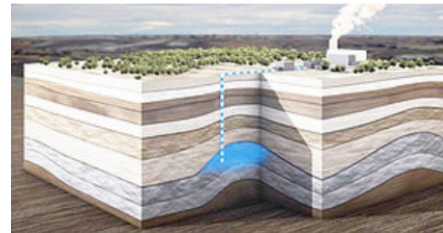
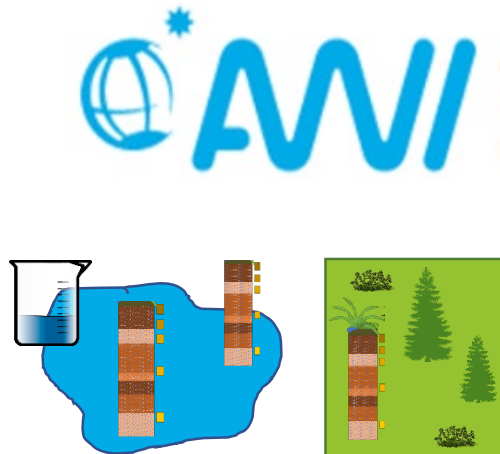


Making your samples FAIR – tools and recommendations from the FAIR WISH Project

FAIR WISH

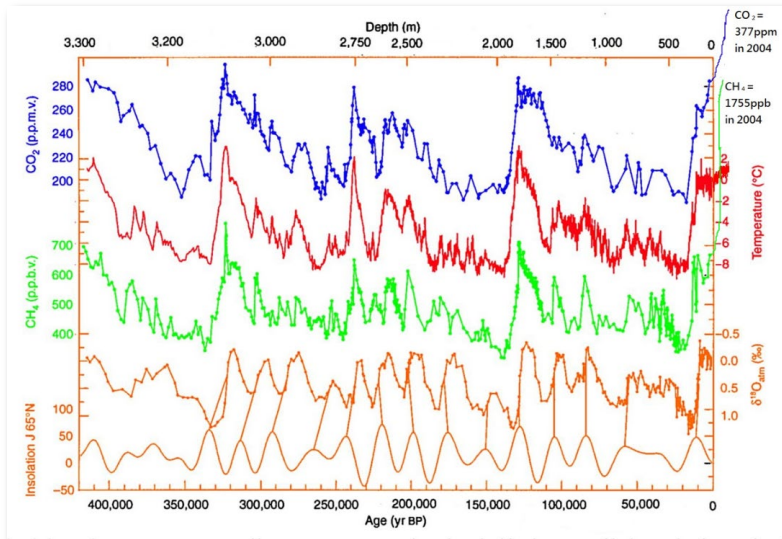
FAIR Workflows to establish IGSN for
Samples in the Helmholtz Association



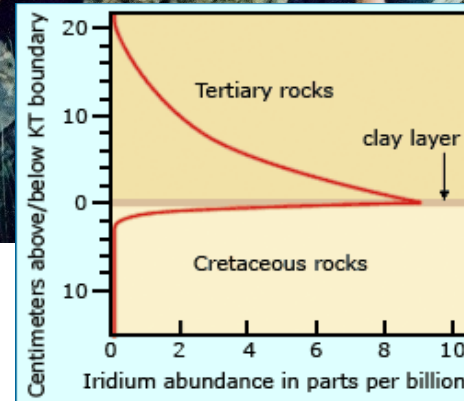
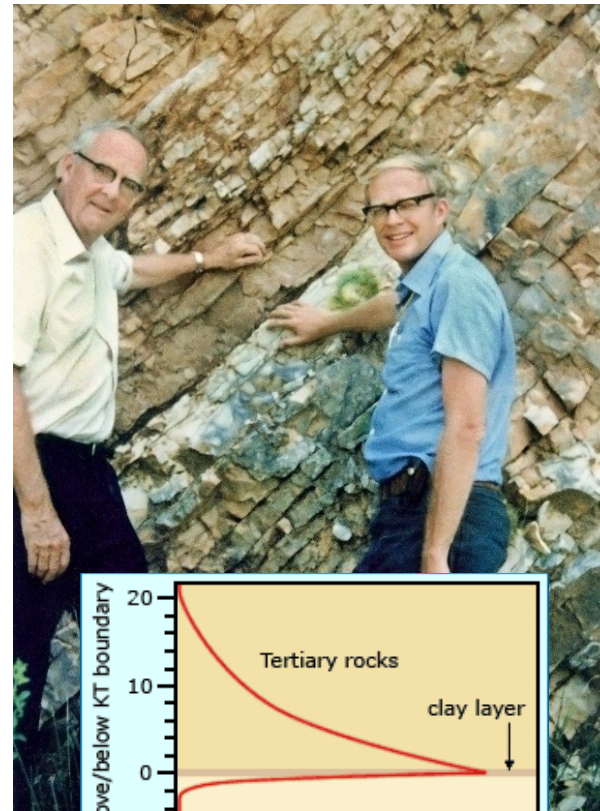
Kirsten Elger, Linda Baldewein, Alex Brauser, Simone Frenzel, Birgit Heim,
Tim Leefmann, Ulrike Kleeberg, Ben Norden, Markeike Wiezcorek, (2021-2023)

Why samples?

