METADATA FOR LEARNING OBJECTS – A CURE FOR INFORMATION OVERFLOW?

Stefanie Sieber and Andreas Henrich
Media Informatics Group
University of Bamberg, Germany

ABSTRACT

Information overflow is one of the downsides that digitalisation brought with it. This obvious fact also applies to the field of e-learning. For this reason, concepts and systems supporting learners in handling the enormous amount of information – such as personal learning environments – have recently gained more attention. It gets more and more important to offer individual capabilities to store, retrieve, and keep track of the different learning contents or learning objects. Of course, metadata can be very useful to organise learning objects. Therefore, we have examined different metadata standards. Based on our insights, we developed a structure of metadata attributes in order to efficiently store learning objects. The results are presented in this paper. Also, thoughts on how to gain this metadata by employing different specific standards are given. To our understanding the use of a reasonable subset of metadata supports tutors and learners at the same time.

KEYWORDS

Learning Objects, Learning Context, Learning Reference, Metadata Standards, Dublin Core, Learning Object Metadata

1. INTRODUCTION

Today, there is neither doubt nor discussion about the justification and necessity of e-learning. E-learning and blended learning are accepted forms of modern learning that are taken for granted. The increasing distribution and use of e-learning – and also the ongoing research in various forms and surroundings – brings with it a great variety of e-learning offers and platforms. Among others, *personal learning environments* (PLEs) recently gained some attention – also due to a rediscovery of the importance of personalisation in learning (Schaffert and Hilzensauer, 2008).

One of the benefits of individual learning environments is the individual support that can be provided to a single learner. PLEs are – because of their individualisation – especially helpful for keeping track on learning and being able to get a personal overview on learning content. Our concept for a PLE particularly heading for this direction is presented in Sieber and Henrich (2009). Since this paper is going to be more about technical than learning-didactic aspects, our preferred term for content being included in PLEs is *learning objects* (LOs). In order to discuss LOs, it is necessary to give an appropriate definition of our understanding first. There are various existing definitions of the term, ranging from very general suggestions to highly specific restrictions. For our purpose we like to choose David A. Wiley's definition of a LO as "any digital resource that can be reused to support learning" (Wiley, 2000). Being more precise, LOs can be resources provided in learning management systems (LMS) and also XHTML web pages or PDF files accessed by a learner when browsing the Web or digital libraries during his or her learning activities.

Typically a lot of LOs will be included in PLEs. Therefore, searching in those LOs and also classifying LOs – and, more technically, storing and retrieving LOs – are two extremely important features for PLEs. In order to be able to accomplish these tasks, different kinds of information need to be obtained. This includes the learning content itself but also metadata information about the LOs. Metadata of good quality can, for example, be used as basis for faceted search facilities as becoming more and more popular in e-shops. Different organisations have developed several formats and standards on metadata. So these standards serve as guidelines for extracting the desired information.

Hence, a brief consideration of the most important metadata standards is the topic of section 2. As a result an appropriate structure on how to store LOs and their metadata is derived in section 3. Thoughts on how to

determine these information units are spent in section 4. This paper ends by concluding with the most important findings for effectively and efficiently storing LOs.

2. METADATA STANDARDS

There is a large number of standards for metadata information. These standards can be categorized in four groups: general metadata standards, learning-specific metadata standards, format-specific metadata standards, and subject-specific standards. The two most important standards are *Dublin Core* (DC)[1] or more precisely the DCMI Metadata Terms, a standard that can be used for various concerns and formats from the group of general metadata standards, and *Learning Object Metadata* (LOM)[2], the most common standard for learning.

However, for reasons of completeness a few other standards need to be mentioned too. Talking about general metadata standards, it is indispensible to consider the standards originating in librarianship. Standards like *Machine Readable Cataloguing* (MARC)[3] or the *Metadata Objects Description Schema* (MODS)[4] have been used for a long time. Also, reference management systems such as *BibTeX*[5] – typically used together with LaTeX – can deliver valuable input on important metadata attributes.

Taking into account our learning context, it also needs to be referred to *IMS Metadata* and the *Sharable Content Object Reference Model* (SCORM). The IMS Global Learning Consortium[6] dedicated itself to "improving education and learning through the strategic application of technology". Therefore, among other standards, they developed IMS Metadata. Furthermore, the Advanced Distributed Learning Initiative (ADL)[7] is also aiming "to leverage the power of newer technologies" and developed SCORM that integrates various standards like DC, LOM, and IMS.

