Geophysical Research Abstracts
Vol. 18, EGU2016-10798, 2016
EGU General Assembly 2016

The IGSN Experience: Successes and Challenges of Implementing Persistent Identifiers for Samples

Kerstin Lehnert and Robert Arko
Columbia University, Lamont-Doherty Earth Observatory, Palisades, United States (lehnert@ldeo.columbia.edu)

Physical samples collected and studied in the Earth sciences represent both a research resource and a research product in the Earth Sciences. As such they need to be properly managed, curated, documented, and cited to ensure re-usability and utility for future science, reproducibility of the data generated by their study, and credit for funding agencies and researchers who invested substantial resources and intellectual effort into their collection and curation. Use of persistent and unique identifiers and deposition of metadata in a persistent registry are therefore as important for physical samples as they are for digital data. The International Geo Sample Number (IGSN) is a persistent, globally unique identifier. Its adoption by individual investigators, repository curators, publishers, and data managers is rapidly growing world-wide. This presentation will provide an analysis of the development and implementation path of the IGSN and relevant insights and experiences gained along its way.

Development of the IGSN started in 2004 as part of a US NSF-funded project to establish a registry for sample metadata, the System for Earth Sample Registration (SESAR). The initial system provided a centralized solution for users to submit information about their samples and obtain IGSNs and bar codes. Challenges encountered during this initial phase related to defining the scope of the registry, granularity of registered objects, responsibilities of relevant actors, and workflows, and designing the registry's metadata schema, its user interfaces, and the identifier itself, including its syntax. The most challenging task though was to make the IGSN an integral part of personal and institutional sample management, digital management of sample-based data, and data publication on a global scale. Besides convincing individual researchers, curators, editors and publishers, as well as data managers in US and non-US academia, state and federal agencies, the PIs of the SESAR project needed to identify ways to organize, operate, and govern the global registry in the short and in the long-term.

A major breakthrough was achieved at an international workshop in February 2011, at which participants designed a new distributed and scalable architecture for the IGSN with international governance by a membership organization modeled after the DataCite consortium. The founding of the international governing body and implementation organization for the IGSN, the IGSN e.V., took place at the AGU Fall Meeting 2011.

Recent progress came at a workshop in September 2015, where stakeholders from both geoscience and life science disciplines drafted a standard IGSN metadata schema for describing samples, to complement the existing schema for registering samples. Consensus was achieved on an essential set of properties to describe a sample's origin and classification, creating a "birth certificate" for the sample. Further consensus was achieved in clarifying that an IGSN may represent exactly one physical sample, sampling feature, or collection of samples; and in aligning the IGSN schema with the existing Observations Data Model (ODM-2). The resulting schema was published online at schema.igsn.org and presented at the AGU Fall Meeting 2015.