ESS-DIVE
Sample ID and Metadata Reporting Format

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

ESS-DIVE BERKELEY LAB NERSC NCEAS DataONE U.S. DEPARTMENT OF ENERGY Office of Science

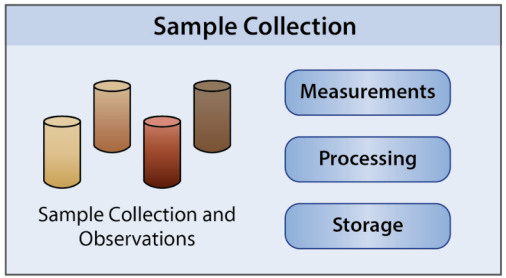
September 2020 Webinar

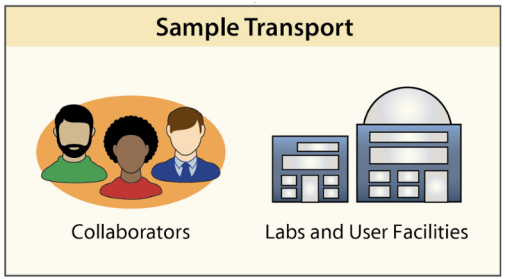
Overview

- Review background work and testing to address specific ESS needs
  - Interoperable metadata for multidisciplinary work
  - Sample relationships, tracking, and linking
  - Efficiency for large sampling campaigns
- Documentation on ESS-DIVE's sample ID and metadata reporting format
- Next steps and feedback

Overall Goal: sample metadata supports FAIR and improve efficiency of sample tracking