- record unique events in history
- are often not reproducible
- are key sources of research data



Samples Help Unravel Past Climate



Samples Record Unique Events in History



Samples are unique



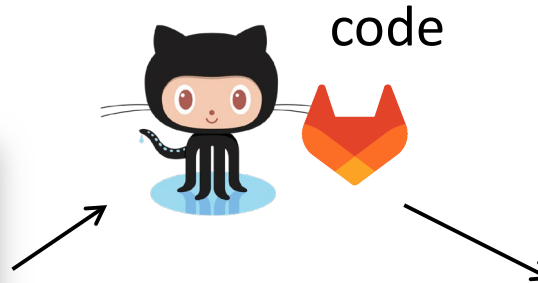
The role of samples for research results

data

| Table S3. Sierra Nevada analyses of plant samples | | | | | Element concentrations (µg/g) | | | | |
|---|------|--------------------------|--|--|-------------------------------|----|----|----|----|
| sample ID | IGSN | brief sample description | | | Al | Fe | Mn | Mg | Ca |

| Table SN1. Sierra Nevada analyses of soil, saprolite, rock, bedload sediment and | | | | | | | | | |
|--|------|---------------------------|-------------------|------------|-------------------------------|------------------|--------------------------------|--------------------------------|------|
| | | | | | Major element oxides (wt%) (c | | | | |
| sample ID | IGSN | sample type | XRF lab | depth (cm) | SiO ₂ | TiO ₂ | Al ₂ O ₃ | Fe ₂ O ₃ | |
| <i>P301 regolith depth profile</i> | | | | | | | | | |
| MW1 | | | | | | | | | |
| MW2 | | | | | | | | | |
| MW3 | | | | | | | | | |
| MW4 | | | | | | | | | |
| mean | | | | | | | | | |
| (2SE) | | | | | | | | | |
| SN01 | * | GFFB1002T | bulk soil | GFZ | 7 | 36.2 | 0.55 | 11.4 | 3.88 |
| SN02 | * | GFFB1002U | bulk soil | GFZ | 20 | 49.3 | 0.71 | 14.9 | 4.84 |
| SN02c | * | GFFB1002U | exchangeable soil | - | 20 | n.a. | <lod | 0.00 | 0.00 |
| SN02r | * | GFFB1002U | residuum soil | - | 20 | n.a. | 0.84 | 14.3 | 5.68 |
| SN03 | * | GFFB1002V | bulk soil | GFZ | 30 | 57.5 | 0.88 | 17.6 | 6.44 |
| SN03c | * | GFFB1002V | bulk soil | GFZ | 39 | 56.6 | 0.93 | 17.7 | 6.85 |
| SN04e | * | GFFB1002R | exchangeable soil | - | 39 | n.a. | <lod | 0.00 | 0.00 |
| SN04r | * | GFFB1002R | residuum soil | - | 39 | n.a. | 0.91 | 15.3 | 6.95 |

code



papers



sample



FAIR WISH



Samples are at the beginning of the research output „food chain“

IGSN International Generic Sample Number

A globally unique and persistent identifier (PID) for physical objects since 2012

- QR code on samples resolves to [virtual sample representations](#) (sample metadata profiles) managed at federated IGSN Allocating Agents
- Guaranteed to be [unique](#) via a centralized control mechanism (unique name spaces)
- Citable in scholarly literature



IGSN Landing Page

General Identifiers

| | |
|---------------|-------------------|
| Program: | SO273 |
| Expedition: | SO273 |
| Type: | Individual Sample |
| Name: | SO273-31D-10 |
| IGSN: | GFSO273N31 |
| Parent IGSN: | GFSO273N21 |
| Release Date: | N/A |

Sampling Location

| | |
|--------------------|------------------------|
| Latitude: | -43.8543 |
| Longitude: | 38.88 |
| Coordinate System: | WGS84 |
| Elevation: | -728 |
| Location Type: | Ridge |
| Location Name: | Southwest Indian Ridge |

Geology

| | |
|----------------------|---|
| Material: | Rock |
| Rock Classification: | Igneous>Volcanic>Mafic |
| Rock Type: | Plag-Basalt |
| Rock Description: | light (most of the primary mineralogy preserved) <ul style="list-style-type: none">• weathering: light (most of the primary mineralogy preserved)• shape: subangular• Manganese crust (mm): 0.1• Lithology: Plag-Basalt• Average grain size: aphanitic• Texture: highly phytic |

Sampling

| | |
|------------------|----------|
| Drilling Method: | dredging |
| Platform Type: | Ship |

Sample Family

- ☐ SO273-31D
- ☐ SO273-31D-1
- ☐ SO273-31D-2
- ☐ SO273-31D-3
- ☐ SO273-31D-4
- ☐ SO273-31D-5
- ☐ SO273-31D-6
- ☐ SO273-31D-7

