

IGSN IDs: Use & Registration

DataCite Training

Rorie Edmunds & Cody Ross

22 September 2022
DataCite Member Meeting 2022



[@datacite](https://twitter.com/datacite)



What is the IGSN ID?

‘The IGSN ID is a globally unique and persistent identifier for material samples.’

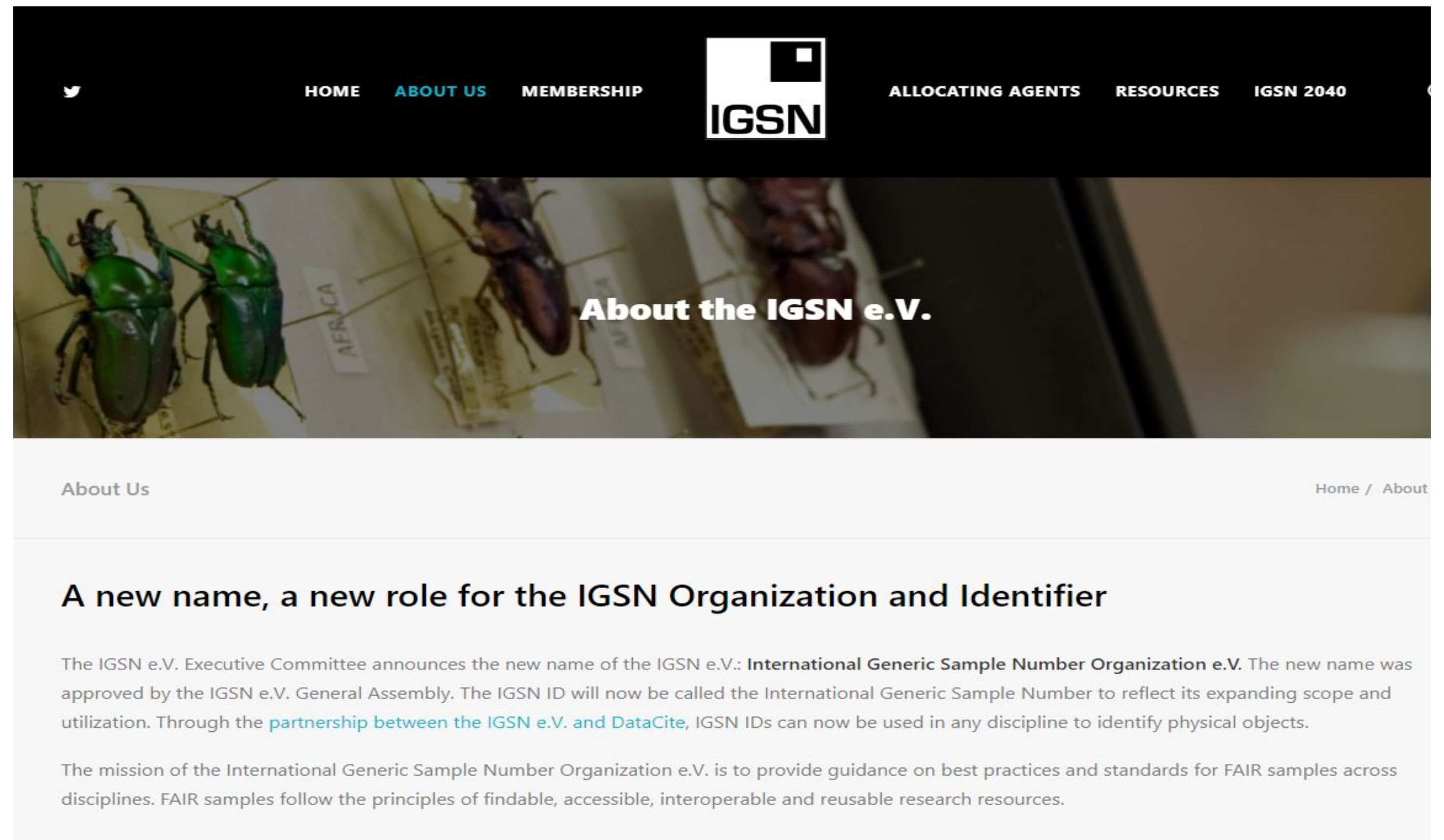


‘The IGSN ID is domain-agnostic. Samples can be any material from anywhere in the universe, not only from Earth.’



