Standardized Metadata for Education: a Status Report

Erik Duval
Departement Computerwetenschappen, Katholieke Universiteit Leuven
Celestijnenlaan 200 A, B-3001 Leuven, Belgium
Erik.Duval@cs.kuleuven.ac.be

Abstract

This paper starts with a brief background to worldwide standardization activities in the field of educational technologies. It then focuses on the specific area of metadata, and introduces the two most relevant approaches there: IEEE LTSC LOM and DC-Education. We review how different communities adopt and support educational metadata for their constituencies. This paper can be considered as a follow-up to [20].

1. Introduction: the role of Standardization

The main aim of standardization is to achieve interoperability between systems or components from diverse origins. Interoperability in this context can be understood as 'enabling information which originates in one context to be used in another in ways that are as highly automated as possible' [1]. In an educational context, this should enable reuse of tools and content across functions (cataloguing, discovery, etc.), levels (simple to complex), semantic and linguistic barriers, and, of course, technology platforms.

Standardization is a requirement for large-scale deployment of learning technologies, as it prevents users from being locked into proprietary systems. Moreover, standards are needed to develop an open base infrastructure that components can be plugged into. This is especially important in an educational context, because of the wide diversity, and the lack of consensus on a universal best solution.

It is important to realise that a standard does *not* impose a particular implementation. Rather, as a common specification, it establishes an opportunity for competition or collaboration by diverse groups.

Very important in the domain of education and training, is that standards can help to establish a base technology infrastructure with *permanency*. This is a high priority need in our field: many impressive results of R&D projects from the last 15 years (and more!) are gathering dust on shelves, because the tools and the equipment that they rely upon no longer exists, or are no longer maintained. Without such permanency of technology, we will not be able to build upon the results of our predecessors, and we will never achieve real *impact*!

Finally, it is important to note that 'real' standards are produced by 'accredited' organizations, that guarantee an open, fair, transparent and inclusive process, and that take care of the maintenance of the standards. Rather different from these 'de jure' standards, 'de facto' standards are based on specifications made by a consortium or company. The proper process (see also figure 1) relies on the development of specifications in consortia, where experience can be gathered on their practical use. When specifications have matured, they can be fed into accredited organizations that can use this input as the basis for 'real' standards.

2. Standardization of Learning Technologies

There are three important accredited standardization organizations in the domain of education and training.

2.1 IEEE LTSC

The aim of the IEEE Learning Technology Standardization Committee (LTSC) is to develop technical standards, recommended practices, and guides for software components, tools, technologies and design methods that facilitate the development, deployment, maintenance and interoperation of computer implementations of education and training components and systems [2]. The LTSC includes a number of working groups, which deal with:

- general issues: architecture, glossary;
- learner issues: learner model, student identifiers, quality, competency;
- content issues: interchange, sequencing, packaging;
- (meta)data: Learning Object Metadata (see section 3.1), localisation, bindings, protocols;
- learning management issues: Computer Managed Instruction, tool-agent communication, platform and media profiles.

The LTSC scope explicitly excludes issues that the organization believes ought *not* to be standardized (evaluation methods, delivery systems, etc.), and issues that it believes ought to be standardized *elsewhere* (multimedia, education standards, cultural adaptation). The overall strategy is to standardize the smallest, useful, doable specification that has technically feasibility, commercial viability, and widespread adoption. The group started in 1996, and holds quarterly meetings.

2.2 CEN ISSS LTWS

The CEN ISSS Learning Technologies Workshop (LTWS) has developed a report that identifies requirements for standards on learning technologies [3]. The main requirements are to improve support for reusability and interoperability, collaboration, metadata, quality, legal issues, multilinguality and multiculturality and accessibility. The report includes recommendations on:

- promotion activities on standards in the educational arena in general, and on accessibility guidelines in particular;
- the development of taxonomies and vocabularies, as these can increase 'semantic interoperability' and enable translations and mappings of (meta)data;
- the elaboration of an 'educational copyright license'.

