A Semantic Framework for Climate Metadata Interoperability
M. Benno Blumenthal, John del Corral, and Haibo Liu
International Research Institute for Climate and Society/Columbia University
Monell Building Palisades NY 10964

Dan Holloway and Nathan Potter
OpEndAP, Inc

The Semantic Web provides a single framework that allows describing datasets according to multiple standards, creating a more complete description of any data model implied in programs that display and manipulate data. Writing Models, Crosswalks, and Objects all with SWRL/IRIDL/IRIDL allows that these data models and metadata standards can be interrelated in a single framework, leading to interoperability.

Crosswalking between different standards can be as simple as two different names for the same variable, but more often the mapping gets more complicated. Frequently, different objects are related conceptually but are very different structurally. Any framework that has both structure and semantic content is needed. A semantic content framework can be thought of as a structure where the semantic content is represented as a set of statements which describe the collaborations of a CF convention (e.g. CF variables define a grid); the objects which define the conceptual objects represented in the convention, e.g. cfsn which describes the more abstract objects (like pre-loaded data) that are being described in the CF convention. XML Schema is a common way to transform CF data to RDF/OWL which allows us to create semantic XML files from RDF information. We have applied this to the WMS schema for example, to extend the needed information for an OpenSearch/WSE service based on RDF extracted from Climatolod files. We also have included controlled vocabularies such as CF-terms, and a more detailed semantic description of datasets and metadata that our data can be found in the corresponding RDF information.

Crosswalks, Models and Objects

Semantic Transformation

Faceted Search

Datasets

IRI RDF Architecture

The RDF architecture is implemented by having a Start Point document that points to enough RDF documents that all the information necessary is found following links. This information is needed and kept virtual to allow additional rules and the results added until a more virtual triples are generated. Queries are then done on that store, resulting in web pages or virtual documents.

Non-RDF documents are imported by applying SWRL Transformations.

Summary

Our working example is composed of datasets and metadata in the IRIDL (LDEO Climate Data Library) and OpenSearch metadata that describe the data. The data is described in a variety of languages, e.g. CF-terms, metadata, and independent variables. These datasets have been provided under distance frameworks that have varying levels of semantic generation. All of the data is described in a variety of languages and frameworks, some way to structure classifications, and important for us to build a faceted search that works across diverse data.