

Managing Metadata with Resource Profiles

Position Paper presented to ADL's Learning Content Registries and Repositories Summit,
Apr. 13-14, 2010

Stephen Downes

National Research Council of Canada
Institute for Information Technology
Moncton, Canada
stephen.downes@nrc-cnrc.gc.ca

Abstract – Existing learning object metadata describing learning resources postulates descriptions contained in a single document. This document, typically authored in IEEE-LOM, is intended to be descriptively complete, that is, it is intended to contain all relevant metadata related to the resource. Based on my 2003 paper, Resource Profiles [1], an alternative approach is proposed. Any given resource may be described in any number of documents, each conforming to a specification relevant to the resource. A workflow is suggested whereby a resource profiles engine manages and combines this data, producing various views of the resource, or in other words, a set of resource profiles.

I. RESOURCES AND RESOURCE DESCRIPTIONS

Educators employ a wide variety of resources. Even in pre-internet days, educators would employ books and notes, classrooms, maps and diagrams, guest speakers, field trips, and more. In the internet era, educational resources can include all of these and more, including online resources, presentations animations, simulations, synchronous events, web quests, online mentoring, multi-user games, and more. Indeed, while some resources may be more or less pedagogically explicit, almost any resource may be used for educational purposes, and when it becomes so used, it becomes (by definition) an educational resource.

Though any resource, including non-digital resources, may be described digitally, using (for example) the Resource Description Framework, it should be apparent from their diversity that a single metadata profile will be inadequate to the task of describing the full range of educational resources. Moreover, even the attempt to encompass all required metadata in a single document produces an unwieldy, and mostly unused, set of elements. Moreover, there is no clear and obvious way to encompass the contributions of multiple authors, especially in the case of matters of opinion. Therefore, resources should not be described with a single document, but rather, with a set of documents, with each document addressing a particular aspect of the resource and authored by the person or entity in the best position to address that aspect.

II. THREE TYPES OF METADATA

For any given resource, a large number of types of metadata may be provided. Existing specifications are illustrative of the types of metadata that may be created. For example, Dublin Core provides documents with bibliographic metadata. VCard and similar formats suggest approaches to author metadata. EXIF and other media-specific data suggest approaches to technical metadata. This list could be extended indefinitely, but in general, there are three major types of metadata, identifiable by the distinct people or entities that author them.

First Party Metadata is metadata related to the creation and nature of the resource itself. It is authoritative metadata authored by the creator or the owner of the resource. One set of examples of first party metadata include bibliographic metadata, which describes the resource authorship, publication data, version sets or editions, and related information. Another set of examples of first party metadata includes technical metadata, describing the authoring tool, technical specifications and formats, appropriate player software, dimensions and size, and related information. A third type of first party metadata is licensing metadata, as described (say) in ODRL or Creative Commons.

Second Party Metadata is metadata related to the use of the resource. Second party metadata is authoritative when generated by the person or entity that actually uses the resource. Interestingly, while second party metadata is widely created and used on the internet, most of it is hidden and stored in proprietary formats. An excellent example of second party metadata is the Page Rank, named for Google founder Larry Page, which is a measure of the times a resource is accessed through a search, the number of times the resource is linked by other resources, and similar criteria. Another type of second party metadata is found in the form of server logs, which reveal access data, page referrers, software and computing environment used, and more. Another example is the Scholarly Works Usage Profile (SWUP) [2]. Second party educational metadata can include context of use (for example, in a course or program) and assessment data relative to the resource.

Third Party Metadata is metadata related to the evaluation, description or classification of resources. Third party metadata is typically authored by an entity or agency independent of both the resource author and potential resource clients or users. Such metadata would typically be created by librarians or reviewers and

is authoritative relative to the assessment board, classification board or archival agency. Content rating metadata, such as PICS, is an example of third party metadata. Classification and indexing data, including Library of Congress and Dewey Decimal classifications, constitute third party metadata (even if authored by the resource creator, as the resource creator may be non-authoritative with respect to classification). Educational metadata, such as semantic density, classification against educational standards or curricula, typical age range, and similar evaluative criteria, constitute third party metadata.

III. DISTRIBUTED METADATA AND RESOURCE IDENTIFIERS

As suggested above, it is expected that the metadata describing a resource may be located in multiple files, and hence, may be located in multiple locations. There is therefore a need to identify a single resource across a number of different files. This need exists independently of resource profiles, and typically one of two major approaches is used: either an identity-based approach, using a registry, such as Purl, DOI or Handle; or a location-based approach, such as URI. Obviously, a combination may be employed, as Handles, etc., can map to URIs [3].

That said, it does not follow that there must be one universal system for resource identifiers. Any given resource may have any number of identifiers, with identifiers created by specific agencies for particular purposes. By analogy, we can consider the case of people, who while they may have non-unique names or titles, may have any number of unique identifiers from specific agencies such as Social Insurance Numbers, driver's license numbers, passport numbers, and more. It is common for publishers to assign their own identifier, and common for repositories to assign unique identifiers for resources acquired from numerous publishers. An identity is just another piece of data, which has as a property a mechanism for accessing the resource it identifies.

IV. EDUCATIONAL METADATA

The specific purpose of an educational standards organization should be to specify that metadata unique to educational purposes. There are three major types of educational metadata: educational standards metadata, educational properties metadata, and educational use metadata.

