Proc. Int. Conf. on Dublin Core and Metadata for e-Communities 2002: 213-215

Abstraction versus Implementation: Issues in Formalizing the NIEHS Application Profile

Corey A Harper
Knight Library, University of Oregon
harpc@ils.unc.edu

Jane Greenberg
University of North Carolina at Chapel Hill
janeg@ils.unc.edu

W. Davenport Robertson, Ellen Leadem
National Institute of Environmental Health Sciences
{robert11, leadem} @niehs.nih.gov

Version 1 of the National Institute of Environmental Health Sciences Metadata Schema

The National Institute of Environment Health Sciences (NIEHS), is an Institute of the National Institutes of Health (NIH), which is a component of the U.S. Department of Health and Human Services (DHHS). As with many governmental organizations, the NIEHS website contains a rich and growing body of important resources for both employees and the general public. The NIEHS library has spearheaded an organization-wide metadata project to provide better access to these resources. Dublin Core was elected for the NIEHS metadata project because it supports semantic interoperability and the potential for data sharing, and because the schema is simple enough to support author-generated metadata. This paper and corresponding poster document issues in formalizing the NIEHS Application Profile, specifically the changes implemented between Version 1 to Version 2, which were influenced by revisions made to the NIEHS metadata form supporting author-generated metadata.

Version 1 was comprised of twenty metadata elements, the majority of which were drawn from the basic Dublin Core Metadata Element Set version 1.1, and the expanded set of Dublin Core Qualifiers (Robertson et al. 2001). NIEHS' Version 1 corresponded quite closely with the Dublin Core semantics. Exceptions included merging the *Creator* element with the *Contributor* element, incorporating an *Audience* element from the Gateway to Educational Materials (GEM) namespace. Version 1 of the NIEHS Metadata Schema was an application profile, in the rough sense of the word, but discrepancies in the formal schema and the public schema, which supported the NIEHS metadata form for author metadata creation, delayed official formalization, until version 2 of this schema.

Version 2: The NIEHS Metadata Schema Becomes an Application Profile

The movement towards an application profile involved resolving discrepancies between the NIEHS' formal schema and the public schema made accessible via the metadata template, identifying the namespaces that elements and qualifiers are derived from, and establishing rules concerning the usage of elements such as the obligation (whether an element's inclusion is required) and cardinality (repeatability of a given element). The most significant questions addresses in defining and documenting a schema were: 1) How many elements are needed? 2) which documents should define element usage rules? and 3) How the documents should be serialized? The following sections correspond to these three issues.

One Element or Two?

The first main step to revising and formalizing the NIEHS application profile was to determine how many metadata elements were needed for a document attribute. Version 1 of the NIEHS metadata schema identified separate elements for *Date Created*

and *Date modified*, as well as for *URL*, *NIEHS Number* and *Other Identifier*; *Alternative Title* was listed as a separate element despite being a standard Dublin Core Qualifier. These six elements could be condensed to *Date*, *Identifier* and *Title* by applying appropriate refinement qualifiers, as was done with *Coverage*, which was refined by *Time* or *Date* qualifiers.

In Version 2 the six elements highlighted above were reduced to four corresponding Dublin Core elements: *Date*, *Coverage*, *Identifier* and *Title* refined by qualifiers. Many of the qualifiers used are defined in the Dublin Core Qualified (DCQ) recommendation. *URL*, *NIEHS Number* and *Other Identifier* are unique to the NIEHS metadata project, although applied to an official Dublin Core element.

Namespace versus Application Profile

The Dublin Core Namespace policy, by defining the Dublin Core Metadata Initiative (DCMI) 'Terms' namespace as containing "DCMI elements and DCMI qualifiers (other than those elements defined in the Dublin Core Metadata Element Set [DCMES], Version 1.1)", sets a precedent for defining new elements *and* qualifiers in a single namespace document (Powell and Wagner, 2001, p. 2). This provided justification for the inclusion of any element, refinement qualifier or encoding scheme *that is completely unique to the NIEHS metadata project* in the NIEHS namespace.

