EXPERIENCES OF A "SEMANTICS SMACKDOWN"

Adam M. Leadbettera,\*

Adam Shepherdb

Robert Arkoc

Cynthia Chandlerb

Yanning Chend

Nkemdirim Dockerye

Renata Ferreiraf

Linyun Fud

Robert Thomasg

Patrick Westd

Stephan Zednikd

1. Marine Institute, Rinville, Oranmore, Galway, Ireland
2. Woods Hole Oceanographic Institution, Woods Hole, MA, United States of America
3. Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, United States of America
4. Tetherless World Constellation, Rensselaer Polytechnic Institute, Troy, NY, United States of America
5. Center for Ocean-Atmospheric Prediction Studies, Florida State University, Tallahassee, FL, United States of America
6. Scripps Institution of Oceanography, University of California, San Diego, CA, United States of America
7. British Oceanographic Data Centre, National Oceanography Centre, Liverpool, United Kingdom

\* Corresponding author: e:mail adam.leadbetter@marine.ie, tel: (+)353 (0)91 387 502

**ABSTRACT**

Within the field of ocean science there is a long history of using controlled vocabularies and other Semantic Web techniques to provide a common and easily exchanged description of datasets. As an activity within the European Union, United States, Australian-funded project “Ocean Data Interoperability Platform”, a workshop took place in June 2014 at Rensselaer Polytechnic Institute to further the use of these Semantic Web techniques with the aim of producing a set of Linked Data publication patterns which describe many parts of a marine science dataset. During the workshop, a Semantic Web development methodology was followed which promoted the use of a team with mixed skills (computer, data and marine science experts) to rapidly prototype a Linked Data publication pattern which could be iterated in the future. In this paper we outline the methodology employed in the workshop, and examine both the technical and sociological outcomes of a workshop of this kind.

**KEYWORDS**

Oceanography; Semantic Web; Linked Data; Development Patterns; Workshop Methodology;

**INTRODUCTION**

Within the domain of ocean sciences, the use of controlled vocabularies to standardise the annotation of data tables and metadata records has been common practice since before the birth of the World Wide Web (Leadbetter et al. 2015). These controlled vocabularies have been variously published as hard-copy manuals (International Oceanographic Commission 1987), text files on File Transfer Protocol sites, and most recently as Resource Description Framework (RDF) eXstensible Markup Language (XML) files on the World Wide Web (Leadbetter et al. 2014). As Linked Data (Berners-Lee 2006) has gained traction as a method of publishing interconnected datasets on the World Wide Web, an international community has emerged that uses ocean science controlled vocabularies as a central resource in the development of "Linked Ocean Data" (Leadbetter et al. 2013).

Several recent reports have highlighted the economic importance of access to research data (RDA Europe, 2014) and the response of research funding agencies increasing requirements for access to research results (US NSF, 2015). There is ever increasing acknowledgement of the need for greater and more efficient access to research data. The work presented in this paper is targeted at supporting this efficient access to research data using Semantic web technology in the context of a global initiative to converge best practices around data sharing in a specific scientific domain. The Ocean Data Interoperability Platform (ODIP) (Glaves et al. 2014) seeks to build on the informal international community to create a more formalised forum for the convergence of data management techniques in the ocean sciences with the goal of building a framework to support greater interoperability across data repositories. As reported by Glaves et al. (2014), one of the cornerstones of ODIP has been the use of controlled vocabularies available to the community, ensuring a common syntax and understanding of the data and metadata curated by the ODIP partners from around the world. In order to further the development of controlled vocabularies and Linked Data within the ODIP context, a workshop was held at Rensselaer Polytechnic Institute’s Tetherless World Constellation during June 2014.

This workshop drew together established researchers from a range of oceanographic institutions with informatics experts from the Tetherless World Constellation and both pre- and post-graduate students from Rensselaer Polytechnic Institute, Florida State University and University of California, San Diego. The aims of the workshop were threefold. The first aim was to drive towards harmonized Linked Data & Semantic Web efforts across related oceanographic data repositories. Secondly, a convergence would be achieved on solutions to a set of common issues emerging in the deployment of Semantic Web applications. Finally, draft best practices for semantic Linked Data for oceanographic data management would be created.

In this paper we introduce the “Smackdown” methodology used within the workshop, discuss the outcomes of the workshop, and reflect on how the “Smackdown” methodology may be improved and reused across the field of Earth Science informatics. We then go on to discuss the technical outcomes of the workshop with reference to the workshops aims, and finally we assess the social outcomes of the process and interactions during the workshop and how they may inform future developments of the Linked Ocean Data concept.