Educational standards metadata describes a resource's relation to an educational standard. An educational standard may be described with curriculum metadata (for example, as documented on the old BECTA curriculum metadata page [4]), course description metadata [5], or competencies metadata, as for example employed by Metadata for Architectural Contents in Europe (MACE) [6]. The purpose of educational standards metadata is to map the current resource to one or more elements in an index or taxonomy, and is therefore typified by the Catalog-Entry fields found throughout IEEE-LOM.

Educational properties metadata describes properties of a resource that may be relevant to the selection of a resource. IEEE-LOM includes a number of educational properties metadata elements under the general Educational metadata heading, including

interactivity type, learning resource type, interactivity level, semantic density, intended end user role, typical age range and difficulty [7]. It should be apparent that these do not exhaust the list of potential educationally-relevant properties. Additionally, the value space provided in IEEE-LOM does not exhaust the possible desired set of value spaces.

Educational Use Metadata describes what is expected or intended to be the context of use of the resource. In IEEE-LOM some educational use metadata is specified in the Relation element. However, additional specifications, such as IMS Simple Sequencing and IMS Learning design, describe educational use in separate documents. Arguably, IMS Content Packaging is an additional form of educational use metadata. The implementation of a learning resource in a specific environment, with the application of specific system tools or resources, as described in Learning Tools Interoperability, also constitutes a source of information about the resource.

While the data produced or used to create these three types of metadata may be suggested or produced by the resource author or through the use of a resource, each of these forms of metadata depends on a third party evaluation of the resource in question. Such metadata therefore constitutes third party metadata, and is not regarded as authoritative if proceeding from a resource author, but rather only if it proceeds (or is verified) by a third party registrar or agency. You can tell the Library of Congress where you think your book belongs, but it is the librarian, not you, who decides.

V. USING RESOURCE PROFILES

The single-document approach to metadata suggests that what can be known about a learning resource can be collected in a single place and used as an a priori form of document indexing in a single repository or library. By this point it should be clear that learning resources need to be described by multiple entities, in multiple ways, using descriptions that are widely varied in nature, authorship and location.

The distributed model of metadata described in this paper is intended to be leveraged to form descriptions that are, first, current and regularly updated, and second, tailored to specific user needs. In this way, the resource profiles approach recognizes that the consumers of learning resource metadata have as many distinct perspectives, interests and needs as the authors themselves.

When a resource is newly created, very little is known about it. While the developer may express opinions about its applicability, difficulty or educational relevance, these are properties that can only be verified through experience. As the resource matures through use and repurposing, its metadata matures as well. If a resource proves to be popular, more and more sources of information become available – it was used here, it was reviewed there, it was linked over there. Thus, the aggregation and storage of metadata describing a given resource needs to be ongoing. This becomes especially important as a resource ages and may fall out-of-date. Recent information will be significantly more relevant than metadata produced on the day the resource was created.

When resource metadata is collected from various sources into a particular repository, specific resource *profiles* of that resource may be created by combining different elements of the metadata files. While a single all-encompassing profile could in theory be produced, by conjoining all elements of all files describing the resource, such a profile is neither anticipated nor desired. Rather, what is expected is that profiles corresponding to specific needs will be created by conjoining only selected elements from different metadata files. A programmer who is implementing a resource in a technical framework will want technical and educational use metadata, while a subject matter expert will be more interested in

second party metadata along with educational properties and educational use metadata.

A resource profiles enabled repository or data store, therefore, will be enabled with a resource profiles engine that selects elements from different files and combines them to form new, and possibly unique, descriptions of a given resource. The definition of a particular type of description is termed, in general, a 'profile'. Each profile may serve a particular purpose, and is composed of a set of one or more rules or procedures for the selection of data values from one or more possible types of input files, and one or more procedures or rules for the presentation of those values to the user.

REFERENCES

- [1] Stephen Downes, Resource Profiles, November 23, 2003. http://www.downes.ca/files/resource_profiles.htm
- [2] Benoit Pauwels, *Exchange of usage metadata in a network of institutional repositories: the case of Economists Online*, Academic Online Resources: Assessment and Usage International Symposium, Lille, 27 November 2009, Slideshare, <http://www.slideshare.net/bpauwels/exchange-of-usage-metadata-in-a-network-of-institutional-repositories-the-case-of-economists-online>
- [3] Andy Powell, Pete Johnston, Lorna Campbell, Phil Barker, *Guidelines for using resource identifiers in Dublin Core metadata and IEEE LOM*, Dublin Core Metadata Initiative, <http://www.ukoln.ac.uk/metadata/dcmi-ieee/identifiers/>
- [4] *Metadata guide for tagging*, Curriculum Online, BECTA, 7 November 2003. <http://industry.becta.org.uk/display.cfm?resID=40282>
- [5] M. Pezeril, *Course description metadata (CDM): A relevant standard for technology-supported learning*, Experience-BVased Quality in European ODL seminar, September 21, 2006, http://www.e-quality-eu.org/pdf/seminar/e-Quality_WS2_MPezeril_article.pdf
- [6] OUNL, with contributions from partners, Integration of Competence Metadata in MACE, 30 May 2008, <http://dspace.ou.nl/bitstream/1820/1764/1/Mace%20Deliverable%205.5%20-%20Integration%20Of%20Competence%20Metadata%20In%20Mace.pdf>
- [7] Wayne Hodgins, et.al., Draft Standard for Learning Object Metadata, IEEE 1484.12, 15 July, 2002, http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf