Geophysical Research Abstracts
Vol. 13, EGU2011-4859, 2011
EGU General Assembly 2011

PANGAEA® and the ICSU World Data System (WDS) - towards a global system of data publishers and data libraries

Michael Diepenbroek
University Bremen, MARUM, PANGAEA, Bremen, Germany (mdiepenbroek@pangaea.de)

In the empirical sciences the number of publications doubles every 10 years, whereby the corresponding data production amounts to a multiple of this increase - in particular in the environmental sciences. The availability of data is one of the key factors for global and complex science approaches („data driven science“). However, increases, volumes, and the heterogeneity of data pose a huge challenge. The lack of a generally available way for archiving and publishing data actually turned out to be a severe structural problem in the empirical sciences. The objective is to organize technically available data in a manageable way that allows to derive knowledge from information whereby preserving the approved quality standards of scientific publishing. Presently, at best a few percent of the last decades data production is available whereby this does not necessarily mean that they are usable. Quality assurance, data and metadata formats as well the interoperability of related systems have to follow global standards and it is the consistency and reliability of data and infrastructure in the end that allows a single scientist to efficiently use data from distributed systems.

The „World Data System“ (WDS) of the „International Council of Science“ (ICSU) is a global federated system of long term data archives and data related services covering a wide spectrum of natural sciences and encouraging interdisciplinary science approaches. Since 2007 the system is under revision with the aim to integrate various data centers and services with different scope. The main focus lies on (1) long term stewardship of scientific data, (2) data and metadata services including data analysis, and (3) data publication services. For the latter strong relationships with science publishers and libraries will be built up. Emphasis will also be given to interlink with the data production side, in particular science projects and programs. Furthermore, the WDS will maintain a common data and information infrastructure and interoperability conform to global standards thus ensuring efficient usage of supplied data and services as well as interlinkage with other networks and programs (see figure).

PANGAEA® (Publishing Network for Geoscientific and Environmental Data) is an information system which represents the ideal functioning of the WDS. Storing more than half a million data sets from all fields of geosciences it belongs to the largest archives for observational 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. Google. All data sets in PANGAEA® are citable, fully documented, and can be referenced via persistent identifier (Digital Object Identifier - DOI) - a premise for data publication. Together with other ICSU World Data Centers 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). The long term goal is a "crossref for scientific data".

Seeing the overall positive impact on the quality and availability of scientific data the PANGAEA® group is offering publication services since several years. They were the main initiators of the "Earth System Science Journal" at Copernicus (http://earth-system-science-data.net/), which is used for standalone peer reviewed data publications. A further milestone was building up a cooperation with Elsevier. A common web service allows to reference supplementary data in PANGAEA® directly from the abstract pages of Science Direct. AGU, Kluwer, Nature, Copernicus, and Thomson Reuters have expressed strong interest and it is likely that similar services are built up until the end of the 2010. The next step with Elsevier is to further integrate the editorial process for the publication of supplementary data with the publication procedures on the Elsevier side, which practically means an extension of the peer review and a synchronization of editorial processes.

PANGAEA® is operated as a long term facility by MARUM at the University Bremen and the Alfred We-

gener 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 (47 International, 46 European and 37 national projects since the last 12 years - www.pangaea.de/projects).