



Data, Samples, Software, Texts - publishing Services at the GFZ German Research Centre for Geosciences

Kirsten Elger, Damian Ulbricht, Roland Bertelmann, and Javier Quinteros

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

There is an increasing international demand for free and open access to publicly funded scientific research products ranging from papers to data and software underlying scholarly publications. To facilitate re-use of data and software, research output should be catalogued and unambiguously identified to track provenance of data, data products or software and to credit the authors. Furthermore, naming samples that are the origin of data unambiguously and citing them with confidence complements data provenance records, creates trust in data and facilitates data synthesis studies.

GFZ is the national laboratory for Geosciences in Germany and provides as part of the Helmholtz Association large scientific infrastructures that create significant volumes of data. Data originate from large monitoring networks, are measured in field campaigns, in laboratories or generated during modelling. To support state of the art research, managing and cataloguing data are an important pillar in GFZ activities.

The data services deployed by the Library and Information Services (LIS) of GFZ are providing support in publication, cataloguing and accessing research output, thus making research products visible. This is in line with the mission of LIS, which is to extend their services to all products of research, including during their whole life cycle: texts, research data, software, samples and outreach products. The FAIR Principles for research data management are guiding principles.

In the framework of GFZ Data Services, the domain repository for geoscientific data, data and software are published with digital object identifiers (DOI). Data and software are complemented with Data Reports that are citable through DOIs and serve as possibility for data descriptions in between classical README files and data descriptor papers in data journals. Furthermore, GFZ is an allocating agent for the International Geo Sample Number (IGSN). IGSN is a globally unique persistent identifier for physical samples that provides discovery functionality of digital sample descriptions via the internet.

A major focus for the data curation is to carefully cross-reference the different research products via their metadata using Persistent Identifiers (PID). In particular, International Geo Sample Numbers (IGSN) are associated to data and written publications when available. Literature and data that are important for understanding are linked to research output through DOIs, Uniform Resource Names (URN) or Handles. IGSNs are also based on the Handle System and may be cross-referenced with data publications via the related identifier type of the DataCite Metadata 4.0 Schema. Furthermore, creators and contributors are linked to research products through Open Researcher and Contributor ID (ORCID).

When PIDs are assigned at GFZ, emphasis is put on choosing an appropriate system. Specifically, the GFZ Seismological Data Archive (GEOFON) assigns DOIs to seismic networks to enable citation of seismological data. ORCIDs have been adopted to identify persons in all metadata instances (e.g. DataCite) whenever possible. EPIC PIDs are used in the context of the EUDAT project to replicate data files to other European data centres and make a minimum set of attributes from the metadata available in order to reach interoperability with other systems.