Metadata Co-Development: A Process Resulting in Metadata about Technical Assistance to Educators

Michael Knapp
Green River Data Analysis
2253 Stage Rd
Guilford, VT 05301, USA
mknapp@greenriver.org

Sara Dexter
University of Nevada
4505 Maryland Parkway
Box 453005
Las Vegas, NV 89154, USA
sdexter@unlv.nevada.edu

Robert McLaughlin
International Graduate Center
4093 Diamond Ruby, Suite 7
Christiansted, St. Croix,
00820, VI
mclaughb@nici-mc2.org

ABSTRACT

Metadata development can be challenging because the vocabulary should be flexible and extensible, widely applicable, interoperable, and both machine and human readable. We describe how we engaged members of organizations in the field of technical assistance to educators in a process of metadata development, and the challenges we faced. The result was a an ontology for the communities of practice that is interoperable and can evolve; it was then used to catalogue resources for dissemination via the Semantic Web.

Categories and Subject Descriptors

K.3.1 [**Computing Milieux**]: Computer Uses in Education; E.0 [**Data**]: General

General Terms

Human Factors, Standardization, Languages

Keywords

Semantic Web, metadata, RDF, resource cataloging, education, technical assistance

1. INTRODUCTION

The Semantic Web offers much promise for users, potentially allowing them to better cooperate with data on computers and thus share and process it so that automation features can be put to work. Metadata to define and link information on the web is essential for the Semantic Web to exist, yet its development can be extremely challenging because the vocabulary should be flexible and extensible, widely applicable, and both machine and human readable.

It is important that computer science educators think creatively about how to efficiently develop metadata. This not only furthers the field of education technically, it also ensurses that knowledge domains not be excluded from the Semantic Web. That is, without an ontology to describe them, knowledge domains could become orphans by not coming onto the Semantic Web. The existing standards for education metadata have been constructed to provide the field of education with resources for instruction and thus refer to considerations like "Interactivity Level" (grade level of intended audience) and "Interactivity Type" (type of interactivity with learning resource) [4]. In contrast with courseware, educational research and technical assistance has yet to develop an ontology. Without metadata, educators will not benefit from the Semantic Web in areas such as technical assistance.

We describe our efforts with non-technical organizations to bring their work onto the Semantic Web. Specifically, we describe our work with numerous individuals and organizations with expertise in technical assistance education topics such as professional development, urban teacher preparation, equity, and school improvement to develop and refine an ontology. The resulting metadata was used to tag resources in a Semantic Web application in order to disseminate articles, reports, conferences and other resources. Here, we report the process of metadata development with these organizations, its outcomes, and the challenges of this work.

WWW2004, May 17–22, 2004, New York, New York, USA.
ACM 1-58113-912-8/04/0005.

2. THE METADATA DEVELOPMENT PROCESS

In order for the educatonal organizations we worked with to use our Semantic Web application to help their members access resources, we had to first work with them to develop and refine metadata. These organizations thought of the specialized collections of resources they wanted to disseminate in terms of their local knowledge. That is, their conceptual frameworks and descriptions had been developed over time as that community socially constructed knowledge together. Sometimes these ways of describing important ideas and resources were not explicit as a framework so they found it difficult at the outset of the metadata development process to upfront engineer a comprehensive and cohesive set of qualifiers and control vocabularies.

Because it was both time-consuming and frustrating for them when we tried to engineer aspects of a metadata vocabulary in advance, we switched to a "grounded" co-development process to generate the metadata. We use the term grounded because we started with the knowledge of these communities of practice and let the metadata vocabulary arise from it, so it could best convey their meanings. Our strategy was to let the organizations rely upon their own community's vocabulary for organizing resources and modify it over time as they saw fit. We wanted vocabularies to also aid the search of the resources by their eventual users, and so it was critical that the ontology we created across the communities allow each one to use descriptors relevant for its group of users. As subject headings would slowly evolved, they had to be combined into a meaningful hierarchy. As we engaged these educators in the process of knowledge representation, we had to refine our cataloguing tool to accommodate the constantly evolving schema.

When initiating development of our ontology in Spring 2001,

we decided to adopt the Dublin Core (DC) metadata set because it appeared to be the most widely recognized metadata standard for online resources as well as the de facto standard element set for the Resource Description Framework (RDF) [5]. At that time, the 15 elements and qualifiers in the DC provided a generalized framework for describing resources [2]. Since we recognized that DC was general in scope and would not provide sufficient detail on which to develop our application, we chose to leverage DC's modularity and extensibility that allows refined metadata to be added within the DC framework.

