T. Hirashima et al. (Eds.) (2011). Proceedings of the 19th International Conference on Computers in Education. Chiang Mai, Thailand: Asia-Pacific Society for Computers in Education

Expanding the Scope of Metadata and the Issue of Quality

Tore HOEL[a] & Jon MASON[b]
[a]*Centre for Educational Research and Development, Oslo and Akershus University College of Applied Sciences, Norway*
[b]*Faculty of Education, Queensland University of Technology, Australia*
tore.hoel@hioa.no
jon@intercog.net

Abstract:
This paper summarises the current context of metadata standards development relevant to Information Technology for Learning, Education, and Training (ITLET). A number of issues are identified that need further discussion in order to harness the potential of ICT in LET. In particular, issues related to assessing quality of standards; how to scope ITLET standards; and how to specify metadata for explanatory content are highlighted.

Keywords: metadata, ICT standards, IEEE LOM, ISO MLR, learning resource description, learning technologies, ITLET

1. Introduction

Metadata is a term that can convey diverse meanings, probably because the root 'meta' has rich semantics and can mean: 'beyond', or 'above' as in metaphysics or metacognition; 'behind' or 'after' as in metaphase; 'change' as in metamorphosis; and, 'together with' as in metabolism. It is a term that has been adopted in recent decades by standards development organizations such as the Institute of Electronics and Electrical Engineers (IEEE) and the International Organization for Standardization (ISO). It also describes the core focus of the Dublin Core Metadata Initiative (DCMI) and is basic to the specifications developed by the IMS Global Learning Consortium (IMS GLC).

While both the IEEE and ISO choose not to define it within their standards that are used by IMS GLC and many other stakeholders worldwide [1, 2], the DCMI uses a very short and commonly accepted definition of metadata as "data about data". The US-based National Information Standards Organization (NISO) provides a longer version as: "structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource" [3]. While it is also noteworthy that Wikipedia declares it to be "an ambiguous term" [4], evidence suggests a common and evolving understanding of this term within the standards development community is that metadata is data that not only describes or identifies other data, but also information resources (digital and non-digital), events, people, and services [5, 6]. Following on from this the discourse on metadata standards for learning is far from over and indications are that the scope of what metadata does within ICT infrastructure is expanding.

It is commonly stated, however, that "the nice thing about standards is that there are so many of them to choose from." This quote, attributed to Andrew Tanenbaum [7] highlights the dilemma of many implementers of metadata standards. Due diligence, particularly in the ISO context, requires that any new standardization project must identify and situate its scope of work in the context of having navigated a complicated standards landscape of the past, relevant current initiatives. For implementers, the challenge is discerning those standards that are best fit for purpose.

IEEE LOM was the first e-learning metadata standard, published in 2002 [1]. Other *de facto* standards at the time, such as the Dublin Core (DC) Metadata Element Set [8], also provided a means of describing learning resources, although without the same degree of specificity. However, this situation unfortunately led to significant confusion by stakeholders in the education community. While both these standards had similarities they were not interoperable, and implementers of systems were left to solve the interoperability issues with workarounds. It was in this context, following the formation within ISO/IEC of SC36 (subcommittee 36, IT for Learning, Education and Training – ITLET), that a project on Metadata for Learning Resources (MLR) was initiated. While there were perceived deficiencies with IEEE LOM, however, national bodies participating in SC36 wanted to produce "a harmonized standard" with LOM and DC. Also, because of industry adoption of LOM as a result of the market success of SCORM [9] they also wanted to "maintain compatibility with the implementation of IEEE 1484.12.1:2002 where practicable". Apart from these initial requirements the consensus about direction of development was not clear and the early years of drafting a framework for specification went through much iteration. This was further complicated as a consequence of significant and evolving work taking place within the Dublin Core Metadata Initiative (DCMI) [8]. As a consequence, SC36 used the rest of the decennium before finalizing their first MLR standard, the framework part, published in 2011 [2].

It might be easy to criticize SC36 on the time span involved in delivering MLR to the ITLET community without understanding the full context – a context which is characterized by unprecedented innovation in information and communications technology (ICT) development and strained relations between a number of the specifications and standardization bodies. Standards development does not happen in isolation, however, and ensuring optimum process and quality outcomes is dependent upon a number of factors. Probably of most importance for the next steps in metadata standards is consideration of the broader context of ICT innovation on the Web and the consequences this has for the specification and use of metadata.

