

# Integrating Data Management Services in Marine Sciences

GEOMAR



Lisa Paglialonga, Pina Springer, Hela Mehrtens, Carsten Schirnick & Claas Faber

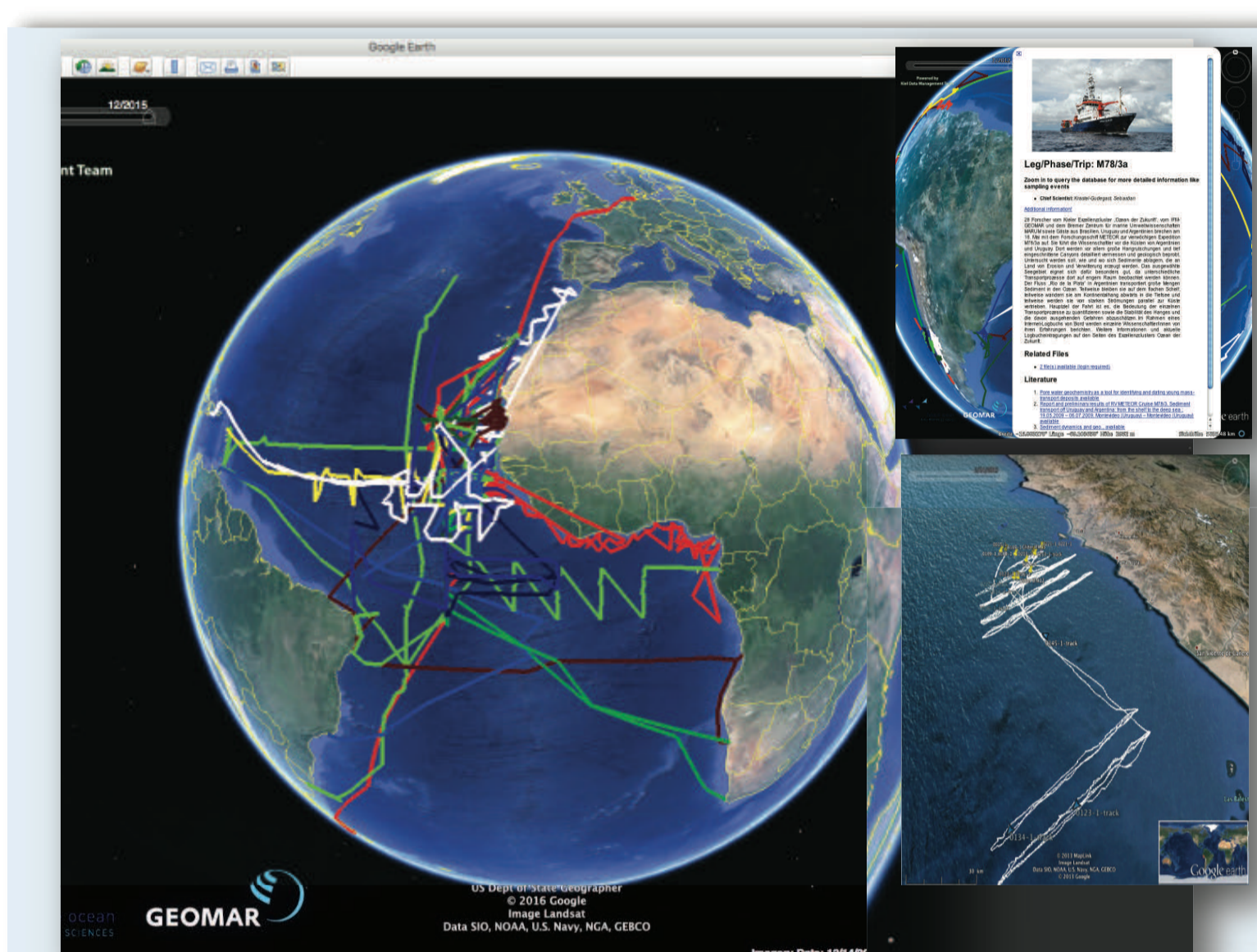
GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany – <https://portal.geomar.de> – [datamanagement@geomar.de](mailto:datamanagement@geomar.de)

