The Evolution of Metadata from Standards to Semantics in E-Learning Applications

Hend S. Al-Khalifa
LTG, ECS, University of Southampton
Southampton, UK
+44 23 8059 8867

hsak04r@ecs.soton.ac.uk

Hugh C. Davis
LTG, ECS,University of Southampton
Southampton, UK
+44 23 8059 8867

hcd@ecs.soton.ac.uk

ABSTRACT

Metadata attempts to describe the content, format, purpose and structure of data. Over the past few years, the IEEE-LOM standard has dominated the metadata world in e-learning applications. However, with the advent of the Semantic Web, e-learning applications are beginning to evolve their metadata representation from these standards by adding semantic structure or by converting entirely to semantic representations of structure. This shift enables the implementation of a range of new tools which can reason over the metadata, providing added value from the stored data. This review paper summarizes this evolution of metadata used in e-learning applications from standards to semantic representation.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]

General Terms

Management, Design, Standardization, Languages.

Keywords

Standard Metadata, Semantic Metadata, E-learning.

1. INTRODUCTION

In the past, metadata was often neglected. But, once computers were in common use for storing data, the need for techniques to retrieve the data became important. Since then the concept of metadata in computer science has evolved, starting from the simple file systems (file names and types) in the early 60s, then to database management systems (to describe database fields) in the early 70s, to the 21st century with the advent of the concept of metadata warehouses [1].

Metadata is a record that consists of structured information about a resource; it can be also defined as information about information or data about data. It is structured in a manner that facilitates the management, discovery and retrieval of resources. Another useful definition for metadata is given by [2] as "*any data which conveys knowledge about an item without requiring examination of the item itself.*"

The importance of metadata has also evolved to include the domain of the Semantic Web. At the heart of the Semantic Web is the idea of adding formal metadata that describes the content, context and/or structure of a web resource [3]. In this paper we are concerned with the e-learning domain and how this transformation has affected the representation, retrieval and manipulation of learning resources. In section 2 we define what we mean by 'standard metadata' from an e-learning context. Section 3, defines what we mean by semantic metadata, in general. Section 4; classifies the different representations of metadata in e-learning applications. Finally, the paper concludes by discussing the significance of the shift to semantic metadata.

2. STANDARD METADATA IN EDUCATION

An educational metadata record extends the scope of a regular metadata. It adds further fields to the metadata record which describes information that has particular educational relevance [4].

A number of organizations are involved in producing metadata standards specifically for learning technology. A list of the major ones includes: ADL, AICC, ARIADNE, CEN/ISSS WS-LT, IEEE LTSC and IMS [5].

Metadata standards are formal specifications used to semantically annotate educational materials of any kind. They have been developed to support both machine interoperability (information exchange) and resource discovery by human users. There are mainly two widely accepted metadata standards in education [6], namely:

1. DC (Dublin Core) educational version, and
2. IEEE-LOM (Institute of Electrical and Electronic Engineers/Learning Object Metadata).

IEEE-LOM [7] record defines 80 fields arranged in a hierarchical structure and adheres to the following categories: 1-General, 2-Lyfecycle, 3-Meta-Metadata, 4-Technical, 5-Educational, 6-Rights, 7-Relation, 8-Annotation and 9-Classification.

One of the main features of IEEE-LOM is its ability to extend and add new data elements as required by applications. This flexibility in the standard has encouraged metadata developers to use IEEE-LOM as the base standard for developing new 'application profiles' (mandatory and optional fields, and extensions) that suit their application needs.

3. SEMANTIC METADATA

Semantic metadata can be defined as "…[linking] *related terms to one another*" [2]. It can be also defined as "…*the process of*

attaching semantic descriptions to Web resources by linking them to a number of classes and properties defined in Ontologies" [8].

Different areas of computer science have different interpretations of what "semantics" mean [9]. For instance, in the domain of databases, metadata is thought of a conceptual schema that describes the structure of a database. The domain of information retrieval, might consider metadata as the set of keywords that describe the main theme of a document, or as a record that confirms to a specific schema (e.g. Dublin Core).

Sheth et al. have described these different depictions of metadata, organizing them into three types of semantics [9]: implicit, formal and powerful. Implicit semantics appear in unstructured text that has loosely defined and less formal structure (e.g. Information Retrieval). Formal semantics appear when the data representation takes a more rigid form (e.g. Knowledge Representation). Finally, powerful semantics imply the combination of simple syntactic structures to represent the meaning of complex ones.