After the determination of element representation (one element or two), the NIEHS metadata team determined that *Author/Contributor* was the only element unique enough to warrant a new element definition. Rather than defining a new element combining the Relation and Source elements, the NIEHS namespace defines two refinement qualifiers ('NIEHS Is original source of' and 'NIEHS Has original source') that function like those defined in the DC Terms namespace. A similar issue arose when determining which Encoding Scheme Qualifiers were unique enough to include in the NIEHS namespace. All the value lists used by the project, except the list applied to the Audience element, were derived from existing value lists. It was decided that altered versions of existing lists should also be defined in the NIEHS namespace document.

Serialization

The third major issue encountered involved the serialization of an element set. XML (Extensible Markup Language) serializations of an element set are useful for providing a machine-readable representation for a local context, such as a template for author generation of metadata. RDF (Resource Description Framework) serializations support semantic interoperability by providing explicit standards for combining varied element sets.

The DCMI registries working group identifies providing "access to DCMI approved domain specific 'application profiles' e.g. the DCMI Education group application profile" as a high priority for the next phase of the DC registry (Heery, 2001, p. 8). In it's current state, this registry contains Resource Description Framework Schemas (RDFS) that define the components of the various DC namespaces. In adherence with this practice, the NIEHS metadata team will encode its application profiles using RDFS. This will ensure that element sets are serialized consistently across initiatives interested in interoperability with NIEHS and vice/versa.

Conclusions and Future Work

Formalizing the NIEHS application profile was both an intellectual and practical undertaking. As part of the process we discovered that elements incorporated into an application profile can be modified in at least four distinct ways: 1) algorithms can automatically supply metadata (e.g., assign default values or extract information from HTML or XML source code), 2) element designations in a formal schema can be modified to facilitate author and searcher (non-expert) understanding in a metadata template or search engine; 3) cardinality and obligation constraints can be provided; and 4) new qualifiers, or qualifiers from alternate element sets can be applied to existing elements in unique ways. We advocate defining new elements and qualifiers in namespaces modeled on RDF.

Among one of the most significant challenges the NIEHS metadata team needs to address now is how to best codify element cardinality and obligation. RDF *does not* provide any mechanism for this need; and although locally defined XML DTD (document type definition) permit documentation, the local nature provides a significant barrier to interoperability. DAML+OIL (DARPA Agent Markup Language/ Ontology Inference Layer) Reference Description (http://www.w3.org/TR/daml+oil-reference), which is evolving into OWL (Web Ontology Language), provides a standardized mechanism for declaring cardinality constraints. Additionally, these languages permit a layering on top of RDFS and adhere to the dumb-down principle used by Qualified Dublin Core. Query engines and processors 'understanding' DAML+OIL, can extract constraints and understand the intended meaning of attributes from a representation, otherwise, they can pass over and interpret the remainder of the representation as long as the RDF syntax is intact. Here in lies a topic of inquiry and experimentation from the next phase of the NIEHE metadata project. In closing, the topics addressed in this paper can help to better inform the future development of application profiles and name-

spaces, working toward an interoperable environment, one that supports the growth of the Semantic Web.

References

Heery, R. 2001, Draft DCMI Open Registry Functional Requirements. Dublin Core Metadata Initiative. [Online} Available at: http://dublincore.org/groups/registry/fun_req_ph1-20011031.shtml

Powell, A and Wagner, H. 2001, Namespace Policy for the Dublin Core Metadata Initiative (DCMI). Dublin Core Metadata Initiative. [Online] Available at: http://dublincore.org/documents/2001/10/26/dcmi-namespace/

Robertson, W.D., Leadem, E.M., Dube, J. and Greenberg, J. 2001, Design and Implementation of the National Institute of Environmental Health Sciences Dublin Core Metadata Schema. In: Proceedings of the International Conference on Dublin Core and Metadata Applications 2001. Tokyo, Japan. National Institute of Informatics. p. 193-199. [Online] Available at: http://www.nii.ac.jp/dc2001/proceedings/product/paper-29.pdf