## Organisation

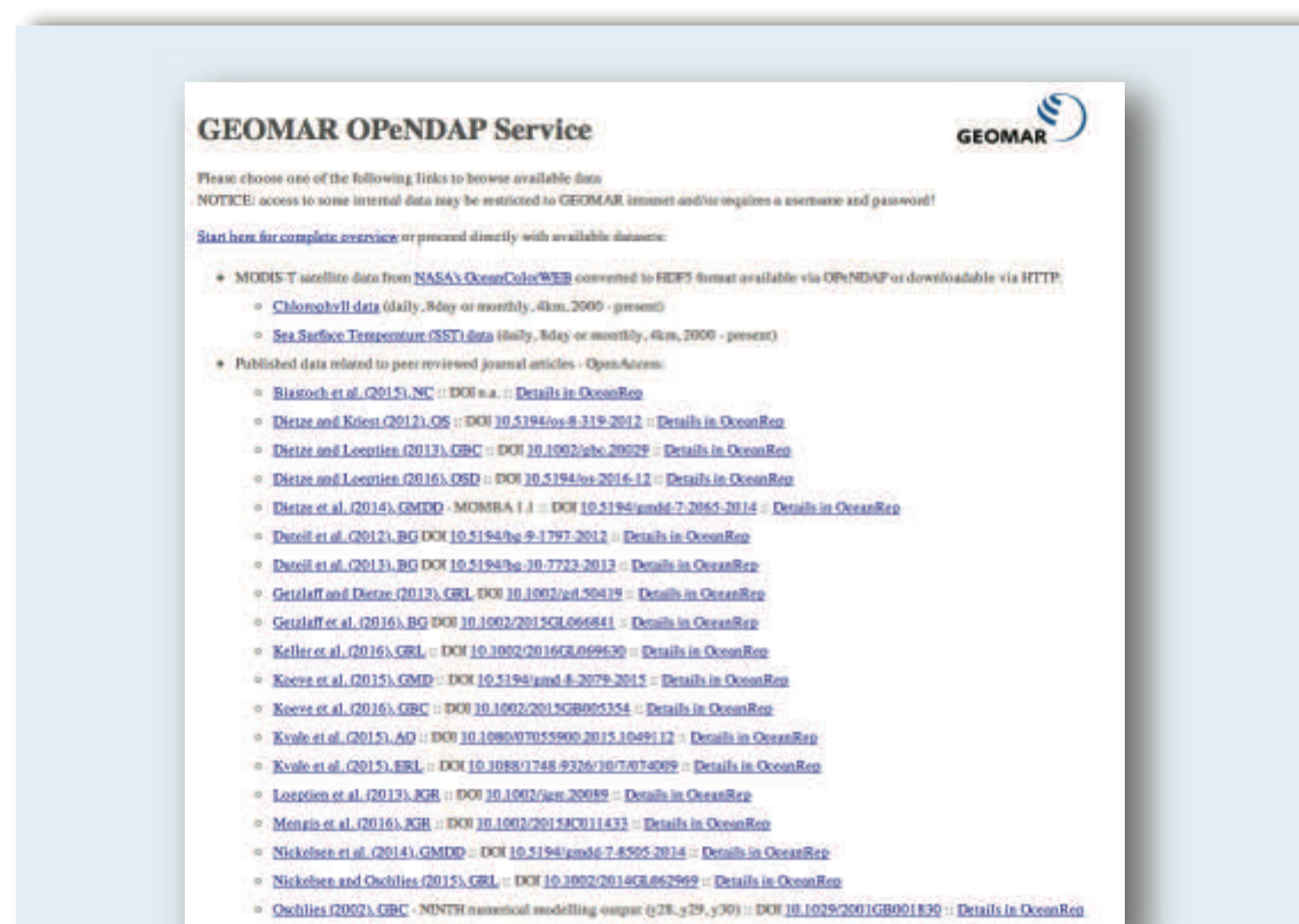
The GEOMAR Data Management Team is a cooperation of Helmholtz Centre for Ocean Research Kiel (GEOMAR) and large-scale research projects, collaborating with other marine research institutions. The coalition has established in Kiel a **single sustainable data management infrastructure (KDMI)** for marine research and supports the entire lifecycle of data description, data exchange, data storage and data archiving.

