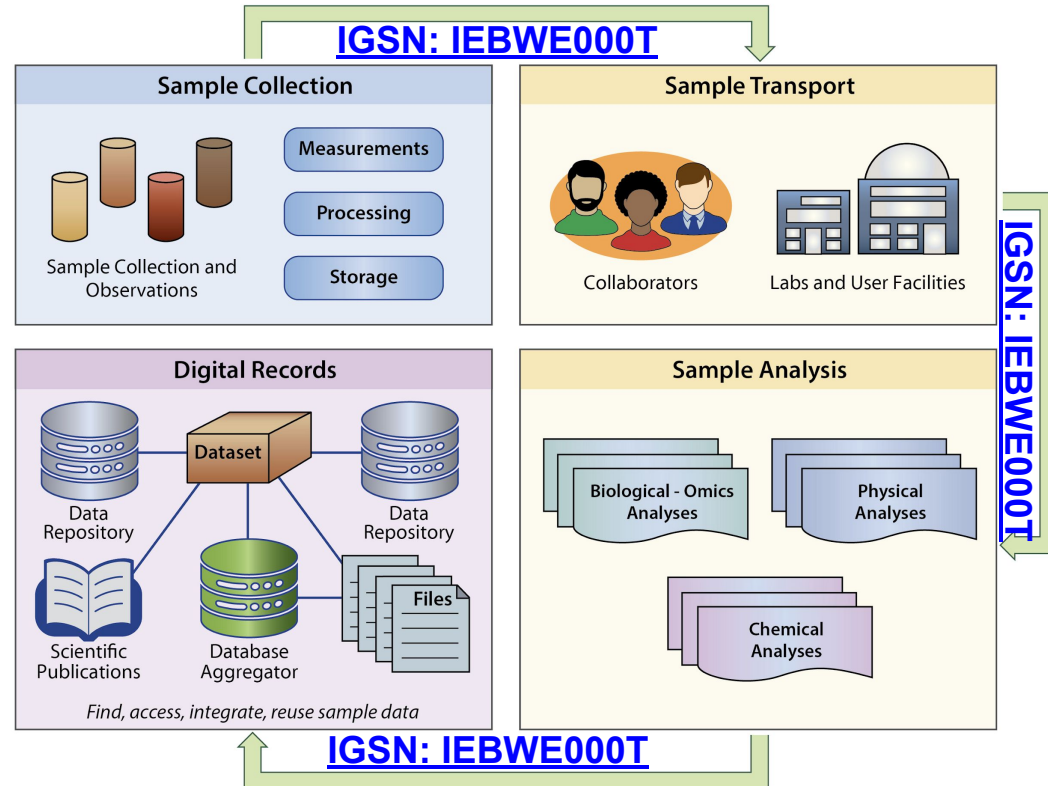
What is Sample Metadata?

Sample Collections Details

- Collector/Chief Scientist*
- Collection Date*
- Collection Time
- Collection Method Description*
- Sample Processing (MIxS)
- 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, MIxS)
- Biome (MIxS)

Sample Description

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

Related Identifiers

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

ESS-DIVE Sample Identifiers and Metadata Reporting Format

Time to Practice!
Access the IGSN-ESS guide and template

1) *ESS-DIVE documentation for samples*

   https://ess-dive.gitbook.io/sample-id-and-metadata/

1) *Instructions* - ***download sample metadata template***

2) *Access metadata guide*

3) *Shared vocabularies (e.g. Material)*

4) *Citation / References*

http://doi.org/10.5334/dsj–2021–011
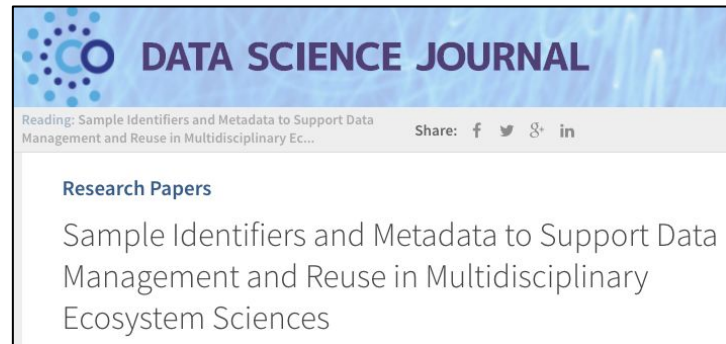
Sampling Planning to Publication Workflow

