Standardizing-Marine-Biological-Data

Brett Johnson

2020-04-08

Contents

Biological data structures, definitions, measurements, and linkages are neccessarily as diverse as the systems they represent. This presents a real challenge when integrating data across biological research domains such as ecology, oceanography, fisheries, and climate sciences.

Chapter 1

Introduction

This is about stacking the right standards for your desired ineroperability with other data types. For example, interopating fish biology measurements with climate level variables. There are a few links neccessary to make this possible. This will permit ecosystem based models.

1.1 Data Structures

The OBIS-ENV Darwin Core Archive Data Structure.

OBIS manual

1.2 Ontologies

An ontology is a classification system for establishing a hierarchically related set of concepts. Concepts are often terms from controlled vocabularies.

From Marine Metadata:

"Ontologies can include all of the following, but are not required to include them, depending on which perspective from above you adhere to:

Classes (general things, types of things) Instances (individual things) Relationships among things Properties of things Functions, processes, constraints, and rules relating to things"

Unified Modeling Language?

1.3 Controlled Vocabularies

There are a number of controlled vocabularies that are used to describe parameters commonly used in a research domain. This allows for greater interoperability of data sets.

- Climate and Format (CF) Standard Names are applied to sensors for application with OPeNDAP web service.
- Device categories using the SeaDataNet device categories in NERC 2.0
- Device make/model using the SeaVoX Device Catalogue in NERC 2.0,
- Platform categories using SeaVoX Platform Categories in NERC 2.0
- Platform instances using the ICES Platform Codes in NERC 2.0
- Unit of measure
- GCMD Keywords
- Geographic Domain/Features of Interest

There are numberous ways to investigate which controlled vocabulary to use and this can be fairly overwhelming. For a simplified overview see here.

Note: To describe a measurement or fact of a biological specimen that conforms to Darwin Core standards, it's neccessary to use the 'Biological entity described elsewhere' method rather than taxon specific.

1.3.1 Collections

1.3.2 Oceanography

Biological and Chemcial Oceanography Data Management Office

Marine metadata interoperability vocab resources

1.3.3 Biology

BioPortal Ecosystem Ontology

1.3.4 NERC Search Interfaces

- SeaDatanet Common Vocab Search Interface:
- SeaDataNet Common Vocabularies:
- SeaDataNet Vocab Library

1.3.5 Geosciences

'UDUNITS' are more common in geosciences

UDUNITS

1.3.6 Eco/EnvO

Environment Ontology including genomics.

1.3.7 Wild Cards

P01 Biological Entity Parameter Code Builder

1.4 Technologies

1.4.1 ERDDAP

ERDDAP provides 'easier access to scientific data' by providing a consistent interface that aggregates many disparate data sources. It does this by providing translation services between many common file types for gridded arrarys ('net CDF' files) and tabular data (spreadsheets). Data access is also made easier because it unifies different types of data servers and access protocols. Here is a basic erddap installation that walks you through how to load a data set.

1.4.2 Semantic Web and Darwin Core

Lessons learned from adapting the Darwin Core vocabulary standard for use in RDF

1.4.3 Resource Description Framework

Darwin Core Resource Description Framework Guide

Chapter 2

Literature

Chapter 3

Methods

We describe our methods in this chapter.

Chapter 4

Applications

Some *significant* applications are demonstrated in this chapter.

4.1 Example one

4.2 Example two

Chapter 5

Final Words

We have finished a nice book.