Geophysical Research Abstracts
Vol. 14, EGU2012-13062-1, 2012
EGU General Assembly 2012

Rolling Deck to Repository (R2R): A "Linked Data" Approach for the U.S. Academic Research Fleet

R. Arko (1), C. Chandler (2), P. Clark (3), A. Milan (4), and J. Mize (5)

(1) Lamont-Doherty Earth Observatory, Palisades, NY USA (arko@ldeo.columbia.edu), (2) Woods Hole Oceanographic Institution, Woods Hole, MA USA, (3) Scripps Institution of Oceanography, La Jolla, CA USA, (4) National Geophysical Data Center, Boulder, CO USA, (5) National Oceanographic Data Center, Stennis, MS USA

The Rolling Deck to Repository (R2R; http://rvdata.us/) program is developing infrastructure to routinely document, assess, and preserve the underway sensor data from U.S. academic research vessels. The R2R master catalog of vessels, instrument systems, operating institutions, cruises, personnel, data sets, event logs, and field reports has grown to over 2,200 cruises in less than two years, and is now accessible via Web services. This catalog is of great value to peer data systems, ranging from large inter/national data centers to small disciplinary data offices, as an aid in quality controlling their own collections and finding related data from authoritative sources.

R2R breaks with the tradition of stovepipe portals built around complex search interfaces tightly bound to backend databases. Instead, we have adopted a Linked Data approach to publish our catalog content, based on the W3C Resource Description Framework (RDF) and Uniform Resource Identifiers (URIs). Our data model is published as a collection of W3C Simple Knowledge Organization System (SKOS) concepts, mapped to partner vocabularies such as those developed by the Global Change Master Directory (GCMD) and the pan-European SeaDataNet partnership, and our catalog content is published as collections of RDF resources with globally unique and persistent identifiers. The combination of exposing our data model, mapping local terms to community-wide vocabularies, and using reliable identifiers improves interoperability and reduces ambiguity. R2R's metric of success is the degree to which peer data systems harvest and reuse our content.

R2R is working collaboratively with the NOAA National Data Centers and the NSF-funded Biological and Chemical Oceanography Data Management Office (BCO-DMO) on a range of Linked Data pilot applications, including production of ISO-compliant metadata and deployment of a RDF Query Language (SPARQL) interface. Our objective is to support a distributed, loosely federated network of complementary systems that collectively manage the vast body of ocean science data. We will present results and lessons learned.