

Prospective FAIR bathymetry data archiving in PANGAEA - Data Publisher for Earth & Environmental Science

Daniel Damaske, MARUM / PANGAEA (Germany), ddamaske@marum.de
Janine Felden, Alfred-Wegener-Institute / PANGAEA (Germany), janine.felden@awi.de
Frank Oliver Glöckner, Alfred-Wegener-Institute / MARUM/ PANGAEA (Germany),
frank.oliver.gloeckner@awi.de
Gauvain Wiemer, DAM (Germany), wiemer@allianz-meeresforschung.de
on behalf of the DAM and PANGAEA Teams

Objective

The PANGAEA data information system (<https://www.pangaea.de/>) is worldwide one of the leading data repositories for Earth & Environmental data hosted by the Alfred-Wegener-Institute, Helmholtz Centre for Polar and Marine Research (AWI), and the Center for Marine Environmental Sciences (MARUM), University of Bremen. Over the last 25 years, PANGAEA archived and published - supervised by scientific data curators – hundreds of thousands of datasets from all over the world from different scientific disciplines and research domains including bathymetry data.

PANGAEA already holds a few hundreds of bathymetry raw datasets containing multi-beam echosounder raw data from entire scientific cruises and as well as processed datasets consisting of vector, ASCII and raster data. The first bathymetry dataset was already stored in 2001.

Since several years, PANGAEA receives an increasing number of bathymetry data submissions, coupled with a constant growth of the data volume. Unfortunately, current data storage and download solutions in PANGAEA are laborious for searching and subsequent downstream processing of bathymetry data. Many datasets do not contain sufficient metadata in order to assure efficient long-term reuse of data. In addition, essential ancillary data, such as sound velocity profiler data (since the sound velocity has profound influence on depth computation) are not yet allocated to most bathymetry datasets stored in PANGAEA. In fact, based on an investigation of all bathymetry related datasets in PANGAEA conducted in September 2019, only eight out of 173,365 files are sound velocity profile (SVP) data files.

To cope with the increasing amount of data to be delivered to PANGAEA, standard operating procedures (SOPs) for bathymetry data, compliant with the FAIR data principles (Wilkinson et al., 2016), is going to be compulsory for all data submissions. Such consistent and efficient data storage concepts are currently developed as part of the “Underway Research Data” project, an initiative of the German Marine Research Alliance (Deutsche Allianz Meeresforschung e.V. (DAM), <https://www.allianz-meeresforschung.de/>). In addition, within the scope of these SOPs, bathymetry data in PANGAEA are automatically processed to generate tracklines and coverage polygons of the surveyed area including PANGAEA metadata and raster previews, which can prospectively be explored on the DAM Data Portal www.marine-data.org/. The visualized data here are also offered as OGC Web Map Services. These efforts and services can contribute to international projects such as the *Seabed 2030* Project from the Nippon Foundation and GEBCO (General Bathymetric Chart of the Ocean), or other bathymetry data collecting data portals relying mainly on pre-existing data streaming services.

Methods

One of the key challenges of bathymetric data curation is the proper amendment of the metadata description in the data curation process. Consistently adding key parameters to each single multi-beam raw data file, e.g. latitude, longitude, date and time - which can be up to thousands of data files per dataset submission - are necessary to meet international standards (e.g. ISO 19139 metadata standards). Since most data owners do not have the capacity or the means to process and provide

these information during data submission, new semi-automated processes and procedures are currently developed and tested for metadata retrieval, mainly relying on the open source bathymetry processing tool MB-System. Similar methods are also developed for processed (gridded) bathymetry datasets and to partially reprocess older bathymetry datasets to the new metadata standard.

Using controlled vocabularies, such as SeaDataNet (NERC Vocabulary Server NVS2.0), ensures further data consistency and interoperability of the published datasets. Furthermore, bathymetry data of the German marine research fleet (predominately of RV Sonne, RV Maria S. Merian, RV Meteor, RV Polarstern) are connected to detailed sensor descriptions in a Sensor Information System <https://sensor.awi.de/> (Koppe et al., 2018), therefore providing the user with additional technical device descriptions.

These measures enhance the possibility to find, reuse and compare data and - at best - judging data quality and getting simultaneously - if available – significant ancillary data (SVP) along the data submission.

Project background

The DAM “Underway Research Data” project, spanning across different institutions, started in mid-2019. The aim of the project is to improve and unify the constant data flow from German research vessels to data repositories like PANGAEA. This comprises multibeam-echosounder and other permanently installed scientific devices and sensors following the FAIR data management aspects (Fig.1). Thus, exploiting the full potential of German research vessels as instant “underway” mobile scientific measuring platforms, provides an added value for scientific research.

The DAM is founded by the German federal government and the northern German federal states of Bremen, Hamburg, Mecklenburg-Western Pomerania, Lower Saxony and Schleswig-Holstein and is currently involving 19 German universities and non-university research institutions with a key focus in marine and climate research. The DAM contributes with its activity to the German effort of the National Research Data Infrastructure (NFDI) and especially NFDI4Earth - the NFDI Consortium for Earth System Sciences.

The “Underway Research Data” project is closely cooperating with the MareHub project of the Helmholtz Association of German Research Centers. Together they develop new visualisation technologies with the aim to offer a web-based map display of marine data of the German marine research vessel fleet, and thus facilitating the discovery of bathymetry data - including a vast package of metadata descriptions - for the national and international scientific community.

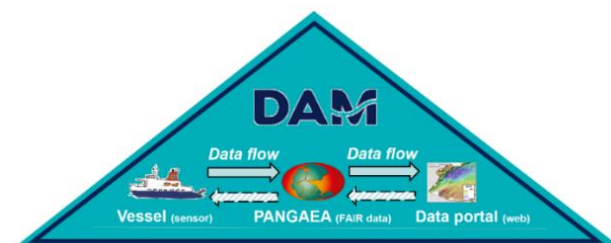


Figure 1: Data flow from research vessels to data archives, FAIR data archiving in PANGAEA and web-based visualisation of bathymetry data. *Data management is accompanied and supported by the German Marine Research Alliance (Deutsche Allianz Meeresforschung e.V., DAM)*

References

Koppe, R. , Gerchow, P. , Macario, A. , Haas, A. , Schäfer-Neth, C. , Rehmcke, S. , Walter, A. , Düde, T. , Weidinger, P. , Schäfer, A. and Pfeiffenberger, H. (2018): SENSOR.awi.de: Management of heterogeneous platforms and sensors , RDA 11th Plenary, Berlin, 21 March 2018 - 23 March 2018

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>