1.1. The Evolving Web

From the very beginnings of the Web metadata has played a key role shaping its ongoing architecture. Whether in the form of HTML META tags (to assist search engines of the time), as hypertext enabling the foundational utility of the Web, or as indexes produced by the sophisticated algorithms developed by Google, metadata has existed in various forms. Hypertext itself represents a powerful instance of executable metadata; Extensible Markup Language (XML) is a key technology that facilitates the interoperability of structured content and provides capacity for very rich encoding of metadata; Web 2.0 tags used in social software applications (such as del.icio.us, Twitter, and Flickr) provide a collective, informal means of classifying and sharing resources [10]; and metadata standards such as the IEEE LOM [1] and the Dublin Core Metadata Element Set [12] have provided standard mechanisms for the description of resources on the Web for many years [5,6].

The recent publication of the ISO standard, *Metadata for Learning Resources – Part 1: Framework*, builds on both these standards and provides an extensible and modular framework for specifying new metadata data elements [2].

While much of the development of e-learning standards started after it became clear that the web would have a huge impact on learning technologies the relational database paradigm continued to influence metadata development. The IMS Content Package specification [13] was developed in tandem with the IEEE LOM and has played a dominant role in moving resources between applications, despite this specification having been modeled on the pre-web CD-ROM media. This legacy contributed to controversy in the development of ISO MLR with regards to "web enabling" the standard, leaving the database area to embrace the web architecture view on metadata.

An environmental scan of related metadata innovations reveals a number of other uses of metadata, particularly in relation to e-learning. For example, the European standards committee CEN has recently standardized Metadata for Learning Opportunities [14], based on the Dublin Core Abstract Model (RDF). This is the first in a series of smaller standards that address "European learner mobility". We also see that syndication technologies such as RSS and ATOM, which play an important role in Web 2.0 technologies, are commonly used in educational settings and are dependent upon metadata (RDF or XML) to function. For the ITLET standards community, however, it is essential to adopt or adapt the work of other communities, e.g., the digital library community. This is because the scope of ITLET embraces *both* standards that need to be specifically developed for ITLET purposes (such as e-learning) as well standards that are useful for it (such as digital library protocols like METS) and promote interoperability with an ever expanding domain of Web technologies [15]. This breadth of scope, however, has also brought alternative specifications and development philosophies to the standardization committee tables. This has often created tensions regarding due process and often politicized or tribalised the consensus-building process required. So, the questions arise: *where to go from here with the application of metadata to ITLET systems and services and how might issues of quality be addressed within the broadening scope of development?*

2. Implementation and the issue of Quality – the top-down perspective

In addressing the quality of a metadata standard a number of guidelines and frameworks aim to improve both the product quality (i.e. the quality of the information models, profiles, bindings etc.), and the process quality (i.e. the modeling technique) [16].

A metadata standard is typically developed once conceptual modeling of a particular domain, e.g., the field of educational technology, is clearly established. Some commentators claim that all metadata schemas have an inherent model, whether this has been made explicit or not [17]. Such a model is also known as an "abstract model", as it describes the abstract overarching information structure the standard relies on for implementation, e.g., the information model or a binding [17]. In the case of DC metadata [8] an abstract model was produced many years after the element set was standardized. The fact that over time it became perceived as necessary to specify is significant, and vindicates the view that was already inherent. Interestingly, the history of the DCMI can be seen as a bottom-up development that responded quickly to the immediate needs of the rapidly evolving Web.

The product quality raises issues of how well the developed models are fit for purpose or give a good description of the domain. This discourse on quality is often based on top-down frameworks [18] which define quality dimensions inspired by e.g. accounting principles, as the Guidelines of Modeling of Schuette and Rotthowe [19]. The guidelines include six principles to improve quality of information models: correctness, clarity, relevance, comparability, economic efficiency, and systematic design – the first three being a more necessary precondition for quality than the other dimensions [16]. Correctness and relevance are principles also found in the semiotic framework (SEQUEL) developed by Lindland *et al.* in 1994 [20], in which the Perceived Semantic Quality dimension is measured by indicators of correctness, relevance, completeness and authenticity. Quality goals of completeness (expressiveness) and validity are not easy to validate empirically. Krogstie *et al.* [21] have revised their SEQUEL framework, taking into consideration the problems participants had making reliable evaluations by means of it. They also supported the criticism of the first version SEQUEL taking a too static view of the domain, leaning towards the "knowledge-as-object" view talking about the model as "externalized knowledge" [21]. Krogstie *et al.* has come up with a concept of "active models" [21], leaving behind the descriptive conceptual modeling ('as-is' modeling) in favour of more prescriptive process models. "The notion of quality for a model is extended by looking at its ability to facilitate learning and action, more then just being a representation of the domain" [21]. This extended quality notion establishes also a new context for assessing metadata standards. The following short case study illustrates why this is much needed.

