

A modular approach to cataloguing marine science data

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

Will Meaney, Marine Institute (Ireland), will.meaney@marine.ie

Elizabeth Tray, Marine Institute (Ireland), elizabeth.tray@marine.ie

Andrew Conway, Marine Institute (Ireland), andrew.conway@marine.ie

Sarah Flynn, Marine Institute (Ireland), sarah.flynn@marine.ie

Tara Keena, Marine Institute (Ireland), tara.keena@marine.ie

Caoimhín Kelly, Marine Institute (Ireland), caoimhin.kelly@marine.ie

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

The ability to access and search metadata for marine science data is a key requirement for answering fundamental principles of data management (making data Findable, Accessible, Interoperable and Reusable) (Wilkinson, *et al.*, 2016). In order to address the requirements of the Findable aspect of FAIR data, a dataset needs to be described by rich metadata in a searchable resource and the dataset must be assigned a clearly labelled persistent, unique identifier. Additional requirements include meeting domain-specific, community defined standards and legislative requirements placed on data publishers, for example maintaining semantic heterogeneity in cross-disciplinary domains, such as Marine Spatial Planning (MSP).

Therefore, in the sphere of oceanographic data management, the need for a modular approach to data cataloguing which is designed to meet a number of requirements has been demonstrated. According to Friddell, *et al.*, 2014, modularity is required in order to represent datasets, projects or programmes and other polar data resources within the catalogue system in other cross-disciplinary topics.

This modular approach has been adopted by the Marine Institute Ireland (Leadbetter, *et al.*, 2020) in developing a data cataloguing system to meet the needs of legislative requirements including the European Spatial Data Infrastructure (INSPIRE) and the Marine Spatial Planning directive. The data catalogue described here makes use of a metadata model focused on oceanographic-domain. It comprises a number of key classes which will be described in detail in the paper, but which include:

- Dataset - combine many different parameters, collected at multiple times and locations, using different instruments
- Dataset Collection - provides a link between a Dataset Collection Activity and a Dataset, as well as linking to the Device(s) used to sample the environment for a given range of parameters. An example of a Dataset Collection may be the Conductivity-Temperature-Depth profiles taken on a research vessel survey allowing the individual sensors to be connected to the activity and the calibration of those sensors to be connected with the associated measurements.
- Dataset Collection Activity - a specialised dataset to cover such activities as research vessel cruises; or the deployments of moored buoys at specific locations for given time periods
- Platform - an entity from which observations may be made, such as a research vessel or a satellite
- Programme - represents a formally recognized scientific effort receiving significant funding, requiring large scale coordination
- Device - aimed at providing enough metadata for a given instance of an instrument to provide a skeleton SensorML record
- Organisation - captures the details of research institutes, data holding centres, monitoring agencies, governmental and private organisations, that are in one way or another engaged in oceanographic and marine research activities, data & information management and/or data acquisition activities

The data model (figure 1) makes extensive use of controlled vocabularies to ensure both consistency and interoperability in the content of attribute fields for the Classes outlined above.

The data model has been implemented in a module for the Drupal open-source web content management system, where ISO19115/19139 compliant XML records can be exported, and made available. Publically available metadata XML records are harvested by a GeoNetwork, which comprises the Marine Institute's public facing data catalogue available at <http://data.marine.ie>. Over 400 ISO19115/19139 compliant XML datasets are published on the Marine Institute's data catalogue.

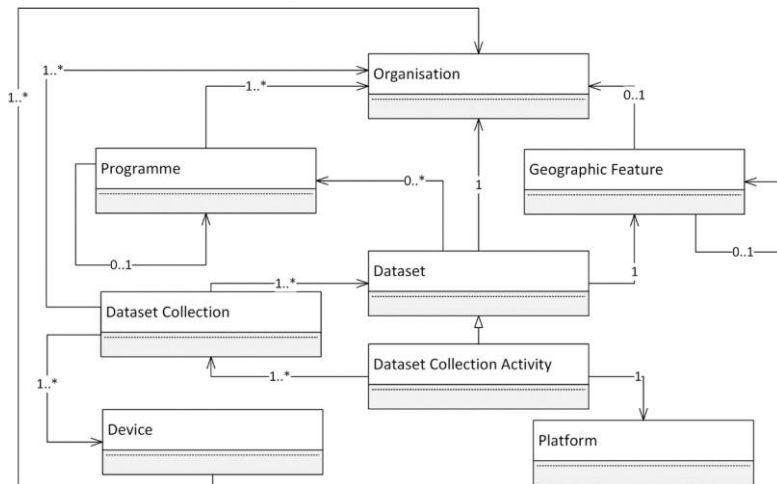


Figure 1: A high-level overview of the data model used in the modular data catalogue approach. The overall class structure is shown in the Unified Modelling Language

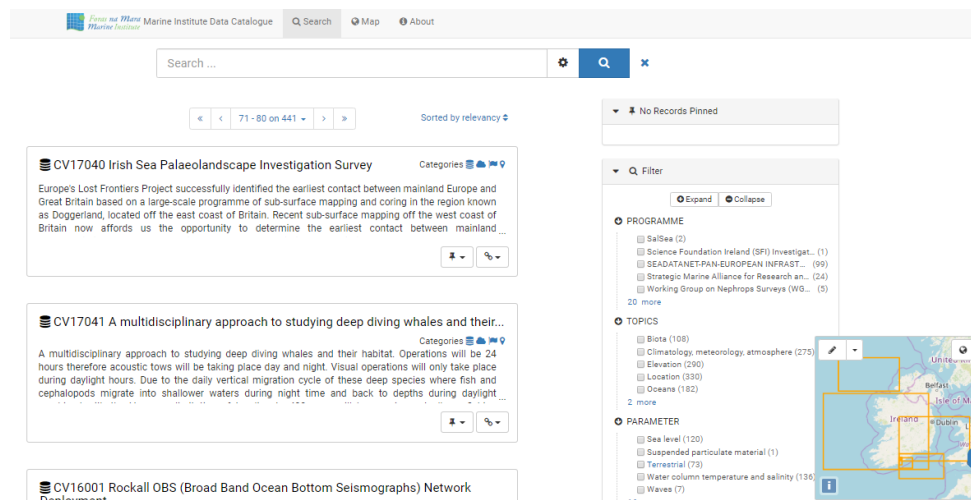


Figure 2: Marine Institute's Metadata Catalogue available at data.marine.ie

References

Friddell JE, LeDrew EF, Vincent WF (2014) The Polar Data Catalogue: best practices for sharing and archiving Canada's polar data. *Data Science Journal*: IFPDA-01

Leadbetter, A., Meaney, W., Tray, E., Conway, A., Flynn, S., Keena, T. Kelly, C. and Thomas, R. (2020). A modular approach to cataloguing marine science data. *Earth Science Informatics*. <https://doi.org/10.1007/s12145-020-00445-w>

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*. <https://doi.org/10.1038/sdata.2016>