Session: C21 Poster: T248A

Semantic framework for climate metadata interoperability

Benno Blumenthal[†]; John del Corral; Haibo Liu; Daniel Holloway; Nathan Potter
[†] Columbia University, USA
Leading author: benno@iri.columbia.edu

The Semantic Web provides a single framework that allows describing datasets according to multiple standards, creating a more complete description than any single standard provides. Going beyond standards, it can explicitly describe the data models implicit in programs that display and manipulate data. Writing Models, Crosswalks , and Objects all within RDF/SemanticWeb means 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 quantity, but sooner or later the mapping gets more complicated. Frequently, different objects are related conceptually but are very different structurally. Our framework thus has both structure and conceptual models: structure models that describe how dataset metadata is written (e.g. cf-att which describes the attributes of a CF convention netcdf file), and conceptual models which describe the conceptual objects represented in the convention, e.g. cf-obj which describes the more abstract objects (like geo-located data) that are being described in the CF convention. XML Schema is a common way to represent structure models for XML files, and we have a translation of XML Schema to RDF/OWL which allows us to create conforming XML files from RDF information. We have applied this to the WCS Schema, for example, to extract the needed information for an OPeNDAP WCS service based on RDF extracted from CF/netcdf files. We also have included controlled vocabularies such as CF standard names or GCMD scientific parameters. Controlled vocabularies are a common way to structure classifications, and important for us to build a faceted search that works across diverse datasets. Our working example is composed of the datasets and some of the metadata in the IRI/LDEO Climate Data Library (http://iridl.ldeo.columbia.edu). These data services enable access and analysis by providing data in a framework which allows format translation, rendering, and application of a variety of analysis functions, including sampling, averaging, regridding, EOFs, and statistical operators. Datasets are both local and remote, allowing a federation of data servers to appear in a uniform space of data access and functionality. Describing the library's contents requires concepts like datasets, units, dependent variables, and independent variables. These datasets have been provided under diverse frameworks that have varied levels of associated metadata. We have created an RDF expression of a taxonomy that forms the basis of a dynamic earth data search interface. The concepts include location, time, quantity, realm, author, and institution. We have also started cross-walking these metadata into various existing metadata schema, so that our data can be found in the corresponding systems. A persistence framework incorporating inference and crawling is used to ingest the metadata information for a specified starting point as well as infer the connections between the diverse data-oriented concepts of the data library and the conceptual framework of the data search. This persistence framework includes inferred crawling and rule construction, OWL/SWRL as well as custom SeRQL construct rule inferencing, and XSL Transform on ingest (both GRDDL/RDFa based, and inferred from RDF information).