2.1 Measuring quality by counting elements in different metadata standards

In the July 2011 issue of IEEE Learning Technology newsletter Pons *et al.* give an update on "e-learning metadata standards" and a comparison between "the most commonly used standard for learning object metadata" IEEE LOM and the new ISO/IEC MLR [22]. They note that the new standard is "based on two basic principles", modularity (since it is a multipart standard) and compatibility (since it "opts for compatibility with LOM and Dublin Core"). Pons *et al.* state that the "usefulness of an educational resource metadata corresponds to data stored about pedagogical features of the educational resource". This leads to a quantitative assessment of which of the two standards that have the highest number of data elements covering the different dimensions of the standard. Pons *et al.* find that MLR "overall incorporates 45% more" of educational information, and "offers much more capacity to include information about intellectual property" (finding 25 such elements in MLR and only 3 in LOM) [22].

This is not the place to discuss what the authors of this IEEE LT newsletter story might have misunderstood about ISO MLR, considering that the parts on educational elements and IPR elements are not yet published. They are not alone in assessing quality of standards in this way. When a European best practice project recently reported on strengths and weaknesses of current specifications and standards it concluded: "ICOPER chose to use LOM instead of DC-Ed as the base standard for Teaching Method/Unit of Learning application profiles because fewer description fields could be mapped to DC-Ed." [23]

The authors of this paper observe that it is a widespread approach to appraise the quality of metadata standards by putting a quantitative metric on expressiveness, assessed by the number of data elements defined in the standard. Our claim is that this implies a static and ontological idealistic view on knowledge, which does not align with an active concept of knowledge and Learning, Education and Training needed in the 21st Century.

3. Quality of standards – the bottom-up perspective

Mendling *et al.* [18] reports on several works on bottom-up metrics related to quality aspects of process models. In summary, this research concluded that "larger models tend to be negatively connected with quality" [18]. Model size is important for understandability of the model. Therefore, a large number of metadata elements will at some level inhibit the quality of a standard, at least when it comes to implementation by diverse communities of stakeholders.

However, expressiveness is not necessarily linked to size of a model, measured by number of data elements. When the CEN WS-LT discussed design principles of the CEN MLO standard the group came up with a declaration stating

"Harmonization efforts should focus on small, simple models based upon existing commonalities that can be expanded upon at national or regional level, rather than all-inclusive monolithic standards." [24]

The key phrase here is "expanded upon". Building on the Dublin Core Abstract Model principles the European experts agreed that *extensibility* was a more important quality criterion than *completeness* (i.e., that the standard covered all aspects of the domain). This lifted the quality discussion to a meta level, from the coverage of the metadata model to the ability of the model language to cope with future need for constructs.

4. Emerging requirements

4.1 Due Process towards a good Product

The LET domain is described as complicated, complex, emergent and adaptive [25, 26] and this highly dynamic context is putting great strain on both the process of standardization and the resulting products. The legitimacy of the activity is under threat [27], and therefore, it is an urgent need to deliver on promises. However, it is impossible to change the process without changing the outcomes. A more agile design process will give a different kind of standards. At the moment, the authors do not observe a strong debate on design principles for ITLET standards. Without an informed discussion on both process and product the tensions in the LET standards community will be disguised as a conflict between organizations and strong personalities rather than between different design approaches.

4.2 Expanding the scope of metadata

No e-learning standard is "pedagogically neutral", as the US Department of Defense (ADL) put it when they first published SCORM [9]. Underlying every standard is an idea of rewarded activity taking part in learning, education or training. Delivery of content has been a primary concern till now. The ubiquitousness of means of *communication* with the evolving web opens up new possibilities for a wider range of pedagogical ideas to be supported by technology. Exchanging units of meaning rather than units of content expands the scope of metadata. We therefore see that recent terminology like "learner-ownable information", "competencies", "learning opportunities", etc. enter the scope statements of ITLET standards. This development has just begun and calls for leadership.

