PRACTICE PAPER

Utilizing the International Geo Sample Number Concept in Continental Scientific Drilling During ICDP Expedition COSC-1

Ronald Conze[1], Henning Lorenz[2], Damian Ulbricht[1], Kirsten Elger[1] and Thomas Gorgas[1]

[1] GFZ German Research Centre for Geosciences, Potsdam, Germany

[2] Uppsala University, Department of Earth Sciences, Uppsala, Sweden

Corresponding author: Ronald Conze (ronald.conze@gfz-potsdam.de)

The International Geo Sample Number (IGSN) is a globally unique persistent identifier (PID) for physical samples that provides discovery functionality of digital sample descriptions via the internet. In this article we describe the implementation of a registration service for IGSNs of the Helmholtz Centre Potsdam – GFZ German Research Centre for Geosciences. This includes the adaption of the metadata schema developed within the context of the System for Earth Sample Registration (SESAR[1]) to better describe the complex sample hierarchy of drilling cores, core sections and samples of scientific drilling projects.

Our case study is the COSC-1 expedition[2] (Collisional Orogeny in the Scandinavian Caledonides) supported by the International Continental Scientific Drilling Program[3] (ICDP). COSC-1 prompted for the first time in ICDP's history to assign and register IGSNs during an on-going drilling campaign preserving the original parent-child relationship of the sample objects.

IGSN-associated data and metadata are distributed and shared with the world wide community through novel web portals, one of which is currently evolving as part of ICDP's collaborative efforts within the GFZ Potsdam and researchers from ICDP's COSC clientele. Thus, COSC-1 can be considered as a 'Prime-Example' for ICDP projects to further improve the quality of scientific research output through a transparent process of producing and managing large quantities of data as they are normally acquired during a typical scientific drilling operation. The IGSN is an important new player in the general publication landscape that can be cited in scholarly literature and also cross-referenced in DOI-bearing scholarly and data publications.

Keywords: sample curation; persistent identifiers; IGSN, geosciences; infrastructure; open data; metadata

Introduction

IGSN – The International Geo Sample Number

The International Geo Sample Number (IGSN) is a globally unique and persistent identifier (PID) for physical samples that reduces (and perhaps even eliminates) problems associated with ambiguous naming of samples. It has been developed to (1) address requirements for reproducibility of sample-based data, (2) ensure discovery, access, and re-usability of samples and data derived from them, (3) recognize sample collection and curation as scholarly contribution to the scientific community, and (4) improve data integration. The

---

[1] SESAR: http://www.geosamples.org/igsnabout.

[2] The ICDP expedition COSC (Collisional Orogeny in the Scandinavian Caledonides): http://cosc.icdp-online.org.

[3] International Continental Scientific Drilling Program (ICDP): http://www.icdp-online.org.

latter especially pertains to scientific drilling projects where samples and subsamples of the same core or core section are analysed in different laboratories and over long periods of time.

IGSN is governed by an international non-profit organisation (IGSN e.V.[4]), which operates the central registration system based on the Handle.Net System (CNRI 2010). The IGSN e.V. aims to develop standard methods for locating, identifying and citing physical samples (Devaraju et al. 2016). Similar to the digital object identifier (DOI), IGSNs resolve to a persistent link on the internet to IGSN landing pages with a virtual sample description, managed by federated IGSN Allocating Agents (e.g. IGSN:ICDP5054EHW1001). The largest collection of registered IGSNs is accessible via the inventor's web portal 'System for Earth Sample Registration', SESAR, at Lamont-Doherty Earth Observatory (Lehnert and Klump 2008). Furthermore, the IGSN provides means for sample citation in the literature and for establishing direct links from specimen to research results and interpretations.

Each member of the IGSN e.V. may become IGSN Allocating Agent and develop an IGSN registration service for their communities. In addition, allocating agents are free to develop individual metadata schemata for the various methodical disciplines using their service as well as to design the IGSN identifier. In addition to the disciplinary metadata schemata, each allocating agent has to provide metadata in the IGSN Description Schema (http://schema.igsn.org/description/). The IGSN Description Schema contains persistent information about registered samples, such as temporal and spatial coordinates of sample acquisition, metadata about involved institutions and sample-requesting scientists, information about the sample material, collection methods, and alternate or related identifiers. This metadata kernel is aligned with the DataCite Metadata Schema 4.0 and will be harvested by the central IGSN metadata catalogue that is currently in development. Where possible, the descriptive metadata schema makes use of vocabulary lists, e.g. the Observations Data Model 2 (ODM2: http://www.odm2.org) and a list of collection methods.