International *Generic* Sample Number

- IGSN ID grew out of the Geological Sciences
- But...it is a PID for all Physical Objects, not only for Geosamples!
- Over 10 million IGSN IDs have been registered thus far
- Used by major research centres, universities, and government geological surveys, and endorsed by scientific publishers



Disciplines & Material Sample Types

- Physical Sciences (Earth, Space, and Environmental Sciences; Chemistry)
 - Rocks/minerals/metals
 - Radioactive elements
 - Soil/sediment (core & individual)
 - Mud
 - Dust
 - Petroleum/oil
 - Condensate/water/Ice cores
 - Air/atmospheric/gas
 - Environmental DNA
 - Astromaterials
 - Other chemical samples (experimental samples?)



Disciplines & Material Sample Types

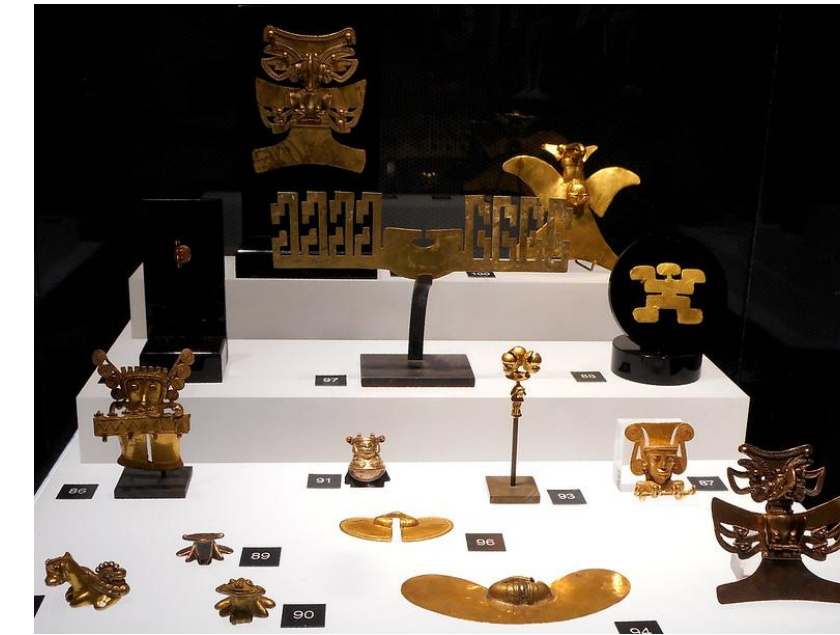
- Life Sciences (Biochemistry, Microbiology, Botany, Zoology, Ecology)

- Biofluids
- Tissues
- Cells
- Genomic
- Animal specimens
- Insect specimens
- Plant specimens
- Fungi/spores
- Microorganisms/biofilms
- Vegetation
- Botanical/herbarium samples