4.3. Beyond Metadata about Learning Content

Conceptions of "learning content" and "learning resources" are in most practical cases semantically equivalent. However, the short history (15-18 years) of metadata standards development relevant to ITLET demonstrates a bias toward "object oriented" conceptions of content. IEEE LOM has been the most successful metadata standard to date that has been specifically developed to support e-learning. Its name gives clear emphasis to the notion of a "learning object". Likewise, in the ISO context, the publication of MLR builds on this approach, although it also implicitly embraces the entity-relationship model of the Semantic Web and the Dublin Core Abstract Model. But in neither of these cases is there any (current) specific support for the discovery of content that is specifically explanatory in nature. Both schemas ultimately privilege content or learning resources that can be described or referenced by factual *information* (derived from the primitives: *who*, *what*, *where*, and *when*). However, *explanation* is more than information; it is often a key to comprehension, understanding, and learning. Clearly, if a metadata specification (such as MLR) purports to support learning in any holistic sense then a means for specifying the varieties of explanatory content and the metadata elements required to do so represents work that could be undertaken.

Mason [28, 29] presents a sense-making model that indicates potential new scope for the application of DC-based metadata in relation to accommodating *explanative* metadata – that is, metadata that can effectively reference *explanatory* as opposed to just *descriptive* content. Mason argues that among the many questions a learner might ask while learning or discovering content *why*-questioning is important as it is closely linked to reasoning and critical thinking. Despite advances in natural language search engines, such as TrueKnowledge [30] and DeepQA [31] there is still a lack of Web-based tools that can support learners asking *why*-questions. This is partly due to the fact that *why*-questioning is often heavily dependent upon context and the variety of *why*-questions possible spans causal, motivational, conditional, and teleological dimensions [32]. Thus, this adds complexity to any information that needs to be rendered into suitable metadata elements.

5. Issues

The following represents a preliminary list of issues that relate to metadata specification and implementation in the field of ITLET. They are listed as a means of stimulating discussion in the context of a workshop:

Issue #1 Quality assessment of specification and standards – are the principles articulated by Schuette and Rotthowe (correctness, clarity, relevance, comparability, economic efficiency, and systematic design) adequate?

Issue #2 Standards development and the tension between innovation and standardization – choosing a good process for a good product and the demands for agility

Issue #3 The expanding scope of opportunities for learning with ICT requires a broadening understanding of stakeholder requirements with regards to standards

Issue #4 How to specify metadata for *explanatory* content, and how to develop an information model of explanatory content?

6. Conclusions

It is not clear if a little more than two decades of development history of ITLET standards should be characterized as a success or a failure. A great number of standards are published, many of those metadata standards. The evolving Web and pedagogical, economical, demographical and other challenges have changed the context for ITLET standards development expecting the standards community to engage in a discourse covering both the quality and design principles of the current work, and the scoping of new and innovative metadata schemes.

This paper has argued that issues of quality, both of process and product, should be addressed. Furthermore, there is scope for an expanded view on metadata in ITLETmoving beyond content objects into the domain of learning opportunities and competencies. It would appear to be timely for the standards community to address metadata for explanatory content in order to assist in stimulating ICT support for the sense-making activities so important for learning. The paper is a contribution to a workshop on Strategic Approaches for e-Learning Standards and as such more concerned to find a strategic entry point to the discussion. However, this aim has limited the authors' possibilities to discuss in depth where to go with metadata development.

7. References

[1] IEEE (2002). IEEE Standard for Learning Object Metadata (LOM) 1484.12-2002, Institute of Electronics and Electrical Engineers.
[2] ISO (2011). ISO/IEC 19788-1:2011 Information technology – Learning, education and training – Metadata for learning resources -- Part 1: Framework. International Organization for Standardization.
[3] NISO (2004). Understanding Metadata. National Information Standards Organization. Available at: http://www.niso.org/publications/press/UnderstandingMetadata.pdf
[4] WIP (2011). Wikipedia entry on Metadata, Available at: http://en.wikipedia.org/wiki/Metadata
[5] Mason, J. (2004). Context and Metadata for Learning, Education, and Training. in R. McGreal (ed.) Online Education Using Learning Objects, London: RoutledgeFalmer.
[6] Mason, J. (2005). Making Sense of Metadata: A Potential Application of Sense-Making Methodology. Paper presented at a non-divisional workshop held at the meeting of the International Communication Association, New York City, Available at http://communication.sbs.ohio-state.edu/sense-making/meet/2005/meet05mason_lit.pdf
[7] Tanenbaum https://secure.wikimedia.org/wikiquote/en/wiki/Andrew_S._Tanenbaum
[8] Dublin Core (DC) Metadata Element Set, online at http://dublincore.org
[9] ADL (2000). Sharable Content Object Reference Model (SCORM), Version 1.0. Advanced Distributed Learning. Available online http://www.adlnet.org