4. CLASSIFICATION OF METADATA REPRESENTATIONS IN E-LEARNING APPLICATIONS

By surveying the literature of metadata used in e-learning applications, we can classify the representation of metadata in e-learning applications into three categories (Figure 1):

1) Standard Metadata,
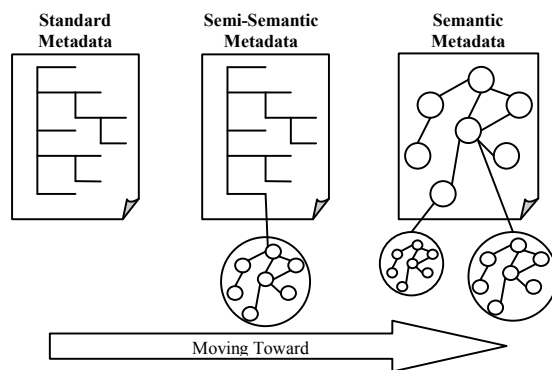2) Semi-Semantic Metadata, and
3) Semantic Metadata.

Figure 1: The different genres of Metadata in e-learning applications

4.1 Standard Metadata

These are applications that use the IEEE-LOM standard or a variation of it (an Application Profile), but, unlike RDF based metadata, the standard only allows for a hierarchical structure which follows a single perspective [10].

Many systems and projects still use IEEE-LOM for describing learning resources. For example, in the UK we have invested in a National Repository for learning objects which makes use of an application profile (UKLOM Core) intended to encourage community sharing of learning resources [11]. Another recent example is the International E-Miage (IEM) project, an Information System Engineering curriculum for Miage distant

teaching which consists of twenty national sites and delivers official degrees. It has used the IEEE-LOM metadata standard to allow for interoperability between the Learning Management System (LMS) deployed within the IEM framework and the Knowledge Pool System of the Ariadne Foundation [12].

4.2 Semi-Semantic Metadata

These are applications that use the IEEE-LOM standard with an extended semantic component. As an example, the Hypermedia Learning Objects System (hylOs) uses IEEE-LOM standard to describe its learning resources. However it extends the relational field in the standard with a semantic net to interconnect different learning objects [10]. Another example is <e-aula> the personalized e-learning application [13]. This application uses a restricted set of LOM categories plus two additional terms: one associated to the pedagogical ontology and the other to the concept domain ontology.

4.3 Semantic Metadata

These are applications that rely completely on domain ontologies to define their metadata. They use RDF as a vehicle to express the semantics of a learning resource. For these applications, using RDF has advantages over the standard metadata approach [14]. First, any RDF data model is based on an "open world assumption" where the metadata is selected from heterogeneous ontologies. On the other hand, the majority of systems that implement LOM take a closed-world approach confining their metadata to that implemented by a particular LMS. Second, RDF allows for the creation of complex statements (i.e. metadata can be further annotated with more metadata). LOM, on the other hand, does not allow for the expression of complex metadata; it only supports extensions through taxonomic classification.

An example is the Edutella P2P network [15]. The project has used RDF-metadata built upon standards like IEEE-LOM/IMS metadata with up to 100 metadata entries that have been complemented by domain specific extensions. Thus, the Edutella infrastructure provides an architecture to connect Edutella Peers based on exchange of RDF metadata.

Another most recent example, is an ongoing project carried out in the laboratories of Advanced Research in Intelligent educational Systems (ARIES), Canada [14]. This project is replacing the IEEE-LOM with more flexible e*cological approach*. This approach sees metadata as the process of reasoning over observed interactions of users with a learning object for a particular purpose.

5. WHICH ONE TO CHOOSE?

The decision to choose between different metadata representations is, to some extent, subjective. It depends on the application scope and needs. Proprietary and closed-world applications (e.g. 'Blackboard' a proprietary Learning Management System) might prefer using standard metadata for its 'coarse grain' semantics, while nonproprietary and open-world applications (e.g. most research work) move toward semantic metadata for its 'fine grain' semantics.

Brooks & McCalla [14] have highlighted three issues related to metadata standards:

- Metadata standards were created with humans being both the consumer and producer of the metadata and learning object

content, so the semantics of metadata fields need only be understood by humans. Furthermore, usually there will be some unfilled metadata fields.

- Metadata standards are inherently centralized. In order to gain the intended interoperability gains, all applications must adhere to the same application profile.
- Metadata standards collect the wrong kind of data. Until now metadata standards have been used to collect the characteristics of the learning object content. However, important information such as usage of the learning objects is often neglected. This kind of information is valuable to show the relevance of the learning object to a particular context, for a particular learner or pedagogical goal.

Therefore, the trend nowadays is to use semantic metadata for its pleasing results in facilitating large scale collaboration [16]. Semantic metadata has the following advantages over standard metadata:

- **Machine Processable Metadata**: since semantic metadata is built using ontologies. This will guarantee that metadata will have well-formed meaning that machines can read, understand and process.
- **Flexibility and Extensibility**: semantic metadata can be further annotated with more metadata without being confined to a fixed template. This also allows for the flexibility of mixing different metadata ontologies.
- **Reasoning**: as metadata is expressed formally, reasoning rules can be defined and new relations can be derived, thus exploiting the use of semantic search.
- **Interoperability**: even though standard metadata promotes interoperability, semantic metadata has the added value of supporting partially agreed ontologies that will enable systems to interoperate much more easily.

