
Towards flexible learning object metadata

Christopher Brooks* and Gord McCalla

Department of Computer Science,
ARIES Laboratory,
University of Saskatchewan,
Saskatoon, Saskatchewan, Canada S7N 5C9
E-mail: cab938@mail.usask.ca
E-mail: mccalla@cs.usask.ca
*Corresponding author

Abstract: This paper outlines the research we are doing in acquiring, describing and using learning object metadata. Instead of the IEEE LOM and other standardised metadata schemes, we argue for a more flexible approach to both defining and associating metadata with learning objects. This approach, which we call the *ecological approach*, sees metadata as the process of reasoning over observed interactions of users with a learning object for a particular purpose. Central to this approach is the notion that Semantic Web enabled computational agents will both provide and consume pieces of actual usage data that have been collected about a learning object in determining the usefulness of this learning object for some new purpose. This is then an evolutionary approach to metadata creation as compared to move traditional prescriptive 'one size fits all' approaches.

Keywords: learning object; Semantic Web; agent; negotiation; e-learning; metadata; standards; ecological.

Reference to this paper should be made as follows: Brooks, C. and McCalla, G. (2006) 'Towards flexible learning object metadata', *Int. J. Continuing Engineering Education and Lifelong Learning*, Vol. 16, Nos. 1/2, pp.50–63.

Biographical notes: Mr. Christopher Brooks received his MSc degree in Computer Science from the University of Saskatchewan in February 2005. Since 2003 he has been working as a research officer within the Laboratory for Advanced Research in Intelligent Educational Systems (ARIES). His current research is aimed at improving the accessibility and usability of learning environments both through the fine-grained modelling of learning object metadata, as well as through the visualisation of activities happening within the environment.

Dr. Gordon McCalla is a Professor in the Department of Computer Science at the University of Saskatchewan. He is also a member of the Laboratory for Advanced Research in Intelligent Educational systems (ARIES) where he carries out research with his students in intelligent instructional planning, evolutionary learning environments, data mining and learner modelling. He is an active member in the International Artificial Intelligence in Education Society (IAIED) and is a principal investigator in the pan-Canadian Learning Object Repository Research Network (LORNET).

1 Introduction

Perhaps the most widely used and accepted learning object specification is the Learning Object Metadata (LOM) standard (Learning Object Metadata Working Group of the IEEE Learning Technology Standards Committee, 2002). This standard identifies 76 different aspects by which a learning object can be annotated and is supported in some way by all major learning object repositories and e-learning platforms. One would think then that learning objects should be rich with a metadata mark-up, and that a content management system could use this data to dynamically assemble a course with little or no human intervention. This, however, is not the case – content management systems are increasingly static, with even relatively simple rule-based sequencing specifications seeing little to no adoption (Brooks et al., 2005).

The ability for a software system to dynamically assemble a course from learning objects is an important goal within the educational technology research community. Such an ability would not only enhance the reusability of learning objects, but it would also allow interesting pedagogical goals to be achieved including the individualisation of a course to a particular learner and the ability to respond to learner impasses and educational opportunities in real-time as they are encountered when the learner is interacting with a learning object. Ideally, these goals could be achieved without the need for much or any human intervention.

Such dynamic assembly is not supported by current e-learning standards: specifications are both too restrictive in the variety of metadata they capture and too lax in how they express the structure of such metadata. Moreover, many learning object repositories support only a few of the fields available and most do not support an external query format (e.g. Simon et al., 2004), which could be used by the computer agents to retrieve objects from the repository. The result is that nearly all learning object-based courses are created directly by instructional designers, who align content they have explicitly hand crafted for a given educational purpose. This purpose generally includes both an educational outcome (e.g. ‘understand relational operators in ECMAScript at a level such that the student can apply them to new situations’) as well as an educational instruction style (e.g. a particular language, background or learning style that a student is assumed to have). This makes dynamic delivery difficult, error prone and time consuming as an instructional designer must create many different versions of a course for the different kinds of purposes he/she hopes to achieve.