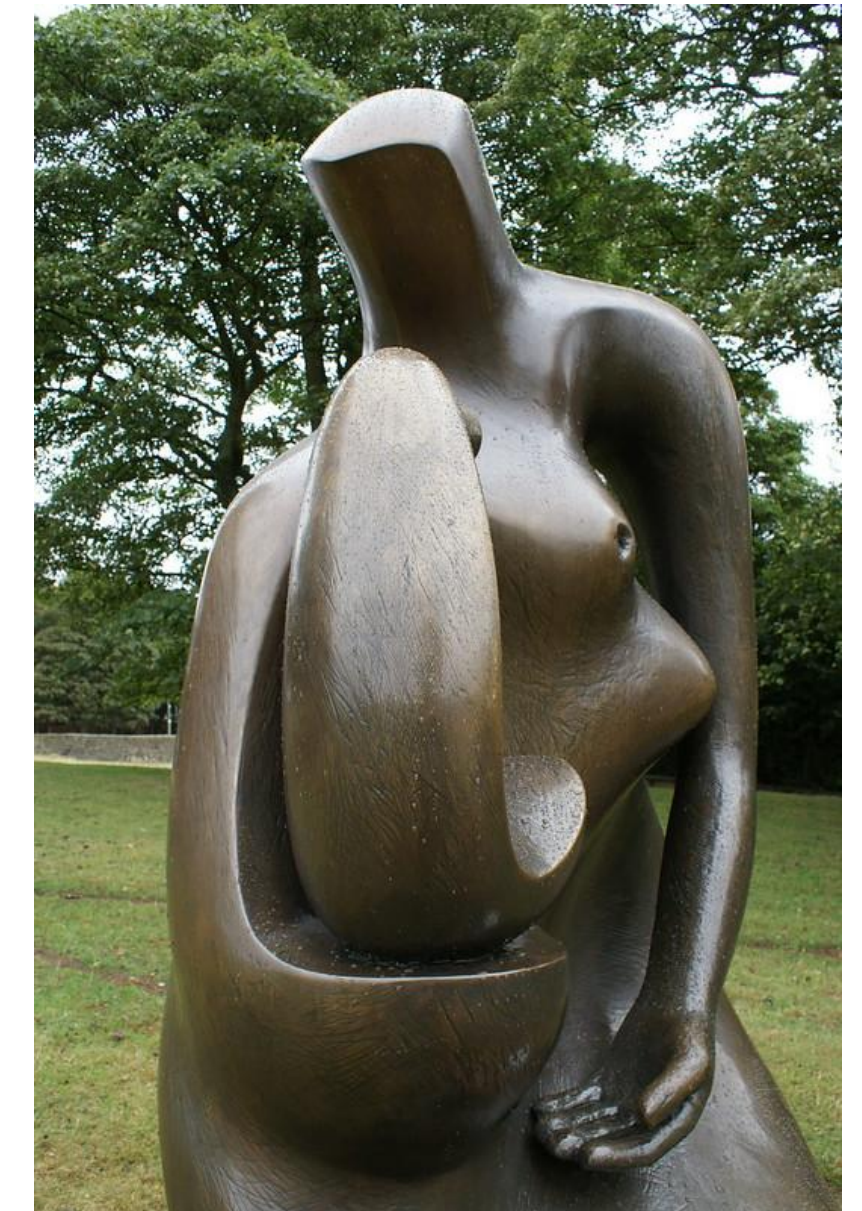
Disciplines & Material Sample Types

- Social Sciences (Sociology, Anthropology, Archaeology, Economics, Human Geography,...)
 - Human/animal/insect/plant remains
 - Fossils
 - Artefacts (tools, weapons, utensils, machines, ornaments, art, buildings, monuments, written records, religious images, clothing,...)
- Applied Sciences (Agriculture, Material Science & Engineering, Medicine)
 - Seed accessions
 - Crop & livestock samples
 - Soil testing
 - Synthetic materials
 - Prototypes



Disciplines & Material Sample Types

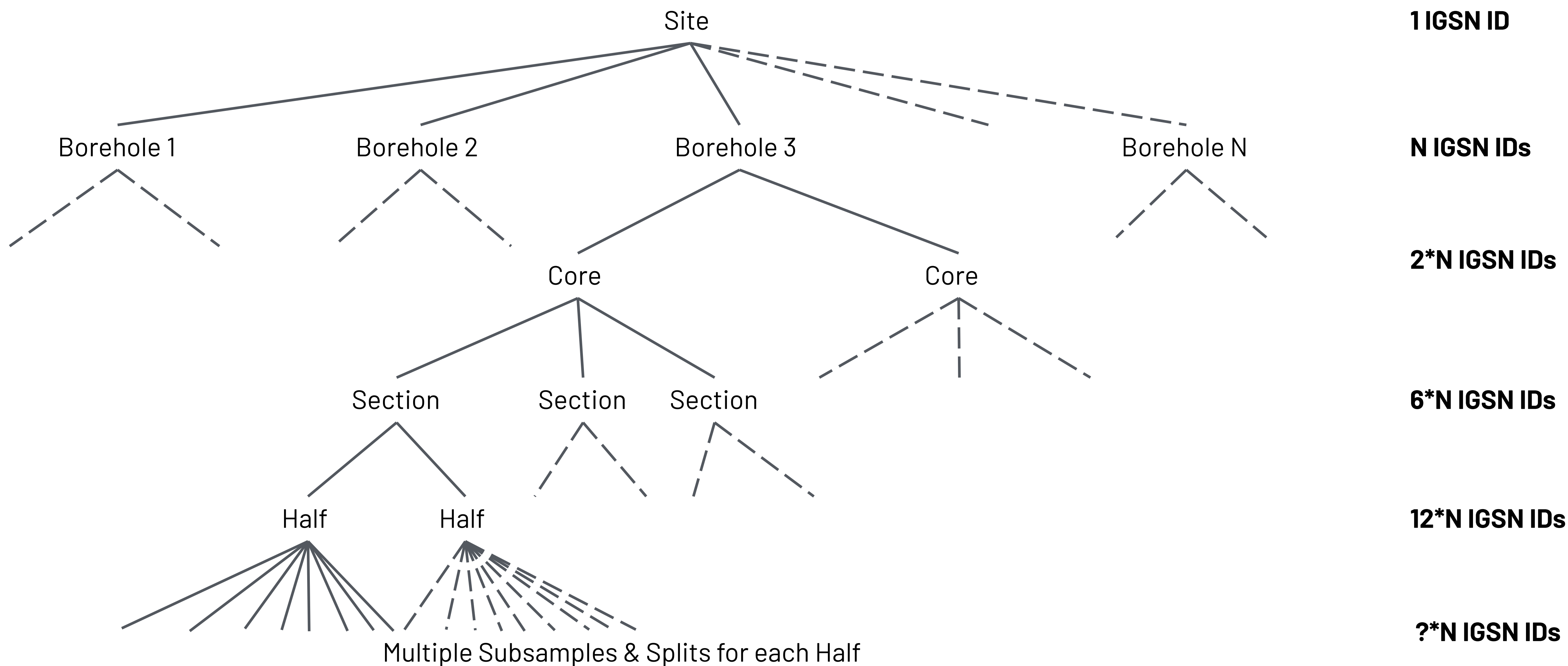
- The Arts
 - Architecture/Ceramics/Sculptures
 - Drawings/paintings/photographs
 - Literary and musical samples
 - Musical instruments
 - Fashion/Costumes/theatre sets
- ...



