Publication Instructions

GFZ Data Services

Publishing data and software is the act of making data available for re-use by others in citable form. This provides a means to gain visibility and credit for one's scientific endeavours and offers increased transparency and reproducibility of the scientific process. A data publication represents a stand-alone and ideally self-describing, research product. Due to the combination of data and metadata and supported by the use of persistent identifiers (PIDs) and controlled disciplinary vocabularies, data publications can be regarded as a best practice for sharing data following the FAIR Principles for research data management. Due to the combination of data and metadata and supported by the use of persistent identifiers (PIDs) and controlled disciplinary vocabularies, data publications can be regarded as a best practice for sharing data following the FAIR Principles for research data management.

Data and software publications with Digital Object Identifiers (DOI) are fully citable in research articles and should be included in the reference lists.

The data repository of GFZ Data Services makes data citable by assigning DOIs to data and scientific software. To foster reuse, emphasis is placed on cross referencing scholarly publications describing the full provenance of a dataset (methods, parameters and how the data were created) with the published datasets. GFZ Data Services increases the interoperability of long-tail data through:

- the provision of comprehensive domain-specific data description via standardised and machine-readable metadata with controlled domain vocabularies via an online Metadata Editor
- complementing the metadata with comprehensive and standardised technical data descriptions or reports, and
- by embedding the research data in wider context by providing cross-references through Persistent Identifiers (DOI, IGSN, ORCID, Fundref) to related research products (text, data, software) and people or institutions involved.

Timing of data publication: For most supplementary data to research articles, the publication of the data will be timely correlated with the publication of the article.

Review Links: Review links show a preview of the future DOI Landing Page with the correct citation (the DOI is already reserved and will not change) and full access to the data and associated documents. The review links were developed for two purposes: (1) proofreading by the author and (2) providing data access for referees during the review process of articles (see below). Please keep in mind: our **review links are static and every change is requiring the generation of a new review link.**

Restricted data access during the review period of articles: Recently, publishers require access to data for referees during the review process of manuscripts. To enable this restricted access to the still unpublished data, we can create a review link.

Data Description Templates

Even with a very rich metadata schema, published data are often not sufficiently described for being reusable. Additional technical descriptions, like the definition of column headers of tabular data, the description of the experiment setup, documentation of software tools used to manipulate the data or an overview of a complex file structure or specific data format, are necessary to fully understand the data. We have further observed that many researchers are unaware what a data publication represents and especially what to include in a data description. This observation, together with the aim to further standardize MSL data, led to the development of Data Description Templates. The templates are provided in commented and "usable" versions. Data descriptions with the templates are fully harmonized, curated by the repository staff and provided as a PDF data description in the data download folder.

Metadata Editor

To be self-describing and re-usable, every data publication should contain rich metadata. Metadata refers to "data about the data" and comprises all information necessary (for humans and machines) to find, read, and interpret the data.

A key tool for metadata generation is the GFZ Metadata Editor which assists scientists to create metadata compliant with metadata schemas popular in the Earth sciences (ISO19115, NASA GCMD DIF, DataCite). Since emphasis is placed on removing barriers, the Metadata Editor is publicly available. Users can fill in their metadata and make a copy to their local hard disk to continue work at a later time. Scientists are not requested to provide information that may be generated automatically. To improve usability, form fields names are translated from the metadata schema to the language familiar to scientists. The Metadata Editor also provides search functionality of structured vocabulary lists. In addition, multiple geospatial references can be entered via an interactive mapping tool, which helps to minimize problems with different conventions to provide latitudes and longitudes.

The Metadata Editor exists in different instances that are tailored to the needs of particular research projects or research infrastructures, e.g. the 'Multi-scale Laboratories' Thematic Core Service of the European Plate Observing System (EPOS) or the Geophysical Instrument Pool Potsdam (GIPP). To access any of these instances or read the Metadata How To, please follow the links below:

- **GFZ Metadata Editor:**
  URL: https://dataservices.gfz-potsdam.de/panmetaworks/metaedit/
- **EPOS Multi-scale Laboratories Metadata Editor:**
  URL: https://dataservices.gfz-potsdam.de/msl/
- **Metadata How To:**
  URL: https://bib.telegrafenberg.de/dataservices/support/metadata-editor-howto
- **Definitions of Metadata Fields used in the Metadata Editor:**
  URL: https://gfzpublic.gfz-potsdam.de/pubman/item/item_5009874

Data Submission Workflow

1. Author: Submits metadata using the GFZ Metadata Editor. Data can be sent via Email or ftp upload link (provided by GFZ Data Services)

2. Curator: Controls submitted metadata and data description for completeness and consistency

3. Curator: (If required) converts and allocates data on GFZ Data Services ftp server

4. Curator: Sends review link for proofreading by author

5. Author: Proofreads and confirms edited data

6. Curator: Publishes dataset and registers the DOI