Identify Community Need: Sample IDs

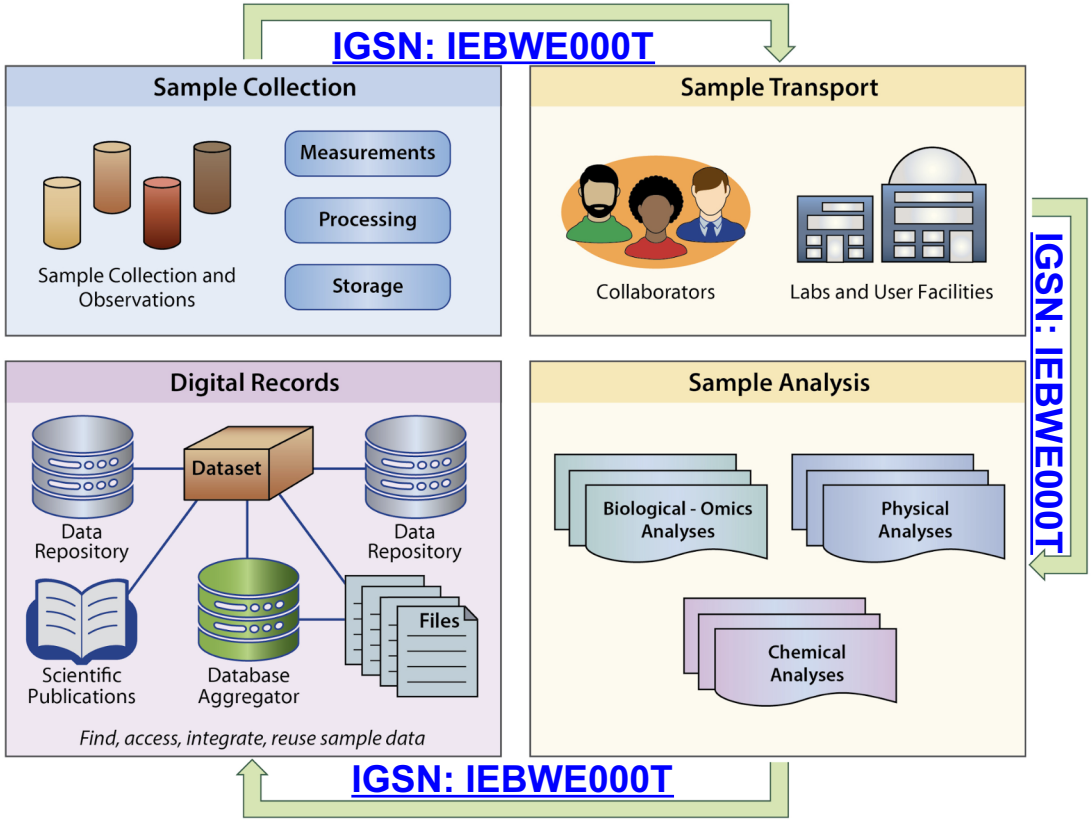




Challenge

Lack of a practical, standardized sample tracking system

Identify Community Need: Sample Standards



Challenge

Lack of a practical, standardized sample tracking system

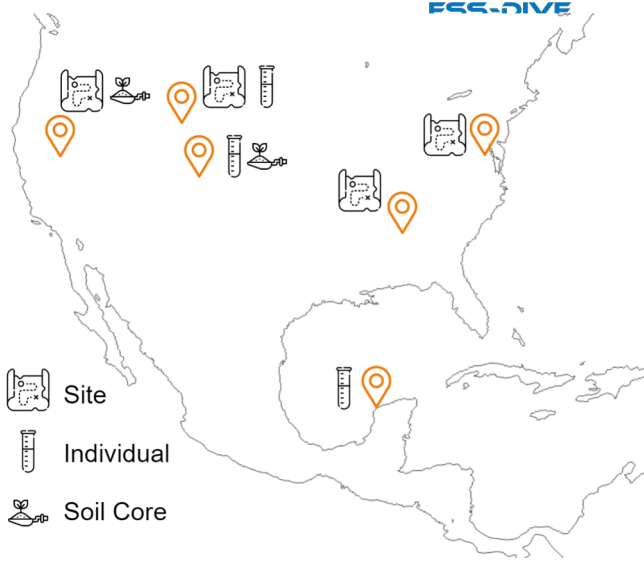
Solution

Global Sample Numbers **(IGSNs)**

Standardized core sample metadata and templates

Multidisciplinary Sample ID and Metadata Test

2,953 Registered IGSNs for variety of sample types



A multidisciplinary approach for sample identifiers and metadata in Environmental Systems Science

DRAFT - Scientific Data

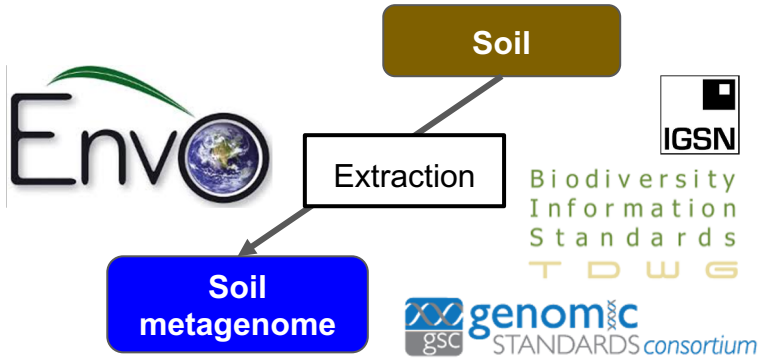
Interoperability of Sample Metadata

IGSN Schema Descriptive Metadata Elements (child	GOLD	MIMS/MIxS Core and Environmental Packages template	ISO19156:201112 (Observations and Measurements) -	Darwin Core (*required fields)
identifier (identifierType)	sample_collection_site			locationID [http://rs.tdwg.org/dwc/terms/locationID]
identifier (identifierType)	gold_id	source_mat_id (maps to materialSampleID in DwC)	SF_Specimen (/spec:SF_Specimen/gml:identifier)	occurenceID* [http://rs.tdwg.org/dwc/terms/occurrenceID] or materialSampleID [http://rs.tdwg.org/dwc/terms/materialSampleID]
name			/spec:SF_Specimen/gml:name	otherCatalogNumbers; catalogNumber [http://rs.tdwg.org/dwc/terms/catalogNumber]; collectionCode [http://rs.tdwg.org/dwc/terms/collectionCode]
alternateIdentifiers (alternateIdentifier, identifierType)				catalogNumber [http://rs.tdwg.org/dwc/terms/catalogNumber]; otherCatalogNumbers [http://rs.tdwg.org/dwc/terms/otherCatalogNumbers]; recordNumber [http://rs.tdwg.org/dwc/iri/recordNumber]
supplementalMetadata (record)			/spec:SF_Specimen/gml:metaDataProperty/gml:GenericMetaData can reference metadata in any dialect for more details.	fieldNotes [http://rs.tdwg.org/dwc/terms/fieldNotes]
geoLocations (geoLocation, geometry, geometryType, sridType)	latitude	lat_lon	SamplingLocation /spec:SF_Specimen/sp mplingLocation/gml:P Line | Polygon	