In this paper, we argue that a more flexible method of associating metadata with learning objects will help in realising the on-demand assembly of courses for different educational purposes. A larger set of ontologies sufficient for particular purposes should be used instead of a single highly constrained taxonomy of values like LOM or Dublin Core (Dublin Core Metadata Initiative, 2001). Further, the ontologies should be marked up in an unambiguous syntax such that they are able to be understood by the software agents. This syntax must take into account the kinds of data types that agents are able to manipulate and must appropriately codify metadata instances to conform to these. Finally, repository and content management software must be able to associate multiple metadata instances with a given learning object, and must allow the agents to pick and choose those instances that fit their needs.

To help achieve these goals, we suggest a new process-oriented view of metadata called the *ecological approach*. This approach focuses on the continual automatic collection of metadata based on end-user experiences with learning objects and having

computational agents reason over these experiences to provide for an adaptable environment. It thus synchronises well with the Semantic Web technologies, which have been used in our initial implementations.

In the remainder of this paper we further discuss issues with metadata standards (Section 2), then motivate and explain the ecological approach in more detail (Section 3). We then show how to use Semantic Web technologies to implement the ecological approach (Section 4) and conclude (in Section 5) with a look at some of the issues raised by this approach as well as a number of directions we are planning to explore in future research.

2 Issues with the metadata standards

With the goal of using learning object metadata in adaptive learning environments, in particular those environments in which software agents are responsible for carrying out the adaptation, we observe three dominant problems with current metadata standards. First and foremost is that metadata formats are typically created with the notion that some human will both be the producer and the consumer of the metadata and the learning object content itself. We believe that such heavy reliance on human intervention is costly and mitigates against real-time adaptivity to individual learner needs. Moreover, when annotating learning objects, humans often do not fill in all of the fields and even when they do, inter-rater reliability is often quite low. We have observed this in some of our other work, in which domain experts (tutorial assistants for a class) were asked to categorise learning objects created from four years of student work (Cooke et al., 2003). The lack of reliability between metadata authors appears to be a general trend regardless of whether the resource being described is a tutorial, web-based discussion thread or other digital artefact. In response, there has recently been some research in how to automatically create metadata directly from learning object content (Cardinaels et al., 2005; Hammouda and Kamel, 2005). In the experiments run by Cardinaels et al. (2005), only a small set of fields (less than 25%) could be data mined from learning object content and even then many of the automatically generated fields disagree with the values set by content experts.

In addition to semantic validity, metadata formats also limit the ability of software tools to accurately express their interactions with a learning object. For instance, while LOM contributors can have a variety of different roles (e.g. author, publisher, instructional designer, etc.), they are represented using the vCard syntax, which was “intended to be used for exchanging information about people and resources.” (Versit Consortium, 1996) This syntax accurately captures contact information for individuals, but fails to provide reasonable references for more automatic processes such as web services or software agents. While the focus on strict machine-readable formats (such as the vCard) is good for software agents, there is minimal validation of the data in actual deployed systems. In a study using five different cross-cultural learning object repositories, Friesen (2004) was unable to find a single conformant piece of vCard information in more than 3000 metadata instances. In addition to this, a number of the metadata fields in both the LOM and the Dublin Core contain open ended or otherwise vague vocabularies. Some of our previous work inspected the version control (referred to as the *lifecycle* element in the LOM) abilities of these two schemes and determined that

they generally lack structure and are unreliable for automatic processing (Brooks, 2005). These vocabularies hamper the abilities of both humans and software agents when trying to make sense of metadata records.

The second main problem with metadata standards is the inherent centralisation of metadata – both during the standardisation effort and in actual deployed e-learning systems. The process of standardisation typically involves gathering together a variety of stakeholders with the expressed purpose of coming to some consensus about what data should be included and how it should be represented. These stakeholders, especially in the area of e-learning, often come with a diverse set of goals they want to accomplish. The resulting standard tends to be either too narrowly focused (such as making version control human readable only, as described previously), or so comprehensive that implementing it is difficult or impossible. Friesen (2004) provides a compelling example of this in a study of 250 learning object metadata records (chosen from five different projects evenly) where only 36% of the elements were used more than half of the time, with many elements never used at all. Further, the elements used often referred to custom or local vocabularies (typically referred to as *application profiles*), a practice that effectively eliminates semantic interoperability. This notion of being overloaded with a broad standard is perhaps expressed best in the Learning Object Metadata Best Practices guide put out by the IMS Global Consortium, Inc., the group responsible for much of the early development of the LOM:

“Many vendors expressed little or no interest in developing products that were required to support a set of meta-data with over 80 elements...[and the] burden to support 80+ meta-data elements on the first iteration of a product is too great for most vendors to choose to bear”. (Anderson and Wason, 2003)

In addition to these issues with the process of metadata standards creation, most deployed software systems typically associate a single metadata instance with a learning object. This limits the ability of e-learning systems to aggregate many individual annotations about a learning object together to obtain a more comprehensive understanding about the nature of a learning object. It should be noted that this limitation is not inherent in the design of the LOM standard, but appears in nearly every deployed system that we have investigated. The sole exception to this seems to be the Standardised Contextualise Access to Metadata framework (SCAM) (Palmér et al., 2004), which acts as a simple learning object repository system.

Consider the case of a content management system trying to find a learning object for a learner that will take that learner no longer than ten minutes to complete. A metadata tag exists in the LOM for this (‘typical learning time’) and is available in a computer-readable format (based on the ISO8601:200 standard – if the metadata editor inputs the data correctly). While the content management system can then search for, obtain and deploy all of those learning objects that typically take less than ten minutes to complete, it is unable to enrich the metadata of the learning object with actual observations based on learners’ experiences since a suitable historical log of these experiences has not been kept. Further, the subtleties as to why the typical learning time for this learning object is ten minutes are lost – Is it an estimate based on the instructional designers’ experience? Is this an aggregation of observed times and, if so, are all of these learners comparable to the one who is currently considering using this learning object?

This leads to the third dominant issue with current metadata standards – the wrong kind of data is being collected. Up until now, the general focus of metadata has been on summarising the characteristics of the learning object content. We argue, that usage information is often as important as (or more important than) content in determining the relevance of the learning object in a particular context, for a particular learner or pedagogical goal. This means that a wide variety of situational metadata, such as characteristics of the learner’s background, characteristics of the task the learner is trying to complete and trace data and outcome information about the learner’s interaction with a learning object, should be collected. Once associated with a learning object directly, such metadata can be seen as appropriate to informing the *process of reasoning over a learning object to determine its suitability for a given purpose and learner*, rather than as a data structure of summarised content.

Key to this perspective is the notion that highly specialised and diverse schemes of data for various purposes and interactions need to be collected. Thus, instead of the current top-down approach where everyone subscribes to the same information about a learning object, we envision learning objects being associated with domain- and task-specific ontologies. When mixed with formal ontology specifications, such as those found within the Semantic Web community, this provides for a much greater expressive power and allows for the use of computational agents that would be impossible with content tags alone.

3 The ecological approach

The ecological approach was first proposed by McCalla (2004) in the context of providing a more flexible and powerful way to mark up and use learning objects to achieve various pedagogical goals. In the ecological approach the e-learning system keeps a learner model for each learner, tracking characteristics of the learner and information about the learner’s interactions with the learning objects they encounter. After a learner has finished with a learning object, the learning object is associated with an instance of the learner model that records the learner’s state(s) and activities while interacting with the learning object. The information in such a learner model instance can include:

- information about the learner, including cognitive, affective and social characteristics and the learner’s goal(s) in accessing the learning object
- information about the learner’s perspectives on the learning object itself, including the learner’s feedback on the content, or the learner’s knowledge of the content (as determined, for example, by a test administered during the learner’s interactions with the learning object)
- information about how the learner interacted with the learning object, including observed metrics such as dwell time, the number of learner keystrokes, patterns of access, etc.
- information about the technical context of use, including characteristics of the learner’s software and hardware environment
- information about the social context of use, including links to the learner model instances attached to learning objects previously encountered by the learner.

