



Linking Environmental Data, Samples and Publications – examples from GFZ Data Services

Damian Ulbricht, Kirsten Elger, and Boris Radosavljevic

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

In recent years scientists, research institutions and funding bodies committed to increase the availability of research products to the public. Research data repositories are important actors in this effort as they persistently archive research data and research software, and make them accessible and discoverable. The provision of the full provenance of research results is supported by cross-referencing these research products with scholarly literature and physical samples. In particular domain-specific repositories facilitate discovery and maximize visibility by indexing the asset with domain keywords, maintaining metadata catalogues and disseminating metadata to external databases and portals.

GFZ is the national laboratory for geosciences in Germany and provides large scientific infrastructures that create significant volumes of data. Data originate from large monitoring networks, field campaigns, laboratories or are generated during modeling. Managing and cataloguing of data and research software are an important pillar of GFZ activities and service for the scientific community and support state of the art research.

The data services developed by the Library and Information Services (LIS) of GFZ provide support in publishing, cataloguing and accessing all kinds of research output encompassing their whole life cycle (texts, research data, software, samples and outreach products) guided by the FAIR Principles for research data management.

GFZ Data Services publishes data and software with DOI to enable citation of research products. During the publication process, cross-linking articles and data reports is emphasised as these are important for understanding and reuse of the data or software. When data are derived from samples, the samples are linked to the data through the International Geo Sample Number (IGSN). Individuals are identified using the Open Researcher and Contributor ID (ORCID). Furthermore, research products are indexed by keywords that are popular in the Earth Sciences, i.e. NASA GCMD Science Keywords, Instruments and Platforms, the ICS International Chronostratigraphic Chart, the CGI Simple Lithology and the INSPIRE GEMET Thesaurus.

XML-based metadata are disseminated to external portals or infrastructures (e.g. B2FIND, EPOS) to improve visibility of research products. The research products are presented via DOI landing pages that retrieve citation information of linked material from public webservices. To maximize visibility and discoverability through the newly developed Google Dataset Search, AJAX scripts embed machine-readable metadata from DataCite's content negotiation services into the HTML pages.