A Bi-directional Semantic Framework for Bibliographic Metadata

John del Corral and Benno Blumenthal

Columbia University, International Research Institute for Climate and Society, Palisades, United States
(jdcorral@iri.columbia.edu)

Often in a research community, like the geo-sciences, there are several different metadata standards used to describe the same type of objects. In this paper we explore using an RDF(Resource Description Framework)/XML framework to address this issue, and create a flexible, reusable solution that can adapt to new metadata standards. We create a semantic framework that contains multiple metadata views of the objects that we wish to describe. The framework is established by creating ontologies for each metadata representation of these objects, and rule-based crosswalks between them so that each object is expressed in all representations, thus all objects can be viewed in multiple systems. For our case study, we chose bibliographic references.

Our search for a geo-science metadata framework began with looking at fixed standards like FGDC, GCMD, and ISO. Soon, questions arose like 'Which standard should we choose?' and 'What happens when we try to describe a dataset or phenomena with characteristics that do not fit into the standard we have chosen?' and 'How do we share metadata with other institutions that may, or may not support the standard we have chosen?'. In many ways, the people who distribute information on the internet are faced with a similar metadata problem. Semantic technologies emerged as a viable solution for us.

The vision of the Semantic Web includes embracing the many ways information appears on the Web, and rendering that information so that, despite a variety of sources, it is understandable to client applications. RDF/XML and Web Ontology Language (OWL) are designed to allow one to express multiple schema in a single framework. Here we create a concrete realization of that vision by using emerging semantic technologies to create a framework where all the objects can be accessed uniformly using two distinctly different metadata structures, independent of which structure is used to initially describe the object.

Most bibliographic references are described using one of two accepted metadata standards, Endnote or BibTeX. The remaining references are usually listed in an ad hoc style. How do we describe bibliographic reference objects in a unified way? We need a mechanism that will keep track of different terms describing the same type of individual object.

OWL ontologies are able to describe bibliographic references with classes, sub-classes, object properties, datatype properties and annotation properties. We build an ontology for each metadata standard that we want to represent. RDF/XML editors, API's, and triple-store databases are used to accomplish the tasks that we need for our case study.

We create an RDF triple-store database where all references can be stored, no matter where they come from. Then we provide multiple ways of viewing and querying these references, which is accessible from a web browser or by an RDF aware client.