3. AN APPROPRIATE STRUCTURE TO STORE LEARNING OBJECTS

What should be obvious is that DCMI Metadata Terms and LOM are the de facto standards for dealing with metadata. Comparing both, LOM is learning-specific and also built upon DCMI Metadata Terms. Therefore the following adaption uses the LOM data elements that can be found in the LOM Draft (IEEE 2002). To refer to DC, reference is also given to the LOM Draft (IEEE 2002, p.44) that provides a table mapping the DCES terms to the LOM data elements.

According to the LOM standard, there are roughly 100 attributes that can be employed to enrich LOs. Unfortunately, lots of these attributes can only be set if explicitly specified by a tutor or manually added by a learner, which is – if possible at all – for both challenging and time-consuming. Naturally this approach would lead to lots of empty values. For this reason, we are focusing on metadata elements that can either be determined automatically or appended with justifiable effort. We derived a reasonable subset of LOM data elements to enrich LOs. These considerations all apply to the general level of LOs; format-specific considerations, standards, and attributes are explicitly excluded at this point. However, a side note to format-dependent issues will follow in section 4.

As mentioned above, LOs are basically any kind of material useful for learning. Therefore, information that can be reliably extracted is limited to general attributes as shown in Figure 1, where the attribute "location" is used as unique identifier for a particular LO. Also, general information like the title, language, description, and format of the LO are stored. Furthermore, the attribute "copyright and other restrictions" is going to be used as support for collaboration scenarios in order to denote different sharing levels for LOs. This is obviously a very individual setting, which is why this kind of metadata cannot be extracted automatically but easily added by using presets.

[1] Dublin Core – http://dublincore.org/
[2] IEEE Learning Technology Standards Committee – http://www.ieeeltsc.org/
[3] Machine Readable Cataloging – http://www.loc.gov/marc/
[4] Metadata Object Description Schema - http://www.loc.gov/standards/mods/
[5] BibTeX – http://www.bibtex.org/
[6] IMS Global Learning Consortium – http://www.imsglobal.org/
[7] Advanced Distributed Learning Initiative – http://www.adlnet.gov/

So far, these considerations only apply to single LOs, whose granularity is simply determined by the given file structure. Thus, to improve the quality and value of LOs, we actually need to go beyond file delimitations and consider content-based features to determine the logical structure of one or more connected LOs. To give an example – a text-based document can be further divided by using the headings included in the document. If done so, the three attributes "structure", "aggregation level", and "relation – kind" need to be part of LO metadata as well. The particular implementation will be subject to further research and build on work already done in this area such as IEEE (2002), Currier and Campbell (2002), and CLEO Collaborative Partners (2003).

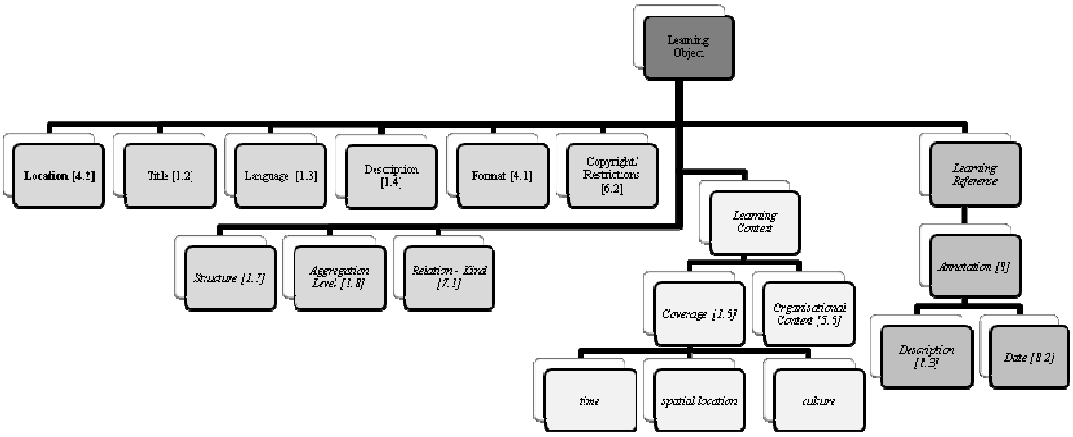

Figure 1. Metadata attributes for Learning Objects

Additionally, we intended to emphasize two different aspects covered in the LOM standard – contextual information and annotations. Therefore two additional units – *Learning Context* (LC) and *Learning Reference* (LR) – have been designed.

What we figured out as one of the most important additional information is contextual information. Therefore contextual information is not directly included into LO metadata but stored in an additional unit. Being more precise, this unit is only meant to link to hierarchical information on different contextual facets such as time, spatial location or organisational units, as shown in Figure 1. Summarizing, especially the organisational context or different aspects on the coverage of a particular LO can be stored. Since an automatic extraction is unlikely, the integration should be achieved by using presets and preferences.

