



This PDF is a simplified version of the original article published in *Internet Archaeology* under the terms of the Creative Commons Attribution 3.0 (CC BY) Unported licence. Enlarged images, models, visualisations etc which support this publication can be found in the original version online. All links also go to the online original.

Please cite this as: Hsu, Y-K. and Edmunds, R. 2026 'Persistent Identifiers: Towards the FAIRness of Archaeological Samples for Scientific Analysis', *Internet Archaeology* 72. <https://doi.org/10.11141/ia.72.4>

Persistent Identifiers: Towards the FAIRness of Archaeological Samples for Scientific Analysis

Yiu-Kang Hsu and Rorie Edmunds

Summary

This paper explores how persistent identifiers (PIDs) and standardised metadata profiles can foster the FAIR (Findable, Accessible, Interoperable, and Reusable) management of archaeological samples and their associated analytical data. We begin by outlining the role of PIDs in ensuring the long-term findability of research outputs, with particular attention to the international generic sample number (IGSN ID) - a PID system specifically designed for material samples and well-suited to archaeological context. We then present the collective work of the IGSN-DataCite Archaeology and Cultural Heritage Community of Practice, which is developing an archaeology-specific metadata profile and a crosswalk recommendation to support the integration of IGSN IDs into existing research workflows. Finally, we illustrate these concepts through a case study at the Deutsches Bergbau-Museum Bochum (DBM), Germany, where PIDs and relevant metadata schemas have been embedded into the laboratory research infrastructure, including electronic laboratory notebooks, service catalogues and data repositories. Collectively, these efforts demonstrate how PID adoption not only strengthens the FAIRness and long-term sustainability of archaeological data but also more firmly positions archaeological research within the global Open Science landscape.

1. Introduction

The FAIRness of research outputs and entities refers to the application of the FAIR principles, making the outputs Findable, Accessible, Interoperable and Reusable (Wilkinson *et al.* 2016). The ultimate goal is to promote 'Open Science', whereby scientific knowledge and the production of that knowledge are freely accessible to the public (UNESCO 2022). The first step towards the FAIR use of research data is to make relevant digital content permanently discoverable, which is achieved through the assignment of a persistent identifier (PID) — a long-lasting entry point that uniquely references a digital resource (CODATA RDM Terminology WG 2023a).

PIDs play a pivotal role in sustaining the long-term discovery of digital research outputs. They are characterised by two key components: first, an actionable, globally unique, identifier that ensures a web-resolvable link (URI) points to the correct digital resource; and second, a service that maintains access to that resource over time, even if the original storage location has changed or become outdated (CODATA RDM Terminology WG 2023b). Unlike typical web addresses (URLs), which are prone to breaking or changes due to server maintenance, restructuring or content relocation, a PID is designed to be a stable designation that ensures permanent access to an associated research asset.

Several PID systems are currently in use (CODATA RDM Terminology WG 2023c), including the Digital Object Identifier ([DOI Foundation](#)), Handle ([Corporation for National Research](#)



[Initiatives](#)), Archival Resource Key ([ARK Alliance](#)), Persistent Uniform Resource Locator ([Internet Archive](#)), and Universal Resource Name ([Deutsche Nationalbibliothek](#)). PIDs have furthermore been assigned to a large variety of research entities (Figure 1). Common examples of their application are ORCID IDs for researchers ([ORCID](#)), ROR IDs for research organisations ([Research Organisation Registry](#)), and DOIs for publications and datasets ([Crossref](#) ; [DataCite e.V.](#)). Practical guidance on the registration and usage of PIDs is clearly stated in the *PID Cookbook* published by PID4NFDI ([PID4NFDI 2025](#)).

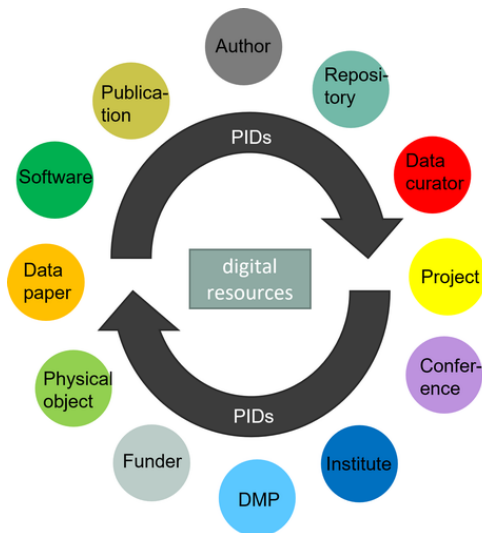


Figure 1: Digital academic resources that are commonly registered with persistent identifiers (PIDs) (image created by Yiu-Kang Hsu, licensed under CC0).

Metadata is another important component of PIDs. Metadata is 'data about data', and when a PID is created, it is registered with metadata that describes and provides context about the related digital content so that it can be more easily discovered, distinguished, interpreted and reused. For instance, a journal article is normally identified by its title, authors and the year of publication. Metadata can be further classified into two types (DataCite e.V. [2025a](#)):

1. PID metadata is general, standardised information stored by the PID registration service (e.g. ORCID) to link a globally unique identifier with a digital resource in a persistent way. Its main purpose is to address primary citation and core discovery of a research output.
2. Catalogue/repository metadata refers to more targeted, disciplinary information that provides a much richer description of the research output. A well-known research metadata profile is Dublin Core (DCMI [2025](#)). The main purpose of such metadata standards is to address domain-specific discovery and reuse of a research output.

The value of PIDs has been recognised by a number of countries. [MoreBrains](#) — a cooperative of consultants who specialise in the values of open research — has conducted cost-benefit analyses of implementing PIDs at a national level for the following nations: Australia (Brown *et al.* [2022a](#)), Czechia (Jones *et al.* [2024](#)), Ireland (National Open Research Forum and MoreBrains Cooperative [2024](#)), and the UK (Brown *et al.* [2021](#), [2022b](#)). In all cases, it has been shown that significant efficiency — and in turn, financial savings — are gained from well-established PID systems. For example, the 2022 report for Australia found that, by investing in PIDs, the country's research sector could save:

- Nearly 38,000 person days per year; the total time cost of rekeying metadata about grants, publications and projects.



- Nearly AUD \$24 million per year (approximately EUR €13.5); the direct financial cost of this wasted effort.

While it is difficult to translate these figures into savings per organisation, it is clear that the effort of an organisation to develop archaeological workflows that incorporate PIDs will ultimately result in significant reductions in resource overheads.