[10] Vuorikari, R. (2009). Tags and self-organisation: a metadata ecology for learning resources in a multilingual context, SIKS Dissertation Series No. 2009-38, Dutch Research School for Information and Knowledge Systems, Available at: http://files.eun.org/riina/thesis_Vuorikari_sm.pdf
[12] ISO (2009). ISO 15836:2009 Information and documentation – The Dublin Core metadata element set. International Organization for Standardization.
[13] IMS Global Learning Consortium Content Package specification, online http://www.imsglobal.org/content/packaging/
[14] CEN (2011). EN 15982 -Metadata for Learning Opportunities- (MLO) Advertising
[15] METS, http://www.loc.gov/standards/mets/
[16] Becker, J., Rosemann, M., Uthmann, C.: Guidelines of Business Process Modeling. In van der Aalst, W., Desel, J., Oberweis, A., eds.: Business Process Management. Models, Techniques, and Empirical Studies. Springer, Berlin et al. (2000) 30–49
[17] Nilsson, M. (2010). From Metadata Interoperability to Harmonization of Standards, 1–233, PhD thesis, Royal Institute of Technology, Stockholm
[18] Mendling, J., Reijers, H. A., & Cardoso, J. (2007). What makes process models understandable? (pp. 48–63). Springer-Verlag.
[19] Schuette, R., & Rotthowe, T. (1998). Lecture Notes in Computer Science. (G. Goos, J. Hartmanis, J. Leeuwen, T.-W. Ling, S. Ram, & M. Lee, Eds.) (Vol. 1507, pp. 240–254). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-540-49524-6_20
[20] Lindland O,, Sindre G and Sølvberg A. (1994). Understanding quality in conceptual modeling. IEEE Software 11(2), 42–49.
[21] Krogstie, J., Sindre, G., & Jørgensen, H. (2006). Process models representing knowledge for action: a revised quality framework. European Journal of Information Systems, 15(1), 91–102. doi:10.1057/palgrave.ejis.3000598
[22] Pons, D., Hilera J. R. and Pagés, C. (2011). E-learning Metadata Standards, Special issue of the IEEE LT Newsletter on Adopting Standards and Specifications for Educational Content, Vol. 13, Issue 3, July 2011, online at http://www.ieeetclt.org/issues/july2011/IEEE-LT-Jul11.htm
[23] ICOPER (2010). D8.6 "Gap analysis report – conclusions of strengths and weaknesses of current specifications and standards", online at http://www.icoper.org/results/deliverables/D8-6
[24] CEN WS-LT (2008). The Athens Declaration
[25] Cooper, A.R. (2010), Key Challenges in the Design of Learning Technology Standards: Observations and Proposals, International Journal of IT Standards and Standardization Research, vol. 8, no. 2, pp 20-29.
[26] Mason, J. and Hoel, T. (2011) The Relevant Question and the Question of Relevance, in T. Hirashima et al. (Eds.). Proceedings of the 19th International Conference on Computers in Education. Chiang Mai, Thailand: Asia-Pacific Society for Computers in Education
[27] Hoel, T. and Hollins, P.A. (2008) Learning technology standards adoption – how to improve process and product legitimacy, ICALT 2008
[28] Mason, J. (2008). A model for sense-making: exploring why in the context of learning and knowing. Proceedings of the 16th International Conference on Computers in Education, (pp. 545-549), Taipei, Taiwan: Asia-Pacific Society for Computers in Education. Available online http://www.apsce.net/ICCE2008/papers/ICCE2008-paper286.pdf
[29] Mason, J. (2009). Knowledge Management and Dublin Core, International Conference on Dublin Core and Metadata Applications, Seoul Korea, pp.41-50 http://dcpapers.dublincore.org/ojs/pubs/article/view/953/950
[30] TrueKnowledge. http://www.trueknowledge.com/
[31] Ferrucci, D. (2011). Smartest Machine: Expert Q&A, PBS Online, WGBH Educational Foundation http://www.pbs.org/wgbh/nova/tech/ferrucci-smartest-machine.html
[32] Evered, R. (2005). A Typology of Explicative Models, in C.C. Lundberg & C.A. Young (Eds.) Foundations for Inquiry: Choices and Tradeoffs in the Organizational Sciences, Stanford, CA: Stanford Business Books.