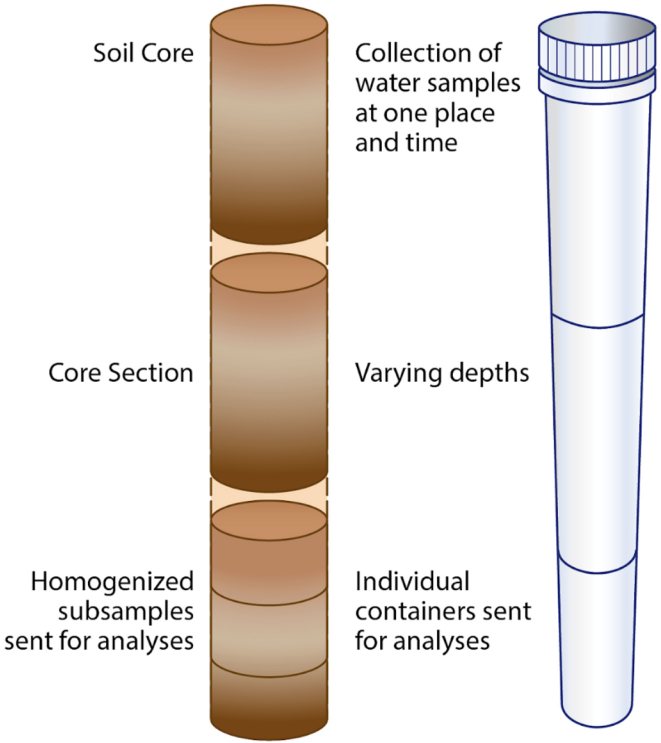
Sample Metadata Translation table

Translation table of related standards and vocabularies - integration with broader community

Interoperability for related geo and bio samples



Assigning IGSNs to Subsamples



Relationship	1. Standard IGSNs	2. Extend primary IGSN	3. Collection or Site ID linked to IGSN
FeatureOf Interest (Parent)	*IGSN:* IEBWE0001	*IGSN:* IEBWE0001	Collection ID: S19S_0006
isSampleOf (Parent/child)	*IGSN:* IEBWE000P IEBWE000Q	*Extensions:* IEBWE00011M IEBWE00012M	*IGSN:* IEBWE000P IEBWE000Q
isSampleOf (child)	*IGSN:* IEBWE000R IEBWE000S IEBWE000T ...	*Extensions:* IEBWE00012MA IEBWE00012MB IEBWE00012MC ...	*Extensions:* IEBWE000PA IEBWE000PB IEBWE000PC ...

Identifier Efficiency and Contextual Information

Project Name: Concise Name for the Project
- Purpose

Collection ID:
Collection of Samples
- Purpose
- Release Date
- Chief Scientist
- Country
- Biome

Site ID: Site or Sampling Feature (e.g. well)
- Location Description
- Latitude
- Longitude
- Coordinate Uncertainty
- Physiographic Feature

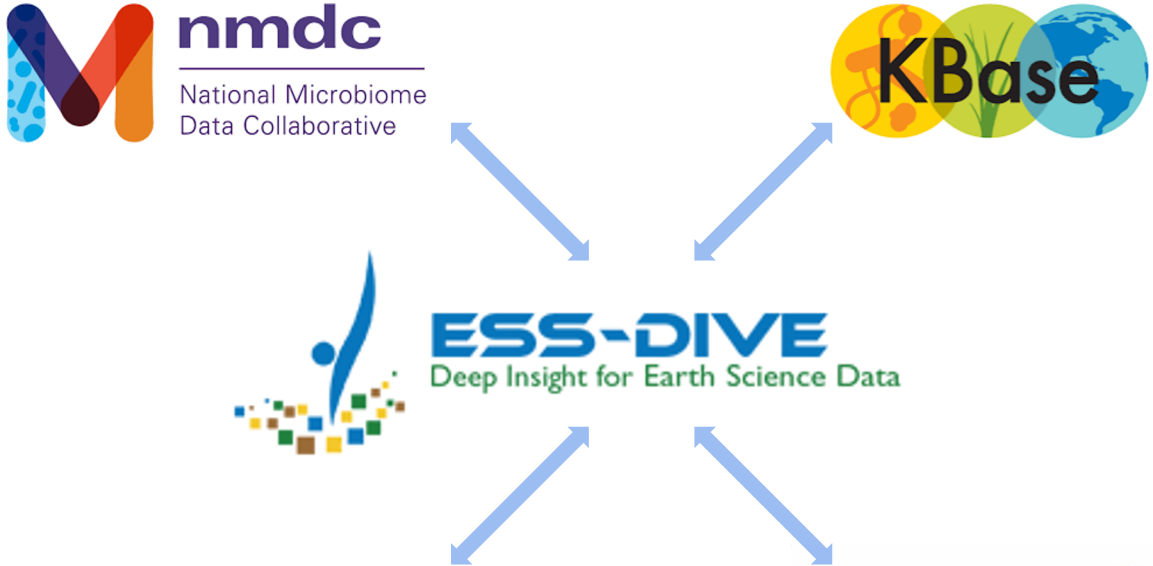
Method ID:
Collection of Samples
- Collection Method Description

Event ID:
Sampling Event
- Collection Date
- Collector

Sample Name and IGSN:
- Material
- Classification
- Sample Description
- Minimum Depth
- Maximum Depth
- Processing Details
- **Project Name**
- **Collection ID**
- **Site ID**
- **Event ID**