IGSN IDs May Also Be Used For

- Collections/aggregates
 - May not want to have a PID for every object, but rather the collection
 - Scale of material samples PIDs can be much larger than for data
- Features-of-interest
 - Collection sites (e.g., borehole, well,...)
 - From ISO 19156:2011 and the SOSA ontology, a specimen is a specialization of a 'feature-of-interest' upon which the sampling activity was carried out
 - A sample is representative of its feature-of-interest, and can be consider a child (i.e., subsample)
- Destroyed/discarded samples
 - The material sample itself does have to be persistent. Samples may be destroyed in an analytical process or discarded
 - Metadata should provide information on the current status of the sample

Scale of Material Samples



Use Cases & Incorporating IGSN IDs into Workflows

Use Case 1 – Linking to the Web

- Why? Open Science
 - Sample exists and information is available
 - Past investment remains useful and supports future research
 - Unique samples preserved long term alongside data related to them
- Digital representation of a sample is its landing page, which displays a description of the sample identified by the IGSN ID
- Presentation of metadata on landing pages may differ among portals and catalogues, but should include elements that improve discoverability
 - Sample images
 - Maps
 - Hierarchical chart outlining relationships
 - QR code of URL to add to sample labels
- Parts of a metadata record may be withheld to protect sensitive information; for example, of vulnerable sites



General Identifiers

Program:	ICDP
Expedition:	ICDP 5054
Type:	Hole
Name:	5054_1_A
IGSN:	ICDP5054EEW1001
Parent IGSN:	N/A
Release Date:	2017-4-1

Sampling Location

Latitude:	63.4063
Longitude:	13.203057
Coordinate System:	WGS84
Elevation:	522
Final Depth:	-1980.8
Location Type:	N/A
Location Name:	Are, Jämtlands lään, Sweden
Location Description:	COSC-1 is located in the vicinity of the abandoned Froea mine
Country:	Sweden
Province:	Jämtlands lään
County:	N/A
City:	Are

Geology

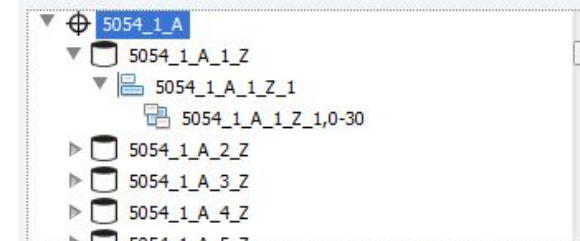
Material:	Rock
Rock Classification:	metamorphic rocks
From Corrected Depth:	102.7
To Corrected Depth:	2502.8
Depth Reference:	meter below ground level
Geological Age:	mid-paleozoic
Geological Unit:	N/A

Drilling

Drilling Method:	Coring>RockCorer wireline diamond coring, HQ and NQ bit size
Operator:	Lund University, Engineering Geology Larsson Drilling Consulting AB
Funding Agency:	Swedish Research Council (Vetenskapsrådet)
Total Length:	2400.1m

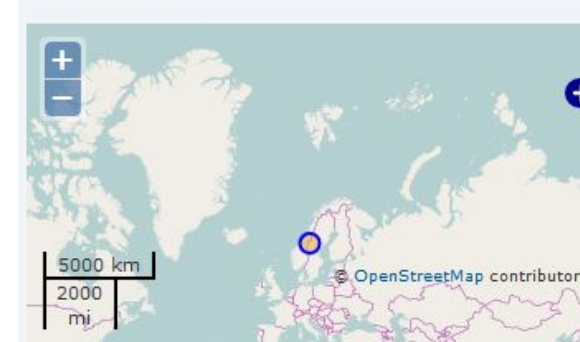


Sample Family



The Sample Family shows a sub-sampling graph. Select entries to navigate samples. Core-Samples are issued to scientists on request. The naming convention for a Core-Sample is: *Expedition_Site_Hole_Core_Section,from-to(cm)*. Hole, Core, and Core-Section are following the same schema respectively.