Drilling Samples and the Drilling Information System (DIS)

The necessity for a meticulous acquisition and documentation of scientific drilling project data and associated samples in a structured and hierarchical way was already described for the German Continental Deep Drilling Program (KTB) (Wächter et al. 1989; Conze et al. 1993; Conze 1995). As one outcome and consequence of the KTB project, the development of a dedicated IT system, the DIS, was initiated. The DIS (Conze et al. 2007, Conze 2016) is based on a relational database with data verification routines and desktop input forms. It is designed for the full documentation of the on-site drilling operations, including the sample material and the acquired primary data. Scientific drilling projects generate a huge amount of sample material. Most prominent are drill cores, where also a number of hierarchical relationships have to be taken care of: the origin of all material of drilling projects are the drill holes, the core runs (a drill core reaching the earth surface), core sections (core runs sub-divided into segments of manageable length), and samples for further analyses. Primary data include core images, multi-sensor core logging data, borehole logging data, lithological descriptions, and so forth.

The data management system **ExpeditionDIS** is dedicated to individual projects and is used in field expeditions, during drilling operations, and for post-drilling examinations of core and sample material in project-associated laboratories. The broad range of different research topics of each ICDP project call for a DIS, which is tailored for each scientific drilling expedition (such as COSC-1) for the inventory of recovered sample material and core samples extracted for research from the drill cores.

In contrast, the **CurationDIS** is used in storage facilities (i.e., core repositories, such as MARUM in Bremen, or the 'Nationales Bohrkernlager für kontinentale Forschungsbohrungen' of the Federal Institute for Geosciences and Natural Resources (BGR) in Berlin-Spandau, Germany[5]). There, sample material from many different expeditions is collected, managed, shared with and distributed to the scientific community. The lifetime of an ExpeditionDIS is limited to the period of the expedition, whereas the CurationDIS is lasting for the operational lifetime of the sample material storage.

Both DIS systems received an update recently that automatically creates an IGSN identifier from the parent-child relationships of the sample material, a prefix individual to the project, and the sample type stored in the DIS relational database. We describe the syntax of the ICDP IGSNs later in this article.

---

[4] IGSN e.V.: http://schema.igsn.org.
[5] 'Nationales Bohrkernlager für kontinentale Forschungsbohrungen' in Berlin-Spandau, Germany: http://www.bohrkernlager.de.

Conze et al: Utilizing the International Geo Sample Number Concept in
Continental Scientific Drilling During ICDP Expedition COSC-1

Art. 2, page 3 of 8

The COSC-1 Project

The COSC-1 project studies mountain building processes by drilling a continuous cored section through the thrust sheets of the Caledonian foreland in Sweden (Lorenz et al. 2015). The Caledonides are an approximately 400 my old mountain belt that originally had Himalayan dimensions.

Research in scientific drilling projects is complex and not limited to the primary goal, which often results in intricate and comprehensive science and sampling programmes. During COSC-1 operations, approximately 2.4 km of drill core were retrieved (the 'Sample Material'), drill mud and mud gases sampled, and diverse screening techniques for microbiology employed. During the project field campaign and subsequent sampling party, inventory-keeping of all drill cores was done routinely using the ExpeditionDIS. Primary core metadata, core scan image files and on-site analytical data (Lorenz et al. 2015) were entered at several DIS client stations immediately upon its retrieval on the drill deck. Obtained core pieces, which are unanimously and uniquely identified to belonging together were treated as a single object and logged as individual 'Core Run' into the DIS. The core run is a 'Child' object of the borehole 'Parent', and does not carry any additional information. IGSNs were assigned immediately to all eligible objects. Off-line IGSN assignment worked flawlessly and without interfering with the researchers' workflow. Analytical data, such as geophysical logs and XRF geochemical analyses, were linked to the respective core sections using adapted data pumps and imported into the DIS. Backups of the database were taken off-site and transferred to ICDP via secure file transfer protocols (sftp).

