Geophysical Research Abstracts
Vol. 15, EGU2013-8738-1, 2013
EGU General Assembly 2013

MaNIDA: Integration of marine expedition information, data and publications: Data Portal of German Marine Research

Roland Koppe and Scientific MaNIDA-Team
Alfred Wegener Institute, Computing and Data Centre, Germany (roland.koppe@awi.de)

The Marine Network for Integrated Data Access (MaNIDA) aims to build a sustainable e-infrastructure to support discovery and re-use of marine data from distinct data providers in Germany (see related abstracts in session ESSI 1.2). In order to provide users integrated access and retrieval of expedition or cruise metadata, data, services and publications as well as relationships among the various objects, we are developing (web) applications based on state of the art technologies: the Data Portal of German Marine Research.

Since the German network of distributed content providers have distinct objectives and mandates for storing digital objects (e.g. long-term data preservation, near real time data, publication repositories), we have to cope with heterogeneous metadata in terms of syntax and semantic, data types and formats as well as access solutions. We have defined a set of core metadata elements which are common to our content providers and therefore useful for discovery and building relationships among objects. Existing catalogues for various types of vocabularies are being used to assure the mapping to community-wide used terms.

We distinguish between expedition metadata and continuously harvestable metadata objects from distinct data providers.
• Existing expedition metadata from distinct sources is integrated and validated in order to create an expedition metadata catalogue which is used as authoritative source for expedition-related content. The web application allows browsing by e.g. research vessel and date, exploring expeditions and research gaps by tracklines and viewing expedition details (begin/end, ports, platforms, chief scientists, events, etc.). Also expedition-related objects from harvesting are dynamically associated with expedition information and presented to the user. Hence we will provide web services to detailed expedition information.
• Other harvestable content is separated into four categories: archived data and data products, near real time data, publications and reports. Reports are a special case of publication, describing cruise planning, cruise reports or popular reports on expeditions and are orthogonal to e.g. peer-reviewed articles. Each object's metadata contains at least: identifier(s) e.g. doi/hdl, title, author(s), date, expedition(s), platform(s) e.g. research vessel Polarstern. Furthermore project(s), parameter(s), device(s) and e.g. geographic coverage are of interest. An international gazetteer resolves geographic coverage to region names and annotates to object metadata. Information is homogenously presented to the user, independent of the underlying format, but adaptable to specific disciplines e.g. bathymetry. Also data access and dissemination information is available to the user as data download link or web services (e.g. WFS, WMS). Based on relationship metadata we are dynamically building graphs of objects to support the user in finding possible relevant associated objects. Technically metadata is based on ISO / OGC standards or provider specification. Metadata is harvested via OAI-PMH or OGC CSW and indexed with Apache Lucene. This enables powerful full-text search, geographic and temporal search as well as faceting.

In this presentation we will illustrate the architecture and the current implementation of our integrated approach.