The Concepts of Metadata in E-Learning

Gholamali Montazer[1], and A. Bahreininejad[2]

Abstract
The rapid changes in the means of information access occasioned by the emergence of the World Wide Web have spawned an upheaval in the means of describing and managing information resources. Metadata is a primary tool in this work and an important link in the value chain of knowledge economies. Yet there is much confusion about how metadata should be integrated into information systems in particular e-learning. Based on the agreement shared by two prominent metadata Initiatives: the Dublin Core Metadata Initiative (DCMI) and the Institute for Electrical and Electronics Engineers (IEEE) Learning Object Metadata (LOM) Working Group, this paper describes the principles (those concepts judged to be common to all domains of metadata and which might inform the design of any metadata schema or application) and practicalities (the rules of thumb, constraints, and infrastructure issues that emerge from bringing theory into practice in the form of useful and sustainable systems).

1. Introduction

Metadata, literally "data about data," is descriptive information about a resource. For example, the card catalog in a public library is a collection of metadata. In the case of the card catalog, the metadata are the information stored on the cards about the Author, Title, and Publication Date of the book or resource (recording, etc.) in question. Metadata allow us to locate an item very quickly without investigating all the individual items through which we are searching.

The Learning Objects Metadata (LOM, 2002; LTSC, 2002) working group is working to create metadata for learning objects (such as Title, Author, Version, Format, etc.) so that people and computers will be able to find objects by searching, as opposed to browsing the entire digital library one object at a time until they find the right one. This paper presents the concepts of metadata and their applications in learning object development by outlining the principles and practicalities of metadata..

[1]. Assis. Prof. of Elec. Eng.
Email: montazer@modares.ac.ir
[2]. Assis. Of Computer Eng.

2. Principles of Metadata

The following sections describe the principles which set out the general truths that may provide a guiding framework for the development of practical solutions for semantic and machine interoperability in any domain using any set of metadata standards.

2.1 Modularity

Metadata modularity is a key organizing principle for environments characterized by vastly diverse sources of content, styles of content management, and approaches to resource description. It allows designers of metadata schemas to create new assemblies based on established metadata schemas and benefit from observed best practice, rather than reinventing elements again.

In a modular metadata world, data elements from different schemas as well as vocabularies and other building blocks can be combined in a syntactically and semantically interoperable way. Thus, application designers should be able to benefit from significant re-usability as they gather existing modules of metadata and stick them together much as individual components. Flexibility should be achievable in the metadata architecture of the Web, allowing application designers to mix a variety of semantic modules within a common syntactic foundation, even though the designers of the modules might not have anticipated a given combination. For example, a discovery metadata module and an instructional management metadata module, expressed in a common syntactic idiom such as XML, should be able to be combined in a compound schema that embodies the functionality of each constituent. In this way, modular sets can be assembled to meet the specific requirements of a given application, meeting domain-specific and local requirements without unduly sacrificing cross-domain interoperability.

2.2 Namespaces and metadata modularity

The notion of namespaces is a fundamental part of the infrastructure of the Web (and particularly XML (NAMES, 1991)), though the concept predates the Web and is familiar to most. Simply put, a namespace is a formal collection of terms managed according to a policy or algorithm. For example, the base protocol of the Web is HTTP, which is a namespace that guarantees that a given URI is globally unique. Any metadata element set is a namespace bounded by the rules and conventions determined by its maintenance agency. Namespace declarations allow the metadata schema designer to define the context for a particular term, thereby assuring that the term has a unique definition within the bounds of the declared namespace. Thus, the declaration of various namespaces within a block of metadata allows the elements within that metadata to be identified as belonging to one or another element set. Examples of metadata declaration are as follows:

1. The Dublin Core Metadata Initiative (DCMI, 2000) element set is defined at a Web location specified by a URI; all Dublin Core elements within the scope of this namespace declaration can be recognized by the prefix *dc:*

2. The IEEE-LOM metadata element set is defined at a Web location specified by a URI; all IEEE-LOM elements within the scope of this namespace declaration can be recognized by the prefix *lom:*

Using this infrastructure, metadata system designers can select elements from suitable existing metadata element sets, taking advantage of the investment of existing communities of expertise, and thereby avoid reinventing well-established metadata sets for each new deployment domain.

