



Enhancing discoverability of the European Directory of Marine Environmental Data (EDMED) with schema.org

Alexandra Kokkinaki (1), Justin Buck (1), Adam Leadbetter (2), Rob Thomas (2), and Mark Hebden (1)

(1) British Oceanographic Data Centre, National Oceanography Centre, LIVERPOOL, United Kingdom (alexk@bodc.ac.uk),

(2) Marine Institute (Ireland)

Domain specific data portals are primarily focused on providing access to datasets that serve the needs of specific scientific communities. While scientists may be able to confidently navigate through their own community's data portals if they are aware of the existence of the portals, discovery of portals and combined searching across those portals is often difficult. The common method of finding information of using a web search engine hasn't previously delivered expected results because environmental datasets have not been marked up in a way for the metadata to be indexed and therefore available in search engine results.

In the European marine domain, SeaDataNet (SDN) is a major operational infrastructure for managing, indexing and providing access to ocean and marine data sets and data products, acquired by European organisations from research cruises and other observational activities in European coastal marine waters, regional seas and the global ocean. SDN was one of the early adopters of publishing standardised metadata on the web using controlled vocabularies (published on the NERC Vocabulary Server) to annotate their datasets and data collection activities as a series of metadata catalogues. These metadata catalogues group data collection activities to describe projects, organisations, cruise summary reports, datasets (European Directory of Marine Environmental Datasets (EDMED)) and observing systems.

In this presentation we are focusing on the datasets that are exposed on the web through the EDMED catalogue. EDMED has long been available online through an individual web based search interface. To enhance and standardise the publication and delivery of EDMED metadata it was made available as Linked Data via a SPARQL endpoint described by the W3C Data Catalog Vocabulary [1] earlier this year. Here we describe how we complemented this effort by publishing EDMED datasets as structured metadata using the schema.org/Dataset standard to describe salient properties of each dataset in order for them to be discoverable by Google's dataset search engine and link to other Google resources. We also discuss the importance of good metadata provided by data publishers because ultimately a search engine's results are only as good as the metadata with which data publishers are able to mark up their datasets for discovery.

References

1. Wood C., Kokkinaki A., Leadbetter A., Thomas R. Exposing the SeaDataNet metadata catalogues via SPARQL endpoints, *Bollettino di Geofisica Teorica ed Applicata*. 57, 37, Proceedings of IMDIS 2018.