**THE SOCIAL ISSUES**

Research into social theory shows that technology is one form of contextual constraint which may be placed on an organization (Ranson, Hinings and Greenwood, 1980). Barley (1990) demonstrates that new technology can be an agent of change in the structure of work practices and actions. This observation lead Orlikowski & Robey (1991) to the conclusion that the development of information technology solutions, and the organizational contexts in which this development occurs, can shape the knowledge and capabilities embedded within the solutions. Therefore, as shown by van de Poel (2000), individuals contribute to the technical outcomes of an organisation through providing their expertise in the design and modelling phases of a technology solution and in the translation of ideas between different elements of a process. From this body of research, it can be concluded that a best practice in a technological project which will lead to organizational change in some way should include representatives of the relevant stakeholder communities who will be engaged in the deployment of the technologies, and experts in the technologies to be deployed.

Thus, we present the development of a Linked Data design workshop, building on the Semantic Web Use Case Methodology of Fox and McGuinness (2008). Van de Poel’s conclusion relate to several elements of Fox and McGuinness’ process: the small team consisting of mixed skills; the analysis; the development of the model ontology; and the scientific review and iteration of early prototypes of the project outcomes.

In the workshop described below, the combined multi-skilled team produced a technologically sound solution to the problem of data sharing in the ocean science domain which can also be understood by users beyond that domain.

The “Smackdown” methodology we describe in this paper is a practical implementation of the earlier models and the presented initiative to create small teams with the required skills and expertise as determined by use case driven scenarios validates the models of van de Poel and of Fox and McGuinness. The “Smackdown” methodology is based around a focused workshop which iterates the Semantic Web Use Case Methodology and provides under- and post-graduate students an opportunity to work alongside experienced practitioners. The short, focused nature of the workshop used to implement the “Smackdown” methodology means that the Semantic Web Use Case Methodology must be adapted, and its steps of reassessing the assembled team and rapid prototype development were dropped for the “Smackdown” methodology. However, the format adopted for the implementation of the “Smackdown” methodology also allowed the introduction of new elements: the opportunity for informal discussions in a coffee shop some distance from the meeting room prior to the timetabled events for each day; broader discussions over lunch at venues away from the main meeting area; and “mental breaks” in the agenda through seminar-like presentations on related topics throughout the course of the workshop.

**THE "SMACKDOWN" METHODOLOGY**

Bearing in mind the history of the use of Controlled Vocabularies and Common Data Models in the marine sciences, the Semantic Web and Linked Data technologies are natural fits in describing and publishing data and metadata in this field. Fernandez Lopez (1999) analysed a range of methodologies for developing ontologies, a formal specification of a conceptual data model (McGuinness, 2003). Fernandez Lopez recommended a life-cycle based approach to ontology building based on rapidly evolving prototypes, known as METHONTOLOGY (Gomez-Perez, 1998). More recently, Fox and McGuinness (2008) have developed a complementary iterative development lifecycle to be used in the creation of Semantic Web applications called the Semantic Web Use Case Methodology. The Semantic Web Use Case Methodology starts with a scientific use case which is an executable action for achieving a certain objective. A team is formed, centered on the use case, who bring expertise from their respective domains. Participants serve the collective by acting in specific roles as outlined by the definition of the Semantic Web Use Case Methodology. The team analyzes and refines the use case to ensure collective agreement on terminology and definitions, scope, and the workflow for accomplishing its execution. Only then, does the team make decisions on the appropriate tools and technology to adopt for execution. Rapid prototyping ensues followed by analysis and iteration through the Semantic Web Use Case Methodology for further refinement.

The “Smackdown” workshop was conceived to provide a focused event for the practical implementation and iteration of the Semantic Web Use Case Methodology. To this end, a small team was created to define a use case; develop a common Linked Data publication pattern to address that use case; and to consider related, but broader, topics in the Semantic Web field. The use case definition and the development of the Linked Data patterns were conducted as moderated discussions amongst the entire assembled team, with a chair nominated from among the senior data managers present. During the moderated discussions, options were sketched out on a PC using a graphics package and projected live onto a screen. Document sharing platforms for collaborative editing were then used to develop solutions throughout the group based on the sketched out solutions visible to the workshop participants.

One of the main aims of the “Smackdown” concept was to provide students (both under- and post-graduate) with an opportunity to mix with experienced practitioners. Therefore, each of the students presented an aspect of either their honours or research projects in the sessions set aside for considering related topics from the Semantic Web field. This enabled them to gain experience of delivering oral descriptions of their work in a ‘safe’ environment, in which feedback was offered from their peers and from more senior members of the community and also their work was then used as a starting point for new discussions.