General Workflow - Sample IDs and Metadata Reporting Format

Step 1
Plan your sample campaign
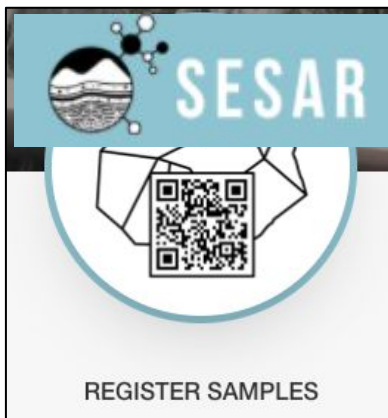
Step 2
Standard sample metadata

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 predominantly montane meadow	Manual>Hammer
			Soil cores that were	

Step 3
Register samples for IGSNs

SESAR

REGISTER SAMPLES

Step 4
Submit sample datasets to ESS-DIVE

ESS-DIVE
Deep Insight for Earth Science Data

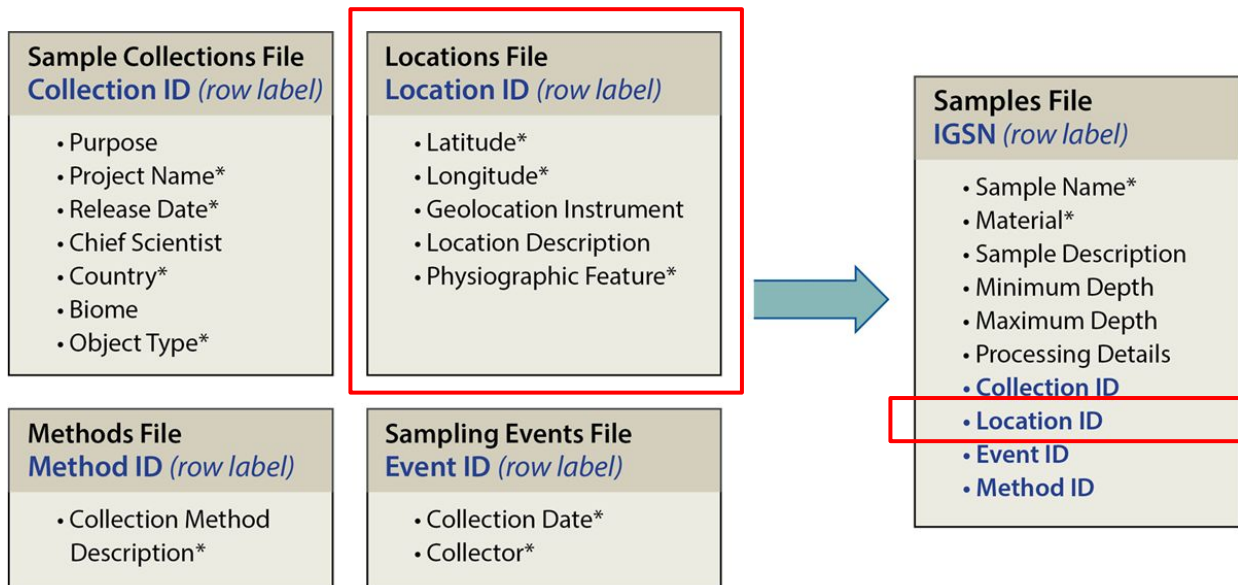Add files to start your dataset

+ Add Files

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)

Sample Collections File
Collection ID *(row label)*

- Purpose
- Project Name*
- Release Date*
- Chief Scientist
- Country*
- Biome
- Object Type*

Locations File
Location ID *(row label)*

- Latitude*
- Longitude*
- Geolocation Instrument
- Location Description
- Physiographic Feature*

Samples File
IGSN *(row label)*

- Sample Name*
- Material*
- Sample Description
- Minimum Depth
- Maximum Depth
- Processing Details
- Collection ID
- Location ID
- Event ID
- Method ID

