

## IGSN Identifiers: A Sustainable PID Strategy for Archaeological and Heritage Samples

|                             |                                                                                                                                                                                                                                                                                                                                                                        |
|-----------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Task Area</b>            | Cross-Area Initiative: TA3 in Collaboration with TA1, TA2, and TA4                                                                                                                                                                                                                                                                                                     |
| <b>Partner Institutions</b> | <p>Lead: Deutsches Bergbau-Museum Bochum</p> <p>Co-Applicants:</p> <p>Participants: Leibniz-Zentrum für Archäologie, Max Planck Institute for Evolutionary Anthropology, Kiel University</p> <p>External members: PID4NFDI, TS4NFDI, National Gallery London (HSDA), Royal Institute for Cultural Heritage (HSDA), Czech Academy of Sciences (CoP), IGSN e.V (CoP)</p> |
| <b>Contact</b>              | Yiu-Kang Hsu (yu-kang.hsu@bergbaumuseum.de)                                                                                                                                                                                                                                                                                                                            |
| <b>TRAILS</b>               | <p>TRAIL 3.2: Poseidon 2.0: Ein Datenspeicher für genetische, bioarchäologische und archäologische Daten</p> <p>TRAIL 3.3: Workflow-Tool für archäologische Experimente und Analytik</p>                                                                                                                                                                               |

## 1. Summary

- Please specify the specific challenges for which solutions are to be developed within the framework of the TRAIL and explain the approaches you are pursuing to achieve this.
- Please briefly summarize the objective of the TRAIL.

Managing the growing number of material samples has nowadays posed a serious challenge within archaeology and cultural heritage research. Samples are often assigned various local identifiers during their transfer, leading to ambiguous naming, loss of provenance, and fragmented data scattered across data sources. Inconsistent local metadata descriptions further hinder the interoperability and linkage of samples with other research outputs.

In response to these issues, IGSN identifiers (IGSN IDs), regulated by the DataCite DOI, provide a means to permanently identify, cite, and locate physical samples. Nevertheless, the usability of IGSN has not yet been fully recognised by the NFDI4O communities. To fill this gap, our TRAIL aims to evaluate the application of IGSN IDs and their metadata profiles to sample-based research within the consortium. Through the expertise and cooperation among Task Areas, we will design a community-driven IGSN strategy by:

1. reviewing, adopting and adapting existing IGSN metadata profiles through community effort.
2. performing crosswalks between generic and community-driven metadata schemas to enhance sample data interoperability.
3. validating IGSN metadata profiles via diverse use cases.
4. producing a Cookbook detailing best practices for launching an IGSN registration service in archaeology and cultural heritage.

By doing so, we lay the solid foundation for the further development of a sustainable IGSN service within the NFDI4O and beyond.

## 2. Relevance

- Please briefly outline which section(s) of the research data cycle are primarily addressed within the TRAIL.
- Which elements of FAIR are particularly relevant? To what extent are the FAIR principles feasible?
- To what extent does TRAIL make a useful contribution to the NFDI as a whole? Please explain the extent to which TRAIL contributes to the goals, measures, tasks, and KPIs from the DFG application of NFDI4O.
- Please explain how the process and content of the preparatory work for Task Area 6 and Task Area 7 are designed.
- Does TRAIL address a real, possibly proven need in the specialist community and show realistic prospects for impact? If so, please explain.

The International Generic Sample Number (IGSN) is a globally unique and persistent identifier (PID), registered through the DataCite PID infrastructure [1]. It is designed for the unambiguous identification of an individual sample, an aggregation of samples, or a feature-of-interest to ensure the reliable citation, discovery, and linkage across the internet. IGSN IDs address several key stages of the research data lifecycle. It begins with the **Collect and Create** phase, where the established workflow ensures that each material sample is given a unique and referable IGSN ID and associated metadata at the point of sample collection or creation. This early registration prevents potential data loss during the sample transfer between curatorial systems and secures essential contextual metadata that connects with relevant digital resources such as analytical data, instrument settings, media, and publications. During the phase of **Store and Manage**, IGSN system builds on a globally trusted PID infrastructure operated by DataCite. It guarantees that IGSN ID will always resolve to its metadata record, even though the original data ceases to exist. At the **Share and Disseminate** stage, metadata crosswalks between community-driven and generic schemas (e.g., DataCite, Dublin Core, and [schema.org](https://schema.org)) encourage a seamless DOI minting process, data harvesting, and linked open data initiative.