More generally speaking, the report emphasizes the need to support multilinguality and cultural wealth. It is explicitly recommended to avoid English as a lingua franca. The workshop has also been working on localisation issues of the IEEE LTSC metadata standard (see section 3.1), in order to ensure that European requirements for cultural diversity and multilingualism are properly met. A French, German, Catalan, Spanish and Italian version of LOM have been developed.

2.3 ISO/IEC JTC1 SC36

Recently, a new subcommittee on 'Information Technology for Learning, Education, and Training' has been set up under the umbrella of the ISO/IEC JTC1 Joint Technical Committee. Its scope is 'standardization in the field of information technologies for learning, education, and training to support individuals, groups, or organizations, and to enable interoperability and reusability of resources and tools' [4]. Working groups are currently being set up on issues such as vocabulary, architecture, learner-related activities, learning object metadata, learning management systems and collaboration technologies. Both LTSC and LTWS have liaison status with ISO/IEC JTC1.

3. Standardized Metadata for Education

In order to promote (re)use of learning content, identification of appropriate learning material needs to be accommodated. This requires searchable descriptions of the material, called metadata, if we want to go beyond simple keyword searches and support queries like 'simulations of a nuclear power plant, that run on the Linux operating system, take students of undergraduate university level 45 minutes to work through and learn about the influence of cooling water temperature on nuclear reaction speed'. As interoperability is key to collecting a critical mass of metadata, standardization efforts in this field are especially relevant. In the remainder of this section, we present and compare briefly the two main initiatives in this area.

3.1 IEEE LTSC LOM

The purpose of LTSC Learning Object Metadata (LOM) is to facilitate search, evaluation, acquisition, and use of learning objects by learners or instructors [5]. The purpose is also to facilitate the sharing and exchange of learning objects, by enabling the development of catalogs and inventories, so that users can create and publish educational material.

This three-year-old initiative has developed an elaborate metadata scheme with a hierarchical structure. Data elements are regrouped under categories (general, lifecycle, metametadata, technical, educational, rights, relation, annotation and classification). Especially relevant is the educational category that includes elements such as

- interactivity type (active versus expositive),
- learning resource type (exercise, simulation, questionnaire, etc.),
- interactivity level (from very low to very high),
- semantic density (idem),
- intended end user role (teacher, author, learner, manager),

- context (primary education to vocational training),
- typical age range,
- difficulty (from very low to very high),
- typical learning time,
- description,
- and language of the typical intended user.

Besides these data elements that are specifically geared towards the domain of education, LOM also includes a rich set of data elements in the other categories. The core specification is now in version 5.0, which will be submitted to a formal ballot. The schema specification seems to be quite mature and stable. The main reluctance from the user community concerns the restricted nature of the vocabularies for certain LOM elements. That is why it was recently decided to change the status of these vocabularies to suggested good practice.