For the workshop under discussion in this paper, the assembled team consisted of a group of eleven individuals with backgrounds as oceanographic data managers, domain experts, and software developers from the US National Science Foundation Rolling Deck to Repository and Biological and Chemical Oceanography Data Management Office programs and the UK’s British Oceanographic Data Centre of the Natural Environment Research Council, and informatics specialists from Rensselaer Polytechnic Institute.

Defining a use case for this sort of development is well documented by Fox *et al.* (2009) and Duerr *et al.* (2015). Fox *et al.* assert that it is key that the use case is developed by domain experts with assistance from the technological project team, and both Fox *et al.* and Duerr *et al.* provide a well-defined use case at the heart of their ontology development work. The central use case for the workshop was to identify a shared pattern amongst oceanographic data facilities for information describing the collection of oceanographic data, its provenance, and storage for future discovery and reuse. This use case is of benefit to end users of marine data as it gives them increased trust (Leadbetter, 2015) in the data through the transparency of information. As it is designed to allow search of datasets from multiple sources it is also of benefit to end users (such as research scientists who are looking for data resources) as it allows a larger knowledge base to be queried. The use case was conceived as part of the ODIP project’s drive to connect marine science data from Europe, the United States and Australia. In trying to reach this goal, discussions by e-mail prior to the workshop and early in the workshop firmed up the use case. In reflecting on the workshop it was generally agreed that progressing through a first iteration of the Semantic Web Use Case Methodology before the workshop would help to ensure time is spent on discussing common language and rapid prototyping. The Semantic Web Use Case Methodology promotes filling out a template, <http://tw.rpi.edu/media/latest/UseCase-Template_SeS>, which was not used in the case of this workshop but would provide a way to virtually collaborate in preparation for the workshop, and to focus discussions in future workshops working in this way.

The use case then became the driving point for the remaining focused discussions within the workshop, with moderators tasked with keeping conversations within the workshop setting to the topic of the use case. Of the iterative steps in the Semantic Web Use Case Methodology, this meeting focused on use case scoping and analysis, developing a common language, use of tools, and review and iteration. Scoping the use case began through discussing participant’s individual goals for the outcomes of executing the use case. As an example, alignment to schema.org and the quality of current geospatial Linked Data efforts were discussed as edge cases of scope. If collectively, it was agreed that pursuit of these edge cases was highly valuable to the shared pattern, they were considered in scope. All topics considered in scope were analyzed and discussed in terms of how we collectively define the concepts within the topics. Only by coming to an agreement on the meaning of an Oceanographic Cruise, Dataset, Instrument, etc. can we begin to align individual implementations to a shared pattern of use. To document our findings the workshop adopted tools that would help to communicate effectively within the time constraints of the workshop. Visual representations of our decisions were created through yEd Graph Editor, projected on a wall, to quickly convey understanding and gather feedback on agreement. After agreeing to a conceptual visualization, Google Docs were used to record sample RDF triples to review the conceptual understanding. Here, the Semantic Web Use Case Methodology iterates as the group analyzes the work, reviews the work against the stated use case, and begins the process over again.

Some steps in the Semantic Web Use Case Methodology were omitted or refined, as the nature of this type of workshop does not allow for reassessing the assembled team. One mitigating strategy to this point was the advantageous location of hosting the meeting at the Tetherless World Constellation Lab at Rensselaer Polytechnic Institute. Students and faculty were invited and encouraged to sit in on as much of the workshop as desired which proved helpful in guiding certain discussions such as the use of the PROV-O ontology. The rapid prototyping step was deferred to be followed up by participants in virtual collaborative environments, such as GitHub, following the workshop and in anticipation of a follow-up physical meeting.

During the final two days of this workshop, an approach to linking data using a generic pattern was discussed. This was driven by the use cases of interoperability between data repositories in the ocean sciences domain and also between ocean sciences data repositories and those from other domains.

