



Enrichment of Data Publications in Earth Sciences – Data Reports as a Missing Link

Kirsten Elger, Roland Bertelmann, Christian Haberland, and Peter L. Evans

GFZ German Research Centre for Geosciences, Potsdam, Germany (kirsten.elger@gfz-potsdam.de)

During the past decade, the relevance of research data stewardship has been rising significantly. Preservation and publication of scientific data for long-term use, including the storage in adequate repositories has been identified as a key issue by the scientific community as well as by bodies like research agencies.

Essential for any kind of re-use is a proper description of the datasets. As a result of the increasing interest, data repositories have been developed and the included research data is accompanied with at least a minimum set of metadata. This metadata is useful for data discovery and a first insight to the content of a dataset. But often data re-use needs more and extended information. Many datasets are accompanied by a small 'readme file' with basic information on the data structure, or other accompanying documents. A source of additional information could be an article published in one of the newly emerging data journals (e.g. Copernicus's ESSD Earth System Science Data or Nature's Scientific Data).

Obviously there is an information gap between a 'readme file', that is only accessible after data download (which often leads to less usage of published datasets than if the information was available beforehand) and the much larger effort to prepare an article for a peer-reviewed data journal.

For many years, GFZ German Research Centre for Geosciences publishes 'Scientific Technical Reports (STR)' as a report series which is electronically persistently available and citable with assigned DOIs. This series was opened for the description of parallel published datasets as 'STR Data'. These are internally reviewed and offer a flexible publication format describing published data in depth, suitable for different datasets ranging from long-term monitoring time series of observatories to field data, to (meta-)databases, and software publications. STR Data offer a full and consistent overview and description to all relevant parameters of a linked published dataset. These reports are readable and citable on their own, but are, of course, closely connected to the respective datasets. Therefore, they give full insight into the framework of the data before data download. This is especially relevant for large and often heterogeneous datasets, like e.g. controlled-source seismic data gathered with instruments of the 'Geophysical Instrument Pool Potsdam GIPP'. Here, details of the instrumentation, data organization, data format, accuracy, geographical coordinates, timing and data completeness, etc. need to be documented.

STR Data are also attractive for the publication of historic datasets, e.g. 30-40 years old seismic experiments. It is also possible for one STR Data to describe several datasets, e.g. from multiple diverse instruments types, or distinct regions of interest.

The publication of DOI-assigned data reports is a helpful tool to fill the gap between basic metadata and restricted 'readme' information on the one hand and preparing extended journal articles on the other hand. They open the way for informed re-use and, with their comprehensive data description, may act as 'appetizer' for the re-use of published datasets.