After completing the field campaign, all sample material was shipped to the core repository in Berlin-Spandau. The project data were subsequently transferred to the BGR CurationDIS. Sampling can continue based on the sampling policies of the project and the storage facility itself. In the case of COSC-1, researchers with accepted and DIS-generated sample requests met for a sampling party to visually inspect the core and mark/list their sampling spots and intervals. Samples were documented in the DIS and so IGSNs automatically assigned, and subsequently physically taken by the BGR curator. The properly DIS-labelled samples and sample lists were then shipped world-wide to the respective researchers. This workflow approach turned out to be successful and efficient as the scientists' time was entirely spent on science, while sampling was performed by the assigned curator in an orderly manner. Thereby also the proper documentation of all sample material and sampling procedures was generated.

IGSN Implementation, Generation and Registration

Inherent IGSN Properties

The most conspicuous requirement for a persistent identifier (PID) in a drilling project context is that it reflects the hierarchy in the sample material. Without satisfying this pre-requirement, the provenance of a referenced piece of drill core could only be tracked by a user who has direct access to the data logged in the DIS.

When persistent identifiers are used, it is quite common to work with prefixes. This allows the separation of responsible domains by different namespaces. For example, the IGSN e.V. assigned the namespace 'ICDP' to ICDP and 'BGRB' for the core archive at BGR. The namespace is followed by an 'Expedition ID' and a report prefix (see **Table 1**). Both create several independent sub-namespaces that could be used in individual DIS systems to do data acquisition independent from each other on remote sites without internet access. An object tag allows for a quick, human-readable identification of the object in question, such as 'Hole', 'Core Run', 'Core Section' or 'Sample'. The IGSN ends with a coded pattern directly derived from primary key values generated by the DIS, and thereby guarantees the uniqueness of the IGSN identifier. **Figure 1** illustrates the IGSN syntax for ICDP and shows links to the ICDP naming convention.

Name Space	Expedition ID	Report Prefix	Object Tag	Coded Pattern
ICDP	5054	E = Expedition R = Repository	H = hole C = core run S = core section X = sample	W1001

