Geophysical Research Abstracts
Vol. 13, EGU2011-13113, 2011
EGU General Assembly 2011

Metadata Standards for Sample-Based Observations

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

Observational data in the Earth Sciences are to a substantial part acquired by description and analysis of physical samples collected from the natural environment such as rock or mineral specimens, sediment cores, soil pedons, and water or gas samples. While metadata requirements and data models for sample-based observations widely overlap with those for sensor-based observations as defined by the OGC Observations and Measurements Encoding Standard (Cox, 2007; http://www.opengeospatial.org/standards/om#schemas), there are additional requirements for sample-based data that stem from the fact that (a) a sample can be analyzed for a wide range of properties in multiple locations (labs) and at different times, (b) a sample can be re-analyzed for the same property after long periods of time, and (c) a sample can be split or sub-sampled to isolate different components of a sample such as mineral phases of a rock, fossil species in a sediment, or suspended particles in a fluid. Metadata schemas for sample-based data also need to ensure that a link to the physical sample (sample archive) is kept.

This presentation will focus on data models and metadata standards for sample-based data that have been developed within the field of geochemistry for data preservation and data exchange. Geochemical data models and data encoding schemas such as the EarthChemXML continue to evolve and improve based on a broadening range of applications and community input. For example, a new effort underway is to align the geochemical data models with the Observation Data Model (ODM) of the Hydrological Information System (e.g. Horsburgh 2008: doi:10.1029/2007WR006392) to allow integration of sample-based data and sensor-based time-series data.

An important aspect of the sample-based data models presented here is the use of a global unique identifier for samples, the International Geo Sample Number (IGSN), which is managed by the System for Earth Sample Registration SESAR (www.geosamples.org). The IGSN facilitates fine-grained access to sample-based data, interoperability between sample-based data systems, and links between data, metadata, and publications, overcoming the problem of ambiguous naming of samples. The metadata profile of the IGSN includes geospatial and descriptive information about a sample as well as information about its physical location (sample archive) and the collection event.