Geophysical Research Abstracts
Vol. 12, EGU2010-11783, 2010
EGU General Assembly 2010

Using an Adiabatic Translation of XML Schema to OWL/RDF to generate OpenGIS Web Coverage Service (WCS) Descriptions

Martin Benno Blumenthal (1), Haibo Liu (1), Nathan Potter (2), and John del Corral (1)

(1) Columbia University, International Research Institute for Climate and Society, United States (benno@iri.columbia.edu),
(2) OPeNDAP Inc., 165 Dean Knauss Dr., Narragansett RI 02882 USA (ndp@opendap.org)

Often in a research community there are several different metadata standards used to describe the same data object. Separate infrastructures are build for each standard, making it difficult to get a unified perspective on the data object. We have chosen to use a RDF (Resource Description Framework)/XML framework to address this issue, to 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 each object can be viewed in multiple systems.

Many existing data description frameworks are expressed in XML Schema. These schema tend to not be the most accurate rendition of the underlying conceptual model, but working with the XML Schema allows us to leverage the considerable human effort that has gone into writing the schema, and to leverage the existing codebase written to handle the data expressed according to the schema.

To take advantage of this prior work, we create a translation where the XML Schema files are translated to OWL/RDF , and the corresponding XML instance files are adiabatically translated to RDF, i.e. the XML can be recreated from the triple store using a fixed recursive query. This stores the multiple versions of the information in a single framework, yet we are able to extract the XML documents corresponding to each individual schema. This framework allows us to use an RDF inference scheme to map between different systems. In particular, we can take the OPeNDAP metadata as represented in RDF, apply inference rules to add the appropriate triples for the WCS representation, and generate the XML Document corresponding to the data as a WCS CoverageDescription.

This mapping from RDF objects to XMLElements is a profound worldview change. Structurally, XML expresses in context. This is manifest in a number of ways. For example, consider the Java JDOM API, a standard way of extracting information from an XML file, where the primary way one finds elements is via the name that the parent element gives it, and there is no concept of a node belonging to two parts of the tree. Thus all information is strictly local to the parent element. There are, of course, conventions to make up for this, i.e. xlink, so that nodes with different parents can be related.

On the other hand, RDF gives every node an id, making it possible to link nodes in many different ways, and RDF thus decontextualizes all the information; identity is global, not related to a parent. This too is a bit of an exaggeration, since blank nodes in RDF are intended to be referenced by their properties: in particular, dumping and restoring a triple-store changes all the blank-node ids, i.e. the ids cannot be used externally under many circumstances.

So this method of extracting XML from an RDF triple-store means we are going to an extraction with context from a RDF triple-store that does not natively have context. This complexity is essential because what we want is to store multiple XML versions of the same thing, and map between them in the RDF context (mapping between requires decontextualization). This means that our RDF store will have many more properties for an object with multiple representations than are relevant for any particular XML extraction, but our extraction schema will pull out the right ones for any particular context.

Here we present the conceptual framework underpinning such a system. A brief overview of the portion of XML Schema that we represent in OWL/RDF, how XML Elements are mapped into RDF objects, and the queries necessary to extract the XML information in context from the context-less representation in the RDF triple store.