## Data Management Services

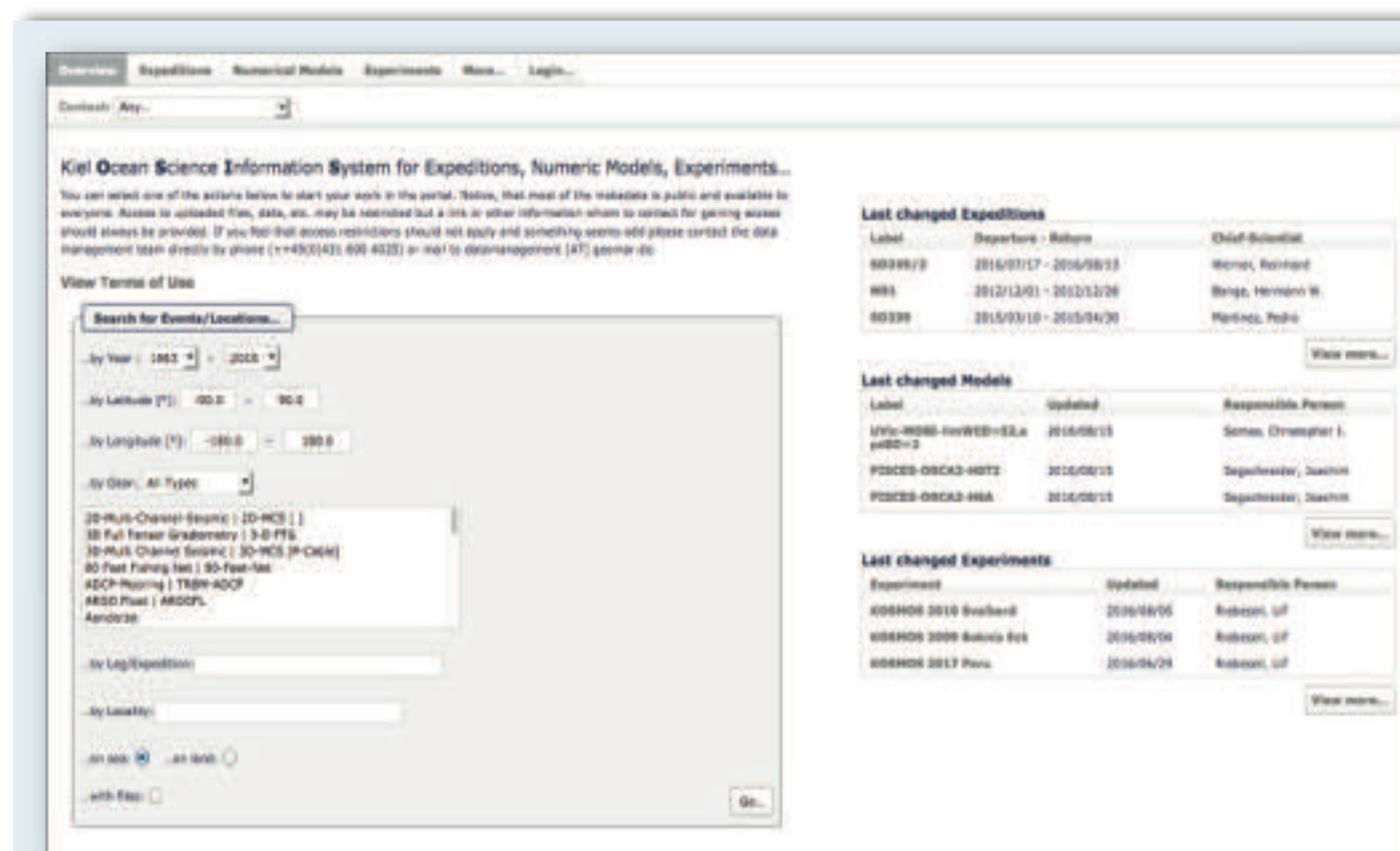
The **Data Management portal** is the starting point for all provided data services within KDMI ([portal.geomar.de](https://portal.geomar.de)). Projects and working groups maintain their public and internal webpages where collaboration tools like wiki, blog, calendar and document exchange are used. User administration for all data services is realised via the portal software (Liferay) and allows guarded **access to data services for internal and external users**. A data curation process has been implemented as service to realize monitoring and adherence to data policy's time schedules (see graphic). It allows setup and surveillance of **deliverables** based on expected or collected research data.



**Visualisation of expeditions via KDMI Geo-Search:** Search cruise tracks, publications, sampling locations, glider tracks, sample descriptions and detailed cruise information as georeferenced information in GoogleEarth. KML files contain network links in order to retrieve and display information on-the-fly.