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

Coordinating on Samples Across Facilities and Data Systems



Persistent Identifier Benefits

1. **Link** and **expand** access pathways

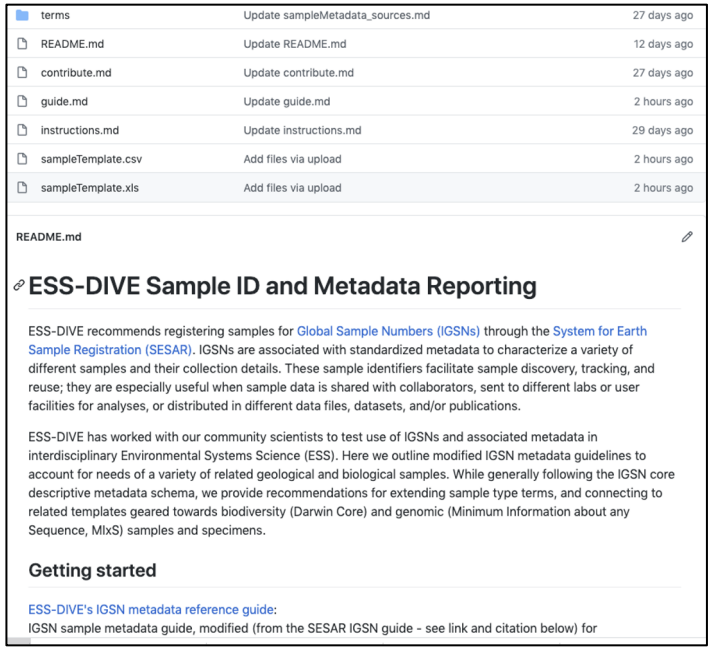
1. **Avoid duplication** of information across platforms

Sample Documentation on Github/Gitbook

ESS-DIVE Sample Documentation

Github: https://github.com/ess-dive-community/essdive-sample-id-metadata

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



terms	Update sampleMetadata_sources.md	27 days ago
README.md	Update README.md	12 days ago
contribute.md	Update contribute.md	27 days ago
guide.md	Update guide.md	2 hours ago
instructions.md	Update instructions.md	29 days ago
sampleTemplate.csv	Add files via upload	2 hours ago
sampleTemplate.xls	Add files via upload	2 hours ago

README.md

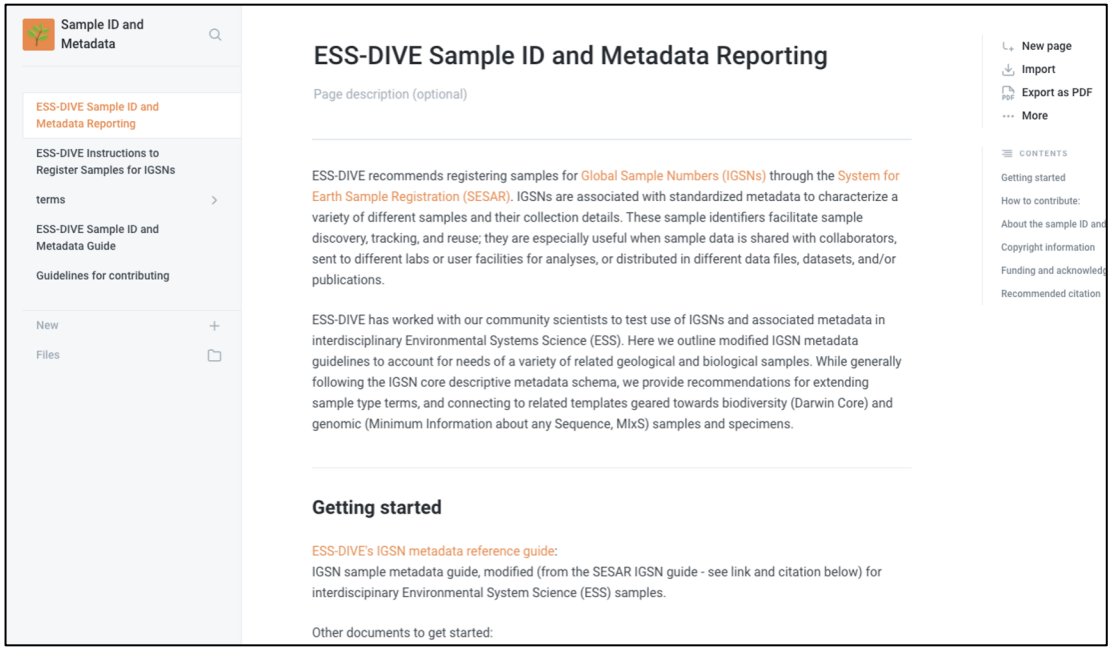
ESS-DIVE Sample ID and Metadata Reporting

ESS-DIVE recommends registering samples for Global Sample Numbers (IGSNs) through the System for Earth Sample Registration (SESAR). IGSNs are associated with standardized metadata to characterize a variety of different samples and their collection details. These sample identifiers facilitate sample discovery, tracking, and reuse; they are especially useful when sample data is shared with collaborators, sent to different labs or user facilities for analyses, or distributed in different data files, datasets, and/or publications.

