Oceans of Linked Data?

Adam Leadbetter (alead@bodc.ac.uk)
British Oceanographic Data Centre, Joseph Proudman Building, Liverpool. L3 5DA

Introduction

The highest quality (Figure 1) published Linked Data has traditionally been represented by the Linking Open Data (LOD) cloud diagram (Figure 2). This diagram gives a high level overview of Linked Data datasets from around the globe, and it often features in presentations and documents on how data published on the World Wide Web can benefit society, especially if those data are linked together. However, at the time of writing, the Linked Open Data Cloud diagram appears to be no longer maintained. In a recent blog post, "The LOD cloud is dead, long live the trusted LOD cloud"[1], Andreas Blumauer of the Semantic Web Company proposed the idea of domain specific, micro-Linked Data clouds. These micro-clouds should consist of resources used again and again by their specific domains due to the specific data or information presented and their highly active maintainers.
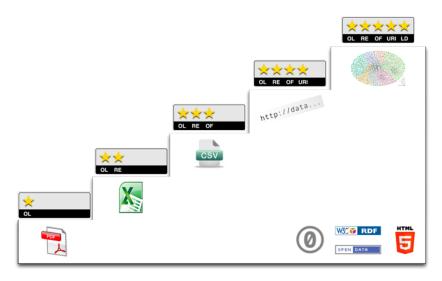

Figure 1 The five star deployment scheme for Linked Open Data (http://5stardata.info/)

There are several existing 5 Star Linked Data resources published in the oceanographic domain. Two resources used to publish definitions of terms relevant to the marine domain are Version 2.0 of the Natural Environment Research Council (NERC) Vocabulary Server and

---

[1] http://blog.semantic-web.at/2013/06/07/the-lod-cloud-is-dead-long-live-the-trusted-lod-cloud/
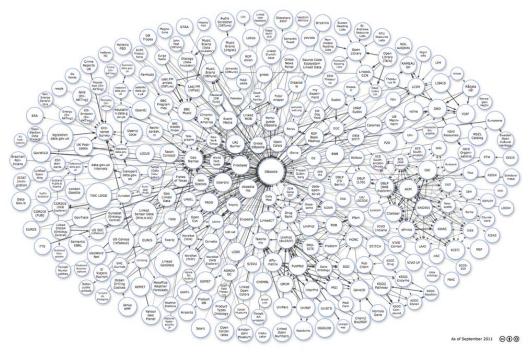
Figure 2 The familiar Linked Open Data cloud diagram

the Marine Metadata Interoperability Ontology Registry and Repository (MMI-ORR). The NERC Vocabulary Server[2] delivers content governed by a number of groups, including SeaDataNet, and a number of de facto standard vocabularies such as the RDF representation of the International Council for the Exploration of the Seas' Platform Codes. The MMI-ORR[3] serves a number of vocabularies created by community and individual efforts, including those used within the United States Integrated Ocean Observing System and the Ocean Observatories Initiative. URIs from these resources are used to define parameters in data files in the Linked Ocean Data Cloud, and such metadata fields as: the sea area in which an observation was made; the data resources; the vessel or platform a measurement was made from; and the instrument used to make an oceanographic measurement. Many of the URIs on these two resources are interlinked, and are used within the 5 Star Linked Data datasets published by the United States National Science Foundation funded projects Rolling Deck to Repository (R2R) and Biological and Chemical Oceanography Data Management Office (BCO-DMO).

---

[2] http://vocab.nerc.ac.uk/
[3] http://mmisw.org/orr/

In the case of BCO-DMO[4] and R2R[5], the two repositories manage related oceanographic data resources, such as research cruises and their datasets, but from different perspectives with different goals. However, these two repositories, by linking their metadata to a common semantic resource such as the NERC Vocabulary Server, can now discover the other repository's related data through links to this common resource. Beyond the obvious advantage of increased resource discovery, this capability has made data management practices such as data validation, much easier to accomplish. For instance, using the SPARQL language, the two repositories can efficiently match their related cruise metadata to the other repository's cruise metadata through their common link to the NERC Vocabulary Server. Not only will a SPARQL query derive matches, but it can also be used to assess the accuracy of shared common metadata values. From these derived matches and assessments, data managers can quickly quality control their own cruise metadata based on the results from the other repository. This work has lead to the creation of the Linked Ocean Data concept[6] (Figure 3).
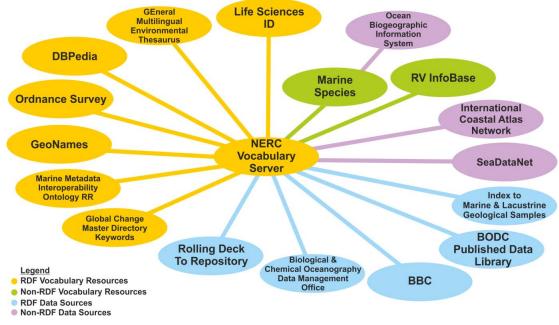

Figure 3 The Linked Ocean Data network

Geo-Spatial Issues

Data collected within the oceanographic realm are by their very nature geospatial: latitude, longitude and depth within the ocean provide vital contextual information allowing the data

to be interpreted correctly and to be assimilated with other data collected in similar oceanographic regions. While we have explored the possibilities of non-geospatial Linked Data, we have not yet progressed to the point of exposing geospatial data in this way. However, the boundaries of sea areas; the trajectories of research vessels and autonomous vehicles and the positions of sampling stations are all (meta)data objects for which we hold geospatial data. Through attending the W3C / OGC workshop we wish to see how the technologies for publishing Linked Geospatial Data are progressing, and learn best practice before continuing down this route. We also wish to meet with other open data providers with whom we can collaborate to produce new information products, linking data from land to sea and bridging the coastal gap using standard technologies.