Introduction

The Biological and Chemical Oceanography Data Management Office (BCO-DMO, http://www.bco-dmo.org/) was created to serve principal investigators funded by the National Science Foundation (NSF) Biological and Chemical Oceanography Sections and Division of Polar Programs Antarctic Organisms & Ecosystems Program as a facility where marine biogeochemical and ecological data and information developed in the course of scientific research can easily be disseminated, protected, and stored on short and intermediate time-frames. The Data Management Office also provides research scientists and others with the tools and systems necessary to discover, work with, and reuse marine biogeochemical and ecological data from heterogeneous sources with increased efficacy.

In the effort to enable better data discoverability, BCO-DMO considers the metadata of these datasets as a valuable asset. The metadata concepts that BCO-DMO captures include funding information such as agency, award, and project descriptions, associated persons, datasets, research cruises, instrumentation, measurements and geospatial and temporal elements. This metadata is semantically modeled with a local ontology. However, this ontology is too local to be effective for discovery and higher-level ontologies have been too obscure for proper implementation and coverage of the metadata concepts. The Data Catalog Vocabulary (DCAT) offers a lightweight pattern for thoroughly modeling BCO-DMO datasets while providing a standard vocabulary for aiding broad discovery. It is the recommendation of BCO-DMO that the DCAT vocabulary become a W3C standard.

Intended Use of DCAT Classes & Properties

BCO-DMO plans to align its local ontology with the DCAT vocabulary modeling over 6,500 unique datasets, each with multiple distributions spanning over 250 different NSF-funded research projects. Every BCO-DMO dataset will be typed as a dcat:Dataset with plans to use its following fields:

    dct:title
    dct:description
    dct:issued
    dct:modified
    dct:language
    dct:publisher
    dct:identifier
    dct:temporal
    dcat:theme
    dcat:keyword
    dcat:contactPoint
    dcat:landingPage
    foaf:homepage
    dcat:distribution

For each BCO-DMO dataset, there exists at least one distribution which will be typed as a dcat:Distribution with plans to use its following fields:

> dct:title
> dct:description
> dct:issued
> dct:modified
> dcat:accessURL (in some cases)
> dcat:downloadURL (in most cases)
> dcat:mediaType
> dct:format

The most attractive aspect of DCAT was the Catalog class. It is desirable for BCO-DMO to present a collection of certain datasets for certain perspectives. Typically, when users search for BCO-DMO data, they often begin those searches with knowledge about one or more of the modeled metadata concepts. For example, a user may know that they are interested in all data generated by a certain investigator or from a certain project, funding source, research cruise or type of instrument. The DCAT Catalog class provides BCO-DMO a way to group related datasets into collections that will make sense and provide a richer discovery experience for users. For these concepts typed as a dcat:Catalog, BCO-DMO plans to use its following fields:

> dct:title
> dct:description
> dct:issued
> dct:modified
> dct:language
> foaf:homepage
> dct:publisher
> dct:spatial (in some cases)
> dcat:themeTaxonomy
> dcat:dataset
> dcat:record

Finally, since BCO-DMO keeps revision history on its datasets, it plans to use the dcat:CatalogRecord class to expose certain BCO-DMO dataset metadata. The CatalogRecord fields planned for use are the following:

> dct:title
> dct:description
> dct:issued
> dct:modified
> foaf:primaryTopic