ESS-DIVE has worked with our community scientists to test use of IGSNs and associated metadata in interdisciplinary Environmental Systems Science (ESS). Here we outline modified IGSN metadata guidelines to account for needs of a variety of related geological and biological samples. While generally following the IGSN core descriptive metadata schema, we provide recommendations for extending sample type terms, and connecting to related templates geared towards biodiversity (Darwin Core) and genomic (Minimum Information about any Sequence, MIxS) samples and specimens.

Getting started

ESS-DIVE's IGSN metadata reference guide:
IGSN sample metadata guide, modified (from the SESAR IGSN guide - see link and citation below) for



Sample ID and Metadata

ESS-DIVE Sample ID and Metadata Reporting
ESS-DIVE Instructions to Register Samples for IGSNs
terms
ESS-DIVE Sample ID and Metadata Guide
Guidelines for contributing
New
Files

ESS-DIVE Sample ID and Metadata Reporting

Page description (optional)

ESS-DIVE recommends registering samples for Global Sample Numbers (IGSNs) through the System for Earth Sample Registration (SESAR). IGSNs are associated with standardized metadata to characterize a variety of different samples and their collection details. These sample identifiers facilitate sample discovery, tracking, and reuse; they are especially useful when sample data is shared with collaborators, sent to different labs or user facilities for analyses, or distributed in different data files, datasets, and/or publications.

ESS-DIVE has worked with our community scientists to test use of IGSNs and associated metadata in interdisciplinary Environmental Systems Science (ESS). Here we outline modified IGSN metadata guidelines to account for needs of a variety of related geological and biological samples. While generally following the IGSN core descriptive metadata schema, we provide recommendations for extending sample type terms, and connecting to related templates geared towards biodiversity (Darwin Core) and genomic (Minimum Information about any Sequence, MIxS) samples and specimens.

Getting started

ESS-DIVE's IGSN metadata reference guide:
IGSN sample metadata guide, modified (from the SESAR IGSN guide - see link and citation below) for interdisciplinary Environmental System Science (ESS) samples.

Other documents to get started:

New page
Import
Export as PDF
More

CONTENTS
Getting started
How to contribute:
About the sample ID and
Copyright information
Funding and acknowledg
Recommended citation

Getting Started

Getting started

ESS-DIVE's IGSN metadata reference guide:
IGSN sample metadata guide, modified (from the SESAR IGSN guide - see link and citation below) for interdiscipinary Environmental System Science (ESS) samples.

Other documents to get started:

- Instructions document: Instructions to register samples for IGSNs through SESAR, and submit related datasets to ESS-DIVE.
- Sample metadata template: Download spreadsheet template with standard fields to register samples for IGSNs.

Instructions: Register for IGSNs

ESS-DIVE Instructions to Register Samples for IGSNs

General information about International Geo/General Sample Numbers (IGSNs) can be found on the System for Earth Sample Registration (SESAR) website.

ESS-DIVE has been testing use of IGSN for the U.S. Department of Energy's Environmental System Science (ESS) community, and has begun outlining recommended modifications to SESAR IGSN terms and some additional metadata fields to include for this community.

Here are instructions on how to register your ESS samples for IGSNs through SESAR, and submit your sample metadata and data to ESS-DIVE.

1. Register and login to SESAR's geopass, and select a user code. The user code should generally be specific to an individual or project. You should use the same user code for all current and future samples. This is important to a.) effectively manage your sample metadata over time (e.g. add links to related publications), and b.) avoid the proliferation of user codes.