Location Map

Drilling Start/End: 2020-03-20 23:06:56+00:00 / 2020-03-21 00:55:02+00:00 *
Latitude: -43.85430 * Longitude: 38.88000 *
Southwest Indian Ridge

Review article: Klump et al., (2021) *Towards Globally Unique Identification of Physical Samples: Governance and Technical Implementation of the IGSN Global Sample Number*. Data Science Journal, <https://doi.org/10.5334/dsj-2021-033>

FAIR WISH



HMC Project Cohorte 2020

Objectives

- (i) develop **standardised and discipline specific IGSN metadata schemes for different samples types** within the research field Earth and Environment (EaE), complementing the core IGSN metadata schema
- (ii) develop **workflows to generate machine-readable IGSN metadata** from different states of digitisation (from templates to databases)
- (iii) develop **workflows to automatically register IGSNs** and



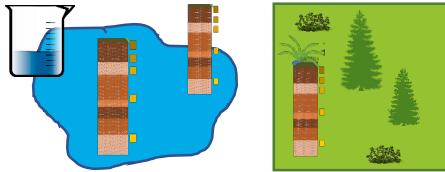
POF:

POF IV 5.3 Natural dynamics of terrestrial Earth surface system

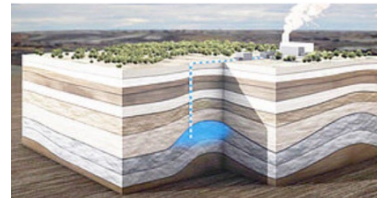
POF IV 8.3 Integrating geoenery and mineral systems

POF IV 4.1 Fluxes and transformation of energy and matter

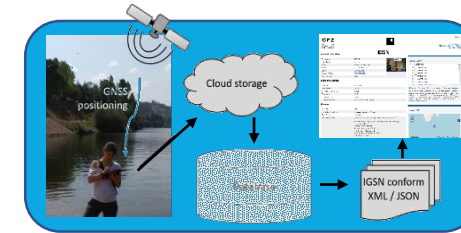
Partners:



Arctic land expedition samples from **Russian-German expeditions**



Rock samples and cores from the **Ketzin pilot site**



Automated IGSN assignment for **Biogeochemical Sample Database**

level of digitalisation

FAIR WISH

FAIR Workflows to establish IGSN for Samples in the **Helmholtz Association**

Major change of IGSN

- [from Jan 2023] IGSN Handle PIDs can only be registered as DataCite DOIs (IGSN-ID) (resource type „physical object“)
- Each DataCite Member or DataCite Consortium Member can assign DataCite „IGSN-IDs“
- IGSN metadata schema is mapped to DataCite schema (Guidelines available, ongoing process)
- All >10 Mio IGSNs registered as Handle PIDs had to be re-registered as DataCite DOIs in Q1/2023

Partnership between IGSN and DataCite

October 28, 2021 Matt Buys and Kerstin Lehnert

<https://doi.org/10.5438/7Z70-1155>

<https://doi.org/10.5438/7Z70-1155>

Earlier this year, DataCite and IGSN announced their roadmap towards a partnership to support the global adoption, implementation, and use of physical sample identifiers. Today, we are pleased to share the announcement of the partnership agreement.

DataCite is a community-led organisation with a vision to connect research and identify knowledge. We have been providing the means to create, find, cite, connect, and use research across 48

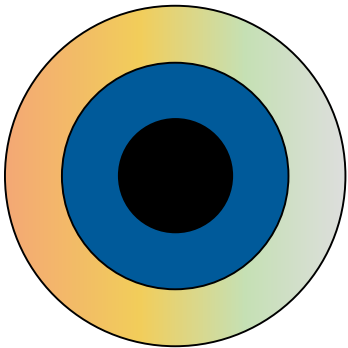
The screenshot shows the DataCite website interface. At the top, there's a navigation bar with the DataCite logo, a feedback icon, and links for Home and All content CC-BY. Below this is a secondary navigation bar with Home, Guides, and API Reference. The main content area is titled 'IGSN ID Metadata Recommendations'. It includes a sidebar with a table of contents: 'CDL's DMPTool: An implementation use case', 'A crosswalk of the DataCite schema to RDA metadata standard for DMPs', 'IGSN IDs', 'About IGSN IDs for Material Samples', 'Using IGSN IDs', 'Registering IGSN IDs', 'IGSN ID Metadata' (highlighted), 'IGSN ID Metadata Recommendations' (highlighted), 'Enriching IGSN ID metadata in the DataCite Metadata Schema', and 'IGSN ID FAQ'. The main text area contains the title 'IGSN ID Metadata Recommendations', a paragraph explaining the partnership and the development of the IGSN-DataCite Crosswalk and Metadata Management Working Group, and a section titled 'IGSN ID recommendations for mandatory properties in DataCite Metadata Schema' which states that IGSN IDs must be registered with at least the six mandatory properties in the DataCite Metadata Schema. At the bottom, there's a URL: <https://support.datacite.org/docs/igsn-id-metadata-recommendations>.

