METADATA CHALLENGES IN INTRODUCING THE GLOBAL IEEE LEARNING OBJECT METADATA (LOM) STANDARD IN A LOCAL ENVIRONMENT

Lars Fredrik Høimyr Edvardsen and Ingeborg Torvik Sølvberg
Dept. of Computer and Information System, Norwegian University of Science &Technology,
Sem Sælands vei 7-9, NO-7491, Trondheim, Norway
{lars.edvardsen,ingeborg} @idi.ntnu.no

Keywords: IEEE LOM, Learning Object Metadata, LOM, Learning Object, LO, Learning Management System, LMS, metadata mapping, crosswalk, metadata challenges

Abstract: The world of closed Learning Management Systems (LMS) is being replaced by open systems for sharing and reusing digital Learning Objects (LOs) between users, courses, institutions and countries. This poses new challenges in describing these LOs with detailed and correct metadata. This information background is needed for querying services to perform accurate queries for LO retrieval. In this paper we present metadata specific challenges when converting from a local LMS with proprietary metadata schema to a global metadata schema. We have uncovered extensive LO description possibilities based on the existing, local LMS, registered metadata, its LO types and the local context. Files can contain extensive metadata descriptions, though require special attention. We have confirmed that technologies developed as crosswalks are valid for usage in this projects for a one-time metadata transferral. However, transferring of all local metadata elements can result in incompatibility issues with other LMSs. This, even when keeping with the global metadata schema.

1 INTRODUCTION

The use of digital Learning Objects (LOs) such as slides, figures, exercises and exams are increasing on all educational levels. This is happening all over the world, in use by both students and teachers. The current generation of Learning Management Systems (LMSs) have had limited, if any LO and Learning Object Metadata (LOM) sharing possibilities. A new generation of LMSs is now emerging which allow sharing and reuse of LOs. Their LOM descriptions are the vital information background needed for querying services to perform accurate queries for LOs. For LMSs this transformation process means converting from a proprietary, local metadata schema to a global schema.

Between intentionally compatible metadata schemas, metadata exchange can be performed lossless. E.g. the national schemas UK LOM Core (Cetis, 2004) (UK) and NORLOM (eStandard, 2005) (Norwegian) are compatibility with the global IEEE LOM (IEEE LTSC, 2005).

For schemas without a pre-intended compatibility, metadata exchange can be more challenging. This is the case for most LMSs. A potential solution is using crosswalks (Chan & Zeng, 2006). Crosswalks are a set of determined equal elements between two schemas. This allow transfer of metadata back and forth between two schema standards, e.g. between Dublin Core and MARC (Library of Congress, 2001). In our work, crosswalks will be used as a one-way tool to transfer existing metadata to the new schema. However, since these schemas are not equal, many-to-one element mappings and many-to-none element mappings can occur. Here the fine-grain metadata schema architecture and existing metadata can get lost when converting. Cases with unequal elements resulting in one-to-many elements need to be addressed.

Metadata mapping is actually an everyday event when converting file formats. Though, it is often

hidden from user sight, like when converting MS PowerPoint slides into Adobe PDF print-outs. How the original metadata elements are converted, updated, excluded or replaced by other metadata is determined by the converting software, such as Adobe Distiller.

If files are to be used as a metadata source, this poses special challenges: There are a range of different file formats in use; many have a proprietary metadata schema. Our studies have uncovered extensive differences in how elements are used. As a result, files need to be given special attention if used as a metadata source.

These are all challenges facing the Norwegian University of Science and Technology (NTNU). Here the Local LMS (LLMS) metadata schema will be converted to NORLOM. The LLMS has a proprietary schema with little resemblance to the destination schema. It uses other element names, which can make discovering of existing metadata sources more challenging. It has extensive use of elements not covered by the IEEE LOM. And it has single elements covering multiple IEEE LOM elements. This results in one-to-many, many-to-one and many-to-none element situations. In addition files are a frequently used LO type, resulting in additional metadata challenges when included as a metadata source.