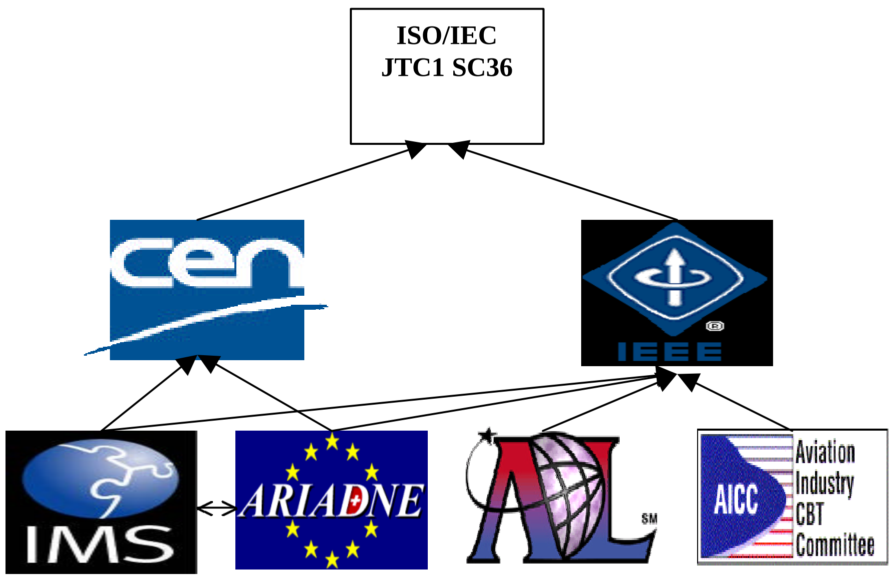

Fig. 1: The main organizations involved in standardization for education and training

3.2 Dublin Core – Education

The Dublin Core (DC) is a set of 15 metadata elements intended to facilitate discovery of electronic resources in a general sense. The objective of the Working Group on Education in the Dublin Core Metadata Initiative is to develop a proposal for the use of Dublin Core metadata in the description of educational resources [6]. At this moment, the Education group recommends to add the following data elements to DC:

- *audience* (both intermediate. like teacher or author, and end users, mostly learners);
- organizational, professional, province/state, national, and international *educational standards*;
- *interactivity type*, *interactivity level* and *typical learning time*: for these, the DC-Education group recommends adopting the corresponding LOM data elements.

3.3 Comparing LOM and DC-Education

LOM was originally developed specifically for the domain of education and training, and is becoming gradually more and more deployed outside of this specific domain. It is rather the other way around with DC-Education, as the Dublin Core metadata element set was originally developed for general resources, and is now being adapted for the specific field of education and training.

It is clear that LOM provides a far richer structure with more detail for ‘semantically interoperable’ metadata on learning objects. The use of namespaces to achieve semantic and syntactic interoperability between DC-Education and LOM is being discussed in a working group with representatives from both organizations.

Awareness in the sector of education and training on the issue of learning technology standardization in general and on metadata specifically, is definitely growing fast. On the other hand, with awareness of the importance of these issues also seems to grow the confusion and misunderstandings. For many interested parties, the difference in status between for instance consortia and accredited standards organizations is not at all clear (see above). And, as has happened so often before with educational technologies, there is a definite danger that expectations are being raised unrealistically high.

4. Adoption in Practice of Educational Metadata

4.1 Adoption in Consortia

There is considerable overlap in the work of the different consortia (ADL, AICC, ARIADNE, IMS, etc. – see figure 1 and below) that work on actual implementations of metadata and other standards. This is quite natural, as many of these consortia contribute to the development of the standards, and then adapt them to the needs of their constituencies, a process referred to as 'profiling' in the standards world. In fact, having several independent implementations of the metadata standards increases the probability that problems or shortcomings will be identified early on.

In this section, we briefly review some recent developments in a number of the more relevant such consortia.