In archaeology, one of the most critical research entities that requires PIDs and rich metadata is the archaeological sample — a small, representative portion of materials or features collected from a larger archaeological site, context or object for the purpose of further examination and scientific analysis. Yet, the consistent use of PIDs for samples and their associated analytical data remains uncommon in many archaeological research laboratories. At the Deutsches Bergbau-Museum Bochum (DBM), Germany, for example, archaeological samples are still labelled with in-house identifiers that carry only minimal metadata. To establish a link between a sample and its original archaeological context, researchers must often contact the sample submitter directly. In addition, samples with in-house identifiers are typically only recorded in spreadsheets and stored in individual laboratory computers, tied to the specific instrument that generated the data. They are neither findable on the Internet (except in published articles) nor reusable by any means.

The DBM is also facing the problem of managing legacy data. Esoteric data management of previous researchers leads to specialised documentation that can only be understood or interpreted by a small group of experts. Even within the same field of study, researchers may utilise different cataloguing systems and metadata frameworks for similar types of material, further impeding data integration, comparability and long-term usability. These inconsistencies highlight the urgent need for a uniform approach to managing archaeological samples through PIDs and standardised metadata profiles.

The following section introduces a particular PID system for material samples, the [international generic sample number](#) (IGSN ID), and its application to the archaeological community in the form of sample metadata profiles. We then present a case study of the research data infrastructure at the DBM, stressing what measures have been considered to achieve the FAIRness of analytical data derived from archaeological samples.

2. IGSN IDs and a metadata profile for archaeological samples

2.1. International generic sample number (IGSN ID)

The IGSN ID (DataCite e.V. [2022](#)) is a PID that may be registered for any type of material sample (i.e. physical object) originating from any discipline, including archaeological samples. Note here that the IGSN ID is assigned to the material sample itself, and that, while the sample does not need to be persistent, it does need to have existed at some point in time (the sample may have been destroyed during analysis). Thus far, IGSN IDs have been assigned to geoscience samples, plant and soil specimens, museum artefacts, synthesised material science samples, and, importantly here, archaeological sites and samples. Over 13 million IGSN IDs [have been registered to date](#) and it is trusted and used by universities, research organisations, and national research infrastructures, as well as being endorsed by publishers (COPDESS [2018](#)) and explicitly included in national PID strategies (Brown *et al.* [2023](#)).

The IGSN ID was founded in 2004 as the International Geo Sample Number, a globally unique identifier for geosciences samples (Klump *et al.* [2021](#)). The additional layer of persistence was added to the global uniqueness of the IGSN ID in 2011, driven by the IGSN organisation ([IGSN e.V.](#)) that was set up to govern the IGSN ID system and managed a



Handle server for the IGSN (P)ID. It was created to solve a problem that is far from exclusive to the geosciences: there are many samples in the world that have the same identifier and there are many samples in the world that have multiple identifiers, so how does anyone know that they are referring to the correct sample? This utility of the IGSN ID beyond the geosciences was understood from its conception, and increased adoption by other disciplinary sample communities for different sample types has come through the formation of a partnership between IGSN e.V. and DataCite in 2021 (Buys and Lehnert [2021](#)). Within this partnership:

- DataCite is providing IGSN ID registration through its services and supporting their ongoing sustainability.
- IGSN e.V is implementing and promoting standard methods for identifying, citing and locating material samples.

Two important, immediate, changes resulted from the partnership. The first was that the IGSN acronym now represents 'International Generic Sample Number', to reflect that the PID can be applied widely across samples from any research domain. The second is that with the IGSN PID infrastructure now under DataCite services, the IGSN ID has been transitioned from a Handle to being functionally a DataCite Digital Object Identifier (DOI), with IGSN ID metadata for samples structured according to the DataCite Metadata Schema (DataCite Metadata Working Group [2024](#)).

To ensure that IGSN ID metadata is of high-quality and complete, a crosswalk recommendation was published by a joint IGSN-DataCite effort (IGSN-DataCite Crosswalk and Metadata Management Working Group [2022](#)), mapping between the registration and discovery metadata schemas developed by the IGSN e.V. and the DataCite Metadata Schema. However, the focus of the IGSN e.V. schema was mainly geoscience samples, and the partnership is looking to better support samples originating from any domain. It is partially for this reason that the IGSN-DataCite Archaeology and Cultural Heritage Community of Practice was established (see [section 2.2](#)).

2.1.1. Using IGSN IDs for archaeological samples

The main reason for employing IGSN IDs is that they provide a digital footprint for an archaeological sample, thus making it part of Open Science. The focus of the FAIR principles and of Open Science principles has been on digital research outputs such as datasets; however, it is equally as vital, if not more so, to connect physical objects to this digital world. By assigning an archaeological sample with an IGSN ID, it exposes the sample, showing that it exists and that information is available about the sample. This helps ensure that the investment put into obtaining that sample continues to be valuable and supports future research. Such long-term preservation of a sample alongside its related data and information is particularly important for many archaeological finds due to their uniqueness, meaning they are impossible to replace.



General Identifiers

Program: ICDP
 Expedition: ICDP 5054
 Type: Hole
 Name: 5054_1_A
 IGSN: ICDP5054EHW1001
 Comment: valid alias: ICDP5054EEW1001
 Parent IGSN: N/A
 Release Date: 2017-3-1

Sampling Location

Latitude: 63.40163
 Longitude: 13.202917
 Coordinate System: WGS84
 Elevation: 522.51
 Final Depth: -1980.29
 Location Type: N/A
 Location Name: Åre, Jämtlands län, Sweden
 Location Description: COSC-1 is located in the vicinity of the abandoned Fröå mine
 Country: Sweden
 Province: Jämtlands län
 County: N/A
 City: Åre

Geology

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

Methods

Downhole Logging: Acoustic televiewer LII, DIP OSG, DLL OSG, ELOG LII, FAC49 OSG, Focused electric LII, Full-waveform sonic LII, Full-waveform sonic OSG, Magnetic susceptibility OSG, Mudparameter OSG, NN, Sidewall density LII, Temperature/Conductivity LII, Temperature/Conductivity OSG, Temperature/Conductivity LII, Three-arm caliper LII

Drilling

Drilling Method: Coring-RackCorer
 wireline diamond coring, HQ and NQ bit size
 Operator: Lund University, Engineering Geology
 Larson Drilling Consulting AB
 Funding Agency: Swedish Research Council (Vetenskapsrådet)
 Total Length: 2400.1m
 Comments: valid alias: ICDP5054EEW1001
 Platform Type: drill rig
 Platform Name: Atlas Copco CT20C
 Platform Description: slimhole wireline coring system
 Chief Scientist: Chris Junlin
 Start Date: 2013-09-05
 End Date: 2014-08-26

Repositories

Current Repository: Bohrkemlager für kontinentale Forschungsbohrungen, Berlin-Spandau, Germany
 Current Repository Contact: Jochen.Ersbacher@bgr.de
 Original Repository: N/A
 Original Repository Contact: N/A