Location Map



Drilling Start/End: 2013-9-5 / 2014-8-26 *
Latitude: 63.40630 * Longitude: 13.20306 *
Are, Jämtlands lään, Sweden

Publications & Datasets

Lorenz, H., Rosberg, J.-E., Juhlin, C., Bjelm, L., Almqvist, B. S. G., Berthet, T., ... Tsang, C.-F. (2015). COSC-1 – drilling of a subduction-related allochthon in the Palaeozoic Caledonide orogen of Scandinavia. *Sci. Drill.*, 19, 1-11. doi:10.5194/sd-19-1-2015

Use Case 2 – Locating

- Why? Collection management
 - Archiving: What do you have? What can be made discoverable?
 - Support for research: Ensure available resources are discoverable internally and externally
 - Knowledge sharing: Often no common, standardized infrastructure for fieldwork and sample management may be rudimentary. Knowledge is with PIs and disappears with them
- Simple ways to link a material sample with its digital representation:
 - Permanently affix a label to the sample
 - Write/engraving the IGSN ID onto the sample or its container
 - Can include the sample's local accession or inventory number
- Convenient to use QR codes/barcodes that encode an IGSN ID as an actionable URI, and offers machine-readable identification
- Ideally, a label will show the QR code, the IGSN ID, and any inventory number in a human-readable way
- Working on best practice for labelling



Use Case 3 – Tracking

- Why? Sample process management
 - Local systems for unique identification of samples have limited scope. Samples names may be ambiguous outside an organization
 - Samples names can often change at each stage of the sample management process
 - Difficult to track samples across institutional and system boundaries
- Most effective way to avoid ambiguity is to apply IGSN IDs as early as possible
 - Use electronic field notebooks to document sampling process and assign IGSN IDs as part of the field sampling activity
 - Encode IGSN ID suffixes to match a specific sampling campaign
- Enables samples to be tracked
 - Through different stages of the life cycle (e.g., sample handling and storage, laboratory analysis, and eventual disposal)
 - If they are moved to different laboratories or repositories
- Unless all different stations/tools/instruments in a lab are connected to a centralized barcode system, one is still relying on people to track samples, and human errors can occur

Example: Optimized Sampling

- A geochemical field campaign by CSIRO in the Nullarbor Desert
 - 6 mins per site (+4 mins travel/refuel)
 - 280 samples on 4 km spacing 80 x 50 km area
 - 5 sample types collected at each site
- In the field: IGSN IDs/FAIMS integrated in tablets and QR code labels
 - IGSN ID sample labels pre-printed with QR codes
 - QR codes read into FAIMS field data app
 - Data capture & backup—no lost/missing sites or matching up photos/bags
- The streamlined process improved the efficiency of sample handling by 100%



<https://faims.edu.au/>



Use Case 4 – Cross-linking & Citation

- Why? Discovery and credit
 - Difficult to unambiguously link samples with associated data and literature
 - Difficult to unambiguously link parents and derived children
 - Citation of samples is not consistent in the literature
- IGSN IDs enable relations to be described between a material sample, and datasets about the sample or publications in which the sample is referenced
 - Samples are cross-linked to other entities by referencing their PIDs and describing the nature of the relationship via `relatedIdentifierType`
 - Integrate in references and datasets as active weblinks to link online sample descriptions, and enable discovery of contextual information
 - Include in dataset metadata, and thus enable discovery in catalogues of research data repositories and portals harvesting metadata
- Working on best practices for citations and data tables

Use Case 5 – Features-of-interest

- Why? Linking collection sites and samples
 - Want to identify related entities closely linked to material samples (e.g., boreholes, mines, outcrops, or other sites)
- IGSN IDs can be used for physical objects that are not 'samples' themselves, but features-of-interest from which a number of samples could have been taken
- Individually identify each feature-of-interest and the samples, which are 'subsamples' of the feature-of-interest. IGSN IDs can then be related to each other using `relatedIdentifier` and mirroring the hierarchical relationships