Table 1: Structure of the extended IGSN for ICDP sample material. The coded pattern is directly derived from the internal object-ID in the DIS, and therefore guarantees uniqueness of the sample.
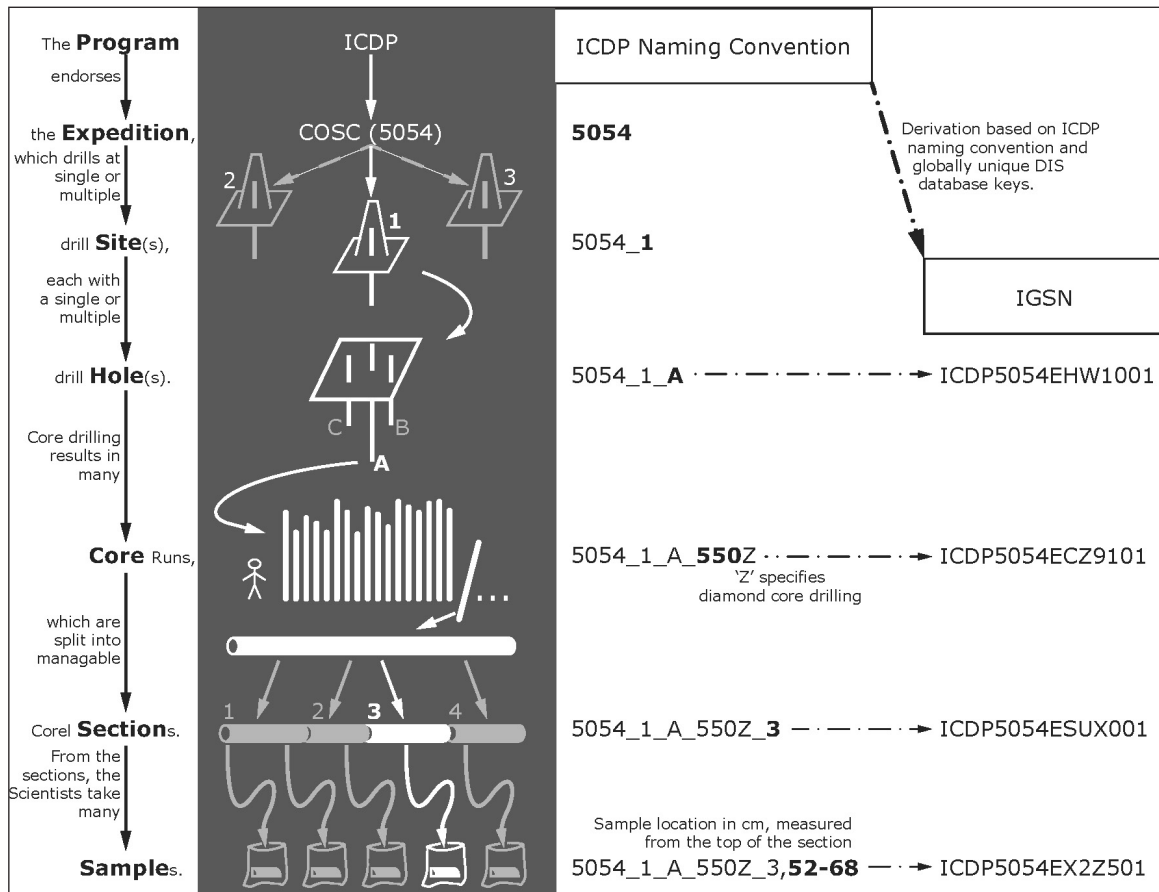
Figure 1: Mapping of sample objects to IGSNs. Example of the data structure and IGSN assignment for drill holes, core runs, core sections and core samples by the ICDP DIS as assigned during the COSC-1 project. The **Expedition** ID (here '5054' for COSC) is defined as unique key value in the ICDP naming convention and is accompanied by a character string that includes indicators for drill **Site** and **Hole** (several boreholes per site and several sites per expedition can exist). Each lower level of the sample hierarchy (**Core** runs, **Sections** and **Samples**) are symbolized by additional characters or numbers to the name of the higher level. The **Hole** is the top level for the IGSN. Each derived object (core runs, core sections, samples, etc.) is related to the borehole through the parent-child relation of each subordinated object/sample.

The ICDP-IGSN Metadata Schema

As discussed before, the IGSN allocating agents have to provide metadata in the IGSN Description Schema. In addition, IGSN-allocating agents typically provide services to their respective topical and geographical communities and often must address the specific and variable needs of these communities. The allocating agent GFZ Potsdam decided to develop a new IGSN metadata schema for the description of scientific drilling projects in the framework of ICDP (ICDP-IGSN Schema). This community-specific schema is loosely based on the original universal SESAR metadata schema, enriched by specific metadata fields representing the drilling process and instruments, analysis methods, extended descriptive elements for the geological and lithological descriptions. In addition, individual metadata fields for the corrected depths in boreholes exist (e.g., IGSN:ICDP5054EXF4601). Furthermore, we put all information necessary for disseminating the IGSN Description Schema into our ICDP-IGSN Schema. The metadata are exported directly from the DIS into XML, and thus prepared for the IGSN registration. This way, we store IGSN metadata in XML format at GFZ Potsdam, which allows retrieving metadata via an Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH) interface.

Conze et al: Utilizing the International Geo Sample Number Concept in
Continental Scientific Drilling During ICDP Expedition COSC-1

Art. 2, page 5 of 8

IGSN Registration

The main technical tasks of IGSN allocating agents are to allow registration of IGSNs for physical samples, guarantee the uniqueness of IGSNs by issuing namespaces to clients, and collect and disseminate IGSN metadata (**Figure 2**). These technical tasks are similar to the tasks which DataCite solves with their corresponding DOI registration infrastructure. As we already modified the DataCite registration software in the past for our data publication activities at GFZ Potsdam (Ulbricht et al., 2016), we used our experiences and modified the program sources of the DataCite Metadata Store to be used as GFZ IGSN registry.