Sample Family

- 5054_1_A
 - 5054_1_A_1_Z
 - 5054_1_A_2_Z
 - 5054_1_A_3_Z
 - 5054_1_A_4_Z
 - 5054_1_A_5_Z
 - 5054_1_A_6_Z
 - 5054_1_A_7_Z

Location Map

Drilling Start/End: 2013-09-05 / 2014-08-26 *
 Latitude: 63.40163 * Longitude: 13.20292 *
 Åre, Jämtlands län, Sweden

Publications & Datasets

Lorenz, H., Roebberg, J.-E., Junlin, C., Bjeltn, L., Almqvist, B. S. G., Berthel, T., ... Tang, C.-F. (2015). COSC-1 - drilling of a subduction-related allochthon in the Palaeozoic Caledonide orogen of Scandinavia. *Scientific Drilling*, 19, 1-11. doi:10.5194/sd-19-1-2015

Lorenz, Henning; Roebberg, Jan-Erik; Junlin, Christopher; Bjeltn, Leif; Almqvist, Björnst; Berthel, Theo; Conze, Ronald; Gary, David G.; Hönigswald, Ineska; Petzold, Christopher; Pedersen, Karsten; Roberts, Nick; Tang, Chinfu (2015). COSC-1 operational report - Operational data sets. GFZ Data Services. <http://doi.org/10.5194/GFZ-SD08-ICDP5054-2015>

Figure 2: A landing page for a geosciences sample (IGSN: ICDP5054EHW1001: <https://doi.org/10.60510/icdp5054ehw1001>) published by the GFZ Helmholtz Centre for Geosciences. Notice that the 'sample family' is incorporated top-right of the landing pages.

The digital representation of an archaeological sample is its landing page (Figure 2), a web-resolvable URL that displays a description of the sample identified by the IGSN ID. Often the presentation of metadata on a landing page will be institutional or community dependent such that this differs among portals and catalogues. However, the metadata should in all cases be as rich and complete as possible, while being within the boundaries of revealing sensitive information — parts of a metadata record may need to be withheld or blurred to protect, for example, a vulnerable site. As well as high-quality, descriptive metadata, it is strongly recommended to include other elements in a landing page that improve discoverability of the archaeological sample, such as images of the sample, maps/diagrams showing the location where it was found, a depiction of the hierarchical relationships among samples (if this exists; see below), and a QR code of the landing page URL that can be added to sample labels.

The last point concerning QR codes brings us to the second major use for IGSN IDs: they enable the inventory or collection management of archaeological samples. It may not be clear or even known what is in a long-term archive and therefore what can be made discoverable. IGSN IDs can support research by ensuring that available resources are discoverable both internally and externally. Furthermore, they enable knowledge sharing; with archaeological sample collection and management very much in the hands of the principal investigators (PIs) of a project, historically there has often been no common, standardised infrastructure for sample management and knowledge sharing. The metadata information about samples is kept in a personal archive and may become inaccessible after the project's lifespan.



IGSN IDs offer a simple way to link an archaeological sample with its digital representation: encode the IGSN ID as an actionable URI within a QR tag or barcode, that is then included in a permanently affixed label on the sample to which the identifier has been assigned (Figure 3). This enables machine-readable identification through handheld devices (even a smartphone) that can form the basis of an inventory/collection management system (or augment an existing one). Such a label can also include any local accession or inventory numbers for the sample, and, as an archaeological sample is a physical entity that is handled by humans, ideally a label will show the QR code, as well as both the IGSN ID and any local identification numbers, in a human-readable way.

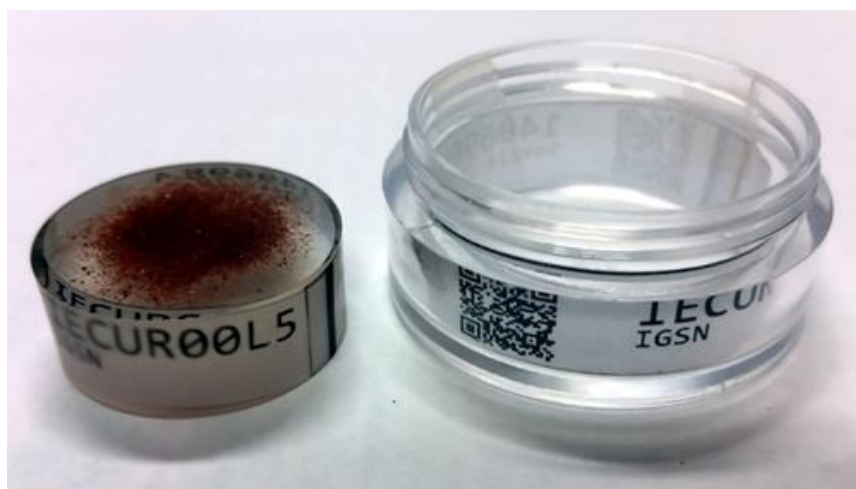


Figure 3: An international generic sample number (IGSN ID) can be encoded as a QR code and affixed as a label to a sample housing. The label should also include human-readable information (image taken by Jens Klump, licensed under CC0).

It should be noted here that the integration of IGSN IDs into existing identification and archaeological sample management systems can be quite straightforward. Because IGSN IDs are functionally DOIs, the identifiers are composed in the usual prefix/suffix configuration of DOIs, whereby the prefix is a fixed number starting with '10.' that uniquely identifies a specific metadata collection on a global level, and the suffix is a customisable string that uniquely identifies the sample within the metadata collection (Figure 4). Although it goes against standard best practice for DOIs (DataCite e.V. [2025b](#)), which recommend the use of a random string for the suffix, there is strong justification in the case of archaeological samples to incorporate local naming conventions already in use as DOI suffixes, and thus to transform locally unique identifiers into globally unique ones (the concatenation of the prefix and the suffix ensure global uniqueness). Hence, one can introduce IGSN IDs without necessitating changes to established working procedures, naming conventions or data systems.



Figure 4: An IGSN ID name is made up of a prefix beginning with '10.' and a suffix, separated by a forward slash. The prefix ensures that the IGSN ID is globally unique. The suffix can be customised to incorporate local identifiers (image created by Rorie Edmunds, licensed under CC0).



As noted in the introduction, the IGSN ID exists because local systems for the unique identification of samples have limited scope. Not only will sample identifiers likely be ambiguous outside an organisation, but these identifiers often can (and do) change at different stages of the archaeological sample management process. As we are again concerned with physical entities, archaeological samples will typically change location, and ambiguity of identifiers makes it difficult to track a sample as it moves across institutional and system boundaries.

