



Rolling Deck to Repository: Opportunities for US-EU Collaboration

Stephen Miller (1), Helen Graves (2), Suzanne Carbotte (3), Robert Arko (3), Cynthia Chandler (4), Dru Clark (1), Aaron Sweeney (1), and Karen Stocks (5)

(1) University of California, San Diego, Scripps Institution of Oceanography, La Jolla, California, United States (spmiller@ucsd.edu), (2) British Geological Survey, Nicker Hill, Keyworth, Nottingham, NG12 5GG, UK, (3) Lamont-Doherty Earth Observatory, Columbia University, Paliades NY 10964-1000, USA, (4) Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA, (5) San Diego Supercomputer Center, UCSD, La Jolla CA 92093-0505, USA

Advances in marine science today are frequently driven by international teams, as they address basin-wide, regional or global problems with multi-disciplinary data aggregations, rather than the more traditional single investigator, single vessel or national approach. In order to serve a broader community of researchers, we are seeking to identify a number of practical steps that can be taken to promote data exchange, encourage collaboration and achieve interoperability. The steps involve the sharing of controlled vocabularies, metadata profiles for cruises and for data objects, and vessel instrument configuration profiles, as well as technology for data harvesting, archiving and dissemination.

A number of separate initiatives in the US, Europe and other regions are currently providing improved access to marine data. We will highlight some of the opportunities for collaboration and interoperability based on recent experiences in the US with the new “Rolling Deck to Repository (R2R)” program, and seek to encourage further discussion.

Launched in September 2009, R2R is a systematic effort to capture, catalog and archive US underway shipboard data. Data are now flowing from 18 ship-operating institutions in University-National Oceanographic Laboratory System (UNOLS) and other contributors. As of January 2011, data from 2,056 cruises on 26 vessels had been submitted, totaling 6,903,782 files (7 TB).

Shipboard data are legendary in their complexity, with a wide range of directory structures, naming conventions and formats across vessels and operators. R2R has developed automated procedures, and testing is well underway to identify, extract and catalog 26 different standard device types, such as multibeam, XBT, ADCP, gravity, magnetics and meteorology. For long-term public access, the selected data sets are transported to US national repositories at the National Oceanic and Atmospheric Administration (NOAA) National Geophysical Data Center (NGDC) and National Oceanographic Data Center (NODC). As the R2R catalog for each cruise is processed it will provide persistent links to archived data. R2R will also create ISO-standard cruise- and dataset-level metadata, based on associated controlled vocabularies, as developed in cooperation with NOAA. The controlled vocabularies and the catalog content will be made available as they are developed via Web services, including cruise- and dataset-level metadata and trackline control points.

R2R creates a standard navigation product for each cruise, with thorough quality control. R2R will manage versions of vessel profiles to track instruments on each vessel, including type, manufacturer, model and shipboard location. Quality assessment is under development, and feedback will be provided to operators. A shipboard event-logger system is being tested, based on ELOG.

The sharing of concerns, expertise, knowledge, resources and technologies will directly benefit a number of current data aggregation activities. International scientific and policy end users may benefit from a broader and deeper search experience as these activities achieve greater alignment. An international group of data managers, programmers, researchers and program officers have been discussing these issues in the last year, laying the groundwork for actual progress in common approaches. The next steps in this collaboration will require a modest investment in networking and development activities. The R2R approaches are intended to be coordinated with marine data initiatives across the US National Science Foundation (NSF) and NOAA, and with European Commission funded projects such as SeaDataNet, Eurofleets, GeoSeas and EMODNET.

The R2R collaborative team is made up of researchers, computer scientists, data managers and programmers. The lead institution is the Lamont-Doherty Earth Observatory, with development and implementation activities at the Scripps Institution of Oceanography, the San Diego Supercomputer Center, the Woods Hole Oceanographic Institution and the Florida State University. Support is provided by the NSF Oceanographic Technical Services program. For further information see <http://rvdata.us>