Identifying shared patterns in information modeling across multiple participants can be an exhausting exercise. To provide mental breaks, related agenda items were scheduled to breakup work on the central use case. These are denoted in Table 1 with asterisks (\*). In Triplestore/SPARQL Review, each data repository explained what RDF store and SPARQL endpoint technology was employed, why it was selected, and discussed known issues or limitations. The agenda item Linked Data at SAMOS was a student presentation on how they planned on implementing Linked Data for the SAMOS initiative. The discussion of PRIZMS & PROV Pingback assessed the usability of the proposed PROV Pingback methodology (Lebo, 2014). The agenda item Linked Data Application Programming Interfaces was a presentation of the Linked Data API for Rolling Deck to Repository (Fu, Arko & Leadbetter, 2014). The agenda item Vocabulary matching in the Rolling Deck to Repository project was a presentation of work on matching terms in the R2R vocabulary to select SeaDataNet vocabularies hosted by the NERC Vocabulary Server (Chen *et al*., 2014). The remaining agenda items address specific points along the Semantic Web Use Case Methodology. Geospatial Linked Data sought to understand the landscape of geospatial representation in Linked Data for potential gazetteers of named ocean locales and seafloor features. Each day of the workshop was planned to start a little later than is traditional in order to allow time for individuals to complete administration tasks prior to the start of the day’s activities, and participants had the opportunity to join informal discussions prior to the main programme for the day. Organising the agenda in this way allowed the small number of participants to all be fully engaged in the workshop programme sessions and for each to contribute in a meaningful way. Similarly, the lunch breaks were timed to be relatively long, and taken away from the main meeting area to allow for discussion of broader topics before more specific, technical discussions in similar areas later.

**THE TECHNICAL OUTCOMES**

Therefore, the approach taken to modelling Linked Ocean Data was to use high level terms from ontologies and vocabularies which are well understood in a range of communities while maintaining the references to domain specific ontologies. An overview of the developed pattern is shown in Figure 1. For example the high level terms of the World Wide Web Consortium’s PROV (Lebo, Sahoo and McGuinness, 2013) ontology and Observations and Measurements, O&M (Cox, 2010), were combined with domain specific ontologies, such as the Ocean Data Ontology, ODO (http://schema.ocean-data.org/odo).

The following section relies heavily on the use of Resource Description Framework namespaces. The namespaces used, and their prefixes as referenced in the following text are summarised in Table 2.

**Cruise**

A cruise is a discrete deployment of a platform (normally a research vessel) which has a trajectory from a port to another, or returning to the port of departure. While the cruise itself is not a geographic feature, its trajectory is, and therefore an intermediate resource for the trajectory must be instantiated to which the geometry can be assigned. The linkage is made using classes and properties from the EarthCube OceanLink ontology (Narock *et* al., 2014, which was current at the time the workshop took place, but which has since been superseded by the GeoLink ontology – see <http://schema.geolink.org/>). For the European partners within the Ocean Data Interoperability Platform, a cruise has been declared as an Environmental Monitoring Activity, and a platform as an Environmental Monitoring Facility. This allows for compatibility with the Environmental Monitoring Facilities (EF) theme of the European Commission's INSPIRE spatial data directive (; Leadbetter & Vodden, 2015).

**Dataset**

During the workshop, dataset entities were deemed to have been generated by cruise activities. However, there are use cases in operational modes (such as in the Shipboard Automated Meteorological and Oceanographic System; http://samos.coaps.fsu.edu/html/) where the cruise (a single port-to-port deployment of a platform, such as a research vessel) is not known. In this case we recommend that a PROV activity be created for a given calendar day's activity on that platform instead of a full cruise instance. This approach allows for subsequent mapping of a day's activity to an R2R cruise instance should that be determined in the future.

The use of prov:wasAttributedTo and odo:fromInstrument to link a dataset to the collecting instrument is optional, is repeatable (i.e. it is a one-to-many relationship) and no rank or instrument order should be inferred from the use of these properties. Similarly, the use of odo:hasObservedProperty to state the discovery level descriptions of the observed properties within a dataset is both optional and repeatable.

**Geometry**

As per the guidance for best practice which emerged from the joint World Wide Web Consortium/Open Geospatial Consortium Linking Geospatial Data workshop (Archer, 2014), geometries are given their own URIs so they may be reused, and are not part of the metadata for their feature (Listing 1). For a cruise, the cruise has a trajectory which is the feature and the track of that trajectory is the geometry.

@prefix olc: <http://schema.oceanlink.org/pattern-name>.

@prefix geo: <http://www.opengis.net/ont/geosparql#>.

@prefix sf: <http://www.opengis.net/ont/sf#>.

<http://foo.bar/cruise> a olc:cruise;

 olc:hasTrajectory <http://foo.bar/cruise\_trajectory>.

<http://foo.bar/cruise\_trajectory> a olc:Trajectory;

 a geo:Feature;

 geo:hasGeometry <http://foo.bar/cruise\_geometry>.

