Semantically supporting data discovery, markup and aggregation in the European Marine Observation and Data Network (EMODnet)

Conference Paper · April 2014

CITATIONS
3

READS
82

2 authors:

Roy Kenneth Lowry
National Oceanography Centre
85 PUBLICATIONS **717** CITATIONS

SEE PROFILE

Adam Marc Leadbetter
Marine Institute
70 PUBLICATIONS **369** CITATIONS

SEE PROFILE

Geophysical Research Abstracts
Vol. 16, EGU2014-13693, 2014
EGU General Assembly 2014

Semantically supporting data discovery, markup and aggregation in the European Marine Observation and Data Network (EMODnet)

Roy Lowry and Adam Leadbetter
British Oceanographic Data Centre, Liverpool, United Kingdom (alead@bodc.ac.uk)

The semantic content of the NERC Vocabulary Server (NVS) has been developed over thirty years. It has been used to mark up metadata and data in a wide range of international projects, including the European Commission (EC) Framework Programme 7 projects SeaDataNet and The Open Service Network for Marine Environmental Data (NETMAR). Within the United States, the National Science Foundation projects Rolling Deck to Repository and Biological & Chemical Data Management Office (BCO-DMO) use concepts from NVS for markup. Further, typed relationships between NVS concepts and terms served by the Marine Metadata Interoperability Ontology Registry and Repository.

The vast majority of the concepts publicly served from NVS (35% of ∼82,000) form the British Oceanographic Data Centre (BODC) Parameter Usage Vocabulary (PUV). The PUV is instantiated on the NVS as a SKOS concept collection. These terms are used to describe the individual channels in data and metadata served by, for example, BODC, SeaDataNet and BCO-DMO. The PUV terms are designed to be very precise and may contain a high level of detail. Some users have reported that the PUV is difficult to navigate due to its size and complexity (a problem CSIRO have begun to address by deploying a SISSVoc interface to the NVS), and it has been difficult to aggregate data as multiple PUV terms can – with full validity – be used to describe the same data channels.

Better approaches to data aggregation are required as a use case for the PUV from the EC European Marine Observation and Data Network (EMODnet) Chemistry project. One solution, proposed and demonstrated during the course of the NETMAR project, is to build new SKOS concept collections which formalise the desired aggregations for given applications, and uses typed relationships to state which PUV concepts contribute to a specific aggregation. Development of these new collections requires input from a group of experts in the application domain who can decide which PUV concepts it is acceptable to aggregate for a given application.

Another approach, which has been developed as a use case for concept and data discovery and will be implemented as part of the EC/United States/Australian collaboration the Ocean Data Interoperability Platform, is to expose the well defined, but little publicised, semantic model which underpins each and every concept within the PUV. This will be done in a machine readable form, so that tools can be built to aggregate data and concepts by, for example, the measured parameter; the environmental sphere or compartment of the sampling; and the methodology of the analysis of the parameter. There is interesting work being developed by CSIRO which may be used in this approach.

The importance of these data aggregations is growing as more data providers use terms from semantic resources to describe their data, and allows for aggregating data from numerous sources. This importance will grow as data become “born semantic”, i.e. when semantics are embedded with data from the point of collection.

In this presentation we introduce a brief history of the development of the PUV; the use cases for data aggregation and discovery outlined above; and the semantic model from which the PUV is built; and the ideas for embedding semantics in data from the point of collection.