

AGU 2011

ED53B-0780

Overcoming Challenges to Making Data Re-Usable: The Example of Geochemical Databases

Rivera, T.A., Lehnert, K., Hsu, L. Johansson, A.

In the early 1990s, the call for systems in which geochemical data could be shared among the research community led to the development of rock-type specific databases, such as PetDB and GEOROC. However, as these and other databases have grown over the last decade, so have the challenges to preserving data integrity, particularly managing of sample metadata. Proper documentation and preservation of metadata are key to qualitative re-use of geochemical data, including the reproduction of the published results. As methodologies advance, and the number of new data-intensive publications increases, the need for documenting and standardizing metadata becomes critical. To date, data managers perform much of the data entry, largely through extracting the geochemical data and associated metadata from publications, as well as performing data quality control and validation. In many cases, especially with legacy data, essential metadata is either missing or becomes a matter of interpretation by the data manager. Following 10 years of data management experience, the Geoinformatics for Geochemistry (GfG) group has recognized four fundamental parameters needed to uphold data reliability: data source, sample information, analytical information, and method-specific information. With the advancement of digital data management and new data policies, the GfG group has begun to solicit the data directly from authors, using templates specifically focused on metadata capture. Once completed, the author uploads the template into the Geochemical Resource Library (GRL), where the data are curated for use by other researchers, educators, and for long-term preservation. From the GRL, a data manager can transfer the data into the appropriate domain database, making them searchable by an expanded audience. Although there are still limitations to the use of the templates, it is an attempt to work more closely with researchers so that the needs for data preservation are communicated and standardized, ultimately leading to the re-use of more meaningful data.