**Model output** and related objects referenced in peer-reviewed papers are made publicly available via GEOMAR's OPeNDAP (THREDDS) Server ([data.geomar.de](https://data.geomar.de)). The data sets are interlinked in Oceanrep and OSIS.



**Ocean Science Information System (OSIS):** Central information hub where scientists upload and share data files in the context of expeditions, models and experiments. Various types of related links connect OSIS to the other services ([portal.geomar.de/osis](https://portal.geomar.de/osis)).

## Data curation process



Measurement/ Experiment

Description in OSIS

Data deliverables

Upload of first data set version (project access only)

6 months later  
Reminder "upload data"

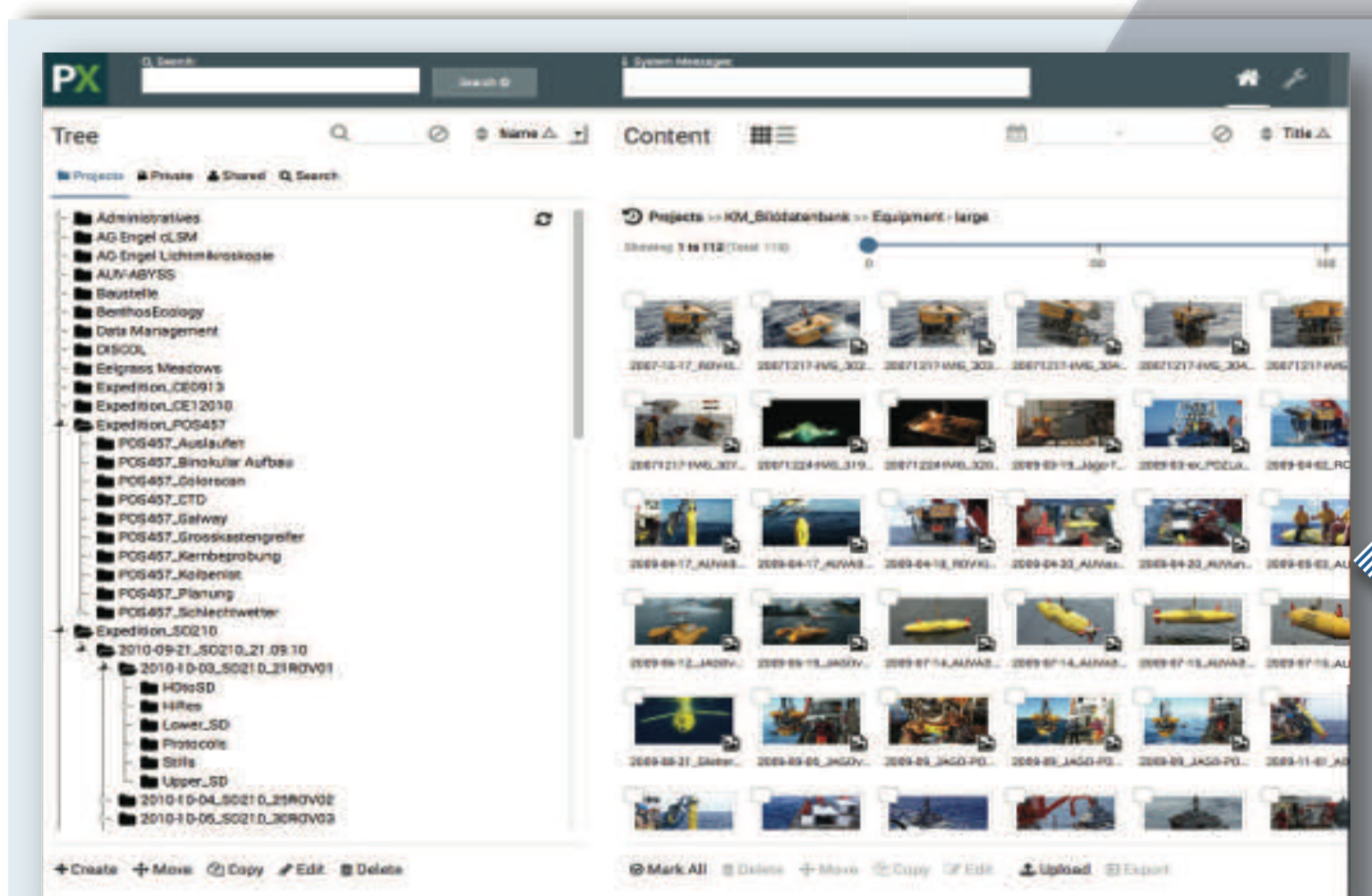
Sharing and use of data within the project

month 7 to month 36

Upload final version

3 years later  
Reminder "publish data"

