Data publication, documentation and user friendly landing pages – improving data discovery and reuse

Kirsten Elger, Damian Ulbricht, and Roland Bertelmann
Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences, Potsdam, Germany (kelger@gfz-potsdam.de)

Research data are the basis for scientific research and often irreplaceable (e.g. observational data). Storage of such data in appropriate, theme specific or institutional repositories is an essential part of ensuring their long term preservation and access. The free and open access to research data for reuse and scrutiny has been identified as a key issue by the scientific community as well as by research agencies and the public. To ensure the datasets to intelligible and usable for others they must be accompanied by comprehensive data description and standardized metadata for data discovery, and ideally should be published using digital object identifier (DOI). These make datasets citable and ensure their long-term accessibility and are accepted in reference lists of journal articles (http://www.copdess.org/statement-of-commitment/).

The GFZ German Research Centre for Geosciences is the national laboratory for Geosciences in Germany and part of the Helmholtz Association, Germany's largest scientific organization. The development and maintenance of data systems is a key component of 'GFZ Data Services' to support state-of-the-art research. The datasets, archived in and published by the GFZ Data Repository cover all geoscientific disciplines and range from large dynamic datasets deriving from global monitoring seismic or geodetic networks with real-time data acquisition, to remotely sensed satellite products, to automatically generated data publications from a database for data from micro meteorological stations, to various model results, to geochemical and rock mechanical analyses from various labs, and field observations.

The user-friendly presentation of published datasets via a DOI landing page is as important for reuse as the storage itself, and the required information is highly specific for each scientific discipline. If dataset descriptions are too general, or require the download of a dataset before knowing its suitability, many researchers often decide not to reuse a published dataset. In contrast to large data repositories without thematic specification, theme-specific data repositories have a large expertise in data discovery and opportunity to develop usable, discipline-specific formats and layouts for specific datasets, including consultation to different formats for the data description (e.g., via a Data Report or an article in a Data Journal) with full consideration of international metadata standards.