From this foundation, IGSN IDs serve as the basis for FAIR principles. They make material samples **findable** by providing globally unique and persistent identifiers. They are documented with the DataCite Metadata Schema and indexed in a searchable system such as DataCite Fabrica and Commons, where metadata can be easily **accessible** through standardised communication protocols (HTTPs). Moreover, IGSN metadata standard and its crosswalks ensure pragmatic **interoperability** for data harvesting and recognition across different systems. The character of persistent and consistent IGSN IDs encourages the **reusability** of research outputs as samples can be cited and referenced with confidence.

Despite the benefits of IGSN implementation, there is a lack of a common PID strategy within communities of NFDI4O. Material samples undoubtedly play a pivotal role across our Task Areas—from excavation and initial documentation in the field (TA1), to museum curation (TA2), laboratory analysis (TA3), conservation research (TA4), and eventual archiving in institutional or national storages (TA5). At each stage, new local identifiers may be generated and associated with diverse metadata standards and formats. Consequently, information about the same sample becomes fragmented, duplicated and difficult to reconcile across repositories. This lack of consistent identification undermines both data reusability and the long-term accessibility of sample-related information. Hence, our TRAIL corresponds to a genuine need for a consistent and sustainable sample PID strategy across all research aspects of NFDI4O.

In relation to the DFG application of NFDI4O, the TRAIL directly supports goals and measures across Task Areas where data standards, protocols, and guidelines are

developed for different research entities (G1.1, G2.2, G3.1, G4.2, M1.1, M2.2, M3.1, and M4.1). Our work also complements measure M5.5 by embedding IGSN IDs in a broader PID ecosystem—comprising publication DOIs, ORCID, Wikidata, and ROR.

Our TRAIL can also contribute to the preparatory work for TA6. IGSN Metadata crosswalks can be potentially integrated into the object data model and the minimal dataset conceptualised by TA6. Meanwhile, guidelines and workflows for IGSN implementation can be formulated as an open educational resource for further dissemination.

### **3. (Further) development of standards and guidelines**

- Is the development and harmonization of metadata standards and ontologies planned within the framework of TRAIL? If so, what specifically is planned? Will relevant standards and reference data (e.g., GND, Wikidata) be used for this purpose?
- Are there plans to develop reusable workflows (e.g., for data integration) within the framework of TRAIL?

We seek to further develop and harmonise existing IGSN metadata profiles for physical samples within disciplines. Specifically, we will review two community-driven sample metadata schemas:

1. [Archaeology and Cultural Heritage Community of Practice](#) (CoP) is designed specifically for archaeological samples from the field. It is a joint effort between DataCite, IGSN e.V., and archaeologists with backgrounds in archaeometry, landscape archaeology, and data management [2].
2. [Heritage Sample Digital Archive](#) (HSDA) is an initiative to develop a standard metadata profile for documenting heritage samples and related entities [3]. The community is supported by the Horizon Europe ECHOES project and focuses on

the standardised documentation of heritage samples from GLAM institutions (galleries, libraries, archives, and museums).

IGSN metadata is designed to complement, rather than replace, existing domain-specific schemas. Its primary functions are to capture the necessary information in alignment with DataCite Metadata Schema for the DOI registration, while allowing a richer set of metadata elements to be displayed on sample landing pages.

A **temporary working group** (TWG) will be established to evaluate schema's usability among Task Areas of NFDI4O and suggest further modification if needed. We will review how well the properties, vocabularies and relational structures can apply to material samples from different research fields within NFDI4O. The next step is to define a core set of harmonised metadata fields across disciplines and controlled vocabularies to populate the selected metadata fields wherever appropriate to enhance curation efficiency and accuracy. The existing terminologies can be applied for material types (Art & Architecture Thesaurus), periods (PeriodO), researchers (ORCID), organisations (ROR), and places (GeoNames and Getty TNG).

Furthermore, we will perform crosswalks between generic schemas (DataCite, Dublin Core, and [schema.org](https://schema.org)) and community schemas to achieve interoperability from diverse data sources. To satisfy the requirement of IGSN registration, we will comply with DataCite [IGSN metadata recommendations](#). The similar mapping assignment will be conducted to annotate community metadata schemas with Dublin Core and [schema.org](https://schema.org).