IGSN IDs in Sample Workflows

- When should one assign an IGSN ID?
 - In the field
 - Pre-assign and batch issue IGSN IDs in 'Draft' state—may have limited metadata or internet connection
 - Encode IGSN IDs in QR codes/barcodes
 - Use field-based tools such as FAIMS
 - When they arrive at the facility
 - During the clean-up process
 - During the analytical process
 - During ingest to a repository for long-term storage
 - Before publication of research results—can incorporate embargo period
- QR code/barcode readers could be installed at each stage to track progress
- Integration in analytical systems such that they have access to samples descriptions
- Incorporate local naming conventions in use to transform locally unique identifiers to globally unique ones. Can be generated without changes to established working procedures, naming conventions, or data systems

**Assign as needed
and bulk register**

Registration: Hands-on Training

Hands-on Overview

During the hands-on, we will:

- Create an IGSN ID repository
- Review an IGSN ID
 - Based on real sample database metadata
 - Using the Fabrica form
 - Using IGSN ID metadata recommendations


DataCite Fabrica Test

[About](#) [Support](#)


[DataCite](#) / [Members](#) / [DataCite Consortium](#) / [Consortium Organizations](#) / [Cooper University](#) / [Repositories](#)


Borehole Repository

[Info](#) [Settings](#) [Prefixes](#) [DOIs](#)

 Set Password

 Update Repository

 Delete Repository

 Transfer

Repository ID ?

HNAJ.EDIUNV

Description

This is an IGSN ID Repository and is only used to register IGSN IDs.

System Email

cody.ross@datacite.org

Domain ?

*

Record created
August 19, 2022 at
19:00:07 UTC

Record last modified

<https://support.datacite.org/>

Creating an IGSN ID Repository

Description (optional)

An IGSN ID repository for sediment samples.

Description of the repository. Please use markdown for formatting.

To create an IGSN ID Repository:

- Create a new repository in Fabrica.
- Enter the words 'IGSN ID' in the repository Description field.
- Email support@datacite.org with the repository ID and prefix of your IGSN ID repository.
- This flags your repository as an IGSN ID repository, helps DataCite and users track IGSN ID registration, and helps us keep in touch with users as we move forward with the IGSN partnership.

Creating an IGSN ID

URL

<https://dataservices.gfz-potsdam.de/igsn/esg/index.php?igsn=GFTH1000K>

Metadata

Summary View 

LC21, shoveled bulk sample, fluvial sediment PhysicalObject

Simon Terweh,

Material Sample published 2018 via BfG (Bundesanstalt f. Gewaesserkunde)

 <https://handle.stage.datacite.org/10.82389/gfth1000k>






Citation

APA 

Terweh, S. (2018). *LC21, shoveled bulk sample, fluvial sediment*. BfG (Bundesanstalt f. Gewaesserkunde). <https://doi.org/10.82389/GFTH1000K>

- Now we're going to go over the metadata of an IGSN ID we've created using the Fabrica form.
- IGSN IDs are functionally DOIs, so you can use the Fabrica form as well as the DataCite REST and MDS APIs to create IGSN IDs.
- This IGSN ID is based on an existing sample in GFZ's sample database.
- We'll go over mainly the mandatory properties according to the recommendations of the IGSN-DataCite Crosswalk and Metadata Management Working Group.

Creating an IGSN ID

		 Physical Sample		Impressum HELMHOLTZ CENTRE POTSDAM GFZ GERMAN RESEARCH CENTRE FOR GEOSCIENCES					
General Identifiers									
Project:	4								
Campaign:	ST_LaCampana_2018								
Type:	Specimen								
Name:	LC21								
IGSN:	GFTTH1000K								
Parent IGSN:	N/A								
Release Date:	N/A								
Sampling Location									
Latitude:	-32.942558								
Longitude:	-71.084417								
Coordinate System:	WGS84								
Elevation:	503.07								
Location Type:	N/A								
Location Name:	LC21_BS								
Location Description:	River deposit								
Country:	Chile								
Province:	Valparaíso								
County:	N/A								
City:	N/A								
Acquisition									
Material:	Sediment								
Sediment Classification:	sediment:fluvial								
Sediment Description	for grain size analysis, shoveled bulk sample from fluvial sediment, coarse fraction sieved in the field, fine fraction taken to lab								
Collection Method:	manual								
Funding Agency:									
Comments:	N/A								
Chief Scientist:	Simon Terweh								
		<div>Sample Family</div> <div>  LC21 </div> <div>  = Specimen The Sample Family shows a sub-sampling graph. Select entries to navigate samples. </div>							
		<div>Location Map</div> <div>  <p>© OpenStreetMap contributors</p> <p>1000 km / 1000 mi</p> </div> <div>Drilling Start/End: 2018-03-14 / 2018-03-14 * Latitude: -32.94256 * Longitude: -71.08442 * LC21BS</div>							

- We strongly encourage you to fill in as many properties as you can to improve discovery.
- You might not have metadata immediately available if you are registering IGSN IDs in the field or in advance of sample collection. If you don't have any metadata about your sample available, you may use draft state or the standard values for unknown information from the DataCite schema. Both of these are described in our Support site documentation.