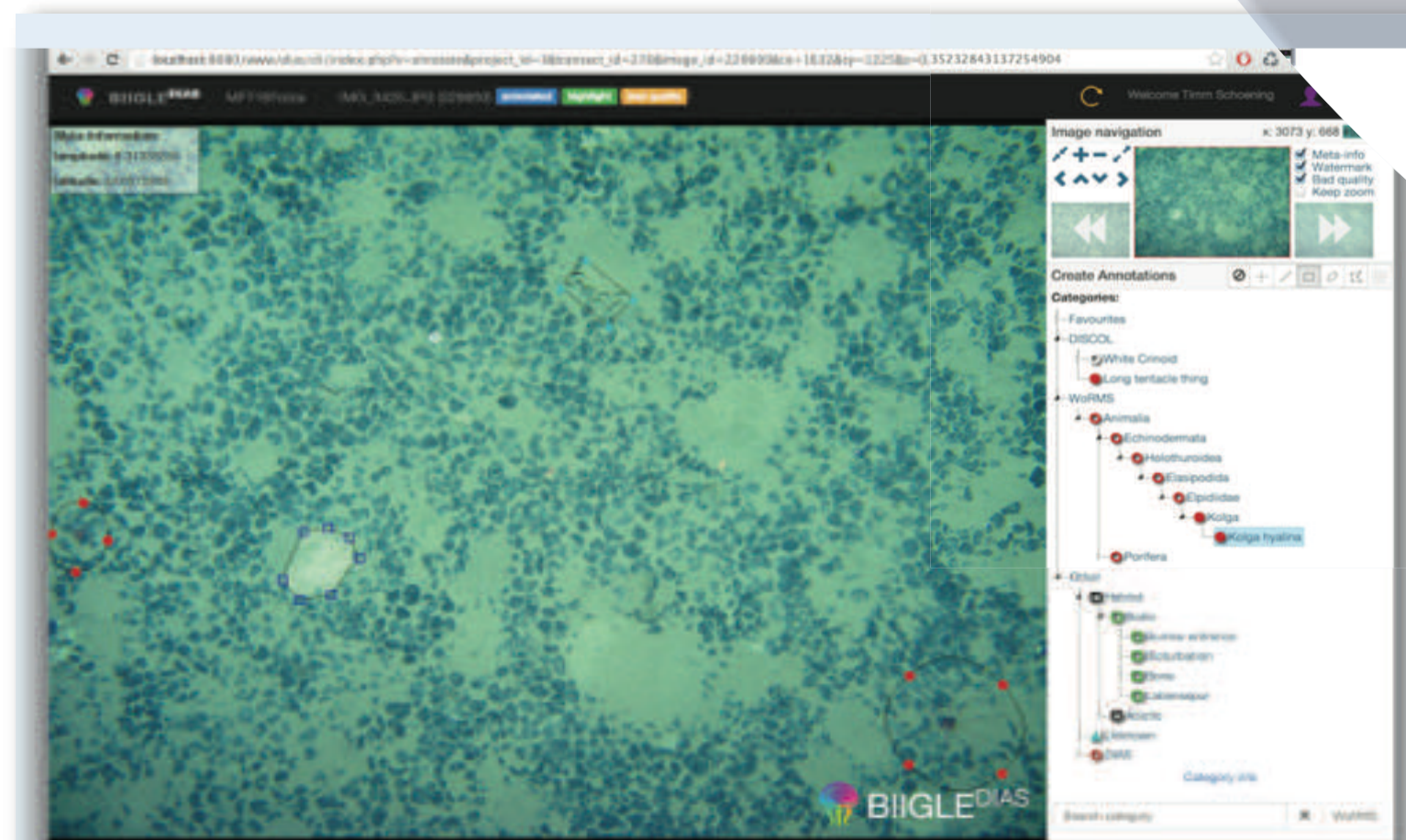
Data publication (open access)