While an IGSN ID may be registered for an archaeological sample at any point in its lifecycle, it is strongly recommended to assign IGSN IDs as early as possible as the most effective way to avoid ambiguity. Ideally, the IGSN ID should be registered for the archaeological sample at its 'birth'; namely, at the point of collection. By adding a permanent, globally unique identifier to an archaeological sample, it enables that sample to be tracked throughout different stages of its lifecycle and as it is moved to different laboratories or archives.

Finally, IGSN IDs have wider applications beyond being registered for individual archaeological samples. In particular, and as intimated previously, IGSN IDs may be assigned to archaeological sites; namely, the location where the archaeological samples were collected. They can, moreover, be assigned at the archaeological sample aggregate level in situations where you want to reference (for instance) multiple fragments that belong to a single object or an assemblage of the same sample type. In both of these cases, IGSN IDs can then be added to individual samples from the site or the aggregate as needed.

From the above, we see that there is often a hierarchical relationship structure whereby a 'parent' site/sample has a number of sample 'children', each of which may then be a parent to their own children, and so on. The same is true when a sample has been subsampled. By adding IGSN IDs to all parents and children, it means that the hierarchical structure can be mirrored in the IGSN ID metadata. Doing this, one can not only take advantage of the fact that some IGSN ID metadata may be inherited by a child from its parent, but also that the topmost parent acts as the 'collection' and may be referenced to represent all of its descendants. For example, if an archaeological site and all samples collected there are registered with IGSN IDs and the parent-child relationships are captured in the IGSN ID metadata, then one may reference the site as the top-level parent, and in doing so effectively refer to all of the samples.

It is extremely helpful for the discovery and distinction of archaeological sites and samples to include a depiction of the hierarchical relationships (the 'sample family') within IGSN ID landing pages (Figure 2).

2.1.2. Benefits of IGSN ID adoption

Many benefits that arise from the adoption of IGSN IDs for archaeological samples have already been alluded to in the previous subsection. However, we would like to explicitly highlight some of the main benefits.

First and foremost, IGSN IDs are functionally DOIs, which are widely recognised across and adopted by a large established community composed of various research disciplines, academic institutes and publishers. They are almost certainly the same type of PID that you are using for the citation of journal articles and likely also for other research outputs, such as datasets. This not only gives archaeological samples with IGSN IDs increased visibility but



also credibility: DOIs are seen as a marker of trust and reliability, having a high-level of authority because of their wide recognition, adoption and support.

Moreover, IGSN IDs are built on the robust international DOI system ([DOI Foundation](#)), ensuring long-term availability and stability. Even if DataCite no longer exists, all IGSN IDs will continue into the future. Having such long-term sustainability means that the IGSN ID can evolve to resolve issues and cater for new practices and trends, with its community providing support and resources for its implementation and maintenance.

Beyond this, IGSN IDs help ensure archaeological samples adhere to the FAIR principles, CARE (Collective benefit, Authority to control, Responsibility, and Ethics) principles (Carroll *et al.* [2020](#)), as well as the principles of Open Science. Specifically, they support:

- Findability: IGSN IDs can be used to locate archaeological samples, ensuring the samples remain discoverable and citable over time.
- Accessibility: IGSN IDs provide a means of accessing archaeological samples distributed across multiple repositories, as well as their descriptive metadata.
- Interoperability: IGSN IDs have standard metadata, which provides a common standard for identifying and describing archaeological samples, and leads to consistencies in data management. It is easier to organise and retrieve information about samples, and to integrate samples and their metadata across systems.
- Reusability: IGSN IDs ensure referencing of archaeological samples with confidence, encouraging their reuse and ensuring proper attribution to the original authors.
- Openness: IGSN IDs expose the existence of archaeological samples and the availability of information; IGSN ID metadata is in the public domain under a CC0 licence.
- CARE: IGSN ID metadata can include Local Contexts' Traditional Knowledge or Biocultural Notices and Labels to ensure indigenous data sovereignty is recognised and upheld.

2.2. IGSN-DataCite archaeology and cultural heritage communities of practice

2.2.1. IGSN community of communities

To scale adoption and use of IGSN IDs across the diverse research communities, the IGSN-DataCite partnership took the decision to establish domain-specific communities of practice (CoPs) in a 'community of communities' type approach. With the IGSN e.V. taking the leading role in creating and managing these CoPs, the aim within disciplinary communities is to:

1. Develop/promote standardised methods for identifying, citing and locating material samples.
2. Provide recommendations on how the DataCite metadata schema be enhanced to better support material samples.

The IGSN-DataCite Archaeology and Cultural Heritage CoP (IGSN e.V. [2024](#); hereinafter the 'Archaeology CoP') was the first of these CoPs, beginning in March 2022 as a pilot: a blueprint for which others might follow. Jointly chaired by IGSN e.V. and DataCite ex official members, the CoP has been meeting monthly since that time, and has brought together a small group of archaeologists with backgrounds in archaeometry, landscape archaeology and data management, from prehistoric to modern eras, all of whom have an interest in material sample collection and management, analytics and publication.



While taking into account the specific challenges for archaeology — in particular, the diversity of sample types, use cases and community practices — the goals of the CoP are:

- Promoting the use of PIDs for archaeological samples.
- Identifying and articulating archaeological sample use cases and workflows within archaeology (sub)disciplines.
- Describing archaeological samples in terms of a minimum metadata profile and using this as a basis for extensions to subdomain samples/edge cases that require more specialised description.

Significant work has been done to meet the first two of these goals, but it is the third that has the most relevance to the subject of this article and is the focus for the remainder of this section.

2.2.2. Archaeological sample metadata profile and CoP crosswalk recommendation

The approach of the CoP was to develop a core metadata profile for describing archaeological samples in a standard way, and then to map this profile to the DataCite Metadata schema such that the archaeology community can more easily register IGSN IDs. In this way, the CoP could also determine any gaps between the descriptive metadata required by an archaeological sample and what can be captured in the DataCite metadata schema, deciding whether these are fundamental and the DataCite schema should be updated to incorporate them, or are highly specific to archaeological samples and thus might be included in recommendations about the content in landing pages (note here that we are talking about the description of only archaeological samples and not of collection sites).