In contrast to the information unit LC, LRs are used to store non-hierarchical information. This unit basically can be used to store any information a learner might want to associate to a LO, without being forced to further classify the desired description. This concept is, of course, derived from the concept of tagging – one characteristic of Web2.0. It seems to be a worthwhile idea to integrate an established concept like that. This way, any information that seems qualified for the learner can easily be integrated as additional information in a controlled manner.

4. HOW TO DETERMINE METADATA

Now that the attributes have been described, a closer look needs to be taken on how to actually determine the desired metadata. Therefore, it is necessary to consider the different formats of LOs and format-specific metadata standards in order to actually extract the metadata. The idea is that the metadata attributes described in section 3 can be automatically extracted for LOs in one of these formats, providing metadata based on one of these format-specific standards. A non-exhaustive overview on possible formats and related standards is shown in Figure 2. Besides format-specific standards there are also subject-specific metadata standards as for example the standard provided by the Text Encoding Initiative (TEI)[8] for the humanities and social sciences

[8] Text Encoding Initiative – http://www.tei-c.org/

or the Categories for the Description of Works of Art (CDWA)[9] for describing artwork. By now, these standards will be intentionally left out because they are too specific. Nevertheless those standards might be helpful in particular cases.

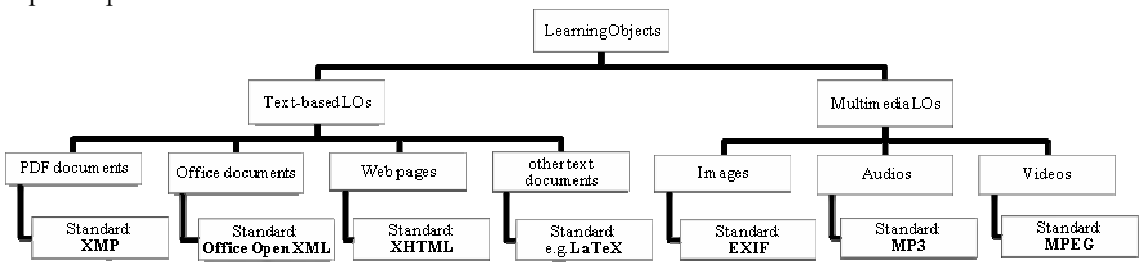

Figure 2. Different kinds of Learning Objects and associated metadata standards

Using the general distinction of text-based and multimedia LOs, we start by considering **metadata extraction for text-based documents**. In order to do so, several standards need to be considered, depending on the particular format of a text-based LO. Glancing over existing formats, there are at least four or five text-based formats that definitely need to be included as formats for possible LOs.

One of the most common text-based formats nowadays is the *Portable Document Format* (PDF). Basic metadata can simply be added through file info and document properties and is – most of the time – already automatically included in PDF documents such that it can be easily extracted. The more powerful metadata standard used in PDF documents is the so-called *Extensible Metadata Platform* (XMP), a XML format for representing metadata. This standard allows the inclusion of almost every desired information as metadata attribute – however by including it manually.

Proceeding, *Microsoft Office* is also one of the formats to be taken care of. The following explanations concentrate on the newest version – Microsoft Office 2007. Since, Office 2007 is XML-based, metadata is automatically included as hidden XML using *Office OpenXML*[10] and also the DCMI metadata terms. Therefore the basic information selected above is typically available.

Next, it is indispensable to talk about *XHTML* documents as possible LOs. Technically, the XHTML Metainformation Module[11] included in the XHTML standard does not define a normative set of properties. However, there is a metadata profile defined by DC[12] standardising the integration of DCMI metadata terms into XHTML pages. Needless to say, the desired information should be available that way.

Other text documents can be of various formats. One of the maybe most important ones among those is *LaTeX*. Generally, the existing metadata is depending on the format LaTeX is transformed to. However, the preamble of a LaTeX document allows the definition of metadata information that can also be transferred to a resulting PDF document – which is supposed to be the most common final format.

Especially when talking about web pages providing learning content, an important differentiation needs to be drawn. On the one hand, there are web pages being true LOs and actually containing learning content. On the other hand, web pages are also hubs providing lots of different LOs like LMS. Since regular XHTML web pages have already been dealt with, the question of how LMS incorporate metadata rises. Unfortunately there is no standardisation at all so far. Just considering the two maybe most important ones – Moodle and Blackboard – a completely different behaviour occurs. While Blackboard allows the enrichment of LOs with metadata while uploading the existing data into the system, Moodle completely lacks of metadata consideration. Also the web pages automatically generated by LMS rarely provide useful metadata.

