



Adding value to semantics through vocabulary extension & mapping

Adam Leadbetter (1), Roy Lowry (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

Over the past thirty years there has been a development of formally controlled vocabularies within the marine sciences domain: from published paper based records through to Semantic Web resources published online (Lowry, 2008). Recently it has been recognised that there has been an explosion in the number of semantic resources available on the internet, and that this is not a sustainable situation (e.g. McGuinness, 2012). In order to both preserve the terms used locally by data managers and to provide better interoperability and sustainability, mappings have been created to a de facto standard semantic resource, in this instance the content of the NERC Vocabulary Server (NVS – available at <http://vocab.nerc.ac.uk/>).

Following the principles of Linked Open Data the vocabularies use dereferencable Uniform Resource Identifiers for their resources which are published in XML using World Wide Web Consortium standards (Resource Description Framework and Simple Knowledge Organisation System). Further, work has been undertaken to map concepts served from the NVS to other semantic resources, again following the Linked Open Data model. Finally, the recent exposure of a SPARQL 1.1 endpoint allows the content to be queried in a standard way, including providing results to a federated (or distributed) query.

Here we demonstrate the results of this work and show:

- How mapping between controlled vocabularies can improve data discovery (for example between the Climate and Forecast Standard Names and the GEneral Multilingual Environmental Thesaurus)
- How mapping disparate resources to the same vocabulary aids interoperability (for example between the Rolling Deck to Repository and Biological Chemical Oceanography Data Management Office projects)
- The use of both standard tools to explore vocabulary contents and custom tools to generate mappings between vocabularies
- Creating impact by mapping dataset discovery terms to media outlets on the Linking Open Data cloud (e.g. television documentaries)

In the European Union, this work has been funded through the Framework Programme 7 projects Open Service Network for Marine Environmental Data and SeaDataNet-II and in the United States by the National Science Foundation funded projects Biological Chemical Oceanography Data Management Office and Rolling Deck to Repository.

References

Lowry, Roy (2008). *25 Years of Controlled Vocabularies in Oceanographic Data Management*. Abstract presented at the American Geophysical Union Fall Meeting, San Francisco, 15th-19th December.

McGuinness, Deborah (2012). *Community Science – The Next Frontier*. Abstract presented at the American Geophysical Union Fall Meeting, San Francisco, 3rd-7th December.