We will then validate the feasibility of the harmonised metadata profile through various use cases. They focus on diverse material samples from different real-world research environments within NFDI4O, including ancient copper alloys and polymer materials from

the DBM, experimental archaeology specimens from the Leibniz-Zentrum für Archäologie, and ancient DNA samples from the Max Planck Institute for Evolutionary Anthropology. Given the established examples, we can further motivate use cases from other Task Areas to consolidate the suitability of the proposed IGSN metadata standard.

#### **4. Development of digital tools and sustainable provision of services and data**

- Are there plans to provide tools for open-source standards as part of TRAIL, and if so, which tools need to be developed or further developed? To what extent will these be included in tool and service catalogues?
- Does TRAIL continue existing activities of NFDI4O? If so, which ones?
- Are the results of TRAIL reusable, and is there concrete potential for continuity? If so, please explain.

The TRAIL is not directly involved in developing any software. Instead, we configure and utilise several open-source tools to facilitate our intended assignments. The [Cordra](#) content management system will be set up to store and manage the developed sample metadata profiles as JSON schemas. Controlled vocabularies will be integrated into Cordra through API connections, and each schema is assigned a persistent Handle ID via Cordra's built-in support system. The collected use cases can also be stored as JSON records through the Cordra's web forms. On the other hand, the detailed documentation of sample schemas and mapping recommendation will be published via MkDocs.

The outputs of this TRAIL lay the groundwork for the further implementation of IGSN infrastructure within NFDI4O by raising awareness of its relevance and benefits in sample-based research. Since IGSN IDs serve as a prerequisite for the Linked Open Data and its integration within the N4O knowledge graph, an IGSN service holds significant potential in the second funding phase of NFDI4O. In parallel, this also

strengthens our position to engage with ongoing international IGSN initiatives, such as the CoP and HSDA.

## **5. Visibility and empowerment of the professional community**

- What measures are planned within the framework of TRAIL to support and train the specialist community (e.g., OER, training courses, onboarding formats, workshops, e-learning, communication measures for tools and standards, etc.)?

The TRAIL will actively engage with relevant Community Clusters to introduce IGSN IDs to the specialist community. It will also participate in the existing international IGSN initiatives (CoP and HSDA) to establish connections between our temporary working group and external audience. These regular interactions provide targeted training opportunities, including webinars on topics such as metadata crosswalks, DataCite DOI infrastructure, and management of JSON-compliant data. Training materials can be designed as open educational resources (OERs) to be submitted to NFDI4O Commons or registered under the [Dalia portal](#).

## **6. Strategic opening, networking, and international connectivity**

- To what extent is cooperation, networking, or collaboration with other NFDI4O task areas planned within the framework of the planned TRAIL?
- In terms of content, at which points is networking with other NFDI consortia planned?
- To what extent is cooperation, networking, or collaboration with national institutions outside the NFDI planned within the framework of the planned TRAIL?
- Does the TRAIL address new technologies, perspectives, communities, or topics that have not yet been covered? If so, please explain.
- To what extent is international cooperation, networking, or collaboration planned in the context of the planned TRAIL?

Building and maintaining a solid user community is extremely important for sustaining our TRAIL activities. To this end, we plan to engage closely with TA1, TA2, and TA4 within NFDI4O, as their research environments—fieldwork, collection, and conservation—represent the key stages of sample generation and management. TRAIL can actively collaborate with the Community Cluster “On-Site Documentation” (TA1), where IGSN IDs could be assigned to archaeological features and finds *in situ*, enabling linked data already at point of primary documentation. Similarly, the clusters “Data Capture and Creation” from TA2 and “Conservation Sciences” from TA4 could also benefit from the usage of IGSN IDs to improve findability and interoperability across the sample lifecycle.

Networking with other NFDI consortia is also explicitly planned, particularly PID4NFDI and TS4NFDI. PID4NFDI will provide guidelines and recommendations to ensure TRAIL’s IGSN ID strategy in alignment with the NFDI PID framework. TS4NFDI could support the technical integration of controlled vocabularies in the sample schemas and offers API-based crosswalk services.

