**Recommendation 4.0 **

Use IGSN for Samples

Recommendation to use IGSN as the standard reference in technical infrastructures to samples where appropriate

Summary

Description

Status: V1.0 28.5.2025

Motivation for this Recommendation:

The International Generic Sample Number (IGSN) is a globally unique and persistent identifier designed specifically for physical samples and related objects. At Helmholtz, we recommend the use of IGSNs to ensure that samples—and other tangible sources from which data are derived—can be reliably identified, referenced, and linked across research workflows. The motivation to use IGSNs lies in their ability to improve traceability, reproducibility, and data integration across disciplines. By assigning a persistent identifier to a sample, researchers can unambiguously connect it to associated datasets, publications, instruments, and collection metadata, supporting FAIR principles and enabling long-term reuse and verification of research outcomes.

Recommendation

It is recommended to use IGSN to identify samples in data infrastructures where appropriate.

[To discuss: What exactly is a sample or for what can we use IGSNs? The term sample may in tis case be interpreted as any object, physical or virtual thing, that is used to derive data from. Hence, even data sets could be treated as samples, if they are the source for new data derived from them e.g. by modeling. Software are not samples in this context and should be referred to by other means.]

For the organization or center management this means:

- it should provide a way to allow personnel to register IGSNs.
- a person or unit should be made responsible to maintain the centres IGSNs.

For data curators this means:

- Enable, train and encourage staff to register IGSN when samples are taken.
- Enable, train and encourage staff to record any parent IGSNs with subsamples.

For researchers it means:

- Record an IGSN with any sample taken.
- Record the parent IGSN with any subsample measured

For data infrastructures:

- record a IGSN to identify samples and parent samples and make this data part of the metadata available for harvesting.
- treat IGSN metadata as the primary source of truth and update your own metadata accordingly.

Also see [3] Baldewein et al. (2023). FAIR WISH D7 -Standard Operating Procedure for automatic IGSN registration. Zenodo. https://doi.org/10.5281/zenodo.10401380

Binding Convention:

	mandatory	conditional	optional
Helmholtz FAIR Principle		X	

Precondition for Implementation:

The institution needs to be a member of Data Cite or needs to partner with a member to be able to register IGSNs.

Related Recommendations

Parent: M0

Dependent: M4.1, M4.2, M4.3, M4.4

Other: none

Contributors

Emanuel Soeding (lead)

Content

[1] Plankytė, Vaida, Macneil, Rory, & Chen, Xiaoli. (2023). Guiding principles for implementing persistent identification and metadata features on research tools to boost interoperability of research data and support sample management workflows. Zenodo. https://doi.org/10.5281/zenodo.8284206

1. Explanation of the Background and Benefits of the Recommendation

About

The International Generic Sample Number (IGSN) is a persistent, globally unique identifier designed to unambiguously reference physical samples and other material objects in the research lifecycle. It enables reliable citation, tracking, and linking of samples to related data, instruments, people, and publications, making them FAIR—findable, accessible, interoperable, and reusable.

History

Originally developed by the geoscience community in the early 2000s, IGSN emerged from the need to manage and cite geological samples across laboratories and institutions. It was formalized through the IGSN e.V. foundation in 2011 and has since evolved into a cross-disciplinary identifier supported by the global research infrastructure. Since 2021, IGSNs have been registered through DataCite, aligning their metadata with other research outputs.

Structure

IGSN records consist of a unique identifier (a prefix-suffix structure similar to DOIs) and a metadata record that captures core descriptive information about the sample: sample type, material, collection method, spatial and temporal context, and links to related entities (e.g., datasets, people, institutions). Metadata can be enhanced to fit domain-specific needs while maintaining a consistent structure for interoperability.

Motivation

Using IGSNs improves sample traceability, ensures reproducibility of results, and supports data integration across disciplines. It allows researchers to explicitly reference the physical basis of data analyses, which is critical for verification, reuse, and credit assignment.

Current Use of IGSN

IGSNs are currently used in a range of domains, including geosciences, environmental sciences, archaeology, and biodiversity research. For example, ocean drilling samples from IODP expeditions, sediment cores, rock specimens, water samples, and even archaeological artifacts have been assigned IGSNs. These identifiers help integrate sample-based research into digital infrastructures and link physical materials to datasets and publications, thus enabling transparent and connected science.

2. Possible alternative solutions

- Internal or Local Identifiers

What: Lab- or institution-specific sample IDs.

Pros: Easy to implement, tailored to local needs.

Cons: Not globally unique, not resolvable, hard to track across systems or publications.

- Handle System / Custom DOIs

What: Using general-purpose persistent identifiers like DOIs or Handles for samples.

Pros: Technically viable; DOI infrastructure is mature.

Cons: Lack of community consensus or metadata model for samples unless built on top of IGSN or similar; harder to ensure consistency and semantic clarity.

- ARK (Archival Resource Key, https://arks.org/)

What: A persistent identifier scheme designed for objects of any type.

Pros: Flexible, openly governed, used by some institutions (e.g., museums, archives).

Cons: Less widely adopted in science, lacks built-in metadata requirements for samples, limited interoperability in research workflows.

Why IGSN?