Following the recommendations of Duval et al., we began the process of developing our Education Reform Metadata Framework (hereafter referred to as EdRef) by selecting relevant elements from existing standards [7]. To DC's 15 basic elements and qualifiers we added the two DC education (DC-Ed) elements "Audience" and "Standard" [1]. We added elements and control vocabularies selected from the Gateway to Educational Materials (GEM) and Institute of Electrical and Electronics Engineers Learning Objects Metadata (IEEE LOM) [3] [4]. For example, our control vocabulary for the element "Resource Type" includes control vocabularies from GEM, IEEE LOM, and DC. Similarly, under "Audience" from DC-Ed, we included the qualifier "Typical Age Range" as derived from IEEE LOM. We added our own vocabulary to the ontology only if we could not locate any existing metadata to meet community needs.

Because we worked with communities from specialty areas of education we postulated that for each, highly specific metadata would be required to describe resources and organize collections according to their conceptual frameworks. We allowed for this specificity through the use of a qualifiers and control vocabularies. For example, we added control vocabulary under the DC-Ed qualifier for "Audience". We termed this control vocabulary "Expertise Values" because it allows organizations to present resources appropriate to the experience and knowledge of the learner.

A central task in building our metadata framework - one that is perpetually under development - is the generation of a list of new subject headings, which serve as control vocabularies for the "Subject and Keyword" elements. This task was necessary because neither Dewey nor Library of Congress headings offered a comprehensive vocabulary pertaining to the content of the organizations with whom we worked.

Technical assistance providers tend to need, and thus borrow, terms from a variety of fields including systems theory, organizational development, politics, finance, cognitive psychology, and equity. Our subject headings needed to combine, span, and integrate these fields of knowledge to create a comprehensive framework. We also had to develop new subject headings for another reason - the vocabularies for technical assistance need to address concepts and terms understood and used not only by researchers but also by practitioners and learners. It was important to organize the new subject headings we developed according to the most credible and widely recognized conceptual framework or set of standards. This approach offered the advantage that content could be organized according to issues of widespread concern to the communities we worked with. Our hope was that this better enables technical resource providers, (or funding agencies, policymakers, professional developers) to assess and target resources and respond to practitioners' needs.

Subject headings were constructed by participating communities with specialized expertise in each topic area. Over time, we used an increasingly limited subset of the metadata. Organizations were interested in adding only the metadata they needed to meet the immediate demands of the learners they supported. We found that the DC elements "Author", "Title", and "Description" were used extensively by the learners visiting the Web Site. The DC-Ed elements of "Standards" and "Audience" were used extensively. Many of the other elements and qualifiers we implemented were ignored or used inconsistently, suggesting they were not of relevance to this community at this time.

3. CONCLUSIONS

We involved organizations that provide technical assistance to educators in metadata development. We built an application that catalogues and presents research-validated recommendations made by these organizations within their own metadata framework. We offered these organizations a flexible framework for organizing information using metadata that they could evolve to keep relevant to the communities of practice that they served. This process contributes to the provisions on the Semantic Web of (1) content about effective professional practices, (2) resource providers available to assist educators in planning for and implementing these practices, and (3) research and evaluation data that attest to the efficacy of these practices and resource providers.

Metadata cannot be rigid and inflexible; they must evolve and be constantly refined [8] [6]. As new stakeholders use our software, they exert pressure to ensure that the metadata evolve. Since it is often easer to be redundant than to collaborate, useful Semantic Web applications must facilitate detecting and addressing duplication of metadata on a continual basis. Effective collaboration between organizations will remain a central challenge to realizing the vision of the Semantic Web. For us, operationally, this has meant the development of software that allows constant refinement of metadata categories to avoid overlap; allows these organizations to make incremental improvements to an integrated ontology; and minimizes the labor of cataloging, combining, and re-cataloging resources.

4. REFERENCES

[1] Dcmi education working group.

[2] Dublin core metadata initiative.

[3] Gateway to educational materials.

[4] Ieee learning technology standards committee.

[5] Resource description framework (rdf) model and syntax specification, Feb 1999.

[6] J. Dron. *Achieving Self-organisation in Network-based Learning Environments*. PhD thesis, 2002.

[7] E. Duval, W. Hodgins, S. Sutton, and S. L. Weibel. Metadata principles and practicalities. *D-Lib Magazine*, 8(4), Apr 2002.

[8] M. Nilsson, M. Palmr, and A. Naeve. Semantic web metadata for e-learning - some architectural guidelines. *W3C 2002 conference*, 2002.