PANGAEA®
Data Publisher for Earth and Environmental Science

Stefanie Schumacher, Amelie Driemel, Hannes Grobe, Rainer Sieger

PANGAEA - Data Publisher for Earth & Environmental Science (www.pangaea.de) is an Open Access data-library aimed at archiving, publishing and distributing georeferenced data from earth system research. The system guarantees long-term availability of its content through a commitment of the hosting institutions Alfred Wegener Institute Helmholtz, Centre of Polar and Marine Research (AWI) and Center for Marine Environmental Sciences (marum).

Observational and analytical data files are stored with their metadata in a relational database. Each data set includes a bibliographic citation with a Digital Object Identifier (DOI). Data are archived as an article supplement or as an independent citable data-publication. PANGAEA is open to any project, institution, or individual scientist to use or to archive and publish data.

PANGAEA actually provides more than 375 000 data set, consisting of >13 billion data points, including collections from national and international research programs. This includes c. 33 000 datasets from national and international Antarctic research, e.g., Cape Roberts Project, ANDRILL, EPICA project, and data of the first and last International Polar Year (IPY 1882-1883 and 2007-2008).

Data Submission
Data and metadata are submitted via a ticket system.

- TAB-delimited text files or excel-format are preferred.
- Binary objects like photos, images, graphics, maps, models, movies or NetCDF files are also accepted.

Data Publication
Data citation with DOI

- Information about:
  - Data Processing
  - Projects

- Sampling position with latitude/longitude

- Parameter list with units, methods and Principle Investigator (PI)

Data Upload
A PANGAEA-Editor processes the data for import into the relational data base. This includes an editorial review process.

Data Access
The PANGAEA Search Engine (www.pangaea.de) bases on Elasticsearch and provides:

- Google-like search-field for author names, parameters, projects, ships etc.
- topic based search fields
- map based search

Facetted Search offers filters to specify the search and outreach.

PANGAEA is furnished with a well developed interoperability framework thus allowing to disseminate metadata and data to registries, data portals, and other service providers.

Archived data are machine readable and mirrored into the PANGAEA data warehouse. The data warehouse allows efficient compilations of huge data amounts from several datasets. The output file can easily be used for further compilations, visualization and models.