While alternatives exist, IGSN is currently the only PID system specifically designed to handle the complexities of referencing physical samples across scientific domains. It combines:

- Global uniqueness and persistence
- A structured, interoperable metadata schema
- Community governance
- Integration with DataCite infrastructure
- Support for linking to related PIDs (e.g., ORCID, ROR, dataset DOIs)

Therefore, for research workflows that require transparent, machine-readable, and citable links between samples and data, IGSN remains the most suitable and sustainable option.

3. Consideration of the advantages and disadvantages of implementing the recommendation

Advantages

Implementing IGSN across data infrastructures, research workflows, and organizational practices significantly improves the quality, traceability, and usability of sample-related metadata.

From an interoperability perspective, IGSNs enable seamless linking between samples, datasets, publications, instruments, and other research outputs. This allows infrastructures to integrate more easily with external systems such as DataCite, ORCID, and disciplinary repositories, fostering a more connected and machine-actionable research ecosystem.

Even when personnel or institutions change, IGSNs preserve the identity and context of physical samples. This also supports consistent discovery and attribution of samples used across multiple studies, which is increasingly important for collaborative and longitudinal research.

Disadvantages and Limitations

One of the main barriers is the initial technical and organizational overhead. Integrating IGSN registration and resolution into local systems may require custom development, workflow redesign, and coordination with external allocating agents. Establishing clear responsibilities—such as assigning a person or unit to manage IGSNs—also requires dedicated resources and long-term commitment.

Another challenge is training and awareness. Researchers and technicians may not be familiar with IGSN or may view it as an extra administrative step. Sustained training efforts, user support, and institutional incentives are needed to ensure consistent and correct usage.

There are also some external dependencies. Institutions must rely on the long-term availability and sustainability of IGSN allocating agencies and infrastructure. If funding or governance of the global IGSN framework becomes unstable, there may be future risks related to service continuity or the need for metadata migration.

Not all research fields have mature practices around sample identification, meaning that IGSN may not yet be a natural fit for every domain. In such cases, complementary identifiers or interim solutions might still be needed. Additionally, keeping local metadata synchronized with authoritative IGSN records can introduce technical complexity, especially if metadata schemas evolve differently.

4. The Recommendation and possible consequences

It is recommended to use IGSN to identify samples in data infrastructures where appropriate.

For organizations this means:

- a person or unit should be made responsible to maintain the centres IGSNs.

For data curators this means:

- Enable, train and encourage staff to register IGSN when samples are taken.
- Enable, train and encourage staff to record any parent IGSNs with subsamples.

For researchers it means:

- Record an IGSN with any sample taken.
- Record the parent IGSN with any subsample measured

For data infrastructures:

- record a IGSN to identify samples and parent samples and make this data part of the metadata available for harvesting.
- treat IGSN metadata as the primary source of truth and update your own metadata accordingly.

Also see [3] Baldewein et al. (2023). FAIR WISH D7 -Standard Operating Procedure for automatic IGSN registration. Zenodo. https://doi.org/10.5281/zenodo.10401380

5. Naming of communities that have already implemented the recommendation

GFZ Data Services

Pangaea

Hereon HCDC (?)

Others?

6. Documentation of the test to validate correct implementation

7. Examples of Instances

IGSN is implemented within the Helmholtz association at AWI, GFZ, and Hereon through the FAIRWish Project [4]. See [3] for more information.

Another implementation is documented at the Kiel University (CAU) [6]

8. Further Information

References

[1] Plankytė, Vaida, Macneil, Rory, & Chen, Xiaoli. (2023). Guiding principles for implementing persistent identification and metadata features on research tools to boost interoperability of research data and support sample management workflows. Zenodo. https://doi.org/10.5281/zenodo.8284206

[2] Klump, J., Lehnert, K., Ulbricht, D., Devaraju, A., Elger, K., Fleischer, D., Ramdeen, S., Wyborn, L. (2021): Towards Globally Unique Identification of Physical Samples: Governance and Technical Implementation of the IGSN Global Sample Number. - Data Science Journal, 20, 1, 1-16., DOI: https://doi.org/10.5334/dsj-2021-033

[3] Baldewein, L., Kleeberg, U., Brauser, A., Elger, K., Frenzel, S., Heim, B., & Wieczorek, M. (2023). FAIR WISH D7 - Standard Operating Procedure for automatic IGSN registration. Zenodo. https://doi.org/10.5281/zenodo.10401380

[4] The FAIR Wish Project: https://helmholtz-metadaten.de/de/inf-projects/fair-wish-fair-workflows-to-establish-igsn-for-samples-in-the-helmholtz-association

[5] IGSN Documentation on forschungsdaten.org https://www.forschungsdaten.org/index.php/IGSN

[6] IGSN Service and Documentation at the University Kiel https://igsn.uni-kiel.de/de

Relevant Community Recommendations

9. History of this document

From:
https://earth-and-environment.helmholtz-metadaten.de/wiki/ - **HMC Earth and Environment Community Wiki**

Permanent link:
https://earth-and-environment.helmholtz-metadaten.de/wiki/doku.php?id=wiki:2_fair-building-blocks:0_pid:3_igsn:start

Last update: **2026/03/30 13:55**