6. CONCLUSION

Although this paper has discussed one commonly used metadata standard (i.e. IEEE-LOM) and its shift towards semantic representation, other e-learning standards such as Question & Test Interoperability (QTI) and Learning Design (LD) also have embraced the semantic approach e.g. [16-18]. The paper also gave some examples of the different emergent genres of metadata representations.

Finally, some might argue that if we have semantic metadata we can not switch back to standard metadata. To answer this argument we say, an application can switch easily from semantic metadata to standard metadata using transformation templates (i.e. XSLT); however, the other way around (i.e. from standard to semantic) is not a straight forward task, and to justify our claim the reader is referred to the unfinished process of RDF binding of IEEE-LOM [19].

7. REFERENCES

[1] Arun, S., *Metadata management: past, present and future.* Decision Support Systems, 2004. **37**(1): p. 151.

[2] Haase, K. *Context for semantic metadata.* in *Proceedings of the 12th ACM International Conference on Multimedia.* 2004. New York, NY, United States: ACM.

[3] Marshall, C.C. and F.M. Shipman. *Which semantic web?* in *Proceedings of the fourteenth ACM conference on Hypertext and hypermedia.* 2003. Nottingham, UK: ACM Press.

[4] Recker, M.M. and D.A. Wiley, *A non-authoritative educational metadata ontology for filtering and recommending learning objects.* Interactive Learning Environments, 2001. **9**(3): p. 255.

[5] Robson, R. *Report on Learning Technology Standards.* in *In The World Conference on Educational Multimedia, Hypermedia and Telecommunications.* 2000.

[6] Stratakis, M., V. Christophides, K. Keenoy, and A. Magkanaraki, *E-Learning Standards.* 2003. p. p. 45

[7] Duval, E., IEEE standard for Learning Object Metadata (LOM). Accessed on May 8, 2006. Available on line http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf

[8] Scerri, S., C. Abela, and M. Montebello. *semantExplorer: A Semantic Web Browser.* In *IADIS International Conference WWW/Internet 2005.* 2005. Lisbon, Portugal.

[9] Sheth, A., C. Ramakrishnan, and C. Thomas, *Semantics for the Semantic Web: The Implicit, the Formal and the Powerful.* International Journal on Semantic Web & Information Systems, 2005. **1**(1): p. 1-18.

[10] Engelhardt, M., A. Hildebrand, D. Lange, and T.C. Schmidt. *Semantic Overlays in Educational Content Networks.* in *TERENA Networking Conference.* 2006. Catania, Italy.

[11] The JORUM repository. Accessed on May 8, 2006. Available on line. http://www.jorum.ac.uk/:

[12] Broisin, J., P. Vidal, D. Marquie, and O. Catteau. *The International E-Miage project: from an Isolated Framework to an Architecture Based on Learning Standards.* in *Advanced Int'l Conference on Telecommunications and Int'l Conference on Internet and Web Applications and Services (AICT-ICIW'06).* 2006: IEEE Computer Society.

[13] Sancho, P., I. Martínez, and B. Fernández-Manjón, *Semantic Web Technologies Applied to e-learning Personalization in <e-aula>.* Journal of Universal Computer Science, 2005. **11**(9): p. 1470-1481.

[14] Brooks, C. and G. McCalla, *Towards flexible learning object metadata.* International Journal of Continuing Engineering Education and Life-Long Learning, 2006. **16**(No.1/2): p. pp. 50 - 63.

[15] Nejdl, W., B. Wolf, C. Qu, S. Decker, M. Sintek, A. Naeve, M. Nilsson, M. Palmér, and T. Risch. *EDUTELLA: a P2P networking infrastructure based on RDF.* in *Proceedings of the 11th international conference on World Wide Web.* 2002. Honolulu, Hawaii, USA: ACM Press.

[16] Millard, D., F. Tao, K. Doody, A. Woukeu, and H. Davis, *The Knowledge Life Cycle for e-learning.* International Journal of Continuing Engineering Education and Lifelong Learning: Special Issue on Application of Semantic Web Technologies in E-learning, 2006. **16**(1/2): p. 110-121.

[17] Knight, C., D. Gašević, and G. Richards, *Ontologies to integrate learning design and learning content.* Journal of Interactive Media in Education (Advances in Learning Design. Special Issue, eds. Colin Tattersall, Rob Koper), 2005. **07**.

[18] Amorim, R., M. Lama, E. Sanchez, A. Riera, and X.A. Vila. *A Learning Design Ontology based on the IMS Specification.* in *In proceedings of the UNFOLD-PROLEARN Joint Workshop.* 2005. Valkenburg, The Netherlands.

[19] DCMI/IEEE LTSC Taskforce, RDF binding of LOM metadata. Accessed on May 8, 2006. Available on line http://dublincore.org/educationwiki/DCMIIEEELTSCTaskforce