2 THE IEEE LOM SCHEMA

The IEEE LOM schema is specially adapted to describe LOs. It divides metadata elements into predefined categories: General, Life Cycle, Meta-Metadata, Technical, Educational, Rights, Relation and Annotation. For other metadata, a 9th category Classification can be used. The initial 8 categories open for LO descriptions containing more than 60 different elements, most of them reusable for multiple registrations. This vastness in numbers and the preciseness of each element poses challenges when moving from a local to this global metadata schema.

The Classification category where created to support a local LO identification schema. It allows creation of local elements within an existing schema structure. Other metadata elements can be included in this category. They are not globally valid, because they only follow a local schema. Re-usage of these metadata can only be performed by the local LMS and other LMSs and services compatible with the local schema.

3 USING AN EXISTING LMS AS METADATA SOURCE

3.1 Discovering potential metadata sources within the LLMS

The LLMS is divided into course-specific sections. Each course has a course-profile with information including: course-name, id, year and semester. The course id includes information about the “course owner”, such as the university department. Each course has predefined users which must log-in to gain course and LO access. Each user has a profile which includes user name, login-information and e-mail address.

The LLMS has functions for distributing course information. Common usage includes sharing of curriculum lists, slides from lectures, presentations of student assignments, e-mail and chat. The legal types of LOs are note, link, exercise, online test, question (chat) session, report and upload file. Each LO type have specific, predefined properties. All the LO types have administrative metadata: publisher name (creator), folder name, date and title.

The LLMS do not control or check uploaded files. Users can upload any file and store it in a course specific section. The most commonly used file formats are MS Office-based, Adobe PDF and JPEG images. These file types have extensive, custom metadata schemas. This is also true for many other used file formats. Hence files can be an uncertain and complicated metadata source.

3.2 Schema mapping

The LLMS has potentially multiple metadata sources: User-, Course-, Institution- and University profiles, and LOs created within the LLMS, as well as uploaded files.

The metadata elements of these sources should now be transferred to the new, global schema. (Zeng & Xiao, 2001) describes 4 relation types: one-to-one, one-to-many, one-to-none and many-to-one.

One-to-one relations are lossless and are used in crosswalks. Here equivalent element types are mapped as they were the same element type. This includes converting between equal schemas with different formatting, e.g. between date formatting: year, month, day vs. day, month and year.

One-to-many elements indicate that the destination schema has finer grain allowing more precise metadata descriptions. Common elements include descriptions of local custom elements.

One-to-none elements indicate a direct loss of metadata from the existing schema. Within any

converting process, an aim would be to avoid losing data. Effort should hence be enforced to avoid this issue.

Many-to-one elements indicate a less grained destination schema. This can result in less detailed metadata descriptions.

3.3 One-to-one elements

The precise definition of the LLMS' LO types, except files, can be used to create crosswalks or one-to-one element relations. This is because of equality between some of the predefined LLMS metadata schema elements and the defined targeting schema elements. Between the two schemas there are equal elements, like shown in Table 1.

Table 1: Title

LLMS metadata	LLMS title = Exercise nr 2
IEEE LOM metadata	1.2 Title = Exercise nr 2

3.4 One-to-many elements

Within the LLMS there is extensive use of local information which is not explicitly described. Moving from a local LMS schema to a global schema will require describing the local schema and its surroundings in the global schema's terms. This includes course specific elements and interpretation of local course characteristics. These can be collected in a course profile allowing LOs created or uploaded to the course to take advantage of the course profile. Candidate course profile elements include course description and its primary user group, as shown in Table 2.

Table 2: Course context

LLMS metadata	LLMS course context = IT3805
IEEE LOM metadata	5.5 Intended End User Role = Learner 5.8 Difficulty = Very difficult 5.11 Language = NO 9.2.2 Taxon = {["Institute", "IDI"]} 9.2.2 Taxon = {["Course", "IT3805"]}

Other candidate elements can be set at a general level for the University as a whole, at Institute and department levels, down to low level, fine grained elements set by individual course lecturers. These profiles can describe practical usage properties of the LMS and all its users, schema name, policy and other politically tuned elements. See Table 3 for an example.

