Facilitating Research in Mantle Petrology with Geoinformatics

Kerstin Lehnert (1) and Jens Klump (2)

(1) Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA, (2) German Centre for Geosciences (GFZ) Potsdam, Potsdam, Germany.

Research in mantle petrology has created a tremendous wealth of data, material collections, and literature, and can be expected to produce even more in the future. Existing and emerging capabilities of Geoinformatics -in the sense of cyberinfrastructure for the Geosciences- can maximize the utility and impact of these resources to the benefit of science and education. Geoinformatics resources include (a) information systems that provide fast and easy access to data, (b) linkages between literature, data and samples that create the potential for new interpretations of the data and materials beyond interpretations already published, and (c) data analysis, visualization, and modeling tools that seamlessly integrate with the data collections.

The EarthChem Deep Lithosphere Dataset (www.earthchem.org) is an example for an information system that provides access to a global collection of petrological and geochemical data of mantle xenoliths. Users can select data based on criteria such as sample name, location, rock type, and author, and retrieve customized datasets of major element, trace element, and isotope ratio data that integrate measurements from any number of publications.

Digital data such as the ones in EarthChem refer to physical samples. Fundamental data sets have been generated from invaluable collections of mantle xenoliths such as the collections at the Smithsonian Institution or at the University of Cape Town. The application and long-term utility of sample-based data is critically dependent on (a) availability of information (metadata) about the samples such as geographical location and time of sampling, (b) links to other data sets derived from individual samples that are dispersed in the literature and in digital data repositories, and (c) access to the samples themselves. Major problems for achieving this include incomplete documentation of samples in publications, use of ambiguous sample names, and the lack of a central catalog that allows finding a sample's archiving location.

The International Geo Sample Number (IGSN) provides solutions. The IGSN is a unique persistent identifier for samples that can be obtained by submitting sample metadata to the System for Earth Sample Registration (SESAR, www.geosamples.org). If data in publications is referenced to the IGSN, the SESAR database will allow access to all sample metadata including the owner and current location. Use of the IGSN in digital data systems allows building linkages between the digital representation of samples in SESAR (sample profiles) and their related data in the literature and in web-accessible digital data repositories.

To make data centers and scientific web portals effective ways of data sharing, scientists need to prepare their data for online publication. As an incentive to authors data publication should have the rank of a citeable publication, adding to his reputation and ranking among his peers.

For data to be citeable it is necessary that they can be referred to in a persistent way. Since the location of internet resources (URL) may easily change, the use of persistent identifiers (e.g. Handle, DOI or URN) as a way to locate the desired dataset reliably via the internet over a long time is a prerequisite for on-line data publication (www.std-doi.de).

9th IKC