Creating an IGSN ID Suffix

* **DOI** The globally unique string that identifies the resource and can't be changed.

10.82389/gfth1000k

- The suffix can be the random string provided by the system, but it can also be edited to be more human readable or more usable to those working with the sample—for example, in the field.
- In this case, the suffix is simply the IGSN ID suffix already assigned to the sample.
- The IGSN e.V. plans to develop disciplinary best practices for DOI suffixes in collaboration with a variety of samples communities as they begin to register IGSN IDs with DataCite.

Creating an IGSN ID URL

*** URL** The location of the landing page with more information about the resource.

`https://dataservices.gfz-potsdam.de/igsn/esg/index.php?igsn=GFTH1000K`

Should be a https URL – within the allowed domain(s) of your repository if domain restrictions are enabled in the repository settings. Http and ftp are also supported.

- For the URL, we've entered the landing page of the sample. This is where the IGSN ID will resolve.

Creating an IGSN ID Creators

*** Creators** The main researchers or organizations involved in producing the resource, in priority order.

Name Identifier

Name Identifier

Uniquely identifies an individual or legal entity, according to various schemas, e.g. ORCID, ROR or ISNI. Use name identifier expressed as URL. The Given Name, Family Name and Name will automatically be filled out for ORCID and ROR identifiers.

+ Add another name identifier

☒ Person ☐ Organization ☐ Unknown

Given Name

Simon

The personal or first name of the creator.

Family Name

Terweh

The surname or last name of the creator.

*** Name (from Given Name and Family Name)**

Terweh, Simon

- For IGSN IDs, a creator could be the sample collector/creator, chief scientist, curator, or even the person who deposited the sample into a repository. We have entered the chief scientist.
- Normally, we would fill in other fields here with name identifiers, affiliations, and more to make our IGSN ID more discoverable.

Creating an IGSN ID

Titles

*** Titles** One or more names or titles by which the resource is known.

LC21, shoveled bulk sample, fluvial sediment

Title Type

Select Title Type ▼

Language

English × ▼

- The title property should include appropriate elements that would help find and distinguish a sample. The exact syntax is at your discretion, but you should consider discoverability when formatting your title.
- In this case, we've included a few recommended items from our documentation: the local sample identifier (LC21), the basic form of the object (shoveled bulk sample), and the material (fluvial sediment).

Creating an IGSN ID Publisher

*** Publisher**

The name of the entity that holds, archives, publishes prints, distributes, releases, issues, or produces the resource.

BfG (Bundesanstalt f. Gewaesserkunde)

This property will be used to formulate the citation, so consider the prominence of the role.

- For IGSN IDs, the publisher should be the organization registering the IGSN ID for the physical sample. Here, we've entered the current repository, BfG.

Creating an IGSN ID

Publication Year

*** Publication Year** The year when the resource was or will be made publicly available.

Must be a year between 1000 and 2022.

- The publication year property should contain the year when the sample was first made available to the research community. This is likely to be the year at the time the physical sample is registered. Here, we've assumed it is the same year as the year of acquisition

Creating an IGSN ID

Resource Type General & Resource Type

* Resource Type General

The general type of the resource.

Physical object

If none of the provided values matches, use Other and specify the resource type in the field below.

Resource Type

Material sample

A description of the resource, the preferred format is a single term of some detail.

- Resource Type General will always be Physical Object for IGSN IDs.
- The Resource Type property can be populated with resource types from external ontologies or shared vocabularies. We recommend the use of either the term 'material sample' or "feature-of-interest' to distinguish between these sampling concepts. In this case, we're dealing with a material sample, so we've entered 'material sample' in the Resource Type.

Creating an IGSN ID

Dates

Dates Different dates relevant to the resource.

2018-03-14



* Date Type

Collected

▼

Type of date.

Date Information


Date Information

Specific information about the date, if appropriate.

- In the Dates property, we've listed when the sample was collected. We've assumed that's the acquisition date.

Creating an IGSN ID

Dates

2018-03-15 

* Date Type

Other ▼

Type of date.

Date Information

Destroyed

Specific information about the date, if appropriate.