- The *ARIADNE* project (1995-2000) recently transformed into a permanent self-sustainable Foundation [7]. One of the most important aims of the Foundation is to organize the further development and exploitation of the Knowledge Pool System, a distributed database of reusable learning components, with associated metadata that describe them. ARIADNE and IMS (see below) jointly submitted their metadata specification to the IEEE LTSC and continue to collaborate on this issue. Within the ARIADNE community, the so-called metadata recommendation makes 23 of the LOM elements mandatory. ARIADNE has developed a taxonomy of science types, disciplines and subdisciplines. These are used to indicate the semantics of the learning object, using the facilities that LOM provides for this purpose. The technological basis for the ARIADNE tool development is XML and databases.
- The *IMS* 'global learning consortium' released version 1.1 of its metadata specification in May 2000. This was a minor update to the previous version, based on LOM version 3.5 [8]. The IMS specification includes suggestions of controlled vocabularies for 12 elements. Often, it lists multiple suggestions, with their origin and suggested applicability. A binding to XML, in the form of a DTD, is also defined.
- The US Department of Defense launched the *Advanced Distributed Learning* initiative in 1997 [9]. In January, ADL released version 1.0 of the 'Sharable Course Object Reference Model (SCORM)'. This reference model applies LOM to raw media (7 mandatory elements), content (15 mandatory elements) and courses (15 mandatory elements). Since January 2000, developers have started to implement the model, which culminated in the so-called plug-fest in May. In parallel, a test suite for SCORM is being developed. The main technology binding is XML.
- *EDNA* [10] relies on DC with additional elements for when the metadata were entered, who approved them, when the learning object is to be reassessed, the user level, categories, conditions, indexing, review, version. EDNA includes its own lists for types of learning objects, user levels and coverage, for the Australian context. The main technology relied upon are HTML META tags.
- The *EUN SchoolNet* initiative regroups most of the ministeries of education in Europe [11]. For its metadata applications, it relies on the Dublin Core metadata element set, with additional elements for rights, approver, release, user level and version .
- The *Gateway to Educational Materials* (GEM) [12] relies also on the Dublin Core element set, with additional elements for DC with additional elements for audience (distinguishing who uses a tool from who benefits from its use), cataloging, duration, essential resources, grade, pedagogy, quality and standards. GEM has also defined controlled vocabulary for audience, format, grade, language, pedagogy (divided in teaching methods, grouping and assessment), relation, resource type and subject (a two level classification system).
- NEEDS [13] and SMETE [14] are build defines its own metadata element set for searching, with title, contributor, publisher, subject heading, affiliates keywords, platform and MIME type.
- The European Joint Research Center has developed a ‘Global Education Multimedia server’ (called GEM, but not related to the Gateway of Educational Materials mentioned above), which should act as a clearinghouse of information, products and services [26,27]. Similar to for instance ARIADNE, a

relational database is used to store the metadata on the server. Access in four languages takes place through a web interface. The GEM system can import from or export to LOM.

4.2 Adoption in Projects and Use in Research

Many smaller scale projects, typically confined to one organisation, are including metadata in their development efforts. Sometimes, the basic idea is to provide a useful service within the own organisation, and sometimes, the main focus is on further research that requires metadata support as basic functionality. Below, we review some of the more noteworthy activities in this domain.

- Work at Carnegie-Mellon investigates how metadata can be used to develop tools that automatically evaluate the consistency of a course [22]. The core idea is that the course is composed of learning objects whose metadata describe the prerequisites and learning objectives. Thus, it is possible to verify automatically that no learning object is presented before the necessary prerequisites have been covered (as learning objectives of preceding material).
- The GESTALT project implemented an extended version of LOM [23]. Tools were developed to match metadata against user profiles, so as to implement personalized educational programs. A CORBA based architecture was developed to access an 'asset metadata server' and user profiles from an LDAP server.
- LOM was adopted as the metadata set for a database of resources for schools in Hawaii [24]. These resources include Ph.D. students, software, weather stations and remotely controlled cameras, courses in computer science and mathematics, telescopes accessible through the web, etc. For this context, the LOM structure was extended (as anticipated in the specification) with more detail on audience, community involvement, discipline, educational objectives and pedagogy. In this process, much earlier work by GEM (see section 4.1) was integrated in the LOM context. These experiences influenced some final developments and fine tuning to the LOM specification.
- At the Darmstadt University of Technology, a LOM editor has been developed in the context of a project on adaptive hypermedia [25]. Learning strategies based on different kinds of relations (is_a, proceeds, part_of, etc.) are supported by dynamically matching metadata of so-called 'media bricks' with a 'concept space' that describes the structure of the application domain.
- At Acadia University, one theme of LOM related work is the use of XML tags and Xlink to connect metadata and the learning objects they describe [15]. Another theme relates to the description of the potential role of the learning object: for that purpose, the concept covered by the object, its context (including prerequisites), ways it can be used in instruction and possible student activities are described [16].
- At the University of Paderborn, granularity of learning objects has been researched, distinguishing between media elements, learning elements, content modules and thematic metastructures [17].
- The Essen Learner Model includes an extension of LOM to provide more detail on didactical methods used with a learning object [18]. Learning sequences are represented as processes that cover the main activities, interactions, information systems and persons involved.
- At the University of New South Wales, the role of learning objects, with associated LOM metadata, in administration, content authoring, learning resource catalogues and delivery has been studied [19]. Basically, the authors argue that these different components should be more tightly integrated.
- Commercial systems are beginning to support LOM, even though, as noted above, the standard is not completely finalized yet. An example of such a commercial systems is imc's CLIX [21], which has been used to build learning sites for companies like Lufthansa, Chrysler-Daimler, and others.

