Geophysical Research Abstracts
Vol. 18, EGU2016-12704, 2016
EGU General Assembly 2016

Federated provenance of oceanographic research cruises: from metadata to data

Rob Thomas (1), Adam Leadbetter (2), and Adam Shepherd (3)

(1) British Oceanographic Data Centre, National Oceanography Centre, Liverpool, United Kingdom (room@bodc.ac.uk), (2) Marine Institute, Ocean Science and Information Services, Oranmore, Co. Galway, Ireland (Adam.Leadbetter@Marine.ie), (3) Woods Hole Oceanographic Institution, Computer and information Services, Woods Hole, USA (ashepherd@whoi.edu)

The World Wide Web Consortium's Provenance Data Model and associated Semantic Web ontology (PROV-O) have created much interest in the Earth and Space Science Informatics community (Ma et al., 2014). Indeed, PROV-O has recently been posited as an upper ontology for the alignment of various data models (Cox, 2015). Similarly, PROV-O has been used as the building blocks of a data release lifecycle ontology (Leadbetter & Buck, 2015).

In this presentation we show that the alignment between different local data descriptions of an oceanographic research cruise can be achieved through alignment with PROV-O and that descriptions of the funding bodies, organisations and researchers involved in a cruise and its associated data release lifecycle can be modelled within a PROV-O based environment.

We show that, at a first-order, this approach is scalable by presenting results from three endpoints (the Biological and Chemical Oceanography Data Management Office at Woods Hole Oceanographic Institution, USA; the British Oceanographic Data Centre at the National Oceanography Centre, UK; and the Marine Institute, Ireland). Current advances in ontology engineering, provide pathways to resolving reasoning issues from varying perspectives on implementing PROV-O. This includes the use of the Information Object design pattern where such edge cases as research cruise scheduling efforts are considered. PROV-O describes only things which have happened, but the Information Object design pattern allows for the description of planned research cruises through its statement that the local data description is not the the entity itself (in this case the planned research cruise) and therefore the local data description itself can be described using the PROV-O model. In particular, we present the use of the data lifecycle ontology to show the connection between research cruise activities and their associated datasets, and the publication of those data sets online with Digital Object Identifiers and more formally in data journals. Use of the SPARQL 1.1 standard allows queries to be federated across these endpoints to create a distributed network of provenance documents.

Future research directions will add further nodes to the federated network of oceanographic research cruise provenance to determine the true scalability of this approach, and will involve analysis of and possible evolution of the data release lifecycle ontology.

References

Nitin Arora et al., 2006. Information object design pattern for modeling domain specific knowledge. 1st ECOOP Workshop on Domain-Specific Program Development.

Simon Cox, 2015. Pitfalls in alignment of observation models resolved using PROV as an upper ontology. Abstract IN33F-07 presented at the American Geophysical Union Fall Meeting, 14-18 December, San Francisco.

Adam Leadbetter & Justin Buck, 2015. Where did my data layer come from?" The semantics of data release. Geophysical Research Abstracts 17, EGU2015-3746-1.

Xiaogang Ma et al., 2014. Ontology engineering in provenance enablement for the National Climate Assessment. Environmental Modelling & Software 61, 191-205. http://dx.doi.org/10.1016/j.envsoft.2014.08.002