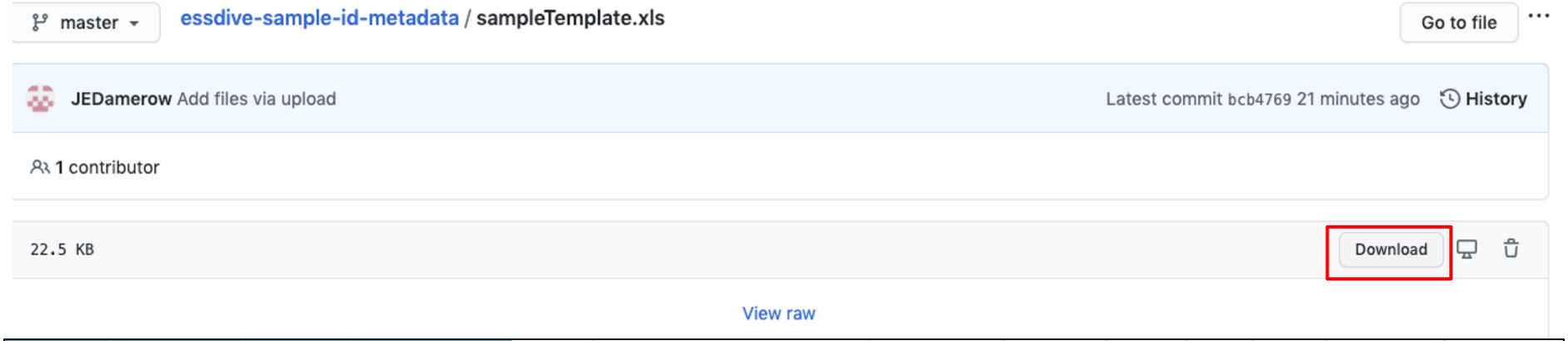
2. Plan the allocation of identifiers for your samples and sample relationships. Within the SESAR IGSN metadata template, sample relationships are captured by specifying the "Parent IGSN" for each IGSN registered. Therefore, parent samples must be registered before the child samples. For ESS-DIVE, we are also now using site IDs, sampling event IDs, and/or sample collection IDs (which should be project-specific and unique IDs). These project-specific IDs can be described in separate files and submitted along with your ESS-DIVE data packages associated with the samples. An example sample "journey map" (documenting sample transport, analyses, and relationships) may help you map out your general sample relationships and decide on the allocation of IDs.

Many projects involved in our testing have expressed a preference to use identifier extensions for highly-related subsamples (e.g. containers sent for multiple analyses that represent a particular sample); You can assign short 1-2 character extensions for subsamples by specifying your own IGSNs and registering those subsamples along with relevenat metadata.

3. Fill out a customized sample template for ESS-DIVE. General fields to describe sample types and locations are required where relevant. Guidance on specific fields were originally developed by SESAR, and modified in some cases by ESS-DIVE. See our sample metadata guide for descriptions and instructions for each metadata field.

Some projects may choose to register samples that have already been collected and analyzed. We are developing R scripts to help facilitate standardization of existing metadata, and will provide these when completed.

Sample Metadata Template Download



Object Type:	Individual Sample	User Code:	IER18									
Sample Name	IGSN	Site Identifier	Sample Description	Material	Collection method description	Longitude	Latitude	Elevation U	Elevation	Field nam	Location Descrip	Collection Date
001-ER18-FO	IER18005A	001-ER18	Foliar leaf samples,Foli	Biology	Collect entire stems where possible and	-106.986098	38.957229	meters	2908.8	Meadow	RMBL Research	2018-06-14
002-ER18-FO	IER18005B	002-ER18	Foliar leaf samples,Foli	Biology	Collect entire stems where possible and	-106.9861	38.957288	meters	2908.8	Meadow	RMBL Research	2018-06-14
003-ER18-FO	IER18005C	003-ER18	Foliar leaf samples,Foli	Biology	Collect entire stems where possible and	-106.98616	38.957368	meters	2908.2	Meadow	RMBL Research	2018-06-14
004-ER18-FO	IER18005D	004-ER18	Foliar leaf samples,Foli	Biology	Collect entire stems where possible and	-106.986105	38.957477	meters	2908.3	Meadow	RMBL Research	2018-06-14
005-ER18-FO	IER18005E	005-ER18	Foliar leaf samples,Foli	Biology	Collect entire stems where possible and	-106.986231	38.957461	meters	2907.8	Meadow	RMBL Research	2018-06-14
006-ER18-FO	IER18005F	006-ER18	Foliar leaf samples,Foli	Biology	Collect entire stems where possible and	-106.984834	38.957868	meters	2915.4	Meadow	RMBL Research	2018-06-14
007-ER18-FO	IER18005G	007-ER18	Foliar leaf samples,Foli	Biology	Collect entire stems where possible and	-106.984753	38.957824	meters	2916.4	Meadow	RMBL Research	2018-06-14
008-ER18-FO	IER18005H	008-ER18	Foliar leaf samples,Foli	Biology	Collect entire stems where possible and	-106.984833	38.957823	m				2018-06-14
009-ER18-FO	IER18005I	009-ER18	Foliar leaf samples,Foli	Biology	Collect entire stems where possible and	-106.984502	38.957937	m				2018-06-14

Completed Sample Metadata Template

Sample Metadata Content - Link to More Details

*Required fields are marked with an asterisk, and indicated in detailed tables below.