- We can also enter a date that the sample was destroyed using date type 'Other' and date information 'Destroyed'.

Creating an IGSN ID

Related Identifiers

Related Identifiers

Identifiers of related resources.

10.21384/GFTH1001K



Must be a globally unique identifier. Visit our support website for [the list of supported unique identifiers](#).

* Related Identifier Type

DOI



The type of the Related Identifier.

* Relation Type

Is part of



The type of the Relation.

Resource Type General

Physical object



The general type of the related resource.

- The first related identifier specifies that the sample is a subsample of another, larger sample with an IGSN ID identifier.
- In this case, the related identifier type is 'DOI' rather than IGSN because the other identifier is an IGSN ID DOI registered with DataCite as opposed to an IGSN ID registered in the previous IGSN registration systems.

Creating an IGSN ID

Related Identifiers



Must be a globally unique identifier. Visit our support website for [the list of supported unique identifiers](#).

* Related Identifier Type

DOI



The type of the Related Identifier.

* Relation Type

Is cited by



The type of the Relation.

Resource Type General

Journal article



The general type of the related resource.

- We know this IGSN ID has been cited in a journal article with a DOI that discusses findings related to the sample. We've specified that relationship in the second related identifier here. That relationship will be displayed in DataCite Commons.

Creating an IGSN ID

Related Identifiers

`https://dataservices.gfz-potsdam.de/igsn/esg/index.php?igsn=GFTH1000K/metadata`



Must be a globally unique identifier. Visit our support website for [the list of supported unique identifiers](#).

* Related Identifier Type

URL

The type of the Related Identifier.

* Relation Type

Has metadata

The type of the Relation.

- In the physical samples community, metadata about a physical sample may be stored in a metadata format specific to a discipline, community, or institution. We can refer to other metadata resources for our sample using the related identifier field in the DataCite schema.
- For related metadata, the related identifier will often be the URL where users can access the related metadata resource. In this case, the related identifier type will be 'URL'.


Creating an IGSN ID

Alternate Identifiers

Alternate Identifiers

An identifier or identifiers other than the primary Identifier applied to the resource being registered.

LC21



* Alternate Identifier Type

Local sample identifier▼

The type of Alternate Identifier.

- In alternate identifier, we can enter other identifiers for the material sample. Here, we’ve entered the local sample name, LC21.

Creating an IGSN ID

URL

<https://dataservices.gfz-potsdam.de/igsn/esg/index.php?igsn=GFTH1000K>

Metadata

Summary View 

LC21, shoveled bulk sample, fluvial sediment PhysicalObject

Simon Terweh,

Material Sample published 2018 via BfG (Bundesanstalt f. Gewaesserkunde)

 <https://handle.stage.datacite.org/10.82389/gfth1000k>

Citation

APA 

Terweh, S. (2018). *LC21, shoveled bulk sample, fluvial sediment*. BfG (Bundesanstalt f. Gewaesserkunde). <https://doi.org/10.82389/GFTH1000K>

- Once we have entered this and any other relevant metadata about our sample, we can create our IGSN ID.
- If more information about the sample becomes available later, we can update the metadata using Fabrica or the REST and MDS APIs.

Creating an IGSN ID

Enhancing Support

URL

<https://dataservices.gfz-potsdam.de/igsn/esg/index.php?igsn=GFTH1000K>

Metadata

DataCite XML

```
<?xml version="1.0" encoding="UTF-8"?>
<resource
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns="http://datacite.org/schema/kernel-4" xsi:schemaLocation="http://datacite.org/sche
    <identifier identifierType="DOI">10.82389/GFTH1000K</identifier>
    <creators>
      <creator>
        <creatorName nameType="Personal">Terweh, Simon</creatorName>
        <givenName>Simon</givenName>
        <familyName>Terweh</familyName>
      </creator>
    </creators>
    <titles>
      <title xml:lang="en">LC31 - shoveled bulk sample - fluvial sediment</title>
```

- We realize that the DataCite Metadata Schema was not specifically designed for IGSN IDs and material samples metadata, and we are working with the IGSN e.V. and the DataCite Metadata Working Group to improve support.
- We welcome feedback and suggestions about how we might make the Schema more effective for materials samples, so please be in touch.



CONNECTING RESEARCH,
IDENTIFYING KNOWLEDGE



info@datacite.org



pidforum.org



datacite.org
blog.datacite.org



support.datacite.org
support@datacite.org



[@datacite](https://twitter.com/datacite)



[DataCite](https://www.youtube.com/DataCite)



[@datacite](https://www.linkedin.com/company/datacite)