2.3 Extensibility

Metadata systems must allow for extensions so that particular needs of a given application can be accommodated. Some metadata elements are likely to be found in most metadata schemas (the concept of *creator* or *identifier* of an information resource, for example). Others will be specific to particular applications or domains (*degree of cloud cover,* for example, in remote sensing data). Metadata architectures must easily accommodate the notion of a base schema with additional elements that tailor a given application to local needs or domain-specific needs without excessively compromising the interoperability provided by the base schema. Another application encountering such extensions should be able to ignore such extensions while making use of any elements understood by both.

2.4 Refinement

Application domains will differ according to the degree of detail that is necessary or desirable. The design of metadata standards should allow schema designers to choose a level of detail appropriate to a given application. Populating databases with metadata is costly, so there are strong economic incentives to create metadata with sufficient detail to meet the functional requirements of an application, but not more. There are two notions of refinement to consider. The first is the addition of qualifiers that refine or make more specific the meaning of an element. *Illustrator, author, composer,* or *sculptor* are all examples of particular types of the more general term, creator. *Date of creation, date of modification,* and *date of acceptance* are all narrower senses of a date attribute. Such refinements might be useful or even essential in a given metadata application, but for general interoperability purposes, the values of such elements can be thought of as subtypes of a broader element.

A second variety of refinement involves the specification of particular schemes or value sets that define the range of values for a given element. Thus, identifying that a metadata value has been selected from a controlled vocabulary or has been constructed according to a particular algorithm may make it much more useful, especially for automated processing. In this way, semantic interoperability across applications can be increased, by relying on a common value set. The encoding of dates and times is an example of the use of an encoding standard to remove ambiguity from the expression of a metadata value. The string *03/06/02* is interpreted as *March 6, 2002* in North America and *June 3, 2002* in Europe and Australia. By using an encoding standard such as the W3C date and time format (W3C-DTF, 1998), a date can be encoded in an unambiguous manner (*2002-03-06*). Specifying the encoding format in the metadata allows unambiguous machine processing as well as improving human comprehension.

2.5 Multilingualism

It is essential to adopt metadata architectures that respect linguistic and cultural diversity. The Web as a global information system is important in that it affords unprecedented access to resources of global scope. However, unless such resources can be made available to users in their native languages, in appropriate character sets, and with metadata appropriate to management of the resources, the Web will fail to achieve its potential as a global information system. Standards typically deal with these issues through the complementary processes of *internationalization* and *localization*: the former process relates to the creation of "neutral" standards, whereas the latter refers to the adaptation of such a neutral standard to a local context. It is important to note that these two processes can sometimes work at cross-purposes.

One of the challenges for global metadata architecture is to assure that the underlying infrastructure can support either strategy equally well, or a mix of the two. Thus, a given application will reflect design choices based on an understanding of this balance and its implications. A basic starting point in promoting global metadata architecture is to translate relevant specification and standards documents into a variety of languages. DCMI maintains a list of translations of its basic documents. Likewise, the European workshop on Learning Technologies is maintaining translations of the LOM specification.

Another essential dimension is ·- include provisions in the metadata for the description of lingual and other cultural aspects of a resource. For example, metadata can describe the language and character set of the resource. The metadata may identify alternative versions of resources, in different languages, as well as the origin of the translations.

On a somewhat more technical level, it is important for global adoption of the standards that both the specifications and the ways these specifications are encoded are as "culturally neutral" as possible. As an example, it would be inappropriate to define the value space of a data element such as educational context in a way that is specific to one national system. Likewise, encodings will often be based on numerical representations of elements or their values, although there is wide practice to use some form of "pseudo-English" as well... (HTML tags are a typical example: the <LI> tag refers to the notion of a "List Item" and is thus somewhat biased linguistically). Multilingualism is one aspect of the broader issue of multiculturalism, which includes, for instance:

- The way in which dates are represented in different calendars,
- The direction in which text is displayed and read,
- Cultural connotations of certain icons and pictograms,
- Standards of practice (name order, collation standards, leading article standards).