Header Rows:
Object Type* | User Code*

Sample IDs and Related Identifiers:
Sample Name* | Other names | IGSN | Parent IGSN | Collection ID | Event ID | Location ID

Sample Description:
Material* | Field name informal classification | Sample Description | Purpose | Size | Size unit | Filter Size | Scientific Name | Sample Remarks

Sample Collection Details:
Collector Chief Scientist* | Collection Date* | Collection Time | Collection Method Description* | Sample Processing | Field Program or Project Name*

Location:
Latitude Coordinate system WGS 84* | Longitude Coordinate system WGS 84* | Coordinate Uncertainty In Meters | Navigation Type | Location Description | Country | Elevation start | Elevation end | Elevation unit | Minimum Depth in Meters | Maximum Depth in Meters | Depth scale | Minimum Distance above Surface in Meters | Maximum Distance above Surface in Meters |

Environmental Context:
Physiographic feature* | Biome

Sample Access:
Release Date* | Current Archive | Current Archive Contact

CONTENTS

Characterize Sample/Object Type

Object type, Material

Environmental context - local/physiographic feature, biome

Object Type

ESS-DIVE Proposed Element	objectType `Required`
Definition	Broad characterization of the nature of a sample or specimen.
Format	SESAR Controlled List. See object type list for revised terms proposed for ESS-DIVE
Additional Instructions	Provide feedback on additional terms or revisions needed.
Examples	Core; Individual Sample; Organism

Object Type Vocabulary

general natural of sample

Any additional terms?

ESS-DIVE Object Type	SESAR IGSN Label	Description
Core	Core	A long cylindrical sample, usually of soil or sediment (for Environmental System Science).
Core Section	Core Section	Segments of a "core", which can be arbitrarily cut or cut to obtain specific depths.
Cuttings	Cuttings	Loose, coarse, unconsolidated material suspended in drilling fluid.
Dredge	Dredge	A group of rocks collected by dragging a dredge along the seafloor.
Experimental Specimen	Experimental Specimen	A synthetic or natural material used during an experiment.
Grab	Grab	A sample (sometimes mechanically collected) from a deposit or area, not intended to be representative of the deposit or area.
Material Sample	Individual Sample	A sample that is an individual unit of material, including soil or sediment samples, biological material, or a bottle of fluid. A physical result of a sampling (or subsampling) event.
Filtrate	NA	A sample that has gone through a filtration process to separate solids from fluids (liquids or gases), using a filter medium through which only the fluid can pass. *Must be associated with a filter size.*
Material Captured in Filter	NA	A material sample captured in filter, for example from a water sample that was filtered. Must be associated with filter size field. *Material options include: "particulate matter" or "particulate matter: organic particulate matter"* .
Organism	NA	A whole specimen of a biological organism. A particular organism or defined group of organisms considered to be taxonomically homogeneous.
Part of organism	NA	A specimen of a portion of a biological organism (e.g. leaf, root, stem).
Collection of organisms	NA	A material entity that consists of two or more organisms, viruses, or viroids.
Site	Site	A place where a sample is collected.

Characterize Sample Material

Material

Proposed ESS-DIVE Element	material `Required`
Definition	Material that the sample consists of.
Format	SESAR controlled list.See ESS-DIVE's proposed material terms from Environment Ontology (ENVO)
Additional Instructions	ESS-DIVE is requesting additional terms for organisms, organic material, and water samples. Please provide feedback on any other terms needed.
Examples	soil; sediment; surface water ENVO:00002042; groundwater ENVO:01001004

Material Vocabulary

Water
Organic material
Plant structure

Is your sample material represented?

...

liquid environmental material: liquid water.	Liquid>aqueous	An environmental material primarily composed of dihydrogen oxide in its liquid form.
liquid environmental material: liquid water: surface water	NA	Water that is found on the surface of an astronomical object.
liquid environmental material: liquid water: underground water: groundwater	NA	Underground water which is located in pore spaces found in rock or unconsolidated deposits such as soil, clay, or gravel.
organic material	NA	Environmental material derived from living organisms and composed primarily of one or more biomacromolecules.
organic material: biofilm material	NA	Material derived from a biofilm, an aggregate of microorganisms in which cells adhere to each other and/or to a surface. These adherent cells are frequently embedded within a self-produced matrix of extracellular polymeric substance (EPS). Biofilm EPS, which is also referred to as slime, is a polymeric conglomeration generally composed of extracellular DNA, proteins, and polysaccharides in various configurations.
organic material: plant litter	NA	Leaf litter is dead plant material, such as leaves, bark, needles, and twigs, that has fallen to the ground.
organic material: soil organic matter	NA	Organic material in soil, which consists of plant and animal residues at various stages of decomposition, cells and tissues of soil organisms, and substances synthesized by soil organisms.
plant structure: leaf	NA	A phyllome (PO:0006001) that is not associated with a reproductive structure.

