SciDataCon 2025



Contribution ID: 39

WORLD DATA SYSTEM

Type: Presentation

## PANGAEA –30 years of publishing data for Earth & Environmental Science

Thursday 16 October 2025 11:11 (11 minutes)

PANGAEA –Data Publisher for Earth & Environmental Science is a worldwide recognised digital data repository that plays a pivotal role in archiving, publishing, and disseminating scientific data related to earth and environmental sciences. As a publicly accessible information system, PANGAEA ensures that high-quality, well-structured, and interoperable datasets are preserved and made available to the scientific community. The platform fosters collaborations across various scientific disciplines, including geology, oceanography, climatology, ecology, and biodiversity by allowing scientists to archive and share georeferenced observational and experimental data. Each dataset is assigned a Digital Object Identifier (DOI), ensuring persistent citation and long-term accessibility. Its commitment to the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles ensures that data curation, management and dissemination align with international best practices, enhancing scientific transparency and reproducibility. In this way, PANGAEA has grown into a widely respected platform serving a diverse range of research communities since its establishment in the early 1990s by the Alfred Wegener Institute –Helmholtz Centre for Polar and Marine Research (AWI) and the Center for Marine Environmental Sciences (MARUM) at the University of Bremen.

PANGAEA plays a vital role in supporting large-scale international research projects and initiatives such as the Intergovernmental Panel on Climate Change (IPCC), the International Ocean Discovery Program (IODP), and the World Data System (WDS). holds a mandate from the World Meteorological Organization (WMO), to host the World Radiation Monitoring Center (WRMC). It is accredited as a World Data Center by the International Council for Science (ICS) since 2001 and has been certified as a trustworthy long-term data archive by Core Trust Seal. The repository integrates seamlessly with other global data infrastructures, ensuring compatibility with frameworks such as the Global Earth Observation System of Systems (GEOSS) and the European Open Science Cloud (EOSC). By linking datasets with scientific publications and ensuring proper attribution, PANGAEA strengthens the credibility and impact of research findings.

The platform's manual data curation process adheres to rigorous standards, involving expert review and validation before datasets are published. Researchers uploading data are required to provide comprehensive metadata compliant to ISO 19115, including descriptions of methodologies, instrumentation, and data provenance for each data set. This meticulous approach minimizes errors and enhances the reliability of published datasets. Additionally, PANGAEA supports a wide variety of data formats, including numerical, textual, image, and geospatial datasets, facilitating diverse applications in scientific research. All data and metadata are compiled in close collaboration between the scientists and trained field experts acting as data editors. Both, data and metadata are checked for completeness and plausibility, ensuring high quality standards according to the FAIR data principles. Semantic interoperability during data curation is ensured through strict application and dynamic evolution of terminologies according to international protocols and standards. All published datasets carry a licence information (CC0 or CC-BY). The structured metadata accompanying each dataset enhances discoverability and usability, allowing researchers to effectively integrate PANGAEA's resources into their work. Currently, PANGAEA provides access to over 434,000 datasets containing over 31 billion individual measurements, including those collected through over 889 national and international projects.

Beyond its function as a repository, PANGAEA offers numerous tools and services. In addition to the classic access to data via the website, an integrative use of data in the form of a DataWarehouse and a set of tools for programmatic data processing are available for this purpose. The two applications written for the scripting languages Python and R, pangaeapy and pangaear, respectively, make use of the well-developed interoperability framework of PANGAEA. This framework allows most effective dissemination of metadata and data to all major internet search-engine registries, library catalogs, data portals, and other service providers, and ensures the optimal findability of data hosted by PANGAEA. The respective web services entail SOAP and REST APIs, a Schema.org/Dataset compliant metadata endpoint and OAI-PMH for various metadata content standards like DataCite, Dublin Core, DIF and ISO 19115 for harvesting. These technical capabilities make PANGAEA an essential resource for interdisciplinary studies addressing complex environmental challenges such as climate change, biodiversity loss, and natural resource management.

Looking forward, PANGAEA aims to further strengthen its role as a cornerstone of global earth and environmental data infrastructure. The rapid increase in data volume and complexity requires to extend PANGAEAs front-office model with trained data stewards all over the world. Efforts to integrate artificial intelligence and machine learning techniques into data curation and retrieval processes are ongoing, promising to further enhance the efficiency and scalability of PANGAEA's operations. By collaborating with established data initiatives and delivering data products for data portals as well as fostering international partnerships, PAN-GAEA will continue to facilitate innovative research in the earth and environmental sciences. As scientific data management evolves, PANGAEA remains at the forefront, providing researchers with the tools and resources necessary to address some of the most pressing environmental challenges of our time.

The presentation will provide an overview of the current status and further perspectives of PANGAEA - data publisher for earth & environmental science.

**Primary authors:** Prof. GLÖCKNER, Frank Oliver (Alfred Wegener Institute - Helmholtz Center for Polar- and Marine Research & MARUM - Center for Marine Environmental Sciences University of Bremen); FELDEN, Janine (Alfred Wegener Institute - Helmholtz Center for Polar- and Marine Research, PANGAEA); Mr SCHINDLER, Uwe (PANGAEA, MARUM/University of Bremen, Germany)

**Presenter:** Prof. GLÖCKNER, Frank Oliver (Alfred Wegener Institute - Helmholtz Center for Polar- and Marine Research & MARUM - Center for Marine Environmental Sciences University of Bremen)

**Session Classification:** Presentations Session 10: Infrastructures to Support Data-Intensive Research - Local to Global

**Track Classification:** SciDataCon2025 Specific Themes: Open research through Interconnected, Interoperable, and Interdisciplinary Data