Clearly, many of these aspects go beyond the immediate context of metadata. However, as mentioned above, it is important that metadata can describe the relevant characteristics, and that it can do so in ways that respect cultural and language differences.

3. Practicalities

The metadata principles may lead to a number of practicalities which are described in the following sections.

3.1 Application Profiles

No single metadata element set will accommodate the functional requirements of all applications, and as the Web dissolves access boundaries, it becomes increasingly important to be able to also cross discovery boundaries. Application profiles will facilitate this by allowing designers to 'mix and match' schemas as appropriate.

An application profile is an assemblage of metadata elements selected from one or more metadata schemas and combined in a compound schema. Application profiles provide the means to express principles of modularity and extensibility. The purpose of an application profile is to adapt or combine existing schemas into a package that is tailored to the functional requirements of a particular application, while retaining interoperability with the original base schemas. Part of such an adaptation may include the elaboration of local metadata elements that have importance in a given community or organization, but which are not expected to be important in a wider context

3.2 Syntax and Semantics

Semantics is about meaning; syntax is about form. Agreements about both are necessary for two communities to share metadata. Two communities may agree about the meaning of the term title or creator or identifier, but until they have a shared convention for identifying and encoding values, they cannot easily exchange their metadata.

It is important, however, to keep syntax and semantics separate as far as possible. The lack of stability in the structured markup realm emphasizes the necessity of maintaining independence between the semantics of metadata elements and their syntactic representation. However, as more information is 'born digital', one expects metadata facilities to be an intrinsic part of the creation and management of the resources, so issues of syntax cannot be ignored even though we are in general more concerned with the meaning of metadata statements rather than how they are exchanged. The IEEE Learning Object Metadata (LOM, 2002) standard provides an example of how this critical need for independence between the semantics of metadata and their syntactical representation can be addressed. LOM will be what is known as a "multi-part standard" where the semantic data model is an independent standard and then each syntactical representation is an independent standard developed as a specific "binding" of the LOM Data Model standard. DCMI also provides recommendations on encoding of Dublin Core metadata in alternative encoding idioms.

3.3 Association Models

There are various ways to associate metadata with resources:

Embedded metadata: resides within the markup of the resource. This implies that the metadata is created at the time that the resource is created, often by the author. Experts differ concerning whether author-created metadata is best or whether it is better to have trained practitioners evaluate and describe resources. As a practical matter, resource description expertise is a scarce and costly commodity, and thus any investment by authors in the description of their intellectual products is likely to be of value. Embedded metadata can also be harvested, and the presumptive increase in visibility that might result is an incentive for creators to assign metadata. Early studies of the efficacy of such metadata are only recently becoming available (GRE-01, 2001).

Associated metadata: is maintained in files tightly coupled to the resources they describe. Such metadata may or may not be harvestable. The advantage of associated metadata derives from the relative ease of managing the metadata without altering the content of the resource itself, but this benefit is purchased at the cost of simplicity, necessitating the co-management of resource files and metadata files.

Third-Party metadata: is maintained in a separate repository by an organization that may or may not have direct control over or access to the content of the resource. Typically such metadata is maintained in a database that is not accessible to harvesters, though the emerging Open Archives Initiative (OAI) Metadata Harvesting Protocol proposes a system that encourages the disclosure of metadata repositories among federated OAI servers (OAI-02, 2001). A given information resource will often have multiple metadata records reflecting the various purposes and perspectives of the organizations that create and manage them. A resource may be created with embedded metadata supplied by the author. A separate record might be created by the issuing organization (an academic department or publisher, for example) and stored in a separate database. A third party (perhaps a library) might create yet another version of metadata, either from scratch or derived from a previous record. In most cases these records will not be managed in a coordinated way, and differences may arise among them that may cause ambiguity or confusion. This may be less than ideal, but must be expected in an environment where various organizations may choose to manage resource descriptions with different objectives.