A further key strength of the TRAIL lies in its international composition through active interaction with IGSN communities. Dr. Jens Klump (IGSN e.V.) and Dr. David Novak (Czech Academy of Sciences) from the CoP have agreed to support the refinement of the community-driven metadata profile for archaeological samples. In parallel, Dr. Joseph Padfield (National Gallery, London) and Dr. Wim Fremout (Royal Institute for Cultural Heritage) from the HSDA are willing to assist in developing the IGSN metadata and PID strategy for heritage samples.

## 7. Deliverables

Specify the deliverables, provide names and specifications for the technical implementation of the project (e.g., shared software, developed interface, implementation and dissemination of knowledge/data, etc.), and state the expected results.

What contributions to the commons and to central offerings of NFDI4O (e.g., open educational resources) will be developed as part of the TRAIL? Relate this to section 5 (see above).

### WP1 Metadata and Crosswalks

- Description

We will begin with a community survey to assess current sample management within disciplines, raise awareness of IGSN IDs among researchers, and identify potential users. Subsequently, a temporary working group will be established to invite interested and representative members of other Task Areas within the NFDI4O. Together we will review, discuss, and adapt the existing IGSN metadata profiles (CoP and HSDA) to the needs of NFDI4O communities. Meanwhile, we will suggest crosswalks recommendation and deliver the outcomes for the acceptance and further comments from the broader archaeological and heritage community via the Community Hub.

- Expected Outcomes

The developed schemas and mapping results will be version-controlled on Github, with stable releases archived on Zenodo. Detailed metadata documentation will be published using MkDocs, while Cordra stores IGSN metadata profiles as JSON schemas.

## WP2 Use Cases

- Description

The Cordra management system will convert stored IGSN schemas into interactive web forms linked to controlled vocabularies. This enables participating institutions to test the suitability of suggested IGSN metadata profiles and submit their use cases as JSON records managed by Cordra. User feedback will encourage necessary adjustment and adaption to iteratively improve the data model and submission workflows.

- Expected Outcomes

This process would not only validate the usability of IGSN and improve its registration workflows, but also for its practical improve the IGSN workflows, but also affirm their suitability for practical use, but also reinforce their adoption and acceptance among NFDI4O communities.

## WP3 Strategy and Community

- Description

The outputs of WP1 and WP2 will be consolidated in this phase by working closely with PID4NFDI, CoP, and HSDA. With their support, we will develop a practical IGSN roadmap tailored to the needs of the archaeology and heritage communities [4][5]. This IGSN strategy will address not only the underlying metadata profiles, but also technical, organisational, and financial components necessary for sustainable PID management. The outcomes will be presented to both NDFI and international communities through workshops, conferences, as well as the Community Hub for further commentaries.

- Expected Outcomes

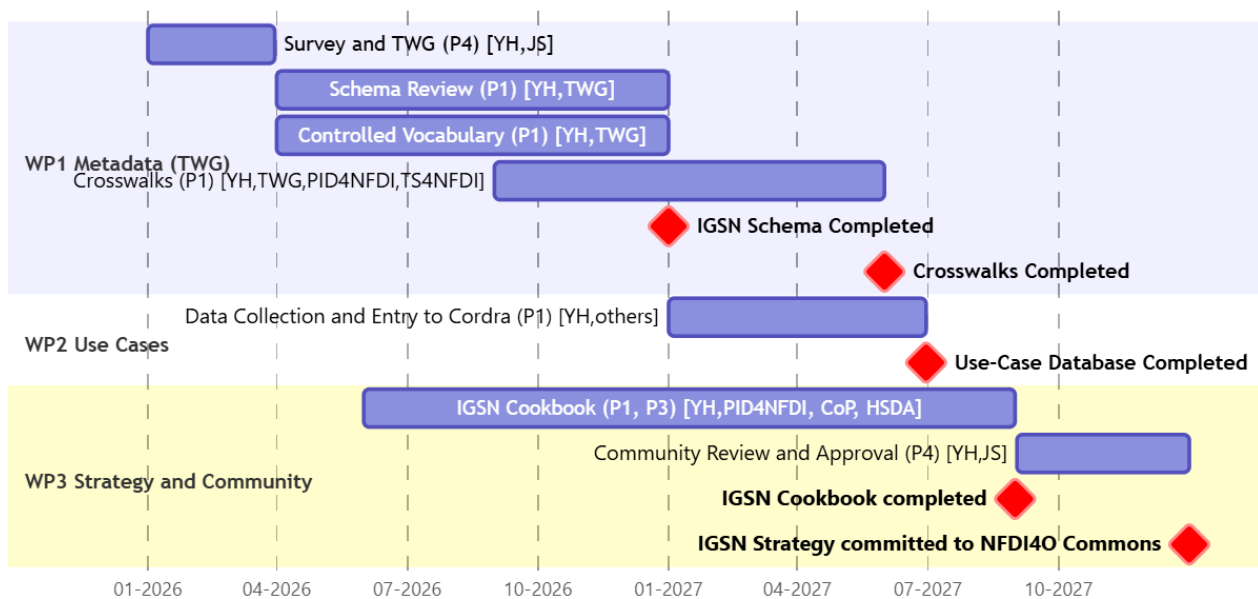
The work package will deliver a Cookbook specifying a practical IGSN strategy for sample-oriented studies in archaeology and cultural heritage. Upon its acceptance within the NFDI4O consortium, the materials will be published via the Commons and MkDocs for wider reuse. Meanwhile, the established communication channels with PID4NFDI, CoP, and HSDA can support further development of IGSN workflows to become a sustainable service in a long run.