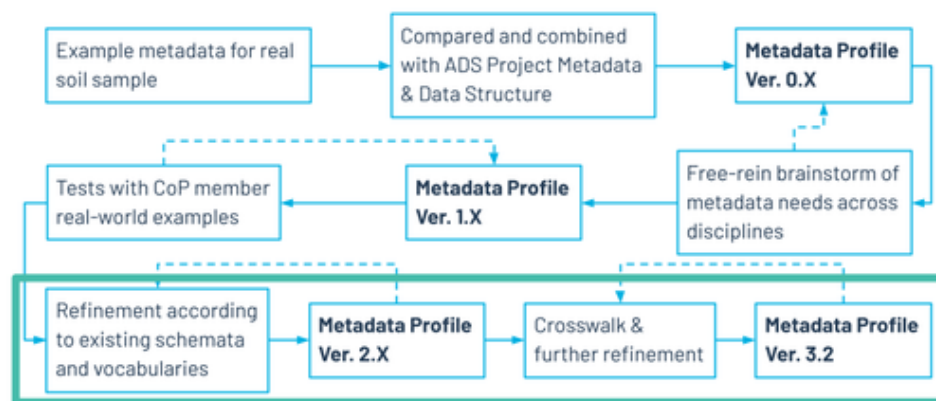


Figure 5: Process undertaken by the IGSN-DataCite Archaeology CoP to develop the Archaeological Metadata Profile and Crosswalk Recommendation. In the latter stages (green box), the CoP was careful to not 'reinvent the wheel' (image created by Rorie Edmunds, licensed under CC0).

The development of the archaeological sample metadata profile is shown in Figure 5. It began in August 2023 with a CoP member providing example metadata for a real soil sample that she had collected in the field. The CoP compared and amalgamated the properties of this example metadata record with those in the project metadata (Niven and Watts 2011) and the [data structure](#) of the Archaeology Data Service in the UK, and led us to a Version 0 of the metadata profile. Taking this as a foundation, the CoP then began iterating Version 0 through a series of brainstorming sessions to form Version 1. The idea behind the brainstorming sessions was to allow free, blue-sky thinking by CoP members as to what properties are ideally needed for describing archaeological samples. The CoP was,



of course, highly cognisant of what already exists, but did not want to be too influenced or bound by the work of others in these initial stages.

The next step was for the CoP to see whether (1) there were any gaps between what would typically be described for an archaeological sample and the metadata profile properties, and (2) anything being asked for was too prescriptive/restrictive for certain sample types (especially in terms of properties considered mandatory). With CoP members coming from a number of archaeology subdomains, to perform this check they were asked to use Version 1 to record the metadata taken from real samples that they work with.

It was at this point that the CoP also started to look at existing metadata schemas, such as the DataCite schema and IGSN descriptive schema (IGSN e.V. [2016](#)). Common community vocabularies were also to ensure standardisation of the description of archaeological samples and avoid reinventing the wheel ([Forum on Information Standards in Heritage](#); [The Getty Research Institute](#)). The CoP was careful when drafting the profile and recommendation to allow for freedom and flexibility, but also to build on and ensure consistency with previous work, and especially the work of the geosciences community that founded the original IGSN ID and IGSN e.V. The combination of these two efforts, the tests and incorporation of existing work, resulted in Version 2 of the profile.

Finally, with the profile somewhat stable at this point, the CoP then moved its focus to mapping the profile to the DataCite schema and thus developing its crosswalk recommendation. This exercise largely followed the process conducted by the IGSN-DataCite partnership metadata working group in developing its recommendation, as well as using the recommendation itself as the basis of the CoP's crosswalk. In performing this mapping and producing the recommendation, the CoP realised that further refinements were needed to the profile. Alongside a final sanity check by selected, external, community members interested in the CoP's endeavours, it brought the CoP to the ready-to-publish Version 3.2 of the profile (<https://doi.org/10.5281/zenodo.17254975>) and Version 0.2 of the recommendation (<https://doi.org/10.5281/zenodo.17446558>). Both of them are published as public Version 0.0.

Table 1: A summary of main properties of the Metadata Profile Version 3.2 developed by the IGSN-DataCite Archaeology CoP. Here, M = Mandatory, R = Recommended, and O = Optional.

#	Property	Obligation	Definition
1	Project	M	The title (and any alternatives) of the site or project.
2	investigationType	R	Lists all investigation types relevant to the resource. Should follow a standard vocabulary.
3	investigationDates	M	Full date range using RKMS-ISO8601 format, indicating when the archaeological project was carried out or the processing dates. For ongoing or multi-annual investigations use the dates of the specific project for which the investigation took place.
4	collectorExcavator	M	Details of the creator(s), compiler(s), funding agencies, or other bodies or people intellectually responsible for the sample collection.
5	Publisher	M	The organisation that is registering the metadata record for the sample.
6	Contributor	R	Other individual(s) or organisation(s) who have contributed to the collecting, analysis, or managing of the sample.
7	sampleTitle	M	Title or name that enables discovery and distinction of the sample.
8	Description	R	Detailed description of the sample. Should include information about the sample and its collection not captured in other properties.
9	Subject	R	Keywords for the subject content of the sample. Keywords should be taken from standard vocabulary.
10	otherIdentifier	M	The project/reference/catalogue/sample number(s) used to identify the sample on site.



11	collectionDate	M	Date when sample was collected following W3CDTF format. May optionally include the time at which the sample was collected.
12	collectionMethod	R	How the sample was collected. Where possible, terminology should follow a standard vocabulary.
13	stratigraphicContext	R	The site or project specific identifier of the stratigraphic or survey unit from which the sample was collected. Considered mandatory in the case of samples collected during surveys and excavations.
14	sampleDepth	R	Depth of the sample within the stratigraphic context.
15	sampleWeight	R	Weight at the point of collection, before analysis.
16	sampleMaterialType	M	The type(s) of material(s) that compose the sample. Should follow a controlled vocabulary.
17	Colour	R	Colour(s) of the sample following a standard vocabulary.
18	absoluteDating	R	Numerical age or range of the sample. Dates and ranges should follow W3CDTF and RKMS-ISO8601, respectively.
19	relativeDating	R	Cultural and user created values for relative dating of the sample. Should follow a standard vocabulary.
20	generalKeywords	O	User-created values to describe aspects of the sample not covered by other metadata.
21	currentStatus	R	Information concerning the current status of the sample. It is recommended to also indicate the date the status was revised.
22	Relations	R	Relationships to other research outputs/entities that are identified by a globally unique identifier.

Table 1 is a summary of the metadata profile developed by the CoP. It contains 22 properties, many with a number of sub-properties, and the obligations of each of these (sub-)properties (i.e. whether they are mandatory, recommended or optional), definitions, and examples. The properties (in order) are loosely grouped as follows:

- Metadata about the investigation that surveyed or excavated the archaeological sample, including about the archaeological site: *1. Project, 2. investigationType, 3. investigationDates.*
- Metadata about the different organisations and people involved in the collection and management of the sample throughout its lifecycle: *4. collectorExcavator, 5. Publisher, 6. Contributor.*
- Metadata concerned with basic discovery of the archaeological sample, including important terms from community-agreed vocabularies: *7. sampleTitle, 8. Description, 9. Subject, 10. otherIdentifier.*
- Metadata about the collection of the sample and the context of the sample when it was collected: *11. collectionDate, 12. collectionMethod, 13. stratigraphicContext, 14. sampleDepth, 15. sampleWeight.*
- Metadata capturing specifics about the archaeological sample to help experts distinguish, understand and potentially reuse the sample: *16. sampleMaterialType, 17. Colour, 18. absoluteDating, 19. relativeDating, 20. generalKeywords, 21. currentStatus.*
- Metadata linking the archaeological sample to any related research outputs, such as its parent or children, any associated data, articles that reference it, and so on: *22. Relations.*

With the exception of the 'generalKeyword' property, which is optional and thus ignored, all of the above are mapped to the DataCite metadata schema in the crosswalk recommendation.