Table 3: University context

LLMS metadata	LLMS University context = NTNU
IEEE LOM metadata	5.6 Context = Higher education 5.7 Typical age range = 18- 9.2.2 Taxon = {["University", "NTNU"]}

Some local elements require usage of multiple global elements to cover the local description. E.g. the LLMS' "Exercise" LO has a range of properties not covered by an individual LO type in IEEE LOM. To fully describe the "Exercise" LO multiple IEEE LOM elements have to be created, as shown in Table 4.

Table 4: LO type description

LLMS metadata	LLMS LO type = Exercise
IEEE LOM metadata	4.1 Format = text/html 5.1 Interactivity type = Active 5.2 Learning Resource type = Exercise 5.3 Interactivity level = High

3.5 One-to-none elements

The issue of one-to-none elements poses a danger of losing data when converting from a local to a global schema. One example is when converting the "Exercise" LO type. It has specific elements specifying if an exercise is mandatory and its delivery date, see Table 5. Such elements are not covered by the IEEE LOM schema.

Table 5: Local elements

LLMS metadata	LO: Obligatory = Yes LO: Final delivery date = 01.10.2006
IEEE LOM metadata	-

For these two exemplified elements and other elements without an equivalent IEEE LOM element, there are two lossless possibilities: Use of an unstructured general description or extend the IEEE LOM schema with custom elements. The first solution results in a many-to-one element situation with loss of precision within the schema as a result. Table 6 shows this scenario by storing the existing element names and entities as a merged text string within the General Description element.

Table 6: Using 1.4 Description for local elements

LLMS metadata	LO: Obligatory = Yes LO: Final delivery date = 01.10.2006
IEEE LOM metadata	1.4 Description = "Obligatory = Yes" 1.4 Description = "Final delivery date = 01.10.2006"

An alternative can be to use the Classification category to extend the global schema. This can result in a lossless schema and metadata coverage, see Table 7 ("NO" referring to language, other string elements refer to element content).

Table 7: Using Classification for local elements

LLMS metadata	LO: Obligatory = Yes
IEEE LOM metadata	9.1 Purpose = Educational Objective 9.2.1 Source = ("NO","NTNU LMS") 9.2.2 Taxon = {["Obligatory", "YES"]}

Use of the Classification category can resolve the missing global elements issue by creating local elements. Simultaneously it looses schema compatibility with other LMSs for these specific elements. One of the intentions of adopting the global schema is then lost. Therefore none of the choices for resolving the one-to-none element situation is perfect. Still we would recommend using the Classification category. This would avoid loosing schema grain and lost metadata. Such a decision would open up for sub-local schema cooperation with other LMSs. This would allow for schema extensions with compatibility between the sub-local LMSs. If the global schema should evolve to include these elements, the local schema could convert to the revised schema at that time.

3.6 Many-to-one elements

Many-to-one elements indicate a less grained target schema, allowing less detailed metadata descriptions. We have not found such elements from LO created within this LLMS. There are, however, multiple elements which are not covered within the IEEE LOM schema which could be mapped to the general description element for a many-to-one scenario.

In such a move the different elements would be merged into one element loosing their initial distinct properties; See Table 6. The metadata can then be stored within the schema, though they would not be accessible as individual elements afterwards. An alternative could be performed with local interpretation of the global schema. This would be in conflict with the global metadata schema. Our recommendation is to use the Classification category for these elements.

3.7 Taking advantage of other metadata sources

3.7.1 Automatically creating relations

There are tasks which a LMS can perform without user interaction. This includes updating metadata records with relations not specified by the user. Such relations can be based on:

- Relations between all LOs within the specific course.
- Folders are frequently used to manage LOs into smaller collections, e.g. for creating a compendium. LOs within the same folder can be given their own, additional relations.
- Two-way relations can be created if the LMS have the targeting LO included.
- Some LO types have included links to external sources, e.g. hyperlinks. Discovered links can be used for creating relations.

3.7.2 Creating keywords

The LMS can be an information provider to other algorithms for creating metadata: A course profile, as described in chapter 3.4, can be used indirectly by submitting background information for e.g. a domain ontology algorithm for generating object keywords. This makes the context analysis a basis for content metadata generation.

