# PANGAEA - Data publisher for Earth & environmental sciences

Michael Diepenbroek

Initiated in the 1990s, PANGAEA[1] has evolved from a paleoclimate-data archive to a multidisciplinary data publisher for Earth and environmental sciences, accredited as a *World Data Center* by the International Council for Science World Data System (ICSU WDS)[2] and as *World Radiation Monitoring Center* (WRMC)[3] within the World Meteorological Organization Information System (WIS)[4].

Even in its earliest stages, data were archived consistently and carefully curated. This involves cleaning, harmonizing, and integrating data, as well as metadata, within PANGAEA's editorial workflow. Consequently, all data sets are annotated including information on how, when, and where they were produced, information about principal investigators, measurement and observation types, sampling and analysis methods, and devices as well as references to literature. In January 2005, the first data sets were registered and minted with a standard-compliant Digital Object Identifier (DOI), which enables proper citation of data and their integration within the publishing-industry workflow and bibliometric analyses. Today, PANGAEA holds around 375'000 citable data sets comprising more than 13 billion data items - numerical and textual data as well as binaries such as images, videos, or files with community specific mime types. Each data item is a georeferenced record including the parameter value, parameter type, and the spatial and temporal coverage; spatio-temporal values themselves are not data items. Over 18% of published data sets include at least one author linked to ORCID (the author identifier of the publishing industry). PANGAEA is operated as an Open Access library and is open to any project, institution, or individual scientist to use or to archive and publish data[5].

As paleoclimate research is the scientific background of PANGAEA's founders, it has a long-lasting relationship with PAGES and also looks back to a long-standing collaboration with the NOAA WDS-Paleo. The recent common focus is on interoperability and findability of paleodata. Both data centers build the archive backbone for paleodata. PANGAEA holds large inventories of all types of paleodata, for example isotope and geochemical data as well as pollen and tree-ring data. An example data collection is the data collected by the PAGES C-PEAT working group[6].

## Editorial
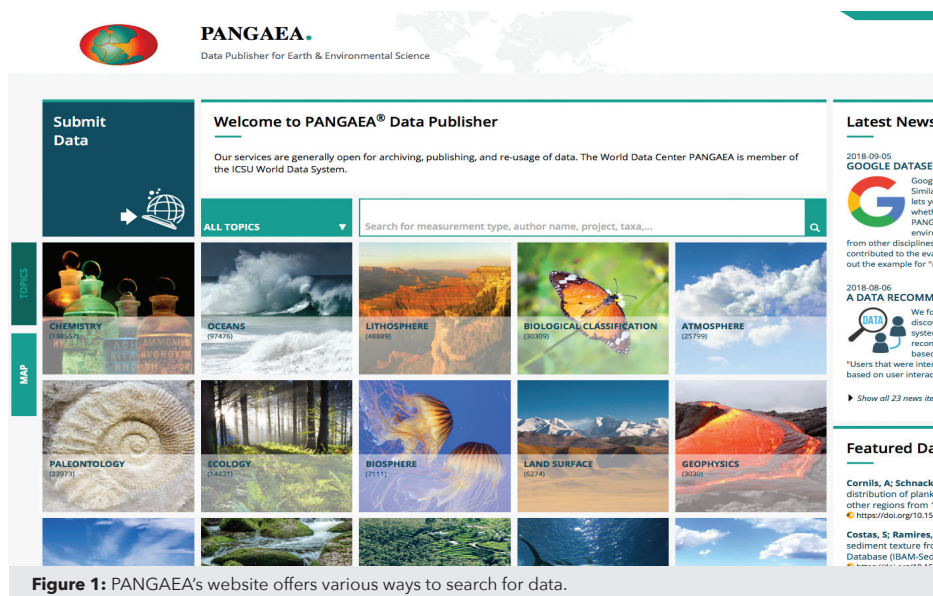PANGAEA is operated by a team of data editors, project managers, and IT specialists[7].



**Figure 1:** PANGAEA's website offers various ways to search for data.

Our editors are scientists with expertise in all fields of Earth and environmental science. They have a deep knowledge of the review and processing of scientific data. The PANGAEA data editorial ensures the integrity and authenticity as well as a high reusability of data. Archived data are machine readable and mirrored into our data warehouse which allows efficient compilations and downloads of data[8].

Data are submitted using a ticket system (Jira[9]) and assigned to an editor who is a specialist in the corresponding data domain. Preparation of the data for import is done with a highly sophisticated editorial system. Data editors check the completeness and validity of data and metadata, reformat data according to the PANGAEA ingest format, and harmonize data and metadata using standard terminologies (Diepenbroek et al. 2017). The editorial review is complemented by inviting authors and external reviewers (e.g. reviewers of articles supplemented by the data) to proofread the data sets. After being accepted, the data sets are archived, provided with a DOI, and registered at DataCite[10].

## Interoperability and findability
PANGAEA is furnished with a well-developed interoperability framework based on internationally accepted standards. All interfaces to the information system are based on web services including map support (Google Earth, Google Maps)[11]. This allows most effective dissemination of metadata and data to all major internet search-engine registries, library catalogues, data portals, and other service providers, and consequently ensures the optimal findability of data hosted by PANGAEA. Scientific data portals supported include DataOne, GEOSS[12], the ICSU WDS[2], GBIF[13] and also the paleo data portal at NCEI[14]. Other infrastructures supported include DataCite[15], ORCID[16], and Scholix[17], which supplies links between scholarly literature and data. Interoperability with ORCID allows users to login with their ORCID ID and link it to their user profile in PANGAEA. In this way, data publications are automatically assigned to matching ORCID IDs.

### AFFILIATION
MARUM - Center for Marine Environmental Sciences, University Bremen, Germany

### CONTACT
Michael Diepenbroek: mdiepenbroek@pangaea.de

### REFERENCES
Diepenbroek M et al. (2017) J Biotech 261: 177-186

### LINKS
[1]pangaea.de
[2]icsu-wds.org
[3]bsrn.awi.de
[4]wmo.int/pages/prog/www/WIS
[5]pangaea.de/submit
[6]pangaea.de/?q=project%3Alabel%3APAGES_C-PEAT
[7]pangaea.de/about/team.php
[8]pangaea.de/tools
[9]en.wikipedia.org/wiki/Jira_(software)
[10]datacite.org
[11]ws.pangaea.de
[12]geoportal.org
[13]gbif.org
[14]ncdc.noaa.gov/paleo-search
[15]datacite.org
[16]orcid.org
[17]scholix.org