

# The place of Schema.org in Linked Ocean Data

**Adam Leadbetter**, Marine Institute (Ireland), adam.leadbetter@marine.ie

**Rob Thomas**, Marine Institute (Ireland), rob.thomas@marine.ie

**Adam Shepherd**, Woods Hole Oceanographic Institution (USA), ashepherd@whoi.edu

**Doug Fils**, Consortium for Ocean Leadership (USA), dfils@oceanleadership.org

**Kevin O'Brien**, NOAA (USA), kevin.m.o'brien@noaa.gov

## Introduction.

The use of Linked Open Data, that is publishing data on the World Wide Web using Web addresses to name things and using structured data standards (Berners-Lee, 2006), in marine data and information systems has been well documented over recent years (Leadbetter, et al 2013; Leadbetter, 2015; Leadbetter et al 2016). However, much of the work has been focused on the use of highly domain-specific Linked Data models (vocabularies, thesauri and ontologies), such as the NERC Vocabulary Server (Leadbetter, Lowry and Clements, 2014), Ocean Data Ontology<sup>1</sup> and GeoLink<sup>2</sup>. However, it has remained difficult to build a comprehensive, global search tool for marine data. In parallel, the major search engine companies have been driving a more generalised ontology known as Schema.org (Guha, Brickley & Macbeth, 2016). Schema.org patterns are well-known to both web developers and web users as they facilitate the display of enhanced search results through “Rich Snippets.” Rich Snippets are related pieces of information displayed alongside a search result (for example when searching for a film or movie the running time; description; aggregated; reviews; related news items – see Figure 1).

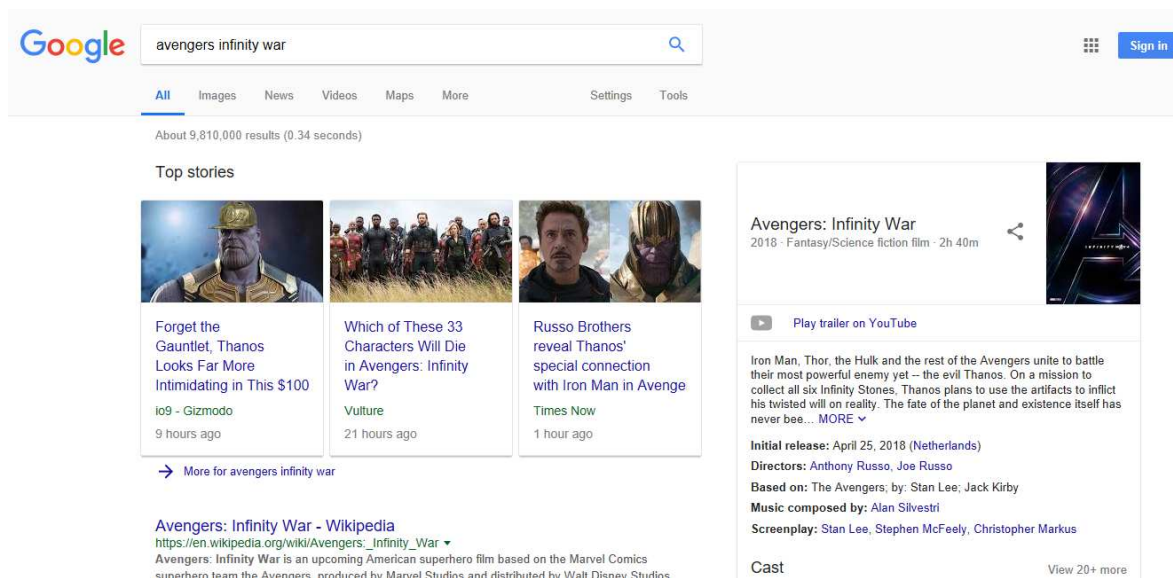


Figure 1: Rich Snippets in search powered by Schema.org. The enhanced search results are shown in the news items and the panel to the right hand side of the search page.