<http://foo.bar/cruise\_geometry> a sf:LineString;

geo:asWKT "<http://www.opengis.net/def/crs/OGC/1.3/CRS84> LINESTRING(-68.7030 37.934, -68.7160 37.92, -67.4650 36.019, -67.3680 35.865, -67.2700 35.715, -67.1720 35.558, -67.0770 35.408, -66.9780 35.257, -66.8880 35.108, -66.7850 34.952, -66.6950 34.802)"^^geo:wktliteral.

*Listing 1. A sample listing of a research vessel cruise track using the W3C/OGC encoding recommendations*

**Observation**

Observations are built from the O&M data model (Cox, 2010), and it is possible to derive the Observed Properties of O&M from the BODC Parameter Usage Vocabulary, the Procedure of O&M from the SeaVoX Device Catalogue and the O&M Result from the relevant data file, which forms part of the dataset resource. Datasets are linked to Observations through both the Result and the Feature of Interest, as shown in Figure 2.

**Temporal entity**

In order to encode temporal entities associated with cruise deployment dates and with project dates, we chose to implement the Temporal Entity objects from OWL Time. An example encoding of a Temporal Entity is shown in Listing 2.

<http://foo/bar/baz#time> a time:TemporalEntity ;
 time:hasBeginning <http://foo/bar/baz#timeBegins> ;
 time:hasEnd <http://foo/bar/baz#timeEnds> .

<http://foo/bar/baz#timeBegins> a time:instance ;
 time:inXSDDateTime "1994-10-31T03:00:00Z"^^xsd:datetime ;
 time:inDateTime <http://foo/bar/baz#timeBeginsDescription> .

<http://foo/bar/baz#timeBeginsDescription> a time:DateTimeDescription ;
 time:unitType time:unitSecond ;
 time:second "00" ;
 time:dayOfWeek time:Tuesday ;
 time:minute "00" ;
 time:hour "03" ;
 time:day "18" ;
 time:month "10" ;
 time:year "1994" ;
 time:dayOfYear "291" ;
 time:timeZone <http://www.w3.org/2006/timezone-world#ZTZ> .

<http://foo/bar/baz#timeEnds> a time:instance ;
 time:inXSDDateTime "1994-10-31T11:02:00Z"^^xsd:datetime ;
 time:inDateTime <http://foo/bar/baz#timeEndsDescription> .

<http://foo/bar/baz#timeEndsDescription> a time:DateTimeDescription ;
 time:unitType time:unitSecond ;
 time:second "00" ;
 time:dayOfWeek time:Monday ;
 time:minute "02" ;
 time:hour "11" ;
 time:day "31" ;
 time:month "10" ;
 time:year "1994" ;
 time:dayOfYear "304" ;
 time:timeZone <http://www.w3.org/2006/timezone-world#ZTZ> .

<http://foo/bar/baz#timeBeginsDescription> a time:DateTimeDescription ;
 time:timeZone <http://www.w3.org/2006/timezone-world#ZTZ> .

*Listing 2. A sample listing of temporal entity encoding*

**In the Wild**

The Linked Data patterns for oceanographic datasets described above were deployed by the British Oceanographic Data Centre at http://linked.bodc.ac.uk to both demonstrate the utility of the patterns and to support the Natural Environment Research Council’s EnviroHack event in February 2015 (Riley, 2015).

The DAnnoTAte team from Envirohack2015 sought to use semantic web technologies to provide added value to datasets by leveraging the pre-existing Characterization of metadata to enable high-quality climate applications and services, or CHARMe, project (Rozum, Raoult and Dee, 2014). CHARMe provides a mechanism for researchers to make annotations about datasets like comments, list related datasets, etc. through the Open Annotation ontology. The Linked Data patterns as employed by the British Oceanographic Data Centre were useful to the DAnnoTAte team for testing its prototype which can be found on Github.com at <http://charme-project.github.io/DAnnoTAte/> (code at <https://github.com/CHARMe-Project/DAnnoTAte>).

**CONCLUSIONS**

In critique of the “Smackdown” methodology applied in the workshop, it could be said that the technologies of the Semantic Web and Linked Data drove the decision making process and that a bias was introduced into the process because the involved parties were already committed to these technologies. However, this is a well-documented issue. As an interpretation of this observation, it should be noted that sharing of data has long been held to be an interaction between organisations (real or virtual) and technology (Markus and Robey, 1988). A structuration perspective highlights the intertwined nature of technological and behavioral choices (DeSanctis and Poole, 1994), and allows technology to be defined as a malleable element of organizational structure (Orlikowski, 1992). Therefore, technology can be seen as a set of rules and resources that enable some actions, while constraining others, and that are in turn shaped by those actions over time. Within this perspective, the technical design of the data sharing patterns within the framework described in this paper is inextricably linked to its implementation and use within the context of an organisation, in this case a small team who continue to develop the patterns as a virtual organisation. This provides a fruitful model for understanding the processes of ongoing learning and growth. Orlikowski (1992) traces a three-way cycle of mutual influence between institutional properties, technology, and the human agents who use and build the tools within the organisation. Evans (1997) suggests a similar cycle of influence between organizational, technological, and policy/planning structures, and the actions people perform within those structures (Figure 3).

