

EGU2020-9750

<https://doi.org/10.5194/egusphere-egu2020-9750>

EGU General Assembly 2020

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



A modular approach to cataloguing oceanographic data

Adam Leadbetter¹, Andrew Conway¹, Sarah Flynn¹, **Tara Keena**¹, Will Meaney, Elizabeth Tray², and Rob Thomas¹

¹Marine Institute, Ocean Science & Information Services, Oranmore, Ireland

²Marine and Freshwater Research Centre, Galway-Mayo Institute of Technology, Galway, Ireland

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) and also in meeting domain-specific, community defined standards and legislative requirements placed on data publishers. 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 can be clearly seen. In this paper we describe a data cataloguing system developed at and in use at the Marine Institute, Ireland 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 focussed 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 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, and the paper will provide details of this application.