4 SPECIAL CHALLENGES REGARDING FILES

Our initial studies have shown that 66% of LOs within the LLMS are files. These can currently be described with a single description element. Though files can have much more they can tell.

4.1 Harvestable file element content

When files are created outside of a LMS and without a predefined document template, the LMS has no power to guide and form the content of the files. This being visual properties of the files or their metadata. If the LMS has information of the file format and its metadata schema, it can harvest metadata from such formatted files. Such collectable metadata is shown in Figure 1. Algorithms for file metadata harvesting has been introduced for specific metadata elements in projects including the AMeGA project (Greenberg et al., 2005), the Greenstone Digital Library (Witten et al., 2003) and in LOMGen (Singh et al., 2004).

```
<title>Slide 1</title>
<!--[if gte mso 9]><xml>
 <o:DocumentProperties>
  <o:Author>Lars</o:Author>
  <o:LastAuthor>Lars</o:LastAuthor>
  <o:Revision>3</o:Revision>
  <o:TotalTime>106</o:TotalTime>
  <o:Created>2006-03-08T11:28:10Z</o:C
  <o:LastSaved>2006-03-08T13:14:33Z</o
  <o:Words>208</o:Words>
  <o:PresentationFormat>On-screen Show
  <o:Company>NTNU</o:Company>
  <o:Bytes>20445</o:Bytes>
  <o:Paragraphs>48</o:Paragraphs>
  <o:Slides>12</o:Slides>
  <o:Version>11.6568</o:Version>
 </o:DocumentProperties>
 <o:OfficeDocumentSettings>
  <o:PixelsPerInch>80</o:PixelsPerInch
 </o:OfficeDocumentSettings>
</xml><![endif]-->
```


Figure 1: Metadata collected from a PowerPoint document

Contrary to the other LO types, the file content is not predefined based on the LLMS' LO types. A file can contain a questionnaire, a list of student names or have any other content. When uploading a file to the LLMS, there are no elements available to determine the LO type of the file contents.

File harvestable metadata opens for extensive metadata collection. Since these files where created outside of the LLMS, there are questions regarding the content of extracted metadata elements. One issue is less informative entities: e.g. in Figure 1 the author element has the entity "Lars". This is a less informative element than the full name collectable from the LLMS. Collectable metadata can also include errors which conflicts the file's metadata schema. Our studies have uncovered examples where file metadata elements have been replacement with advertisements.

Other elements can give more descriptive and precise metadata descriptions than elements created within the LLMS. This includes the element for document language; the LLMS do not have a dedicated element for LO language, whereas many text based documents contain registration of the actual language used.

LMSs must be maintained in order to recognize and take advantage of the currently used file formats.

4.2 One-to-none elements

Similar to the LLMS' other LO types; files can contain metadata which are not covered by the global metadata schema. These issues and solutions are equal to the LLMS' LO types, though the amount of elements with missing global elements can increase. We have discovered missing IEEE LOM elements for a file's number of pages, slides or spreadsheets, paragraphs, lines, words, characters, notes and creator- and producer application. For multimedia content there are missing elements for:

- Image: Resolution (dpi), number of pixels, colour depth
- Sound: Number of channels, bit-rate, actual content playing time
- Multimedia: Frames per second, image and sound metadata

In order to cover these elements lossless within the IEEE LOM schema extensive use of the Classification category would be required.

4.3 Many-to-one elements

When including files as a metadata source, this increases the number of candidate elements sources within the LLMS. Selecting the best candidate element can then be more challenging. For example we want to give a LO the correct title. The title element is specified in the LLMS and in the metadata for many file formats. Many documents can have a harvestable visual title. See the example in Table 8. Here we can choose from four element sources, but IEEE LOM gives room for only one title element. In order to determine the best candidate metadata source, when multiple sources are available, we need techniques to assist in this process.

Table 8: Multiple title sources

LLMS metadata	LLMS title = Exercise nr 2 File metadata title = IT3805 exerc. 2 File name = IT3805exec2 version 1 Visual title = Exercise 2 – Metadata
IEEE LOM metadata	1.2 Title = ?