Contributing Feedback

Getting started

ESS-DIVE's IGSN metadata reference guide:
IGSN sample metadata guide, modified (from the SESAR IGSN guide - see link and citation below) for interdiscipinary Environmental System Science (ESS) samples.

Other documents to get started:

- Instructions document: Instructions to register samples for IGSNs through SESAR, and submit related datasets to ESS-DIVE.
- Sample metadata template: Download spreadsheet template with standard fields to register samples for IGSNs.

How to contribute:

ESS-DIVE is a community-focused data repository. Our sample reporting format must fit the needs of ESS data contributors and users, and we need your feedback. See our guide on how to contribute to ESS-DIVE's sample metadata reporting format.

Submit an Issue
...or
Contact us

ess-dive-support@lbl.gov

Issue templates:

- New Term
- Change Term
- Change Documentation

Guidelines for contributing

Here's how you can contribute to ESS-DIVE's proposed sample ID and metadata reporting format:

- Want to make a change to the reporting format? Copy and paste one of the **issue templates** below and submit a new issue.
- Have a question? Contact us at ess-dive-support@lbl.gov.

Citation

These issue templates were modeled from that provided by Darwin Core:

Darwin Core maintenance group, Biodiversity Information Standards (TDWG) (2014). Darwin Core. Zenodo. https://doi.org/10.5281/zenodo.592792

Issue templates

New term

Copy and paste this template into your new issue if you want to create a term that does not exist yet.

```
## New term

* Submitter:
* Justification (why is this term necessary?):

Proposed definition of the new term:

* Term name (in lowerCamelCase):
* Definition of the term:
* Format of the term:
* Additional Instructions (recommendations regarding content, etc.):
* Examples:
* Refines (name of the broader term this term refines, if applicable):
* Replaces (name of the existing term that would be replaced by this term, if applicable):
```

Write Preview

Attach files by dragging & dropping, selecting or pasting them.

Styling with Markdown is supported

Submit new issue

ESS-DIVE

Summary of Sample ID & Metadata Resources

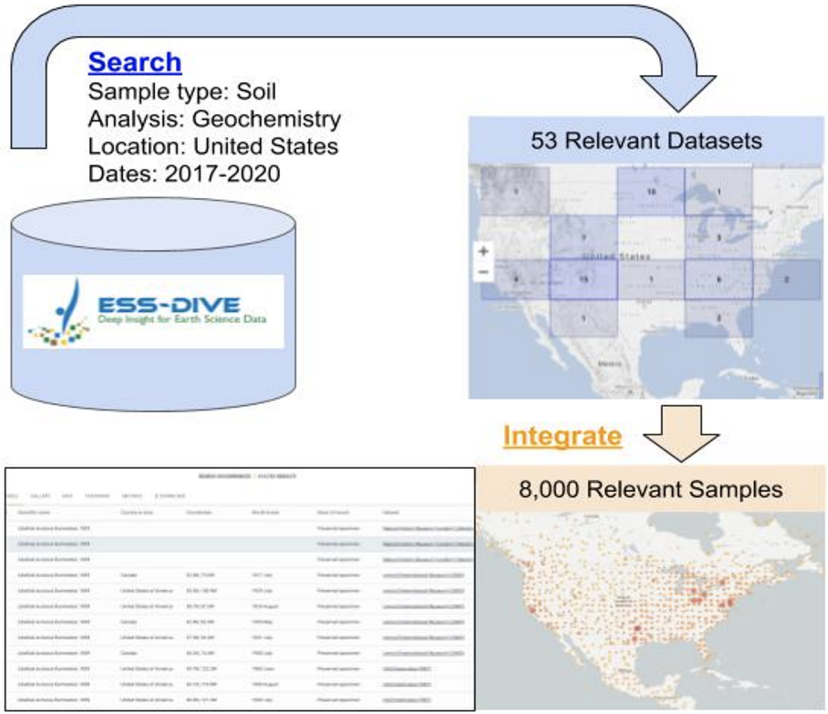
Github: https://github.com/ess-dive-community/essdive-sample-id-metadata

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

Primary Content

- Metadata reference guide
- Instructions to register IGSNs
- Sample metadata template
- Guidelines for contributing - provide feedback, revisions

Summary of Benefits



- Supports sample tracking
- Enables future community sample tools (e.g. sample portal, fusion database)
- Promote reuse: long-term preservation, discoverability, integration of sample data
- Multidisciplinary sample documentation - global search

Next Steps

- Help test our new reporting format for usability: email JoanDamerow@lbl.gov
- Contribute by submitting issues on Github

ESS-DIVE

- Tools to help standardize and quality-check metadata
- Future: Fusion database for sample search and integration

PollEv.com/joandamerow220