In conclusion, we have seen that the short and intense workshop, with a tight technological focus, has provided a significant furthering of the modelling of marine science datasets as Linked Data. One major benefit of this approach was the mixture of computer science (i.e. technical) experts and marine science (domain specific) experts involved in the workshop who could ensure a consistent approach to the use of semantics, but also a meaningful representation of the data in the scientific context. As an aim of the National Science Foundation funding for the Ocean Data Interoperability Platform was capacity building, it is important to note that four of the participants in the workshop were Bachelors, Masters or PhD students. This is consistent with the vision of Urban and Boscolo (2013) in using scientific meetings to enhance the development of early-career scientists. Whilst Urban and Boscolo suggested organizing short workshops either before or after major conferences, the workshop described in this paper, developed using the “Smackdown” methodology, was sufficiently long to warrant its own calendar slot and also had funding available for travel and catering. However, the workshop did also facilitate the early-career researchers meeting more senior peers and giving them experience in presenting their work and discussing future approaches to the issues at hand. The “Smackdown” methodology also introduced the concepts of broader discussions in informal venues (coffee shops, restaurants) during the workshop which proved a successful way of assisting the early-career researchers with integrating into the group of experience practitioners who were already a close-knit community. Similarly, the seminar-like presentations encouraged the early-career researchers to learn, contribute and to distribute their knowledge throughout the workshop.

It has also been shown that the Linked Data patterns generated by the workshop have been successfully deployed by the British Oceanographic Data Centre in response to the needs of the EnviroHack event, and these Linked Data could be incorporated into applications such as JellyStrike, a system to identify and track likely jellyfish hotspots using atmospheric and oceanographic data; Energy Cast, a weather-forecast and machine learning informed green energy app to help match domestic energy supply with energy consumption; or CHARMe - a self-learning data system to allow users to enrich environmental data by adding annotations, use cases and key metrics to make more sense of our data-rich world. Finally, the driving force of the workshop, extending the application of Semantic Web techniques within the ocean sciences domain, has been shown to allow practitioners and users of the technical outcomes to be involved in the development process allowing for easier uptake of the results. Use of the patterns to publish resources demonstrates that the technology has impacted organisational activities. Availability of these patterns through an organisational repository also led to the Natural Environment Research Council sponsoring the EnviroHack event and the policy decision that other data centres in its network should pursue similar publication mechanisms.

**ACKNOWLEDGEMENTS**

The work described in this paper was funded in part through European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 312492, and in part by supplemental funding from the National Science Foundation to the R2R program (NSF OCE 0947877, 0947822, 0947828, 0947784). The workshop was kindly hosted by the Tetherless World Constellation at Rensselaer Polytechnic Institute. At the time of the workshop, Adam Leadbetter represented the British Oceanographic Data Centre.

**REFERENCES**

Archer P (2014) Report of the Joint World Wide Web Consortium/Open Geospatial Consortium workshop Linking Geospatial Data, 5th - 6th March, 2014, Campus London, Shoreditch, London, UK. World Wide Web Consortium. https://www.w3.org/2014/03/lgd/report Accessed 27 May 2015

Barley S. R. (1990). The alignment of technology and structure through roles and networks. Administrative Science Quarterly 35, 61-103.

BCO-DMO (2013) Ocean-Data-Ontology. Biological and Chemical Oceanography Data Management Office, Woods Hole Oceanographic Institution. https://github.com/BCODMO/Ocean-Data-Ontology Accessed 10 April 2015

Berente N, Howison J, King JL, Lyytinen K, Wilkins-Diehr N (2012) Virtual Organizations as Sociotechnical Systems: Exploring How Organization Scientists and Virtual Organization Leaders Can Collaborate. Virtual Organizations as Sociotechnical Systems Workshop. Case Western Reserve University, Cleveland, Ohio, May 14-16. http://managingcenters.net/wp-content/uploads/2013/02/VOSS-Report.pdf Accessed 5 May 2015

Berners-Lee T (2006) Linked Data - design issues. World Wide Web Consortium. http://www.w3.org/DesignIssues/LinkedData.html. Accessed 10 April 2015

Chen, Y., Shepherd, A., Chandler, C., Arko, R., & Leadbetter, A. (2014, May). Ontology Based Vocabulary Matching for Oceanographic Instruments. In EGU General Assembly Conference Abstracts (Vol. 16, p. 12909).

