Chapter 4
Linked Ocean Data 2.0

Adam Leadbetter
Marine Institute, Ireland

Adam Shepherd
Woods Hole Oceanographic Institution, USA

Michelle Cheatham
Wright State University, USA

Rob Thomas
British Oceanographic Data Centre, National Oceanography Centre, UK

ABSTRACT

Within the theme of sustainable development, it is not desirable to either have data siloed in one location where it cannot be reused for purposes beyond which it was originally collected, or in a state where it cannot be integrated into a holistic view of the marine environment. As such, the links between datasets should be formally documented and exploited as best as possible. Given this, the use of Semantic Web technology and information modelling patterns are explored in this chapter with reference to the marine domain. Further, new strategies for adding semantic annotation to data in real-time are discussed and prototyped.

INTRODUCTION

Within the domain of marine data management, there is a history, beginning in the 1980s, of using controlled vocabularies to define the content of metadata fields and data file channels (UNESCO, 1987; Lowry, 2003; Merati & Burger, 2004; Lawrence et al., 2009; Schaap & Lowry, 2010). The availability of such vocabularies as Semantic Web resources through such platforms as the Marine Metadata Interoperability Ontology Registry and Repository (Graybeal et al., 2012) and the Natural Environment Research Council Vocabulary Server (Leadbetter et al., 2014) has led to the use of Linked Data (Berners-Lee, 2006) patterns to publish metadata and data concerning the marine domain. The Linked Ocean Data concept has been shown to facilitate distributed eScience through validation of chained web processing services and dynamic discovery and aggregation of datasets alongside increased impact of datasets through formalised links to the acquisition methodologies and associated publications interpreting the data (Leadbetter, 2015).

When considering sustainable development of the oceans, it is most important to be able to combine data from a range of producers in a single application. For example, the Marine Renewable Energy Portal

DOI: 10.4018/978-1-5225-0700-0.ch004

delivered by the Sustainable Energy Authority of Ireland and the Irish Marine Institute allows potential energy developers access to live dashboards of current and forecast wave, wind and tidal conditions using data from the Marine Institute, the Commissioner of Irish Lights and the United States' National Oceanic and Atmospheric Administration. However, these data are not connected in a formal Linked Data sense and moving in this direction would better allow users to incorporate other data sources, generate their own dashboards, and enquire of the data in new ways to generate new information and knowledge. Further, these principles can be extended to the creation of new decision support tools for planning operations on board research vessels, or at remote bases when considering the maintenance of missions of deployed Autonomous Underwater Vehicles.

In this chapter we will expand on the existing Linked Ocean Data research to include the emerging concepts of heterogeneous data integration through the use of ontology design patterns, and publishing Linked Data from sources of real-time observations. This will include:

- Revisiting the current state of the Linked Ocean Data cloud to show how it has expanded since it was first proposed in 2013
- Conceptualising the Linked Ocean Data cloud using advances in interchange visualisation techniques (Zeng et al., 2013; Krzywinski et al, 2011) to give new insights into the interconnectedness of the data and information represented within the cloud
- Examining how the emergence of the concept of ontology design patterns has been successfully applied in the ocean science domain
- Considering the relationships which are forming between Linked Data and Big Data and how these have begun to be explored in marine science
- Why the Linked Ocean Data paradigm is important for a sustainable approach to the exploration and development of the ocean

THE STATE OF THE LINKED OCEAN DATA CLOUD

The Linked Ocean Data concept was introduced by Leadbetter *et al.* (2013) and was refined by Leadbetter (2015). The initial Linked Ocean Data cloud consisted of 18 nodes, and by the final publication listed above a further two nodes were incorporated. However, the visualizations used in these descriptions of the Linked Ocean Data cloud have not shown any quantitative information concerning the linkages between the nodes of the cloud, solely the qualitative fact that nodes are linked. However, prior to the introduction of the Linked Ocean Data cloud, Leadbetter and Lowry (2012) experiment with visualisations of connections between controlled vocabularies using circos plots (Krzywinski 2009) which were originally designed for use in visualizing genomics data, but have since been repurposed to show customer flow in the motor industry, volume of courier shipments, database schemas, and presidential debates. The 2012 mappings from the NERC Vocabulary Server to other external controlled vocabulary services is shown in Figure 1. While providing some qualitative information about both the number and directions of the connections in the space occupied by the Linked Ocean Data cloud, this approach is not intuitively interpreted by those users who are unfamiliar with the circos concept. More recently, Zeng et al. (2013) have developed the circos plot concepts to show interchange patterns in the movements of passengers through public transport networks. From the 2015 state of the Linked Ocean Data cloud, it could be seen that there are several nodes where an interchange of data links may occur. Therefore an