Original IGSN Metadata Levels (Handle PIDs)

Registration Metadata

- IGSN
- ResourceURI
- Registrant ID
- timeStamp
- status

mandatory



Descriptive Metadata

- names
- relatedIdentifier
- description
- Registrant
- collector
- contributors
- geolocation
- resourceTypes
- materials
- collectionMethods
- collectionTime
- sampleAccess
- supplementalMeta

**domain-specific
(FAIR WISH)**

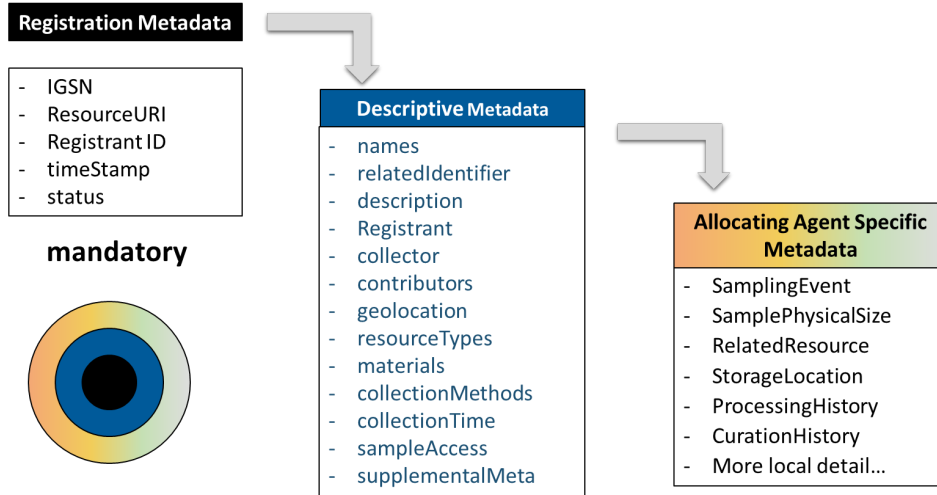
Allocating Agent Specific Metadata

- SamplingEvent
- SamplePhysicalSize
- RelatedResource
- StorageLocation
- ProcessingHistory
- CurationHistory
- More local detail...

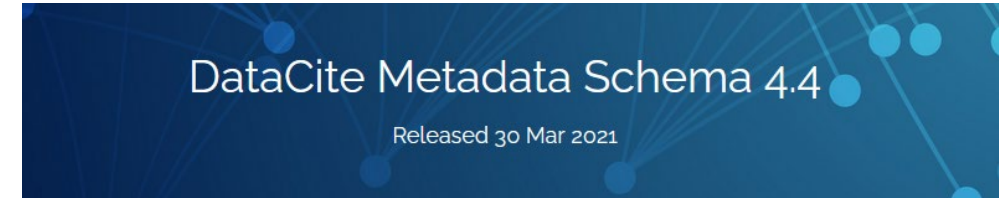
Common Kernel: recommended for data discovery

IGSN Metadata → DataCite

Original IGSN Metadata Levels (Handles)



Modular IGSN metadata allowing for description of many different sample types; rich and very sample-specific properties



```
<xs:include schemaLocation="include/datacite-funderIdentifierType-v4.xsd"/>
<xs:include schemaLocation="include/datacite-descriptionType-v4.xsd"/>
<xs:include schemaLocation="include/datacite-nameType-v4.xsd"/>
<xs:include schemaLocation="include/datacite-numberType-v4.xsd"/>
<xs:element name="resource">
  <xs:annotation>
    <xs:documentation>
      Root element of a single record. This wrapper element is for XML implementation only and is not defined in the
      DataCite DOI standard. Note: This is the case for all wrapper elements within this schema.
    </xs:documentation>
  </xs:annotation>
  <xs:complexType>
    <xs:all>
      <!--REQUIRED FIELDS-->
      <xs:element name="identifier">
        <xs:annotation>
          <xs:documentation>
```

DataCite's mandatory metadata properties:
identifier (DOI), creator, title,
year, publisher, resource type plus
several properties **recommended for discovery**

Challenges: IGSN metadata are more specific for sample descriptions than DataCite metadata. How to compose a „title“ for samples? How complete could/should IGSN metadata be mapped? How can IGSN e.V. ensure consistency of IGSN-IDs across disciplines, agents...

FAIR WISH Results

Full documentation of IGSN metadata for the allocating agent GFZ



IGSN Description Metadata is intended to describe the core elements of a specimen. The set of attributes is seen as the specimens and should not contain stateful attributes, where possible. The base document for the development are the [Kernel Workshop](#) held in Los Angeles in September 2015.