Cox S (2010) Observations and Measurements. OGC Abstract Specification Topic 20. Open Geospatial Consortium document 10-004r3. http://portal.opengeospatial.org/files/41579. Accessed December 8, 2014.

DeSanctis, G., & Poole, M. S. (1994). Capturing the complexity in advanced technology use: Adaptive structuration theory. Organization science, 5(2), 121-147.

Duerr R., McCusker J., Parsons M., Khalsa S., Pulsifer P., Thompson C., Yan R., McGuinness D., & Fox P. (2015). Formalizing the semantics of sea ice. Earth Science Informatics 8(1): 51-62.

Evans, J. D. (1997). Infrastructures for sharing geographic information among environmental agencies (Doctoral dissertation, Massachusetts Institute of Technology).

Fernández López M. (1999) Overview Of Methodologies For Building Ontologies. Proceedings of the IJCAI-99 workshop on Ontologies and Problem Solving Methods (KRR5). Stockholm, Sweden, August 2.

Fox P, McGuinness D. (2008) TWC Semantic Web Methodology. Tetherless World Constellation, Rensselaer Polytechnic Institute. http://tw.rpi.edu/web/doc/TWC\_SemanticWebMethodology. Accessed 10 April 2015

Fox P., McGuinness D., Cinquini L., West P., Garcia J., Benedict J. & Middleton D. (2009). Ontology-supported scientific data frameworks: The virtual solar-terrestrial observatory experience. Computers & Geosciences, 35(4): 724-738.

Fu L., Arko R., Leadbetter, A. (2014, May). Rapid Deployment of a RESTful Service for Oceanographic Research Cruises. In EGU General Assembly Conference Abstracts (Vol. 16, p. 1335).

Glaves H, Schaap D, Arko R, Proctor R (2014) Ocean Data Interoperability Platform (ODIP): supporting the development of a common global framework for marine data management through international collaboration. In EGU General Assembly Conference Abstracts 16: 14366.

Gómez-Pérez, A. (1998). Knowledge sharing and reuse. Handbook of applied expert systems, 10-11.

INSPIRE (2013). D2.8.II/III.7 Data Specification on Environmental Monitoring Facilities - Technical Guidelines. Accessed February 01, 2016. http://inspire.ec.europa.eu/documents/Data\_Specifications/INSPIRE\_DataSpecification\_EF\_v3.0.pdf

International Oceanographic Commission (1987) A General Formatting System for Geo-referenced Data. International Oceanographic Commission, Paris, France.

Leadbetter A, Arko R, Chandler C, Shepherd A, Lowry, R (2013) Linked Data: An oceanographic perspective. J. Ocean Tech. 8(3):7-12,

Leadbetter A, Arko R, Chandler C, Shepherd A, Lowry, R (2015) Loose integration of local information to generate collaborative marine science knowledge. In Diviacco P, Fox P, Pshenichy C, Leadbetter A (eds) Collaborative Knowledge in Scientific Research Networks. IGI Global, Hershey, PA, pp 238-261.

Leadbetter A, Lowry R, Clements D O (2014) Putting meaning into NETMAR - the open service network for marine environmental data. Int. J. of Digital Earth, 7(10):811-828. doi: 10.1080/17538947.2013.781243

Leadbetter A, Vodden P (2015) Semantic linking of complex properties, monitoring processes and facilities in web-based representations of the environment. Int. J. of Digital Earth, online only. doi: 10.1080/17538947.2015.1033483

Lebo T, Sahoo S, McGuinness, D eds. (2013) PROV-O: The PROV Ontology. World Wide Web Consortium. http://www.w3.org/TR/2013/REC-prov-o-20130430/ Accessed October 17 2014,

Lebo, T., West, P., & McGuinness, D. L. (2014) Walking into the Future with PROV Pingback: An Application to OPeNDAP Using Prizms. In Provenance and Annotation of Data and Processes. Springer International Publishing, 31-43.

McGuinness, D. L. (2003). Ontologies come of age. Spinning the semantic web: bringing the World Wide Web to its full potential, 171-196.

Markus, M. L., & Robey, D. (1988). Information technology and organizational change: causal structure in theory and research. Management science, 34(5), 583-598.