In **Figure 2** we outline how the GFZ IGSN registry fits into the federated structure of the IGSN e.V. and its allocating agents. While an internet browser interface to the registry software exists, metadata and URLs to IGSN landing pages (**Figure 3**) are expected to be registered by machines through web-service APIs, which are feeding corresponding data through existing databases into the web portal. The GFZ IGSN registry is designed as a proxy to the global IGSN registry that mints handles which resolve to landing pages.

Since COSC-1 has produced over 4460 IGSNs so far, it made full use of the programming interface of the GFZ IGSN registry software. The registration of URLs via web-portal landing pages and metadata was accomplished with a set of shell and python scripts. Since the IGSN Description Schema makes use of vocabulary lists, this information had to be adapted to the ICDP-IGSN schema. In particular, we had to ensure to include information about the appropriate ODM2 term for materials, specimen and feature type and the correct term for collection methods, which originates from SESAR. However, the material (rock), the resource type (hole, core, core section, core sample), and the collection method (rock corer) could be easily integrated into a conversion stylesheet.

The ICDP-IGSN schema for samples is designed as superset of the IGSN description schema that has to be available for data dissemination through the OAI-PMH. Furthermore, the ICDP-IGSN schema is used to generate the web portal landing pages through the *Extensible Stylesheet Transformation* (XSLT).
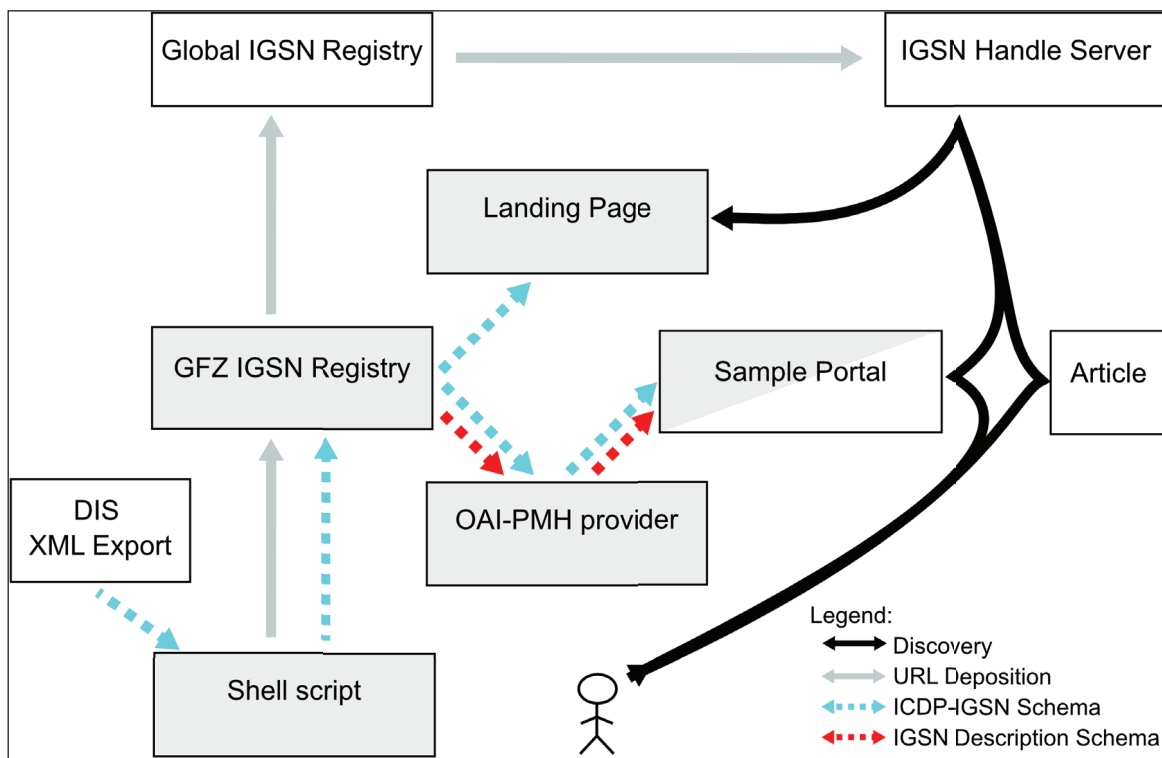

Figure 2: Visualisation of the registration infrastructure. The associated XML metadata flows (dotted lines and arrows) at the GFZ IGSN Allocating Agent during registration of ICDP samples are shown. Building blocks of the IGSN allocating agent are highlighted in grey boxes. In additon, the steps of an exemplary sample discovery through a web portal or a research article are shown as black arrows.