Now speaking about **multimedia LOs**, there are basically three different media types that need to be taken into account – images, audio, and videos. Hence, existing standards for these different formats need to be examined. At this point, the important standards are only named briefly but of course they need to and will be considered in more detail for the actual implementation of the overall structure for LOs.

[9] Categories for Description of Works of Art – http://www.getty.edu/research/conducting_research/standards/cdwa/
[10] Office Open XML – http://msdn.microsoft.com/en-us/library/aa338205.aspx
[11] XHTML Metainformation Module – http://www.w3.org/TR/2002/WD-xhtml2-20020805/mod-meta.html
[12] Expressing Dublin Core in HTML/XHMTL meta and link elements - http://dublincore.org/documents/dc-html/

Starting with images, there are the *IPTC Photo Metadata Standards*[13] (Information Interchange Model, IPTC Core and Extension) that allow the integration of information on photographs into the image files, or the *JEITA Exchangeable Image File Format* (EXIF)[14] for storing details on shooting the photo, and the *NISO Metadata for Images in XML* (NISO MIX, ANSI/NISO Z39.87)[15] to raster digital images.

Looking at audio and video files the ISO standard ISO/IEC 15938 also known as Multimedia Content Description Interface or *MPEG-7*[16] provides possibilities to enhance several formats with metadata information. A more specific possibility is including ID3[17] data containers into MP3 – one of the most common audio formats – to deliver information with an audio file.

Besides those typical formats there are other forms of information that can be taken care of – like geographical data. The standard that should be mentioned here is the *Content Standard for Digital Geospatial Metadata* (CSDGM, ISO 19115)[18] that is maintained by the Federal Geographic Data Committee (FDGC).

5. CONCLUSION

To conclude, we like to summarize the future challenges resulting from the findings evolved above. We developed a reasonable subset of LOM data elements that can easily be automatically extracted or at least appended with justifiable effort. Of course, limiting metadata extraction to these few attributes might also cause a loss of important information. Nevertheless, to our understanding an indeed small but well-tended set of metadata is much more useful than a bigger set coming with the catch of lots of empty values.

As already mentioned, one of the next steps needs to consider the format-specific standards more intensely. Secondly, future work will also deal with a logical fragmentation to LOs in order to maximize information included in our structure that can easily be used for reasoning and recommendation later, as already described in Sieber and Henrich (2009).

Finally, what should not be forgotten is that all these findings still rely on e-learning authors providing at least basic metadata when creating e-learning content. Mostly, basic information can be included by simply configuring the system and programs used for producing the e-learning units correctly. However, this is a fact that cannot be avoided; therefore we should all remember this when developing e-learning content ourselves.

REFERENCES

CLEO Collaborative Partners, 2003. *CLEO Extensions to the IEEE Learning Object Metadata*, Version 1.1, CLEO White Paper CLEO001, http://www.cleolab.org/CLEO_LOM_Ext_v1d1a.pdf, last visited 2010-01-25.

Currier, S. and Campbell, L., 2002, Evaluating learning resources for reusability: the "dner & learning objects" study. *Winds of Changing in the Sea of Learning, Proceedings of the 19th Annual Conference of the Australian Society for Computers in Tertiary Education (ASCILITE).* Auckland, New Zealand, pp. 139-148.

IEEE, 2002. *Draft Standards for Learning Object Metadata.* http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf.

Schaffert, S. and Hilzensauer, W., 2008, On the way towards Personal Learning Environments: Seven crucial aspects. In *In eLearning Papers* (9), http://www.elearningeuropa.info/files/media/media15971.pdf, last visited 2010-01-25.

Sieber, S. and Henrich A., 2009, Effectively Managing and Processing Personal Learning Content. *Advances in Web Based Learning, Proc. of the 8th Intern. Conference on Web Based Learning (ICWL).* Aachen, Germany, pp. 363-366.

Wiley, D., 2000, Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. *In The Instructional Use of Learning Objects: Online Version 2000.* http://www.reusability.org/read/chapters/wiley.doc

[13] International Press Telecommunications Council Photo Metadata Standards – http://iptc.org/cms/site/index.html?channel=CH0089
[14] Japan Electronics and Information Technology Industries Association Exchangeable Image File Format – http://www.jeita.or.jp/
[15] National Information Standards Organization Metadata for Images – http://www.loc.gov/standards/mix/
[16] Multimedia Content Description Interface – http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm
[17] ID3 – http://www.id3.org/
[18] Content Standard for Digital Geospatial Metadata – http://www.fgdc.gov/metadata/csdgm/