Conclusion

It is striking to note that, at present, there is little activity on interoperability between independently developed systems for metadata management. Some demonstrator work has been done between the ARIADNE and GESTALT consortia (see section 4.1), but larger scale developments are rare at this moment. We believe that it is now time to start this interoperability development, as the specifications and implementations are maturing. Only then will we be able to amass a critical quantity of learning material, and will the standard indeed serve its ultimate goal.

References

[1] http://www.indecs.org/

[2] http://ltsc.ieee.org/
[3] http://www.cenorm.be/isss/Workshop/lt/Default.htm
[4] http://jtc1sc36.org/
[5] http://ltsc.ieee.org/wg12/index.html
[6] http://purl.org/dc/groups/education.htm
[7] http://www.ariadne-eu.org/
[8] http://www.imsproject.org/
[9] http://www.adlnet.org/
[10] http://www.edna.edu.au/EdNA
[11] http://www.eun.org/
[12] http://geminfo.org/
[13] http://www.needs.org/
[14] http://www.smete.org/
[15] Duane Currie and Craig Place. *Learning Object Containers: a suggested method of transporting metadata with a learning object.* Proceedings of EdMedia2000, pp. 1265-1266.
[16] Gary Hepburn and Craig Place. *Learning Objects: Communicating the Pedagogical Potential.* Proceedings of EdMedia2000, pp. 1330-1331.
[17] Leena Suhl and Stephan Kassanke. *OR-World – Using Learning Objects in a Hypermedia Learning Environment*. Proceedings of EdMedia2000, pp. 1367-1368.
[18] Heimo H. Adelsberger, Markus Bick and Jan M. Pawlowski. *The Essen Learning Model A Step Towards a Standard Model of Learning Processes.* Proceedings of EdMedia2000, pp. 83-88.
[19] Tony Koppi, Lisa Hodgson and Jason Bayly. *The Often Missing but Essential Component for Online Learning: a Learning Resource Catalogue.* Proceedings of EdMedia2000, pp. 502–506.
[20] Robby Robson. *Report on Learning Technology Standards*, Proceedings of EdMedia2000, pp. 936-941.
[21] http://www.im-c.de
[22] Peter Brusilovsky. *Concept-Based Courseware Engineering for Large Scale Web-based Education.* Proceedings of WebNet2000, To be published.
[23] Konstantopoulos M., Darzentas J.S., Koutsabasis P., Spyrou T., Darzentas J. *Towards Integration of Learning Objects Metadata and Learner Profiles Design: Lessons Learnt from GESTALT.* To be published in the special issue on educational metadata of the Journal on Interactive Learning Environments, 2001.
[24] D. D. Suthers. *Using Learning Object Metadata in a Database of Primary and Secondary School Resources.* To appear in Proceedings of International Conference on Computers in Education, Nov. 21-24, 2000, Taipei, Taiwan.
[25] A. Steinacker, C. Seeberg, K. Reichenberger, S. Fischer and R. Steinmetz. *Dynamically Generated Tables of Contents as Guided Tours in Adaptive Hypermedia Systems.* Proceedings of EdMedia99.
[26] C. Best, P. Shiels and M. de Paola. *Global Education Multimedia Server – GEM.* Proceedings of EdMedia99, pp.1727-1728.
[27] http://gem.jrc.it/