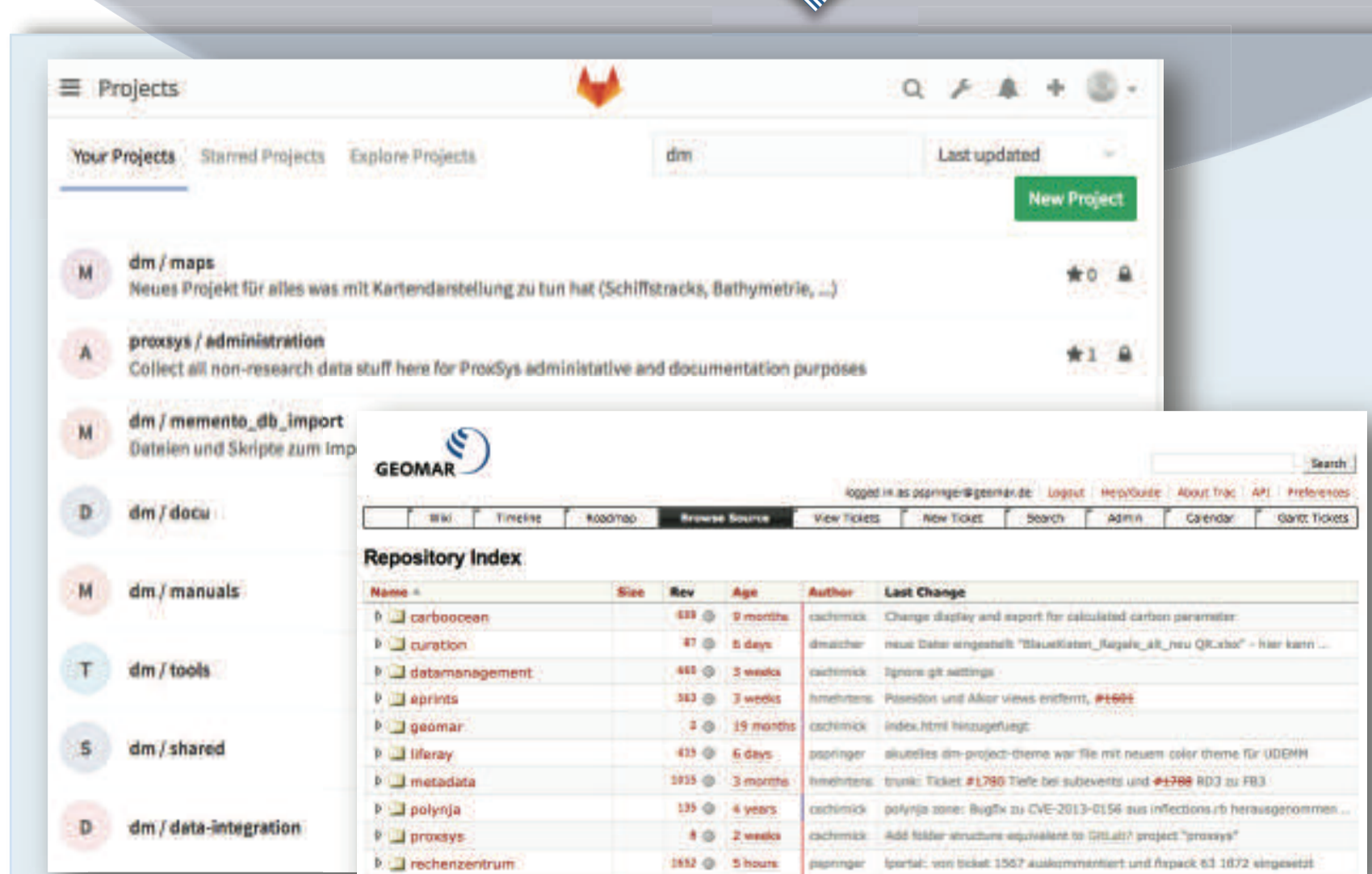
**Media Database (ProxSys):** Allows to store, describe and search images and videos e.g. acquired during research dives of remotely operated and autonomous underwater vehicles.



