



Use-cases and services for unique identifiers in the geological sciences

Jens Klump (1) and Kerstin Lehnert (2)

(1) German Research Centre for Geosciences, CeGIT, Potsdam, Germany (jens.klump@gfz-potsdam.de, 49 331 2881703), (2) Lamont-Doherty Earth Observatory, Columbia University, New York, NY, USA

One of the main drivers towards the inception of the world wide web was the ability to link scientific sources over the internet. The Uniform Resource Locator (URL) used to locate resources on the web soon turned out to be ephemeral in nature. A more reliable way of addressing objects was needed, a way of persistent identification to make digital objects, or digital representations of objects, part of the record of science. The challenge here is not to design a system of persistent identifiers for a particular resource, which could be journal articles, authors, geological specimens, or instruments, the real challenge is the long-term operation and governance of such a system.

With their high degree of centralization, the scientific publishing houses were able to quickly implement and adopt a system for unique and persistent identification, the Digital Object Identifier (DOI) [®]. At the same time, other identifier systems exist alongside DOI, e.g. URN, ARK, handle [®], and others. While the DOI system was initially implemented exclusively for journal articles, its application has been extended to the identification of data, thus making data citable. Persistent identification is needed in science beyond the identification of journal articles and data. Other unique identification systems have been developed for samples (IGSN), authors (e.g. ORCID, VIAF, ResearcherID), and institutions to unambiguously locate and link resources.

Persistent identifiers are now available for literature, data, samples, and authors. More applications, e.g. identification of methods or instruments, scientific cruises and expeditions, will follow. In conjunction with semantic web technology, the application of unique and persistent identifiers in the geosciences will aid discovery both through systematic data mining, exploratory data analysis, and serendipity effects.

An example for advanced discovery of resources through unique identifiers is the application of persistent identifiers for geological samples and their linking to DOIs for data and journal articles. As many data in the geosciences are derived from samples, it is crucial to be able to uniquely identify the samples from which a set of data were derived. Incomplete documentation of samples in publications, use of ambiguous sample names have been major obstacles for synthesis studies and re-use of data. Access to samples for re-analysis and re-appraisal has been limited due to the lack of a central catalogue that allows finding a sample's archiving location. The International Geo Sample Number (IGSN) now provides solutions to the questions of unique sample identification and discovery. Use of the IGSN in digital data systems allows building linkages between the digital representation of samples in sample registries, e.g. SESAR, and their related data in the literature and in web accessible digital data repositories.

This poster will discuss existing and emerging applications for persistent identifiers in the geological sciences.