Art. 2, page 6 of 8

Conze et al: Utilizing the International Geo Sample Number Concept in
Continental Scientific Drilling During ICDP Expedition COSC-1

Figure 3: Example of an IGSN Landing Page for a Core Sample of the COSC-1 project (IGSN: ICDP5054EXF4601). The left part contains the full sample description thematically grouped into General Identifiers (for the drilling project and the IGSN hierarchy), Sampling Location, Geology, and Methods used to produce primary borehole data, Drilling (details on the drilling method, instrumentation, PIs, and drilling dates) as well as location of the sample (Repositories). The top right box allows to browse through the sample hierarchy. Different icons indicate the type of sample (Hole, Core, Core Section, Core Sample). The map below shows the geographical location, whereas the lower right part highlights publications that are related with the sample (here the initial scientific publication in Scientific Drilling Journal[6] and the Operational Report of COSC-1, Lorenz et al. 2015a and 2015b, and the master thesis and the data publication of Hierold 2016 and Hierold et al. 2016).

---

[6] Scientific Drilling Journal: http://www.scientific-drilling.net/.

Conze et al: Utilizing the International Geo Sample Number Concept in
Continental Scientific Drilling During ICDP Expedition COSC-1

Art. 2, page 7 of 8

IGSN Landing Pages

To provide easy web-access to information about IGSN registered scientific drilling sample material and samples, a specific landing page (**Figure 3**) was developed for the COSC-1 drill hole A (5054_1_A), whose top- level IGSN 'ICDP5054EHW1001' can be resolved using the URL http://hdl.handle.net/10273/ICDP-5054EHW1001. Landing pages comprise the complete descriptive data of the IGSN-registered sample material and allow the navigation through the hierarchical 'Parent-Child' data structure. Additionally, related publications and data sets are listed as DOI-referenced sources on this landing page and in the metadata. As soon as an IGSN is registered by the allocating agent, its sample description (metadata) is made available to the public domain via this web portal system.

Discussion und Outlook

The development of IGSN is an important tool to increase the visibility and access of physical samples. It is complementary to other text and data publication formats, including journal articles, data publications, reports, etc. For a full overview it is essential to carefully cross-reference all participating publications via the metadata. DataCite metadata is offering a broad range of 'relation types' in their 'related identifier' package, not only to tie different publications to a dataset or a report, but also to classify the related material in (dataset) documentation, supplement to a journal article and material for further reading. The relevance of IGSN as PID for physical samples is also reflected in the newly released DataCite Metadata Schema 4.0, where IGSN is added as new 'relation type' option (DataCite 2016). In addition, it is already possible to cite IGSNs (including the link to the landing pages) in articles of certain journals (e.g. Lloyd et al. 2014; this paper).

For COSC we put efforts in presenting the whole sample family. However, it is difficult to control what happens after a sample leaves the core repository. Scientists can generate their own samples and hand over subsamples to colleagues, which are then out of reach for the core repository manager. Hence such subsamples are impossible to integrate in the CurationDIS. This may often set the lower limit of the IGSN hierarchy for drilling projects.

The content and provision of descriptive data of drilling projects is highly relevant beyond the reaches of continental scientific drilling, although it is not directly coupled to the IGSN. Similar problems persist in the ocean research drilling community and at any other organisation that holds relevant information about boreholes and drill cores that should be made accessible to the entire scientific community. A relevant example is the on-going work within the EU Horizon 2020 project European Plate Observing System (EPOS)[7] EPOS is a multi-disciplinary e-Infrastructure for Solid Earth Science in Europe. It integrates the distributed national Research Infrastructures (RIs) by harmonising existing service and component interfaces, which are co-developed by IT specialists and the geoscientific communities. Since no disciplinary metadata standards exist for the drilling community, available practices, such as the here developed metadata schema for the IGSN registration of a scientific borehole, can be considered as first steps towards these.