3.4 Identifying and naming metadata elements: tokens against labels

The global scope of the Web URI (Uniform Resource Identifier) namespace means that each data element in an element set can be represented by a globally addressable name (its URI). Invariant global identifiers make machine processing of metadata across languages and applications far easier, but may impose unnatural constraints in a given context. Identifiers such as URIs are not convenient as labels to be read by people, especially when such labels are in a language or character set other than the natural language of a given application. People prefer to read simple strings that have meaning in their own language. Particular tools and applications can use different presentation labels within their systems to make the labels more understandable and useful in a given linguistic, cultural, or domain context.

3.5 Metadata registries

Metadata registries represent an important topic of digital library research at this time. As the number of metadata and application profile schemas designed to meet the needs of particular discourse and practice communities increases, the importance of the management and disclosure roles of registries will similarly increase. The expectation is that registries will provide the means to identify and refer to established schemas and application profiles, potentially including the means for machine mapping among different schemas. In addition, it is expected that such registries will contain, or link to, important controlled vocabularies from which the values of metadata fields can be selected. Such registries will assume the characteristics of an electronic dictionary, available for consultation by:

- Application designers, who will be able to consult registries to identify existing metadata schemas and schema components that might meet their needs or to identify extensions to those schema that other application designers have developed to meet a given local need.
- Creators and managers of metadata, who can consult a registry to ascertain the definition or usage statements concerning an element or the available or preferred candidate value sets to be used to populate particular elements.
- Applications, which can resolve URIs associated with a schema, an element, or a value set in order to compare or evaluate elements or their values in a set of metadata.
- End users, who might consult a registry to better understand definitions or context of metadata terms, and thereby improve their search or processing effectiveness.

Thus, registries will provide the means to manage and disclose metadata schema declarations, application profile declarations, and value space declarations. As any given metadata schema or application profile evolves, registries will maintain the relationships among that schema's various versions in order to promote semantic and machine interoperability over time (HEE-00, 2000).

The DCMI Registry Working Group is exploring some of these issues through the explication of functional requirements for a multilingual DCMI metadata registry and vocabulary management system. Initial prototypes for this system can be accessed at (DC-REGISTRY, 2002). It is likely that registries will vary in the depth of their functionality with some being simple links to schema declarations while others may be richly functional databases. Some registries will be managed by namespace authorities and will hold the canonical copies of schema and value space declarations while other registries will harvest those declarations from such authoritative sources and thereby make them available in a more distributed manner (HEE-00, 2000).

3.6 Completeness of description

There is a strong inclination on the part of creators of metadata to 'fill in all the blanks.' If an element is available, people want to use it in a description. Applications should be designed to make evident that not every available element is necessarily appropriate for every resource type. Similarly, applications should provide assistance where possible in selection of an appropriate value for a particular element. To the extent that metadata creation facilities are built into content creation applications, the application can identify values for some elements more reliably than the user. Ultimately, the richness of metadata descriptions will be determined by policies and best practices designated by the agency creating the metadata, and those policies and practices will be guided by the functional requirements of services or applications. However, detailed metadata descriptions may improve searching precision; require higher investment in creation of metadata; and make it more difficult to promote consistency in creation of metadata, while simple descriptions are easier and less costly to generate; may result in more false results or more effort on the part of searchers to identify most relevant results; and improve probability of cross-disciplinary interoperability.

3.7 Mandatory against optional elements

Designing metadata standards for a global, cross-disciplinary information environment requires a high degree of flexibility. An element that is essential in one domain may not even be sensible in another, hence few, if any, elements in a general metadata set should be thought of as mandatory. On the other hand, it is entirely reasonable within a given application or even an application domain, to require particular elements. Thus, communities of practice should be encouraged to further specify standards of practice for a given metadata standard that will encourage uniformity of descriptions within a given domain. This can be done in the form of an application profile as described earlier, and shared with others within a community of practice in order to promote convergence and thereby increase interoperability.

3.8 Subjective and objective metadata

Metadata is broadly defined as structured data about data. However, the process of creating metadata can involve both subjective and objective input. Some metadata is clearly objective: assertions of fact about authorship, date of creation, version, and other attributes are generally able to be determined in an objective way. This objective metadata can also be machine generated in most instances, such as the "properties" metadata generated when creating a file in a word processor or spreadsheet application.