IGSN Descriptive Metadata Elements

| ID | Element | A/C | Occ | Definition | Description |
|-----|----------------|--------------|---------------------|----------------------------|--|
| | resource | Root element | | | IGSN "birth" of sample, associated with |
| 1 | identifier | C | 1 | string | The Identifier identifies a resource. GeoSample Number (IGSN member) "10273/foo" |
| 1.1 | identifierType | A | 1 | include/identifierType.xsd | currently only type=IGSN is supported |
| 2 | name | | 1...1, not nillable | string | Text string for people to understand what is identified. What would typically be presented in a user interface. Collector's or contributor's local/field name used to name the specimen; not globally unique but typically unique within a set of specimens. |

IGSN Descriptive Metadata GFZ Sample Elements

<https://dataservices.gfz-potsdam.de/igsn/schemas/description-ext/1.3/sample.xsd>

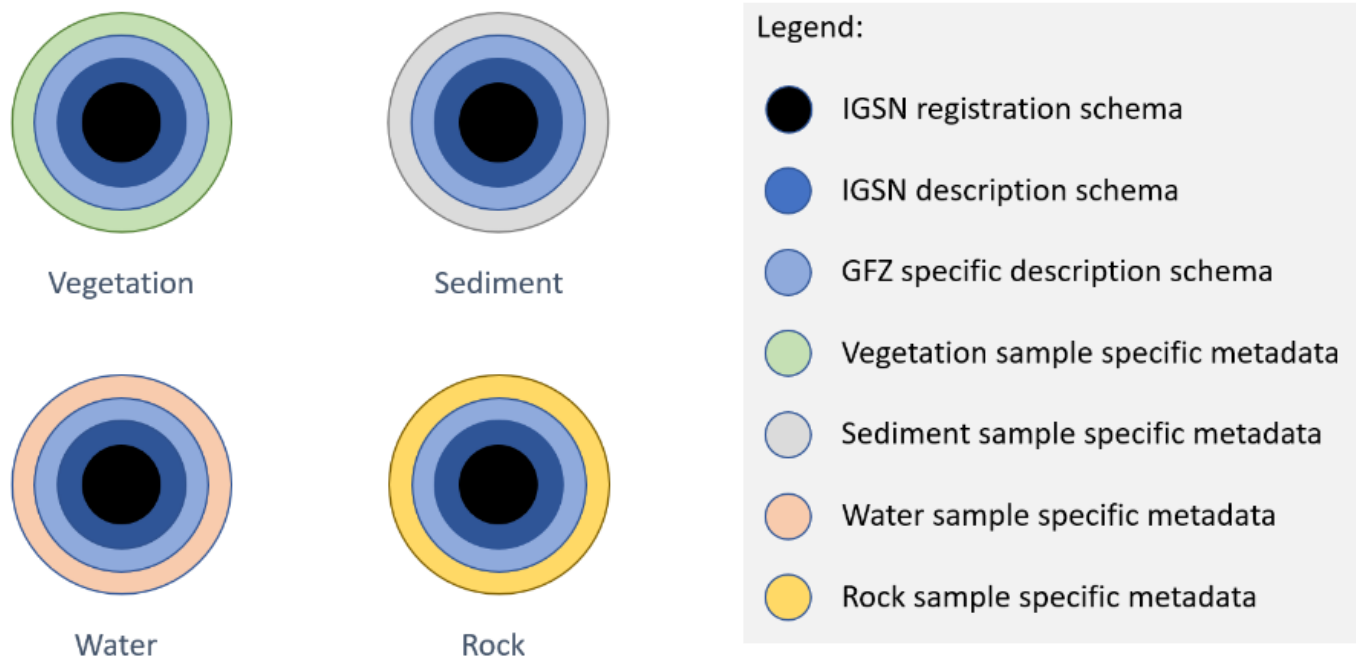
| ID | Element | A/C | Occ | Definition | Description and instructions |
|----|-------------|--------------|-----|--------------------|---|
| | sample | Root element | | | resource (17.) collection |
| 1 | user_code | | | String | SESAR user code, five characters e.g. IESES. Three-character user codes are legacy user codes. They are still supported. |
| 2 | sample_type | | | include/sample.xsd | also called object type, controlled list restriction: table 2 |
| 3 | name | | | String(token) | Name of the sample |
| 4 | igsn | | | String | If being user-assigned, the nine-character IGSN of the sample, leave line out to get automatic assignment (preferred) See detail at http://www.geosamples.org/aboutigsn |

FAIR WISH Results

D1 - List of identified linked open data vocabularies to be included in IGSN metadata

(<https://doi.org/10.5281/zenodo.6787199>)

Sample-specific metadata recommendations



- **Linked data vocabularies (RDF/SKOS)** are key elements for machine-actionable sample description metadata.
- We have identified several of these to be used for describing material properties for different sample types
- They will be included in the GFZ specific IGSN metadata schema to improve the standardisation of sample descriptions

FAIR WISH Results

D1 - List of identified linked open data vocabularies to be included in IGSN metadata

(<https://doi.org/10.5281/zenodo.6787199>)

Example: Biome Vocabulary

JOURNAL ARTICLE

Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity ^{FREE}

David M. Olson, Eric Dinerstein, Eric D. Wikramanayake, Neil D. Burgess, George V. N. Powell, Emma C. Underwood, Jennifer A. D'amico, Illanga Itoua, Holly E. Strand, John C. Morrison ... [Show more](#)

[Author Notes](#)

BioScience, Volume 51, Issue 11, November 2001, Pages 933–938, [https://doi.org/10.1641/0006-3568\(2001\)051\[0933:TEOTWA\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2)

Published: 01 November 2001

From: [https://doi.org/10.1641/0006-3568\(2001\)051\[0933:TEOTWA\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2)

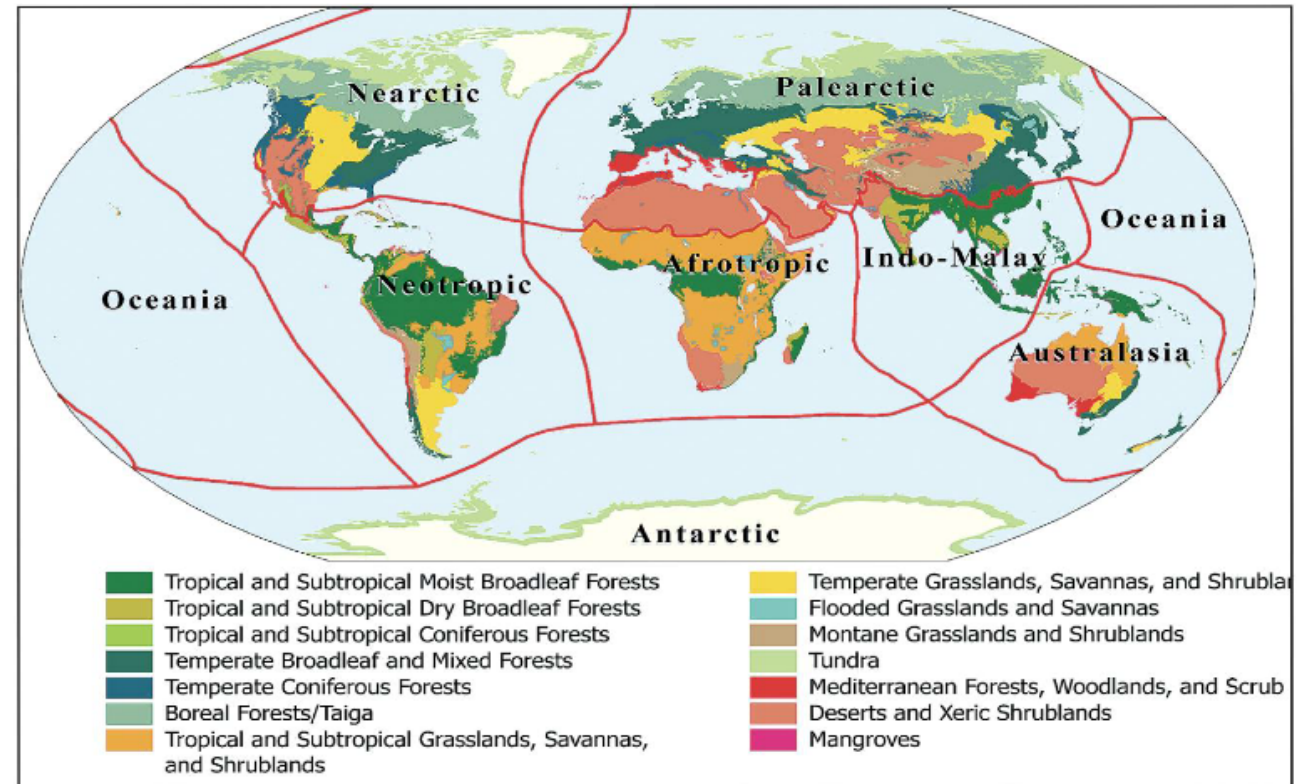
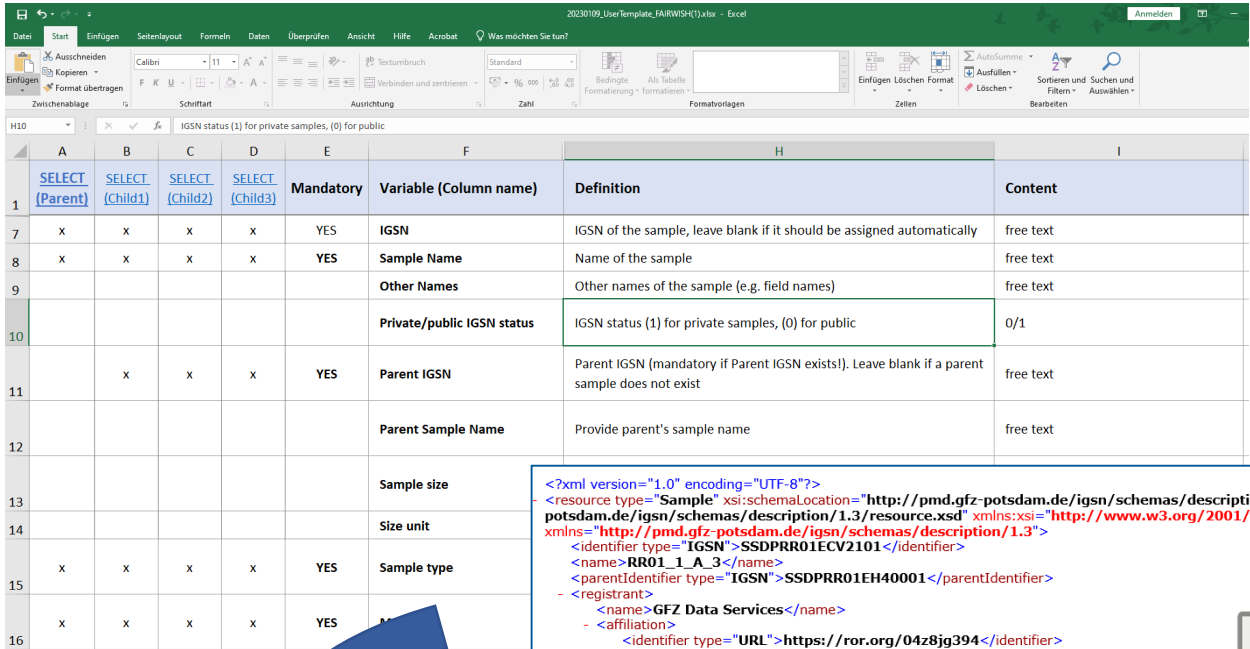


Figure 1. The ecoregions are categorized within 14 biomes and eight biogeographic realms to facilitate representation analyses.


FAIR WISH Results

D2 – D4 - FAIR Samples Template and tutorials

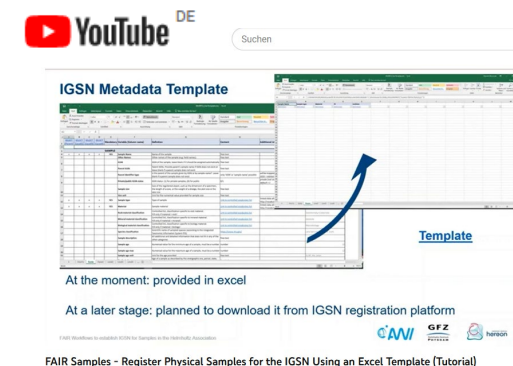


| | A | B | C | D | E | F | G | H | I |
|----|-----------------|-----------------|-----------------|-----------------|-----------|----------------------------|---|---|-----------|
| 1 | SELECT (Parent) | SELECT (Child1) | SELECT (Child2) | SELECT (Child3) | Mandatory | Variable (Column name) | | Definition | Content |
| 7 | X | X | X | X | YES | IGSN | | IGSN of the sample, leave blank if it should be assigned automatically | free text |
| 8 | X | X | X | X | YES | Sample Name | | Name of the sample | free text |
| 9 | | | | | | Other Names | | Other names of the sample (e.g. field names) | free text |
| 10 | | | | | | Private/public IGSN status | | IGSN status (1) for private samples, (0) for public | 0/1 |
| 11 | X | X | X | X | YES | Parent IGSN | | Parent IGSN (mandatory if Parent IGSN exists!). Leave blank if a parent sample does not exist | free text |
| 12 | | | | | | Parent Sample Name | | Provide parent's sample name | free text |
| 13 | | | | | | Sample size | | | |
| 14 | | | | | | Size unit | | | |
| 15 | X | X | X | X | YES | Sample type | | | |
| 16 | X | X | X | X | YES | | | | |

```
<?xml version="1.0" encoding="UTF-8"?>
<resource type="Sample" xsi:schemaLocation="http://pmd.gfz-potsdam.de/igsn/schemas/description/1.3 http://pmd.gfz-potsdam.de/igsn/schemas/description/1.3/resource.xsd" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns="http://pmd.gfz-potsdam.de/igsn/schemas/description/1.3">
  <identifier type="IGSN">SSDPRR01ECV2101</identifier>
  <name>RR01_1_A_3</name>
  <parentIdentifier type="IGSN">SSDPRR01EH40001</parentIdentifier>
  <registrant>
    <name>GFZ Data Services</name>
    <affiliation>
      <identifier type="URL">https://ror.org/04z8jg394</identifier>
      <name>GFZ Potsdam</name>
    </affiliation>
  </registrant>
  <collector>
    <identifier type="ORCID">0000-0003-2776-0846</identifier>
    <name>Christopher Juhlin</name>
    <affiliation>
      <identifier type="URL">https://www.ror.org/048a87296</identifier>
      <name>Uppsala University</name>
    </affiliation>
  </collector>
</resource>
```



The **FAIR Samples Template** is a customisable template to provide standardised sample descriptions. **Users can identify metadata elements from the full schema** that align with their samples. The template is the **source for machine-actionable XML metadata** used for IGSN registration.



