SFB754 - data management in large interdisciplinary collaborative research projects: what matters?

Hela Mehrtens, Pina Springer, Carsten Schirnick, Christiane Schelten, GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany

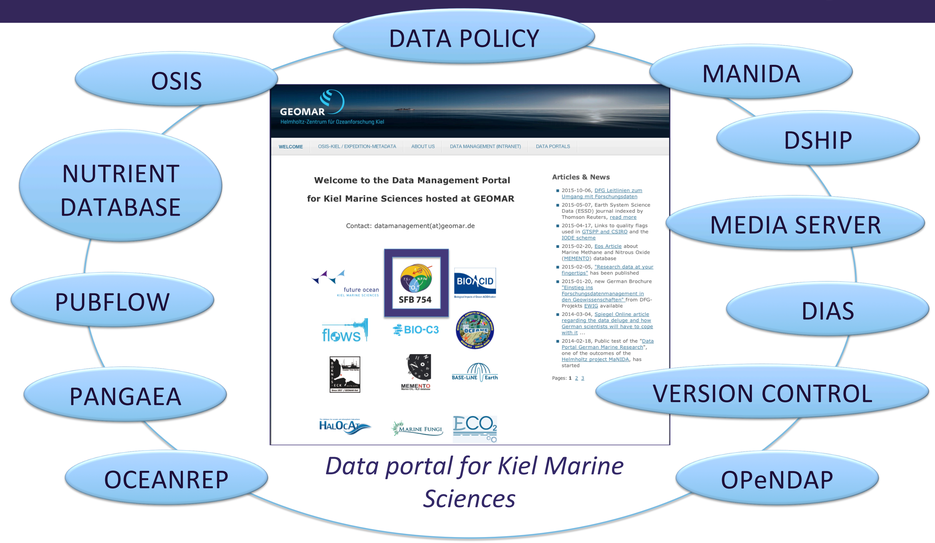
SFB 754

Major Objectives

- A sustainable and shared data management infrastructure for marine sciences in Kiel (KDMI)
- Support entire data life cycle: data capture, data description, data storage, data sharing and data publication with archival
- Continuous enhancement by new features developed in close cooperation with scientists and data centres
- Provide on-site personal support and training for scientists during entire data life cycle
- Ensure long-term availability, open access and citability of data sets

Why?

Collaborative research essentially depends on a supportive data management approach. To manage the data life cycle from measurement to publication, scientists need various tools and technical assistance depending on the type of acquired data (expedition, model, experiment).
Transparency of what data are available, where they can be found, how they may be used, how they have to be cited and the warranty of long-term availability is provided by the Kiel data management infrastructure (KDMI) in collaboration with international data centres.



Data portal for Kiel Marine Sciences

How?

Institutional tools like version control system, media server and OPeNDAP server and specific applications such as a nutrient database assist scientists in their daily work. The SFB 754 data manager assists as data curator when data are to be published (e.g. PANGAEA) in order to warrant long-term archival and access to the data. The cooperation with a world data centre will make the data globally findable while links to the data producers will ensure citability and provide points of contact for the scientific community.
The concept of a single data management infrastructure initiated by large-scale projects and participating institutions has proven to be very successful. We have experienced a snowball-like propagation among marine researchers at GEOMAR and Kiel University, they continue to engage data management services well known from collaboration with SFB 754.

Data curation process

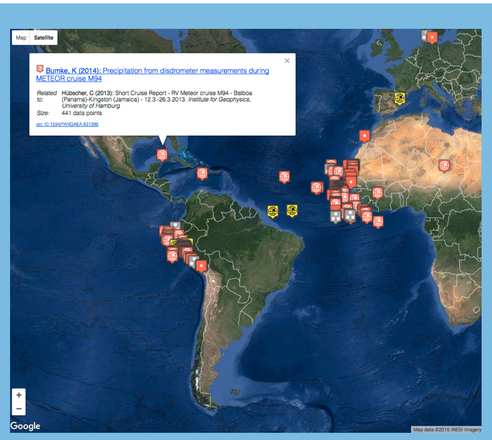
Measurement/Experiment → *Data deliverables* (6 months later) → *Reminder "upload data"* (month 7 to month 36) → *Reminder "publish data"* (3 years later)

- Description in the data portal
- Upload of first data set version (project access only)
- Sharing and use of data within the project
- Upload final version
- Data publication (open access)

Where?

The KDMI provides a web-based **portal** solution including the Ocean Science Information System (OSIS) that allows scientists to upload and share their data files. It joins the SFB 754 project's public web site and its internal collaborative sites providing access to various information and databases. A **data policy** has been proposed and the SFB 754 board as well as the SFB 754 members agreed on its application. A new and now widely appreciated **service** has been implemented realizing monitoring and adherance to the data policy's time schedules. It allows setup and surveillance of **deliverables** based on expected or collected research data (see graphic).

Interlinking OSIS and data sets in SFB 754 publication list



PANGAEA map of SFB 754 data

What?

The data portal (OSIS) allows structured data upload and project internal **exchange** in the context of SFB 754 expeditions, numeric models and field or lab experiments. Metadata of what, when, where and who document the file's origin and are always public. OSIS provides an **overview** of available data and links them to related publications in the institutional repository OceanRep. Both OSIS and OceanRep **interlink** each other and other data centres in realtime.
According to the SFB 754 data policy data are **published** in a data centre within 3 years after the end of the expedition, mainly at the world data centre PANGAEA. More than 300 data sets are already published, listed on the **public website** via RSS feed.

Contact
datamanagement@geomar.de
http://portal.geomar.de

DFG Deutsche Forschungsgemeinschaft
GEOMAR
CAU Kiel University Christian-Albrechts-Universität zu Kiel