Other metadata may be subjective, either because such elements are subject to differing points of view (assignment of keywords, summarization of content in an abstract), or because they are specifically intended to represent a subjective evaluation (a review of a book or a presentation). Even more formal metadata elements become subjective when used within a cultural or domain context that is subject to local interpretation. For example, a pedagogical characteristic that is dependent on a particular educational philosophy may be important within a given context, but will have no meaning outside that context. The requirement for metadata design is, as far as possible, to make that context explicit so that applications can more easily recognize when a given element is constrained by such context as opposed to being more broadly applicable.

3.9 Automated generation of metadata

Most resource discovery metadata prior to the Web was created by humans in the labor-intensive activity of library cataloging. Cataloging metadata remains the most successful standard for resource discovery of books and periodicals, but it is costly to create and impractical for many materials available on the Internet. Web search engines harvest and index a significant portion of the Internet and provide low cost index access to it, generally in an advertiser-supported model. Such indexing can be thought of as a kind of metadata, and for many information needs, it provides a surprisingly cost effective solution to resource discovery. Between these two extremes lies a broad range of metadata creation that can be automated to some degree, and which can be expected to grow in importance as advances in such areas as natural language processing, data mining, profile and pattern recognition algorithms become more effective.

4. Conclusions

Metadata is a key part of the information infrastructure necessary to help create order in the chaos of the Web, infusing description, classification, and organization to help create more useful stores of information. Sources of metadata, like the sources of the resources themselves, will be of different quality and organized around different purposes to reflect the different objectives and business models of information providers. The social policies, organizational priorities, and market forces that shape the information spaces of the Web will undoubtedly create unforeseen opportunities and niches.

This paper has expressed some common understandings about metadata principles and practicalities that two metadata communities, DCMI and IEEE (LTSC) agree to be at the heart of their work towards e-learning. It is worthy of note that these commonalities did not emerge by design or intentional agreement, but rather are the expressions of years of independent work and the development of community practices.

5. References

1. DCMI. (2000). *Dublin core metadata initiative education working group Website* [Online]. Available: http://purl.org/DC/groups/education.htm.
2. DC-REGISTRY, (2002). The Open Metadata Registry Prototype [Home Page]. http://wip.dublincore.org:8080/registry/Registry.
3. Duval, E., Hodgins, W., Suuton, S., and Weibel, S.L. (2002). Metadata Principles and Practicalities, Available:

http://www.dlib.org/dlib/april02/weibel/04weibel.html

4. GRE-01, (2001). Greenberg, Jane, M. Pattuelli, B. Parsia, W. Robertson, Author-generated Dublin Core Metadata for Web Resources: A Baseline Study in an Organization, *Journal of Digital Information,* volume 2 issue 2 http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Greenberg/.
5. HEE-00, (2000). Heery, Rachel and Manjula Patel, Application Profiles: Mixing and Matching Metadata Schemas, *Ariadne,* Issue 25 http://www.ariadne.ac.uk/issue25/app-profiles/intro.html
6. IMS. (2000a). *Instructional Management Systems Project Website* [On-line]. Available: http://imsproject.org/
7. IMS. (2000b). *Instructional Management Systems Project Website* [On-line]. Available: http://imsproject.org/imMembers.html
8. LOM (2002). *LOM Working Draft v6.2* [On-line]. Available: http://ltsc.ieee.org/.
9. LTSC. (2002). *Learning Technology Standards Committee Website* [On-line]. Available: http://ltsc.ieee.org/
10. NAMES, (1991). *Namespaces in XML,* World Wide Web Consortium, Editors: Tim Bray, Dave Hollander, and Andrew Layman, http://www.w3.org/TR/REC-xml-names
11. OAI-02, (2001). *The Open Archives Initiative Protocol for Metadata Harvesting,* Protocol Version 1.1 of 2001-07-02, Document Version 2001-06-20, http://www.openarchives.org/OAI/openarchivesprotocol.htm.
12. W3C-DTF, (1998). *Date and Time Formats* [Technical Note] Misha Wolf and Charles Wicksteed 1998-08-27, http://www.w3.org/TR/NOTE-datetime.
13. W3C, (2002). XML Documentations, [Online] Available: *http://www.w3c.org*