## Data Management Portal for Marine Sciences



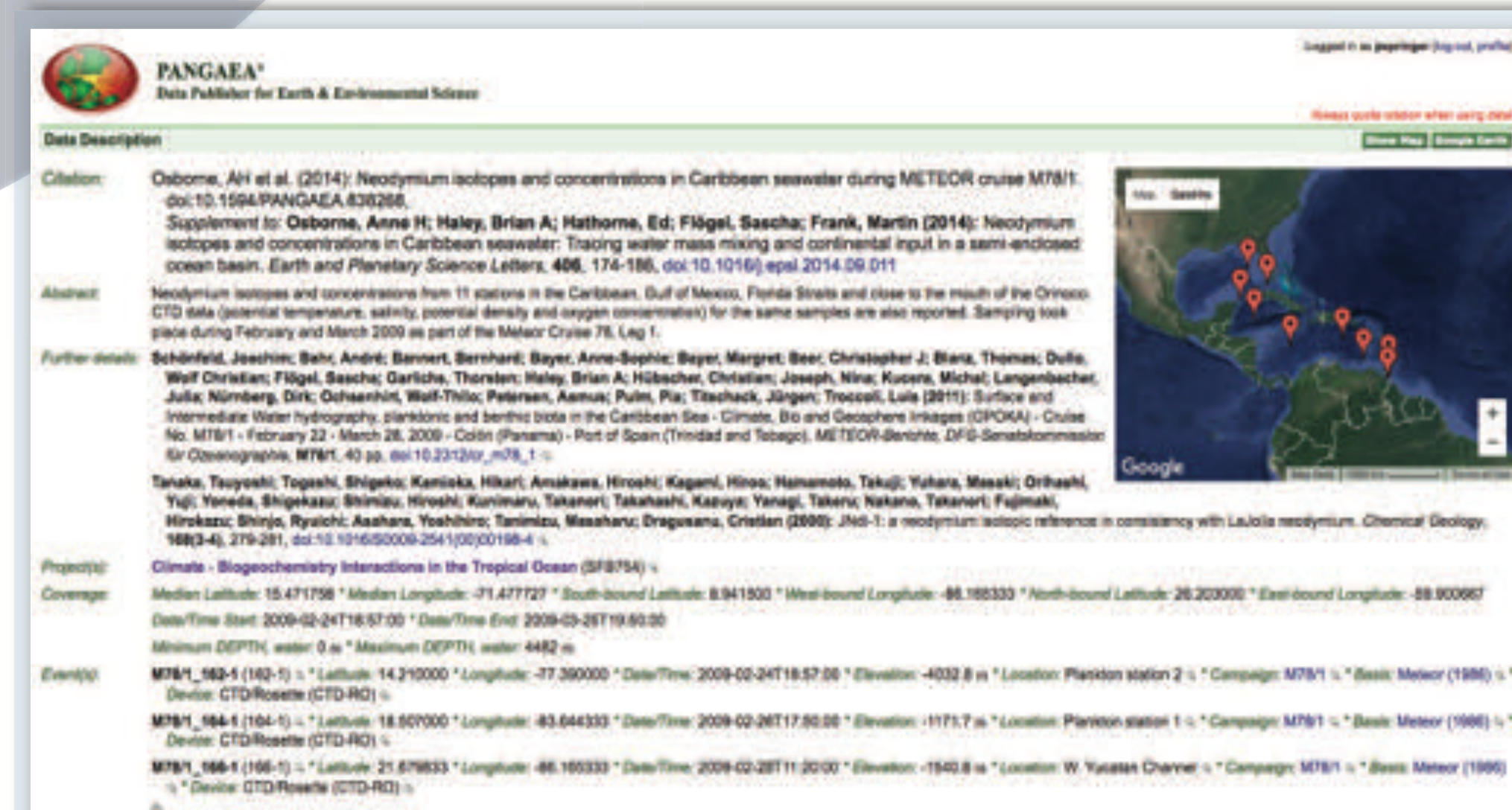
**DIAS (Image and video annotation system):** Uses webservice to access images and videos stored in the media database and provides them for all projects members. It avoids extensive network traffic, facilitates cooperation and reduces obsolete interchange.



**Version control systems (Subversion and Git):** Used for code development, documentation, data analysis and exchange. The TRAC Wiki and ticket system linked to Subversion and Git in the near future is used for documentation, improved collaboration and dissemination.



**OceanRep: An institutional repository for full text print publications** is maintained by the GEOMAR library and also used for specific publication lists, e.g. projects, numerical models or platforms ([oceanrep.geomar.de](https://oceanrep.geomar.de)). It is linked to OSIS allowing linkage between authors, their data and publications. An automated data link routine adds icons when additional external data are available.



**Data Publication in World Data Centres (e.g. PANGAEA):** The data management team assists as data curators when data are to be published in WDC's in order to warrant **long-term archival** and **access to the data** with persistent identifiers. This cooperation will make the data globally searchable while links to the data producers will ensure citability and provide contacts for the scientific community ([www.pangaea.de](https://www.pangaea.de)).