

Integrating data management services in marine sciences

Lisa Paglialonga, GEOMAR Helmholtz Centre for Ocean Research, Kiel (Germany),
lpaglialonga@geomar.de

Hela Mehrtens, GEOMAR Helmholtz Centre for Ocean Research, Kiel (Germany),
hmehrtens@geomar.de

Pina Springer, GEOMAR Helmholtz Centre for Ocean Research, Kiel (Germany),
pspringer@geomar.de

Carsten Schirnack, GEOMAR Helmholtz Centre for Ocean Research, Kiel (Germany),
cschirmack@geomar.de

Claas Faber, GEOMAR Helmholtz Centre for Ocean Research, Kiel (Germany),
cfaber@geomar.de

Organisation

The joint GEOMAR Data Management Team is a cooperation of Helmholtz Centre for Ocean Research Kiel (GEOMAR) and several large-scale research projects including collaboration with other marine research institutions. The coalition has established in Kiel a single sustainable data management infrastructure (KDMI) for marine research and supports the entire lifecycle of data description, data storage and data archiving. The platform is used to prepare data for peer-reviewed publications, exchange data within a project and for publication in longterm data archives. It is continuously improved by extensions developed in close cooperation with scientists and data centres.

Data management services

KDMI provides a web-based portal including OSIS (Ocean Science Information System) <https://portal.geomar.de/kdmi> which allows scientists to upload and share data files in the context of cruises, expeditions or models with allowance for any file format and content structure (Fig. 1). Access control for a file is primarily based on the community context it was uploaded within, but may be further restricted by the file's owner e.g. for a limited time of exclusive research (e.g. moratorium). Metadata of what, when, where and who are always public for institutional use and document the file's content. Visualisation of metadata is provided by KML formatted files which contain the appropriate links in order to retrieve and display e.g. the events of a cruise and availability of files or publications on-the-fly in tools such as GoogleEarth.

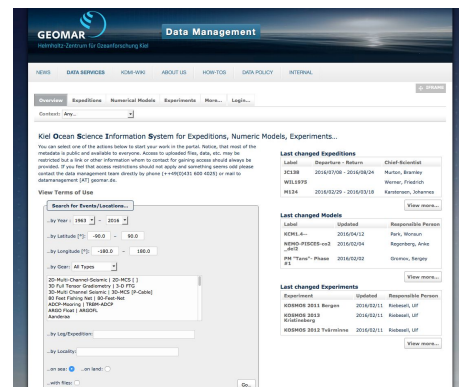


Fig. 1: Ocean Science Information System (OSIS)

A new and now widely appreciated service has been established realizing data management plan time schedules. It allows setup and surveillance of deliverables based on expected or collected research data. OSIS now also includes a numerical model database which provides an overview of existing model runs linked to their output results and to publications. Files (e.g. configuration files) can be uploaded as well and aid scientists in documenting their procedures and their outcome enriched by a minimal set of common metadata. Model output referenced in peer-reviewed papers is made publicly available via GEOMAR's OPeNDAP Server <https://data.geomar.de>.

Another data management service is the media archive (ProxSys, commercial software) (Fig.2). It allows to store, describe and search images and videos acquired during research dives of remotely operated and autonomous underwater vehicles equipped with up to four high-resolution cameras (stills or movies). An annotation system is currently developed (Fig. 3) which uses webservices to access images and videos and provides them for all projects members. This system avoids extensive network traffic and facilitates cooperation and reduces obsolete interchange.

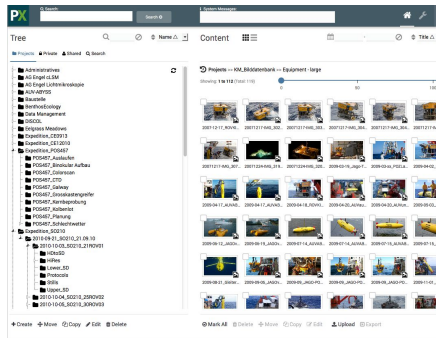


Fig 2.: Media storage (ProxSys)

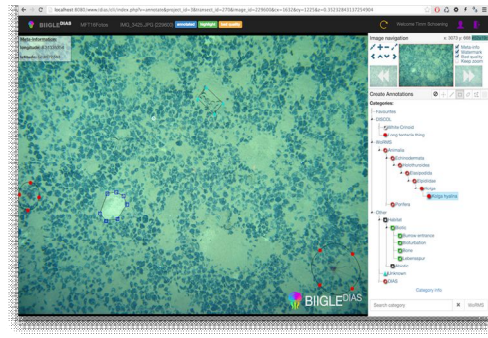


Fig 3.: Annotation System

We also offer version control systems (Subversion and Git), which are not only used for code development and documentation but also for data analysis and exchange. The TRAC Wiki and ticket system linked to Subversion and Git in the near future is used for documentation as well as improved collaboration and dissemination.

Integration

The data management portal (<https://portal.geomar.de>) is the starting point for all provided services. It allows to maintain public and internal webpages where collaboration tools like wiki, blog, calendar and document exchange can be used. User administration for all data management services is realised via the portal software and thus independent from institute's identity management systems while still allowing guarded access to data services for internal and external users.

An institutional repository for full text print publications is actively maintained by the GEOMAR library and also used for project publications (OceanRep: <http://oceanrep.geomar.de/>). The repository is linked with the data management portal allowing linkage between authors, their data and publications. Linkage between data and print publication is available in both systems: publications in the repository provide links to data and vice versa. The repository is also used to auto-generate publication lists on-the-fly and to present them on personal or a projects' homepages. An automated data link routine adds icons when additional external data are available related to a publication.

The data management team assists as data curators when data are to be published in World Data Centres (e.g. WDC-MARE, <http://www.pangaea.de>) in order to warrant long-term archival and access to the data (Fig.4). This cooperation with a world data center will make the data globally searchable while links to the data producers will ensure citability and provide points of contact for the scientific community. PANGAEA is harvested by several international portals, e.g. GEOSS (<http://www.geoportal.org>).

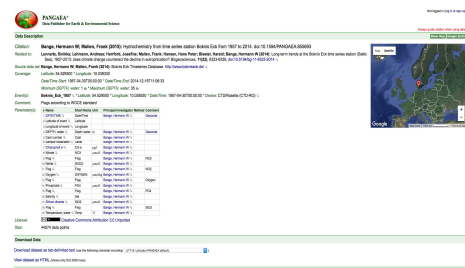


Fig. 4 : PANGAEA dataset

The concept of large-scale projects joining a single approach to establish one data management system has proved to be very successful. We experience propagation especially at GEOMAR and Kiel University to use the services known from the collaboration for new expeditions and research projects. We also observe an ongoing need to train new and young scientists and to include data output of new instrumentations. One of the essential data management experiences is the need for on-site personal contact and training of researchers during the entire data lifecycle. The high efficiency of data publications related to peer-reviewed journal publications still relies on on-site data stewardship conducted by a dedicated permanent data management team.