Video
Tutorial

Template: <https://doi.org/10.5281/zenodo.7520015>

Documentation: <https://doi.org/10.5281/zenodo.7147531>

New version for hierarchical samples available very soon!

FAIR WISH Results

D5 – Mapping of database metadata to machine readable IGSN metadata (Use Case Hereon) (<https://doi.org/10.5281/zenodo.7788653>)

1.3. The Hereon expedition database

The Hereon expedition database has been developed since 2016 within the coastMap project (Baldewein et al., 2018). Its aim is to allow querying and downloading of campaign data collected during ship and land-based sampling campaigns at the Hereon Institute of Carbon Cycles and Hereon Institute of Coastal Environmental Chemistry.

The fully normalised relational database has been continuously expanded to fully describe the highly diverse metadata associated with biogeochemical campaign samples (long-tail data according to Heidorn, 2008). It currently consists of 34 tables, storing mostly metadata and the relationship between metadata elements. Eleven of these tables directly or indirectly store the sample metadata needed for IGSN registration. This includes tables storing the sample, information on the campaign, as well as those storing information on associated scientists, projects, and related publications.

6.1. Database mapping for sample sites

| IGSN field name | Variability | Database field | Example |
|---------------------------|-------------|---|---|
| igsn | variable | station.IGSN | GFHER63BAE |
| parent_igsn | constant | | |
| name | variable | station.STATION_NAME | BOHAISEA2018Nov_Stat_Qingjinghuang_Drainage_Canal |
| sample_other_names | variable | station.STATION_LABEL | Qingjinghuang Drainage Canal |
| is_private | constant | | 0 |
| sample_type | constant | | Site |
| material | constant | | NotApplicable |
| description | constant | | |
| collection_start_date | variable | station.station_start_time | 2018-11-01 10:24:00.0000000 +08:00 |
| collection_end_date | variable | station.station_end_time | 2018-11-01 10:24:00.0000000 +08:00 |
| collection_date_time_zone | variable | Time zone info of station.station_start_time | UTC+08:00 |
| collection_date_precision | constant | | time |
| depth_min | constant | | |
| depth_scale | constant | | |
| collection_method | constant | | |
| collection_method_descr | constant | | |
| cruise_field_prgm | variable | campaign.campaign_code | CE17013 |
| patform_type | variable | If vessel.VESSEL_NAME starts with RV, then 'Ship', else leave blank | Ship |
| platform_name | variable | vessel.VESSEL_NAME | RV Celtic Explorer |

FAIR WISH Results

IGSN Registration for all use cases



new Datacenters

→ AWI: Polar Terrestrial Environmental Systems
→ Earth Shape SPP
→ Expedition database Hereon
→ GFZ Potsdam
→ Geothermal Energy Systems
→ High Latitude Lakes
→ ICDP
→ Medusa

Sample Type

core
core half round
core sample
core section
core whole round
ctd
cuttings
dredge
grab
hole
individual sample
other
site
specimen

Found 35051 samples.

GFBNO7002EXM4001 Individual Sample

Material: Rock
Classification: undefined

GFBNO7002EXL4001 Individual Sample

Material: Rock
Classification: undefined

GFBNO7002EXH4001 Individual Sample

Material: Rock
Classification: undefined

GFBNO7002EXI4001 Individual Sample

Material: Rock
Classification: undefined

GFBNO7002EXG4001 Individual Sample

Material: Rock
Classification: undefined



The screenshot shows the DataCite Commons interface. At the top, the search bar contains 'client.uid:gfz.igsan'. Below the search bar, there are tabs for 'Works', 'People', 'Organizations', and 'Repositories'. The 'Works' tab is selected, showing '36,442 Works'. On the left, under 'Creators & Contributors', a list of names and counts is shown: Juhlin, Christopher (11,650), Sutherland, Rupert (2,321), Jackson, Marie D. (1,696), Turowski, Jens M. (829), Calner, Mikael (114), and Buchwal, Agata (3). Below this, under 'Publication Year', a list shows years and counts: 2023 (832), 2022 (24,362), 2021 (897), 2020 (479), and 2019 (1,792). On the right, two sample entries are displayed. The first is 'IGSN GFHER22229 (p_HE511_Stat_HE511/80_25) Individual Sample: Liquid>aqueous from HE511 expedition' by Justus van Beusekom, published in 2022. The second is 'IGSN GFHER2222C (p_HE511_Stat_HE511/42_16.5) Individual Sample: Liquid>aqueous from HE511 expedition' by Justus van Beusekom, also published in 2022. Both entries include a DOI link and a 'Physical Object' label.

<https://dataservices.gfz-potsdam.de/igsan-new/>

- **December 2022:** IGSN registration for >14500 samples and sites from all use cases (Hereons Expedition Database, German-Russian Expeditions of AWI, GFZ)/ handles
- **January - March 2023:** Metadata mapping and Re-registration of all IGSNs of GFZ to DataCite IGSN IDs

Project Team