The versioning of the public profile and recommendation are important, as they indicate that these are living documents. The archaeology community is strongly encouraged to adopt them both, but they are also seen as a starting point for the community to enhance over



time. In this regard, the profile has been released with a [GitHub repository](#) that enables people to suggest changes for future versions of the profile (and, in turn, revisions to the recommendation). The CoP will work with the wider community to review and update the profile and recommendation on a regular basis, with the process governed by the IGSN e.V.

The release of the profile and recommendation is expected to impact the archaeology community by promoting and reinforcing best practices. In particular, the two CoP outputs advocate the:

1. utilisation of standard (ideally FAIR) community vocabularies and formats as a norm by archaeologists when describing their samples;
2. assigning and referencing of PIDs for all research outputs and entities, not only for archaeological samples.

3. The FAIR strategy at the research laboratory of the DBM

This section presents a brief description of how PIDs and associated metadata frameworks are planned to be implemented in an archaeometry laboratory. Since 2023, DBM has been developing a research infrastructure for the management of analytical data, particularly those derived from physical objects/samples related to ancient mining and metallurgy. This initiative arises from a collaborative effort to establish a nationwide research data infrastructure for the material remains of human history, known as [NFDI4Objects](#). One of the DBM's core missions is to provide a range of data services to the general public, enabling the description, deposition and dissemination of analytical results generated by scientific instruments in laboratory settings. To this end, a work plan focused on the analysis of archaeological samples has been formulated. This incorporates four key programmes: (1) the development of metadata profiles for analytical services, devices and techniques; (2) data management through electronic laboratory notebooks; (3) search portal for analytical services; and (4) data repository for scientific results (Figure 6). The PIDs of different types are considered throughout these programmes.

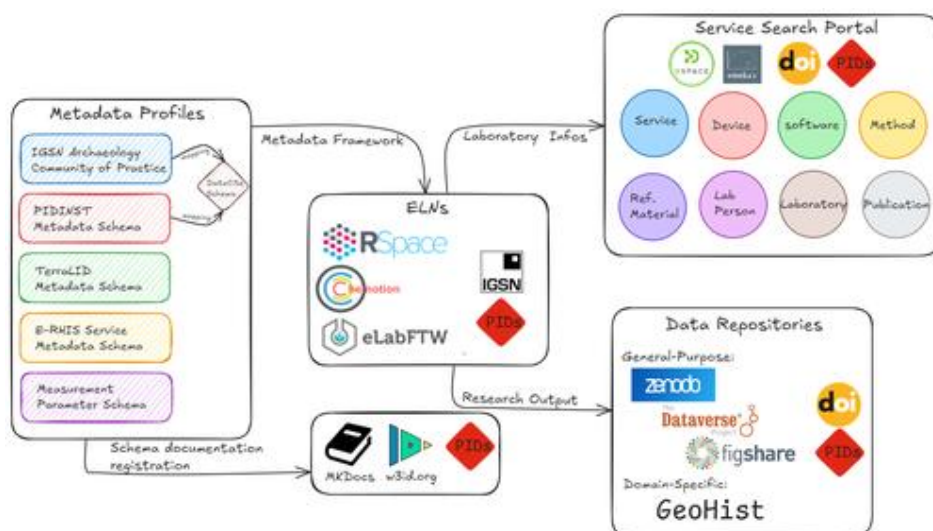


Figure 6: A FAIR research infrastructure for analytical data at the research laboratory of the DBM (image created by Yiu-Kang Hsu, licensed under CC0).

3.1. Metadata profiles



The metadata profile that describes the underlying analytical results serves the core of the entire infrastructure. It provides a rich description of a physical sample, allowing researchers to unambiguously identify the data they seek in a comprehensive and universally recognisable manner. In this regard, we are adopting existing community-driven metadata schemas to meet our objectives. The IGSN CoP has been selected to record the fundamental archaeological information. The profile is well mapped to the DataCite schema (DataCite Metadata Working Group [2024](#)) thereby facilitating the future implementation of IGSN IDs in electronic laboratory notebooks and data repositories.

While IGSN CoP recommends general metadata information about archaeological samples, [TerraLID](#) with its prototype GlobaLID (Klein *et al.* [2022](#)) provides a domain-specific metadata profile for lead isotope data through active community engagement. The profile contains multiple modules, including sites, assemblages, objects, samples, analyses and materials. Properties in each module are registered with PIDs to facilitate linkage between sample records.

Apart from samples, it is equally essential to manage the scientific devices that generate the associated analytical data. For this purpose, we opt for the metadata standard created by the Persistent Identification of Instruments Working Group ([PIDNST](#)). It is an initiative within the Research Data Alliance (RDA) and seeks to establish a community-driven standard for the globally unique identification of measuring instruments. The PIDNST schema specifies the key information required to link an active instrument with the data it generates, including instrument types, measurement properties, laboratory practitioners, manufacturers, models, landing pages and identifiers. [B2INST](#) complies with such a data standard and allows for the registration of PIDs for instruments. Meanwhile, the PIDNST profile is already mapped to the DataCite Schema, thereby facilitating the interoperability within the DOI system.

As a research laboratory may need to showcase a wide range of analytical services to the public, it is essential to establish a metadata profile to regulate such activities. The Metadata Working Group within the European Research Infrastructure for Heritage Science (E-RIHS) has published a [knowledge base](#) that specifies detailed descriptions of services, devices, software, standard methods, techniques, researchers and organisations (<https://github.com/E-RIHS/schema>). The instances (i.e. metadata records from partner laboratories) are stored and managed in a Cordra system, with the minting of Handle IDs for registration. The system enables users to complete the necessary information through a user-friendly online form. In addition, the E-RIHS metadata profile is complemented by a set of [controlled vocabularies](#).