Narock, T., Krisnadhi, A., Hitzler, P., Cheatham, M., Arko, R., Carbotte, S, Shepherd, A., Chandler, C., Raymond, L., Wiebe, P. & Finin, T. (2014, October). The OceanLink Project. In 2014 IEEE International Conference on Big Data (pp. 14-21). IEEE.

Orlikowski, W. J. (1990). Information technology and the structuring of organizations. Information Systems Research, 2(2), 143-169.

Orlikowski, W. J. (1992). The duality of technology: Rethinking the concept of technology in organizations. Organization Science, 3(3), 398-427.

Ranson S., Hinings B., & Greenwood R. (1980). The structuring of organizational structures. Administrative Science Quarterly, 25, 1-17.

RDA Europe (2014) "The Data Harvest: How sharing research data can yield knowledge, jobs, and growth". An RDA Europe Report. December, 2014. <http://europe.rd-alliance.org/documents/publications-reports/data-harvest-how-sharing-research-data-can-yield-knowledge-jobs-and>

Riley, E. (2015) Tackling environmental issues at EnviroHack 2015 – Part 1. Digital Catapult Centre. http://www.digitalcatapultcentre.org.uk/tackling-environmental-issues-at-envirohack-2015-part-1/ Accessed June 8 2014.

Rozum, I., Raoult, B., & Dee, D. (2014). Assessing variability in climate data: a significant event viewer tool. In EGU General Assembly Conference Abstracts 16: 13327.

United States National Science Foundation (2015) "NSF’S PUBLIC ACCESS PLAN: Today’s Data, Tomorrow’s Discoveries; Increasing Access to the Results of Research Funded by the National Science Foundation". US NSF 15-52. Arlington, VA, USA. March 2015.

<http://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf15051>

Urban Jr, E. R., & Boscolo, R. (2013). Using Scientific Meetings to Enhance the Development of Early Career Scientists. OCEANOGRAPHY, 26(2), 164-170.

Van De Poel, I. (2000). On the role of outsiders in technical development. Technology Analysis & Strategic Management, 12(3), 383-397.

|  |  |
| --- | --- |
| **Agenda Order** | **Item** |
| #1 | \*Triplestore, SPARQL Review  |
| #2 | Geospatial Linked Data |
| #3 | Schema.org Dataset |
| #4 | \*Linked Data in the Shipboard Automated Meteorological and Oceanographic System (SAMOS) |
| #5 | Harmonizing Use of Common Vocabularies for Ocean Data |
| #6 | \*Prizms & PROV Pingback |
| #7 | \*Linked Data Application Programming Interfaces |
| #8 | Semantic Entity Matching |
| #9 | O & M / PROV / EarthCube-OceanLink / Ocean Data Ontology |
| #10 | Beyond Linked Ocean Metadata (links to the data values) |
| #11 | \*Vocabulary matching in the Rolling Deck to Repository project |
| #12 | Persistent identifiers for organisations and people |

*Table 1. Outline workshop agenda*

|  |  |
| --- | --- |
| **Prefix** | **Namespace URL** |
| arpfo | http://vocab.ox.ac.uk/projectfunding# |
| dcat | http://www.w3.org/ns/dcat# |
| dcterms | http://www.purl.org/dc/terms/ |
| ef | http://eidc.ceh.ac.uk/onto/ef# |
| foaf | http://xmlns.com/foaf/0.1/ |
| geo | http://www.opengis.net/ont/geosparql# |
| obs | [http://def.seegrid.csiro.au/isotc211/iso19156/2011/observation#](http://def.seegrid.csiro.au/isotc211/iso19156/2011/observation) |
| obs-prop | http://environment.data.gov.au/def/op# |
| odo | http://ocean-data.org/schema/ |
| ol-cruise | http://schema.oceanlink.org/cruise# |
| prov | http://www.w3.org/ns/prov# |
| rdf | http://www.w3.org/1999/02/22-rdf-syntax-ns# |
| sf | http://www.opengis.net/ont/sf# |
| time | http://www.w3.org/2006/time# |

*Table 2.* *RDF Namespaces used in this document*



*Figure 1. A high-level overview of the Linked Data pattern developed during the workshop. Linked Data publication Patterns were created for each box in the diagram; arrows show the direction of relationship between the patterns.*



*Figure 2. The Linked Data publication pattern for an oceanographic observation developed during the workshop*



*Figure 3. Mutual influence of technology, organization, and policy/planning on people and actions (Evans, 1997). Arrow (a) represents technology first influencing action; which leads to (b) action influencing the organisation; (c) organisation in turn influencing action; (d) action influencing technology; (e) policy & planning influencing action; (f) action influencing policy & planning; and finally technological and organisational change.*