



## **Data portal based on Open Archives Initiative Protocols and Apache Lucene**

**U. Schindler** (1), M. Diepenbroek (1)

(1) MARUM, University of Bremen, Leobener Str., D-28359 Bremen

The World Data Center for Marine and Environmental Sciences (WDC-MARE) with its information system PANGAEA ([www.pangaea.de](http://www.pangaea.de)) provides data portals for several EU projects (EUR-OCEANS, CARBOOCEAN) to disseminate data and metadata for international data networks. We present a generic portal system architecture suitable for geoscientific data portals. The portals harvest data providers with Open Archives Initiative (OAI) protocols using metadata in DIF or ISO-19139 format. Current implementations of OAI only support Dublin Core metadata, the new Java based portal software will support any XML format and makes them searchable through Apache Lucene without any other database software. The open architecture makes it possible to define searchable fields in several data formats by XPath allowing not only full text queries, even ranges are retrievable. The metadata of all providers are stored in separate indices which makes it possible to combine them in several different portals. The web service interface allows to support custom front-ends for users and additional visualization in maps. The software will be made freely available through the open source concept.