Finally, we are formulating a metadata profile for instrument parameters that control analytical experiments, may influence subsequent interpretations and/or are potentially important for data comparison and discovery. These parameters vary according to different analytical techniques, sample types, and models of devices. To address this, a temporary working group has been formed to focus on the most relevant analytical techniques conducted in archaeometry laboratories in Germany (Hsu *et al.* [2024](#)). The current draft includes metadata profiles for Raman spectroscopy, Fourier-transform infrared spectroscopy, X-ray fluorescence, X-ray diffraction, scanning electronic microscopy, and inductively coupled plasma mass spectrometry. In the future, these profiles will be published on an online documentation platform, with each concept/term linked to a PID through [w3id.org](#).

3.2. Electronic laboratory notebook

An electronic laboratory notebook (ELN) is a digital platform for recording research activities and outcomes in a laboratory setting. It supports the management of digital assets linked to



scientific workflows, including samples, analytical protocols, instruments, consumables, reference standards, researchers and publications.

ELNs offer several advantages. For example, the metadata profiles mentioned earlier can be integrated into an ELN, helping to standardise the documentation of associated digital contents. They also allow users to collaborate more effectively by organising and sharing protocols and research outcomes. In this sense, an ELN serves as an internal knowledge base and networking hub for analytical data generated within the laboratory. Moreover, the ELN is ideally the first point of entry for minting PIDs, particularly for physical samples intended for scientific investigation. Finally, some ELNs provide direct export of data to repositories such as Zenodo, Dataverse and Figshare, thereby fulfilling the entire research data lifecycle.

A variety of ELN products are available on the market, including both commercial and open-source solutions. The [ELN Finder](#) offers a comprehensive overview of current software options. At the DBM, we are currently testing three open-source tools for laboratory management: [RSpace](#), [Chemotion](#) and [eLabFTW](#).

With regard to PIDs, [LabArchives](#) supports the assignment of DOIs to different levels of digital resources, including an entire notebook, individual folders, specific pages, and single entries. Notably, RSpace is currently the only ELN that provides native support for minting IGSN identifiers. This functionality is integrated into its inventory module, enabling users to draft, edit metadata, and publish IGSN IDs directly for samples.

In parallel, a [focus group within PID4NFDI](#) (Persistent Identifiers for the German National Research Data Infrastructure) is actively discussing how PIDs can be implemented across open-source ELNs.

3.3. Catalogue of services

While ELNs manage internal laboratory data workflows, service catalogues describe available analytical services for interested users. By collecting service information from its partner laboratories, E-RIHS has designed a [catalog of analytical activities](#) for a funding programme, allowing applicants to select services suitable for analysing their samples.

Several other online platforms offer similar functionalities, providing technical specifications of services and instruments. For example, [openiris.io](#) serves as a single-entry point for researchers to share, access and manage resources related to laboratory instruments. [LabInfrastructure@Geo.X](#) specifically covers laboratory facilities within the research network of geosciences in Berlin and Potsdam. More recently, [AnalyteMe](#) has emerged as a global initiative to create a comprehensive laboratory database. The project aims to compile an extensive instrument registry, provide detailed documentation of configurations and applications, and offer informative listings of available analytical services.

The DBM is also building up a similar search portal aimed at the archaeometry community. The initiative currently involves five laboratories in Germany — DBM, CEZA (Curt-Engelhorn-Zentrum Archäometrie), LEIZA (Leibniz-Zentrum für Archäologie), Rathgen-Forschungslabor, and Science and Archaeometric Laboratory at the Hochschule für Bildende Künste Dresden — each contributing information about their analytical facilities to a centralised database. The data framework follows the previously described metadata profiles, encompassing detailed information on services, software, instruments, reference materials, researchers and organisations.



For the first prototype, we are testing the infrastructure using the open-source platforms [DSpace](#) and [Omeka S](#). DSpace provides a robust framework for documentation, while Omeka S is particularly well suited for museum collections and exhibitions. In addition to search functionalities, the portal will assign PIDs to digital resources and enable authorised authors to upload, edit and publish their own digital content.

3.4. Data repositories

To ensure long-term public access to research outputs, electronic laboratory notebooks (e.g. RSpace) provide the option to export data to widely used general-purpose repositories, such as Zenodo and Dataverse. Alternatively, researchers can explore [re3data.org](#), a registry of research data repositories, to search for suitable discipline-specific platforms. The digitalisation team at the DBM is currently developing a specialised repository named GeoHist ([Georesources in Human History](#)). This platform curates digital resources on the (pre)historical, ancient, pre-modern and industrial use of raw materials across archaeological, archaeometric, anthropological and historical studies. Moving forward, our objective is to integrate laboratory data into this system, creating a single access point that links instruments, analytical results, and archaeological information.

4. Concluding remarks

PIDs and interoperable metadata standards provide the foundation for making archaeological samples and their associated analytical data discoverable, accessible and reusable over the long term. By adopting IGSN IDs, archaeology can move beyond isolated, in-house cataloguing systems and into globally interoperable frameworks that strengthen research integrity and sustainability. The work of the [IGSN-DataCite Archaeology and Cultural Heritage Community of Practice](#) illustrates how discipline-based collaboration can produce practical solutions — such as the community's archaeology-specific sample metadata profile, which is grounded in international standards.

A case study at the DBM demonstrates how these principles can be put into practice at the laboratory scale. By embedding PIDs into metadata profiles, electronic laboratory notebooks, service catalogues and repositories, archaeologically generated data can be fully integrated into the broader linked open data ecosystem (Schmidt *et al.* [2022](#)). PID-based approaches are therefore far more than technical compliance measures: they represent a transformative step towards ensuring the long-term usability of archaeological samples, improving research efficiency, and opening new avenues for interdisciplinary and community collaboration.

Bibliography

Brown, C., Simons, N., Bangert, D., and Sadler, S 2023 *RDA National PID Strategies Guide and Checklist (1.0)*, Zenodo. <https://doi.org/10.15497/RDA/00091>

Brown, J., Jones, P., Meadows, A. and Murphy, F 2022a *Incentives to Invest in Identifiers: A Cost–Benefit Analysis of Persistent Identifiers in Australian Research Systems*, Zenodo. <https://doi.org/10.5281/zenodo.7100578>

Brown, J., Jones, P., Meadows, A. and Murphy, F 2022b *Revised Cost–Benefit Analysis for the UK PID Support Network*, Zenodo. <https://doi.org/10.5281/zenodo.7356219>



Brown, J., Jones, P., Meadows, A., Murphy, F. and Clayton, P 2021 *UK PID Consortium: Cost–Benefit Analysis*, Zenodo. <https://doi.org/10.5281/zenodo.4772627>

Buys, M. and Lehnert, K 2021 'Partnership between IGSN and DataCite', [online] DataCite e.V. <https://datacite.org/blog/igsn-and-datacite-partnership/>

Carroll, S., Garba, I., Figueroa-Rodríguez, O., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J., Anderson, J. and Hudson, M 2020 'The CARE principles for Indigenous data governance', *Data Science Journal* **9**, 43. <https://doi.org/10.5334/dsj-2020-043>

CODATA RDM Terminology WG 2023a 'CODATA Research Data Management Terminology: PID System', [online] CODATA. <https://terms.codata.org/rdmt/pid-system> [Last accessed: 5 September 2025].

