



Enabling discoveries in earth sciences through linking literature and data

J. Klump (1) and R. Huber (2)

(1) GeoForschungsZentrum Potsdam, (2) MARUM, Universität Bremen
(jens.klump@gfz-potsdam.de/Fax +49-331-2881703)

Intensive research in the earth sciences over the past decades has created a tremendous wealth of literature, data, and material collections. So far, literature, data and sample collections have been separated. Information technology and the internet, in particular the new cyberinfrastructures for the earth sciences, offer ways to interlink literature, data and samples, creating the potential for new interpretations of the data and materials beyond the interpretation already published in the literature. To achieve this, technical, editorial and custodial issues need to be resolved.

A key technical issue is the use of persistent identifiers for literature, data and sample collection objects. Past experience has shown that URLs are transient, but systems of persistent identifiers (e.g. DOI, URN) already exist and can be used to reference these objects. Far more challenging is the task ahead of interlinking literature, data and samples with as little editorial work as possible. Keeping the amount of work small is essential to allow the indexing of the existing back catalogue of already existing works.

A key technology to solve this task is the automatic creation of ontologies. Systems such as Google Scholar are able to create ontologies automatically, and also to import ontologies and attach these to other ontologies, thus expanding the network of interlinking references. Ideally, the literature should already reference the materials used and the data derived from these. Since this is not yet done, repositories publishing data and tracking sample material record the literature based on these data and samples in their databases. These entries can then be used to create ontologies, which can then be connected to the ontologies generated from reference lists.

This presentation will look at existing systems for data publication

(STD-DOI project, <http://www.std-doi.de>), sample identification (SESAR project, <http://www.geosamples.org>) and for the management of interconnected literature, data publications and sample collections (TaxonConcept, <http://taxonconcept.stratigraphy.net>), and how these systems can be used to enable new discoveries in the earth sciences.