Geophysical Research Abstracts
Vol. 16, EGU2014-2101-1, 2014
EGU General Assembly 2014

Linked Ocean Data

Adam Leadbetter (1), Robert Arko (2), Cynthia Chandler (3), and Adam Shepherd (3)

(1) British Oceanographic Data Centre, Liverpool, United Kingdom, (2) Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York, United States of America, (3) Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, United States of America

"Linked Data" is a term used in Computer Science to encapsulate a methodology for publishing data and metadata in a structured format so that links may be created and exploited between objects. Berners-Lee (2006) outlines the following four design principles of a Linked Data system:

- Use Uniform Resource Identifiers (URIs) as names for things.
- Use HyperText Transfer Protocol (HTTP) URIs so that people can look up those names.
- When someone looks up a URI, provide useful information, using the standards (Resource Description Framework [RDF] and the RDF query language [SPARQL]).
- Include links to other URIs so that they can discover more things.

In 2010, Berners-Lee revisited his original design plan for Linked Data to encourage data owners along a path to "good Linked Data". This revision involved the creation of a five star rating system for Linked Data outlined below.

- One star: Available on the web (in any format).
- Two stars: Available as machine-readable structured data (e.g. An Excel spreadsheet instead of an image scan of a table).
- Three stars: As two stars plus the use of a non-proprietary format (e.g. Comma Separated Values instead of Excel).
- Four stars: As three stars plus the use of open standards from the World Wide Web Commission (W3C) (i.e. RDF and SPARQL) to identify things, so that people can point to your data and metadata.
- Five stars: All the above plus link your data to other people's data to provide context

Here we present work building on the SeaDataNet common vocabularies served by the NERC Vocabulary Server, connecting projects such as the Rolling Deck to Repository (R2R) and the Biological and Chemical Oceanography Data Management Office (BCO-DMO) and other vocabularies such as the Marine Metadata Interoperability Ontology Register and Repository and the NASA Global Change Master Directory to create a Linked Ocean Data cloud. Publishing the vocabularies and metadata in standard RDF XML and exposing SPARQL endpoints renders them five-star Linked Data repositories. The benefits of this approach include:

- increased interoperability between the metadata created by projects;
- improved data discovery as users of SeaDataNet, R2R and BCO-DMO terms can find data using labels with which they are familiar
- both standard tools and newly developed custom tools may be used to explore the data; and using standards means the custom tools are easier to develop

Linked Data is a concept which has been in existence for nearly a decade, and has a simple set of formal best practices associated with it. Linked Data is increasingly being seen as a driver of the next generation of "community science" activities. While many data providers in the oceanographic domain may be unaware of Linked Data, they may also be providing it at one of its lower levels. Here we have shown that it is possible to deliver the highest standard of Linked Oceanographic Data, and some of the benefits of the approach.