CODATA RDM Terminology WG 2023b 'CODATA Research Data Management Terminology: PID Service', [online] CODATA. <https://terms.codata.org/rdmt/pid-service> [Last accessed: 5 September 2025].

CODATA RDM Terminology WG 2023c 'CODATA Research Data Management Terminology: Persistent Identifier', [online] CODATA. <https://terms.codata.org/rdmt/persistent-identifier> [Last accessed: 5 September 2025].

COPDESS 2018 'Commitment Statement in the Earth, Space, and Environmental Sciences', [online] Coalition for Publishing Data in the Earth and Space Sciences (COPDESS). <https://copdess.org/enabling-fair-data-project/commitment-statement-in-the-earth-space-and-environmental-sciences/> [Last accessed: 5 September 2025].

DataCite Metadata Working Group 2024 'DataCite Metadata Schema for the Publication and Citation of Research Data and Other Research Outputs', Version 4.6, [online] DataCite e.V. <https://doi.org/10.14454/mzv1-5b55> [Last accessed: 5 September 2025].

DataCite e.V 2022 'IGSN IDs', [online] DataCite e.V. <https://support.datacite.org/docs/igsn-ids> [Last accessed: 5 September 2025].

DataCite e.V 2025a 'DOI Basics', [online] DataCite e.V. <https://support.datacite.org/docs/doi-basics> [Last accessed: 5 September 2025].

DataCite e.V 2025b 'Harmonizing DataCite Schema Metadata and Disciplinary Sample Metadata', [online] DataCite e.V. <https://support.datacite.org/docs/harmonizing-datacite-schema-metadata-and-disciplinary-sample-metadata> [Last accessed: 5 September 2025].

DCMI 2025 'DublinCore: DCMI Metadata Terms', [online] Dublin Core Metadata Initiative (DCMI). <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/> [Last accessed: 5 September 2025].

Deutsche Nationalbibliothek 2019 'URN Service', [online] Deutsche Nationalbibliothek. https://www.dnb.de/EN/Professionell/Services/URN-Service/urn-service_node.html [Last accessed: 5 September 2025].



Forum on Information Standards in Heritage 2025 *FISH Terminologies*, [online] Forum on Information Standards in Heritage. <https://heritage-standards.org.uk/fish-vocabularies/> [Last accessed: 5 September 2025].

Hsu, Y.-K., Jan, S., Löffler, I. and Stöllner, T 2024 *NFDI4Objects Temporary Working Group: Metadata Profiles for Analytical Methods in Archaeometry*, Zenodo. <https://doi.org/10.5281/zenodo.10990411>

IGSN e.V 2016 *IGSN Descriptive Metadata Version 1.0* [online] IGSN e.V. <https://github.com/IGSN/metadata/wiki/IGSN-Descriptive-Metadata-Version-1.0> [Last accessed: 5 September 2025].

IGSN e.V 2024 'Archaeology', [online] IGSN e.V. <https://ev.igsn.org/communities/archaeology> [Last accessed: 5 September 2025].

IGSN–DataCite Crosswalk and Metadata Management Working Group 2022 'IGSN ID Metadata Recommendations', [online] DataCite e.V. <https://support.datacite.org/docs/igsn-id-metadata-recommendations> [Last accessed: 5 September 2025].

Jones, P., Brown, J., Meadows, A. and Murphy, F 2024 'A cost-benefit analysis for PID implementation in Czechia', [online] National Library of Technology. <https://doi.org/10.48813/x4vd-yj13>

Klein, S., Rose, T., Westner, K.J. and Hsu, Y.-K. 2022 'From OXALID to GlobalID: Introducing a modern and FAIR lead isotope database with an interactive application', *Archaeometry* **64**(4), 935–950. <https://doi.org/10.1111/arcm.12762>

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**, 33. <https://doi.org/10.5334/dsj-2021-033>

National Open Research Forum and MoreBrains Cooperative 2024 'Efficiency and Insight: A Cost-Benefit Analysis for a Central Service to Support Persistent Identifier Implementation in Ireland', Text [Type], Digital Repository of Ireland [Publisher] [Depositor]. <https://doi.org/10.7486/DR1.nz80kt123>

Niven, K. and Watts, J. 2011 *Guides to Good Practice: Project Metadata*, [online] Archaeology Data Service/Digital Antiquity. <https://archaeologydataservice.ac.uk/help-guidance/guides-to-good-practice/the-project-lifecycle/project-metadata/> [Last accessed: 5 September 2025].

PID4NFDI 2025 'PID4NFDI Cookbook: Getting Started with PIDs', [online] PID4NFDI. <https://pid4nfdi-training.readthedocs.io/en/latest/> [Last accessed: 5 September 2025].

Schmidt, S.C., Thiery, F. and Trognitz, M 2022 'Practices of linked open data in archaeology and their realisation in Wikidata', *Digital* **2**(3), 333–364. <https://doi.org/10.3390/digital2030019>

UNESCO 2022 *Understanding Open Science. Online Open Access programme and meeting document*, UNESCO. <https://doi.org/10.54677/UTCD9302>



Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hoof, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J. and Mons, B 2016 'The FAIR Guiding Principles for scientific data management and stewardship', *Scientific Data* **3**(1), 160018. <https://doi.org/10.1038/sdata.2016.18>

Online Resources

ARK Alliance. <https://arks.org/> [Last accessed: 5 September 2025].

Handle.Net Registry, Corporation for National Research Initiatives. <https://www.handle.net/> [Last accessed: 5 September 2025].

Crossref <https://www.crossref.org/> [Last accessed: 5 September 2025].

DataCite e.V., Connecting Research, Advancing Knowledge <https://datacite.org/> [Last accessed: 5 September 2025].

DOI Foundation <https://www.doi.org/> [Last accessed: 5 September 2025].

Getty Research Institute, Art and Architecture Thesaurus® Online <https://www.getty.edu/research/tools/vocabularies/aat/> [Last accessed: 5 September 2025].

IGSN e.V <https://ev.igsn.org/> [Last accessed: 5 September 2025].

Internet Archive, PURL Administration <https://purl.archive.org/> [Last accessed: 5 September 2025].

Morebrains Consulting <https://www.morebrains.coop/> [Last accessed: 5 September 2025].

ORCID: Connecting Research and Researchers, ORCID, Inc. <https://orcid.org/> [Last accessed: 5 September 2025].

Research Organization Registry (ROR) <https://ror.org/> [Last accessed: 5 September 2025].