## Schema.org for the Oceanographic Domain.

Over the past year activity has been focussed on delivering dataset descriptions in Schema.org compliant forms. This has been done in a number of projects with slightly different approaches, but each activity has been done in liaison with Google, who are developing a dataset portal from the encoded metadata.

<sup>1</sup> <http://ocean-data.org/schema/>

<sup>2</sup> <http://schema.geolink.org/>

As part of Ireland's Digital Ocean Programme (Leadbetter, O'Grady and Burke, 2016), the Erddap data server software package published by NOAA (Simons, 2017) has been extended to output Schema.org Dataset metadata from its dataset catalogue landing pages. Additionally, the Drupal based Marine Institute data catalogue, which has also been developed under the Digital Ocean programme, has the same Schema.org dataset output driven from its internal database. Additionally, the Biological and Chemical Oceanography Data Management Office (BCO-DMO) is also embedding Schema.org dataset metadata from their own Drupal content management system.

The EarthCube Architecture Project 418 is harvesting JSON-LD embedded in web pages across national data repositories in the U.S., looking for datasets that are described using Schema.org and other external vocabularies.

Under the Linked Data developments for the SeaDataNet catalogue structure, the various catalogues have been mapped through to Schema.org representations from the other Linked Data flavours in which they have been represented. Various, this includes the Schema.org dataset pattern for the EDMED dataset and CDI data directory; Schema.org organisation for the EDMO organisation directory; and Schema.org Event for the Cruise Summary Report directory. Similarly, to address the issue of unique Linked Data identifiers for individual people, a Schema.org person profile of the IODE's OceanExpert repository has been developed under the Ocean Data Interoperability Platform and SeaDataCloud.

### **Conclusions.**

While the development of Linked Ocean Data has been valuable in gaining a richer understanding of our datasets and their structure, it would now seem that the time is right to invest in mapping these efforts to the more general Schema.org patterns. This is not without its issues, as there are patterns which have not yet been fully developed in Schema.org that are valuable in Linked Ocean Data. The Schema.org developers have called the vocabulary a core vocabulary which should be extended where richer semantics are required. As well, the referencing of standard vocabularies, such as those in the NERC Vocabulary Server, for naming parameters continues to be vital for interoperability but is not well defined in Schema.org. However, by agreeing to vocabulary referencing and by pursuing developments with the Schema.org governance bodies, these hurdles can be overcome and a truly global Linked Ocean Data profile can be achieved. In the face of continued difficulties in building successful marine data discovery systems, such a globally linked profile will enhance general discoverability through its inherent connections with well-known and widely used search tools.

### **References**

- Berners-Lee, T. (2006). Design issues: Linked Data. URL <http://www.w3.org/DesignIssues/LinkedData.html>.
- Guha, R. V., Brickley, D., & Macbeth, S. (2016). Schema. org: evolution of structured data on the web. *Communications of the ACM*, 59(2), 44-51.
- Leadbetter, A. (2015). Linked Ocean Data. *Studies on the Semantic Web volume 20.: The Semantic Web in Earth and Space Science. Current Status and Future Directions*, 11-31.
- Leadbetter, A., Arko, R., Chandler, C., Shepherd, A., & Lowry, R. (2013). Linked Data: An oceanographic perspective. *Journal of Ocean Technology*, 8(3), 7-12.
- Leadbetter, A., Cheatham, M., Shepherd, A., & Thomas, R. (2016). Linked Ocean Data 2.0. *Oceanographic and Marine Cross-Domain Data Management for Sustainable Development*, 69-99.
- Leadbetter, A. M., Lowry, R. K., & Clements, D. O. (2014). Putting meaning into NETMAR—the open service network for marine environmental data. *International Journal of Digital Earth*, 7(10), 811-828.
- Leadbetter, A., O'Grady, E., & Burke, N. (2016). Ireland's Integrated Digital Ocean. *Bollettino di Geofisica teorica ed applicata* 57(supplement), 224-226.
- Simons, R. (2017). ERDDAP. <https://coastwatch.pfeg.noaa.gov/erddap> Monterey, CA: NOAA/NMFS/SWFSC/ERD.