## **8. Work Plan**

Submit a resource-based milestone plan (in the form of a Gantt chart, if applicable) for the planned work phase (6-24 months). The plan should contain a realistic time frame with clearly defined milestones and concrete deliverables. Please take into account the reflections from the TRAIL evaluation and the SC decision on the strategic pillars.

The TRAIL is planned for 24 months, starting on 01.01.2026 and concluding on 31.12.2027. The primary tasks regarding metadata profiles, mapping, and community building are conducted by Yiu-Kang Hsu (YH) from the DBM, while Jan Sessing (JS) assists in coordination work with other task areas. Other researchers from participating institutions contribute their use cases and feedback to the IGSN workflows. PID4NFDI, CoP, and HSDA support the formulation of the IGSN strategy. The specific assignments and milestones in relation to the TRAIL timeline and strategic pillars (P) are illustrated by the following Gantt charts.

## P1: Standards/Guidelines | P3: Community | P4: Networking



## 9. Funding / Own contributions

The TRAIL will build on the existing institutional resources, expertise, and collaborative networks to perform the planned activities. Fifty percent of Yiu-Kang Hsu's (DBM) position will be dedicated to this project, with Jan Sessing (DBM) assisting in community building and engagement. Contributions from other participating institutions, the intended temporary working group, and consortia are rooted in a strong community-driven and collaborative initiative. Among these contributors are Katrin Westner (DBM), Ivan Calandra (LEIZA), Mattis thor Straten (Kiel), and Stephan Schiffels (Max Planck Institute), who will provide representative use to validate the IGSN metadata workflows. PID4NFDI, TS4NFDI, and our international partners (National Gallery London, Royal Institute for Cultural Heritage, Czech Academy of Sciences, and IGSN e.V) will support us to shape a sustainable IGSN strategy within research fields of the consortium.

## 10. Additions/Additional information

Yiu-Kang Hsu is a founding member of two ongoing IGSN initiatives – the Community of Practice for Archaeology (CoP) and the Heritage Sample Digital Archive (HSDA). Through these collaborative networks, the TRAIL can gain access to appropriate expertise and resources essential for implementing the planned work packages.

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

[2] Edmunds, R., Klump, J., Bartholdy, B. P., Crook, P., Corns, A., Hsu, Y.-K., Keller, C., Novák, D., Plomp, E., Rose, T., Ross, S., & Sessing, J. (2025). Archaeological Sample Metadata Profile (1.0). IGSN e.V. <https://doi.org/10.5281/zenodo.17254975>.

[3] E-RIHS (2025). *HSDA*. GitHub repository. <https://github.com/E-RIHS/HSDA>.

[4] Hagemann-Wilholt, S., & Kahlert, T. (2024). D4.2 Concept for Sustainable PID Registration Workflows. Zenodo. <https://doi.org/10.5281/zenodo.14267446>.

[5] Tatscheck, J., Manske, A., & Rettberg, N. (2025). 5 Steps to Base4NFDI Service Integration – A Guide for NFDI Consortia. Zenodo. <https://doi.org/10.5281/zenodo.17475101>.