Conclusion

This article summarizes the state-of-the art sample and data curation in a successful drilling project, exemplified by COSC-1 and co-sponsored by ICDP in consortium with multiple additional academic and industry partner institutions and agencies. COSC-1 provides a superb 'Blueprint' for future ICDP projects, i.e., how to plan, organize, conduct and finalize a typical ICDP drilling project from start to finish. COSC-1 has demonstrated the feasibility and applicability of modern geoscientific technologies by merging old and well-proven techniques and methods (e.g., core image scanning as part of acquiring 'primary data') with novel developments in database management, data publication and dissemination. This article highlights the ICDP's DIS in conjunction with the rapidly evolving IGSN system and associated web-portals for providing open-access of drilling-generated data and metadata.

Acknowledgements

---

[7] European Plate Observing System (EPOS): https://www.epos-ip.org/.
[8] smartcube GmbH, Berlin, Germany: http://www.smartcube.de.

Art. 2, page 8 of 8

Conze et al: Utilizing the International Geo Sample Number Concept in
Continental Scientific Drilling During ICDP Expedition COSC-1

Competing Interests

The authors have no competing interests to declare.

References

CNRI 2010 Technical Manual, Handle.Net Version 8.1. Corporation for National Research Initiatives CNRI, December. Retrieved from: http://handle.net/tech_manual/HandleTool_UserManual.pdf.

Conze, R 1995 Electronic Data Processing at KTB. KTB Report 95-2, pp. F1–F13.

Conze, R 2016 Drilling Information System DIS and Core Scanner. JLSRF, 2: A63. DOI: https://doi.org/10.17815/jlsrf-2-130

Conze, R, Häner, R and **Yazici, A** 1993 Part C: Technical aspects and future developments-The KTB Information System. KTB Report 93-2, pp. 535–541.

DataCite Metadata Working Group 2016 DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. Version 4.0, DOI: https://doi.org/10.5438/0012

Devaraju, A, Klump, J, Cox, S J D and **Golodoniuc, P** 2016 Representing and publishing physical sample descriptions. *Computers&Geosc.*, 96: 1–10, DOI: https://doi.org/10.1016/j.cageo.2016.07.018

Hierold, J 2016 Analysis of element behavior in mylonites of the Seve Nappe of the Scandinavian Caledonides using different core scanning methods. Master Thesis, Scientific Technical Report STR; 16/07. DOI: https://doi.org/10.2312/GFZ.b103-16070

Hierold, J, Körting, F, Kollaske, T, Rogass, C and **Harms, U** 2016 Analysis of element behavior in mylonites of the Seve Nappe of the Scandinavian Caledonides using different core scanning methods (Datasets). *GFZ Data Services*. DOI: http://doi.org/10.5880/ICDP.5054.001

Lehnert, K and **Klump, J** 2008 Facilitating Research in Mantle Petrology with Geoinformatics. 9th International Kimberlite. Conference Extended Abstract No. 9IKCA00250.

Lloyd, A S, et al. 2014 NanoSIMS results from olivine-hosted melt embayment: Magma ascent rate during explosive basaltic eruptions. *Journal of Volcanology and Geothermal Research*, 283: 1–18. DOI: https://doi.org/10.1016/j.jvolgeores.2014.06.002

Lorenz, H, et al. 2015a COSC-1 operational report – Scientific data sets, DOI: https://doi.org/10.1594/GFZ.SDDB.ICDP.5054.2015

Lorenz, H, et al. 2015b COSC-1 – drilling of a subduction-related allochthon in the Palaeozoic Caledonide orogen of Scandinavia. *Scientific Drilling*, 19: 1–11, DOI: https://doi.org/10.5194/sd-19-1-2015

Ulbricht, D, Elger, K, Bertelmann, R and **Klump, J** 2016 PanMetaDocs, eSciDoc and DOIDB – An Infrastructure for the Curation and Publication of File-Based Datasets for GFZ Data Services. *International Journal of Geoinformatics*, 53: 25, DOI: https://doi.org/10.3390/ijgi5030025

Wächter, J and **Friese-Haug, M** 1989 I. KTB Oberpfalz VB – Datenverarbeitung. KTBase KTB database – der Kern eines wissenschaftlich/technischen Informationssystems: Hardware-Konfiguration und Struktur der Datenbank. KTB Report 89-2, pp. I6–I20.