5 SUMMARY AND FUTURE WORK

Converting a local LMS' metadata schema to a global schema requires extensive information about both the local and global schemas, the elements they contain and the intentions behind each element:

- The local LO types, their properties and how they can be used
- The local setting in which the LOs are created or published
- The "hidden knowledge" not explicitly present within the local schema or the LO, though available through local knowledge of the LMS, the LOs and the local educational system

- Available data sources and their potential metadata element sources, and
- The targeting metadata schema, its available elements and their intended usage.

Within the LLMS there is a potential to create rich IEEE LOM metadata records, where the data collection can be based on multiple data sources. This opens up for creation of descriptive metadata records with many finely grained elements enabling precise LO queries.

The technologies developed as crosswalks for a 2-way metadata transferral between schemas, have shown validity for this project. We have uncovered extensive schema mapping possibilities where:

- Single local elements described multiple IEEE LOM elements
- Local elements without a direct equivalent within the IEEE LOM schema
- Multiple local elements describing a single entity IEEE LOM element
- Reduced reliability caused by LO elements containing error-full metadata.

We have discovered that the file LO type is the prime candidate in order to locate Many-to-one elements. Files have shown to be a less reliable metadata source.

There are unresolved issues regarding how to deal with elements that are not covered by the current IEEE LOM version. Excluding these elements results in lost data. Using the Classification category results in elements not understood by other LMSs and services using the global schema.

In future work we will analyze the content of discovered metadata sources. This includes LO files collected from the LLMS in the Adobe PDF, MS Word, MS PowerPoint, MS Excel and JPEG file formats. We will analyze elements which have shown to contain entities and comparing elements where there are multiple candidate sources. This includes elements for title and author name. We will compare the results between the different file formats and the other LLMS' LO types.

By doing these efforts we will show which metadata sources that are available based on the LLMS, which metadata sources that should be used and which, if any, metadata sources to give priority.

REFERENCES

CETIS, 2004, *UK LOM Core v 0.2*, http://www.cetis.ac.uk/profiles/uklomcore/uklomcore_v0p2_may04.doc

Chan, L. M., Zeng, L. M., 2006, *Metadata Interoperability and Standardization – A Study of Methodology Part I - Achieving Interoperability at the Schema Level*, D-Lib Magazine, June 2006, Volume 12 Number 6, ISSN 1082-9873

eStandard, 2005, *Norsk LOM-profil – NORLOM. Versjon 1.0*, http://www.estandard.no/norlom/v1.0/NORLOM_v1_0_mars_2005.pdf

Greenberg J., Spurgin, K., Crystal, A., Cronquist, M., Wilson, A., 2005, *Final Report for the AMeGA (Automatic Metadata Generation Applications) Project*, UNC School of information and library science, http://www.loc.gov/catdir/bibcontrol/lc_amega_final_report.pdf

IEEE LTSC, 2005, *IEEE P1484.12.3/D8, 2005-02-22 Draft Standard for Learning Technology - Extensible Markup Language Schema Definition Language Binding for Learning Object Metadata*, WG12: Related Materials, http://ltsc.ieee.org/wg12/files/IEEE_1484_12_03_d8_submitted.pdf

Library of Congress, 2001, *Dublin Core/MARC/GILS Crosswalk, Network Development and MARC Standards Office*, 2001-03-12

Singh, A., Boley, H., Bhavsar, V.C., 2004, *LOMGen: A Learning Object Metadata Generator Applied to Computer Science Terminology*, National Research Council and University of New Brunswick, Learning Objects Summit Fredericton, NB, Canada, March 29-30, 2004, http://www.cs.unb.ca/agentmatcher/LOMGen/paper/LOMGen-29Mar2004.ppt

Witten, I. H., Bainbridge, D., Boddie, S., Don, K. J., McPherson, J. R., 2003, *Greenstone Digital Library – Inside Greenstone Collections*, http://www.greenstone.org/docs/inside_greenstone.pdf

Zeng, M. L., Xian L., 2001, *Mapping metadata elements of different format.* E-Libraries 2001, Proceedings, May 15-17, 2001, New York: 91-99. Medford, NJ: Information Today, Inc.