

# Interoperability of institutional data management systems

Hela Mehrstens), [hmehtens@geomar.de](mailto:hmehrstens@geomar.de)

Pina Springer, [pspringer@geomar.de](mailto:pspringer@geomar.de)

Carsten Schirnack, [cschirnack@geomar.de](mailto:cschirnack@geomar.de)

Claas Faber, [cfaber@geomar.de](mailto:cfaber@geomar.de)

Doris Maicher, [dmaicher@geomar.de](mailto:dmaicher@geomar.de)

Lisa Paglialonga, [lpaglialonga@geomar.de](mailto:lpaglialonga@geomar.de)

Sören Lorenz), [slorenz@geomar.de](mailto:slorenz@geomar.de)

All authors: , *GEOMAR Helmholtz Centre for Ocean Research, Kiel (Germany)*

## Motivation

The knowledge about FAIR data requirements is rising. As an institutional data management team, we implement, develop and maintain infrastructures, tools and services at GEOMAR for more than a decade now. We have systems to make data findable, accessible and reusable but interoperability is the remaining challenge. Here we describe the infrastructure, challenges and next steps to come closer to this aim.

## The GEOMAR data management systems

In 2009 GEOMAR started to establish an institutional data management platform enabling collaborative data management in research projects by integration of several components (Fig. 1): (1) OSIS, (2) OceanRep, (3) Liferay portal, (4) THREDDS / OPeNDAP, (5) Gitlab. The challenge was to come up with data management solutions that were fast to implement, easy to use and highly accepted by users. (1) The Ocean Science Information System **OSIS** ([osis.geomar.de](http://osis.geomar.de)) is our metadata and data exchange system for research data, which was developed in-house. It serves both as the instrument to collect the necessary metadata and also as an information system, allowing the search and dissemination of metadata and data. For several projects it is used to collect data of expeditions, experiments and model experiments and to display and remind for deliverables such as to upload or publish data in a domain specific data centre according to the timeline of an agreed data policy. It serves as our central data information hub. (2) **OceanRep** is the institutional repository at GEOMAR, based on the EPrints software. It contains quality controlled entries of the research output at GEOMAR, links to data and software repositories and project information, shows collaboration with research partners and can be harvested via OAI-PMH. (3) We use the **portal** software Liferay as our central data management platform. It comes out of the box with a lot of useful collaboration features. It also regulates the individual or project based access for researchers to a series of data management applications. More than 100 separate sites serve researchers as internal workspaces for their project collaboration. It also offers public websites and our services can be integrated into these sites (project specific views of OSIS, publication lists from OceanRep). (4) An **OPeNDAP** server provides access to model data referenced in peer-reviewed papers or papers in discussion. OPeNDAP is an established protocol for sharing access to large gridded datasets and allows subsetting on the server side. We are currently using a THREDDS server to provide our OPeNDAP service. (5) Software is increasingly regarded as integral part of sustainable research and as a research output in its own right. To enable collaborative software development in research projects, we operate **Gitlab** as a service within GEOMAR and also for external project partners. Our Gitlab instance on premise not only offers revision control for software code but also collaboration tools (e.g. issue tracker), a container registry as well as development tools (e.g. Continuous Integration pipelines).

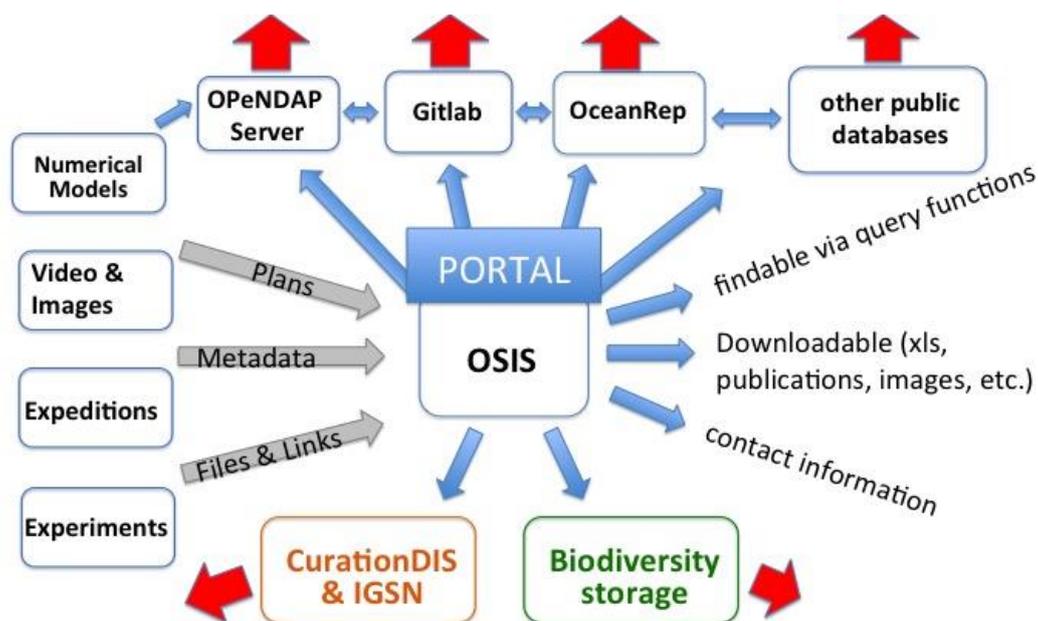


Figure 1: Relation of data and sample management structures

### Current challenges

To open our collaborative data management system to the community, we face organisational and technical challenges. The integration and interoperability of the components on the institutional level is established, as shown in Fig. 1. However, to reach interoperability for the marine community, locally established workflows and vocabularies have to be opened to discussion within the community and adapt to external requirements. Local operational constraints need a technical redesign. The institutional data management service portfolio has, at least partly, to be transformed into a community service portfolio. To achieve these efforts, we plan: (1) Machine readable provision of metadata has to be implemented on an early stage of data acquisition. (2) Design tools for researchers to allow them to provide accurate and relevant attributes for their data from the beginning. (3) Enhancing the use of persistent identifiers along the data and sample collection and managing cycle.

### Next steps raising interoperability

The given institutional implementations of data management infrastructures and services are diverse. They typically have discipline specific and technical specialisation that fit perfectly for the institution. This is true for example at GEOMAR for its established workflows and interoperable components (1. to 5. described above) which foster collaboration and exchange of the most diverse expertise. This strength is counteracted by a weak interoperability with research data management (RDM) environments of other institutional or community-specific systems. On the long run, and to preserve productive diversity, only a federated community infrastructure will meet the requirement of interoperability. This requires trusting and trusted services by the providing institution and a negotiation among the community on a federated RDM service portfolio. GEOMAR aims at providing a collaborative RDM platform for marine projects as a community service. How to achieve this on the operational level is currently developed jointly with other marine research institutions in community projects, e.g. within the DAM (German Alliance of Marine Research) and MareHub, the marine-oriented initiative within the Helmholtz research area Earth and Environment which is reaching out to other disciplines, for example atmosphere and earth and even beyond. We will present our ideas on how to meet the challenges and a roadmap towards them.