



PANGAEA[®] - Data Publisher for Earth & Environmental Science - Research data enters scholarly communication and big data analysis

Michael Diepenbroek (1), Uwe Schindler (1), Morris Riedel (2), and Robert Huber (1)

(1) MARUM, Universität Bremen, Germany, (2) Jülich Supercomputing Centre, Forschungszentrum Jülich, Germany

The ICSU World Data Center PANGAEA is an information system for acquisition, processing, long term storage, and publication of geo-referenced data related to earth science fields. Storing more than 350.000 data sets from all fields of geosciences it belongs to the largest archives for observational earth science data. Standard conform interfaces (ISO, OGC, W3C, OAI) enable access from a variety of data and information portals, among them the search engine of PANGAEA itself (www.pangaea.de) and e.g. GBIF. All data sets in PANGAEA are citable, fully documented, and can be referenced via persistent identifiers (Digital Object Identifier - DOI) - a premise for data publication. Together with other ICSU World Data Centers (www.icsu-wds.org) and the Technical Information Library in Germany (TIB) PANGAEA had a share in the implementation of a DOI based registry for scientific data, which by now is supported by a worldwide consortium of libraries (www.datacite.org). A further milestone was building up strong co-operations with science publishers as Elsevier, Springer, Wiley, AGU, Nature and others. A common web service allows to reference supplementary data in PANGAEA directly from an articles abstract page (e.g. Science Direct). The next step with science publishers is to further integrate the editorial process for the publication of supplementary data with the publication procedures on the journal side.

Data centric research efforts such as environmental modelling or big data analysing approaches represent new challenges for PANGAEA. Integrated data warehouse technologies are used for highly efficient retrievals and compilations of time slices or surface data matrixes on any measurement parameters out of the whole data continuum. Further, new and emerging big data approaches are currently investigated within PANGAEA to e.g. evaluate its usability for quality control or data clustering.

PANGAEA is operated as a joint long term facility by MARUM at the University Bremen and the Alfred Wegener Institute for Polar and Marine Research (AWI). More than 80% of the funding results from project data management and the implementation of spatial data infrastructures (more than 160 International to national projects since the last 15 years - www.pangaea.de/projects).