Methods File
Method ID *(row label)*

- Collection Method Description*

Sampling Events File
Event ID *(row label)*

- Collection Date*
- Collector*

EESA20-060

Consistently use identifiers to manage and link related metadata across files

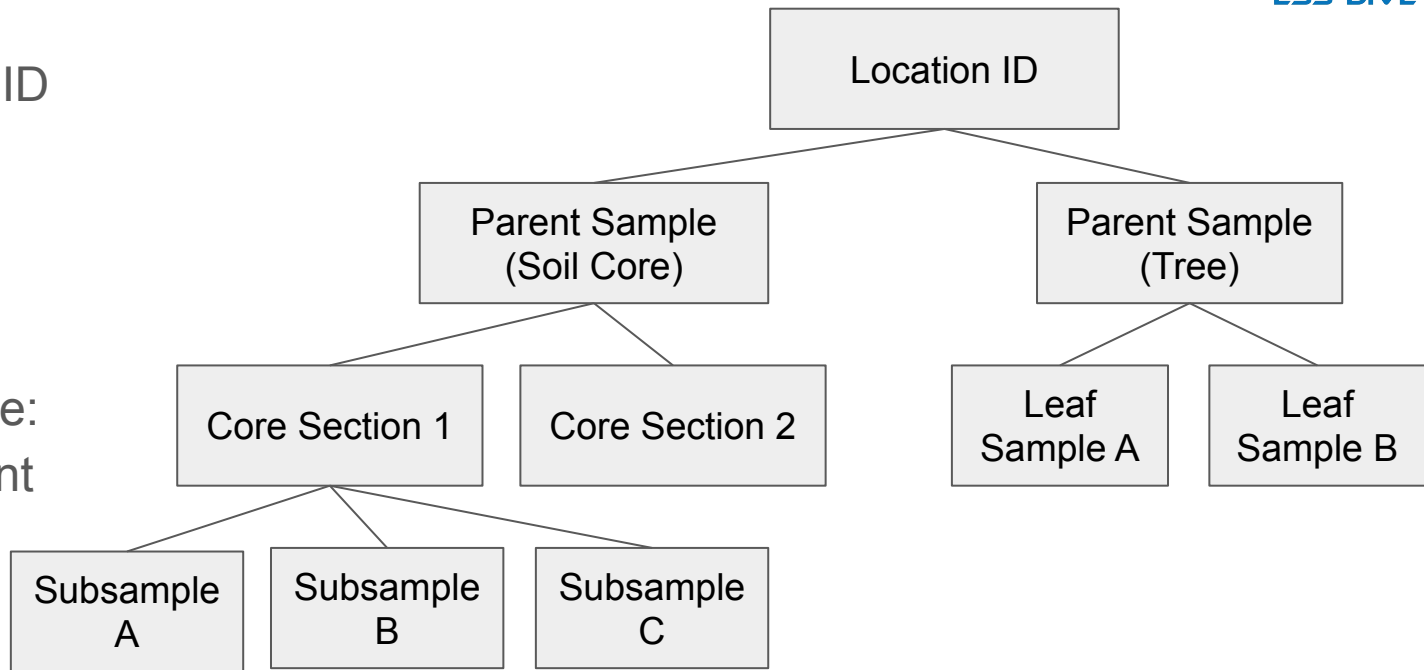# Link related samples using identifiers

Sampling Event ID

Location ID

Parent Sample

- Each sample: record parent

LOCATIONS FILE

locationID	Other name(s)	Description	Latitude	
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
ER-CPR1				
ER-EAQ1				
ER-EBC0				
ER-TTL1				

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:000
Rock_2					
Copper					
EBC_2					

SUBSAMPLE FILE

Sample Name	Parent IGSN	IGSN	locationID	Collection dat	
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

13

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 SESAR and ESS-DIVE vocabularies to characterize them. Links to shared vocabularies are listed in the guide.

[Link to material terms](#)

Do you have sample types that are not represented in the shared vocabulary lists?

International Geo/General Sample Numbers: How and why to use them

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

Demo of IGSN batch sample registration

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

17

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: pdate your SESAR IGSN sample metadata with related urls (dataset DOI, url type, url description)

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**

10

Questions?