Over time, each learning object thus slowly accumulates learner model instances that collectively form a record of the experiences of all sorts of learners as they have interacted with the learning object. The collected learner model instances can then be inspected for patterns about how learners interacted with the learning object, for example, that learners whose knowledge has been evaluated as weak did not have long dwell times or that learners with certain cognitive characteristics did well. The sequence of learner model instances for a particular learner forms a ‘learning trail’ through the learning object repository and this trail can also reveal interesting patterns of success and failure for the learner.

There are an enormous number of patterns that can be found when inspecting actual learner behaviour. The key to find meaningful patterns is the *purpose* (in the sense of Vassileva et al. (2002)) for which the patterns are sought. Each such purpose places its own particular constraints on what patterns are meaningful, how to look for these patterns and how to use what these patterns reveal to achieve the purpose. Thus, determining whether to recommend a specific learning object to a particular learner may require comparing this learner to other learners on important characteristics and then looking at how similar learners have evaluated (or been evaluated by) the learning object (and moreover the characteristics considered to be important are themselves determined by the learner’s own goals). On the other hand, determining whether a learning object is now obsolete may require an examination of all learners’ evaluations of the learning object, trying to extract temporal patterns in the evaluations that show how recent learners like or dislike the learning object. The key point is that it is the purpose that determines what information to use and how it is to be used. An ideal goal for a real-time e-learning system is that this determination be made *actively* (in the sense of McCalla et al. (2000)) at the time the purpose is invoked, so that no *a priori* interpretation needs to be given to the information; however, time constraints on executing the data mining algorithms may mitigate against such real-time computation in many circumstances.

In sum, then, the ecological approach promotes the notion that information gradually accumulates about learning objects, the information is about the use of the learning object by real learners and this information is interpreted only in the context of end use. The approach is ecological because over time the system is populated with more and more information and algorithms emulating natural selection (based on purposes) that can determine what information is useful and what is not.

There are many possible pedagogical goals supported by the ecological approach to e-learning. This approach could underline the design of:

- a study aid, for example, to retrieve for a learner relevant papers from a cache of such papers for a graduate student trying to learn about an area of research (e.g. Tang and McCalla, 2003)
- a recommender system, to recommend some content to a learner that is relevant to his her current task (e.g. Recker and Wiley, 2001)
- an instructional planner, to plan out a sequence of content pages of relevance to a learner, sort of an individualised curriculum of study
- a group formation tool, to suggest to the learner a group of other learners relevant to solving a particular task or learning about a particular subject (e.g. Winter and McCalla, 2003a,b)

- a help seeker, to find another learner who can help the learner solve a problem he/she has encountered (e.g. the I-Help system (Greer et al., 1998))
- a reminder system, to keep a learner updated with new information that is relevant to the learner's goals
- an evaluation tool, to allow learners' interactions with educational content to be studied by instructional and cognitive scientists, in particular to look at the experiences of all learners or particular types of learners with some educational content
- an end-use tagging system, to automatically derive educational content tags from pre-established ontologies based on the experiences of the actual users of the content and that can be parameterised by end-use variables such as the type of learner, success/failure of the educational content for each type of learner, etc. A variant of this possibility is the ability to refine, modify or change pre-assigned metadata based on inferences from end use
- an 'intelligent' garbage collection system to determine the ongoing relevance of educational content and, if necessary, to suggest modifications or even that it be deleted as no longer being useful to learners (e.g. as discussed by Bannan-Ritland et al. (2000)).

The ecological approach, through its focus on learners and on learners' end use of learning objects, through its ability to capture metadata either automatically or at least more naturally than through manual attachment of standardised metadata and through its support for many different kinds of end applications, has motivated many of the approaches we are currently exploring in our research group. However, there is still much work to do to fully flesh out the approach, to prove that the approach is tractable and to demonstrate effectiveness in actual application systems. In the next section, we discuss one direction we are exploring: the integration of the ecological approach and the Semantic Web.

4 A Semantic Web approach to supporting the ecological approach

As described more fully by Brooks et al. (2005), the Department of Computer Science at the University of Saskatchewan has developed a set of e-learning applications,¹ which includes both a discussion forum system (asynchronous and synchronous) and a learning object-based content management system. Each of these systems collects significant amounts of data from the interactions of learners with their learning environment. These data are then transferred to applications (or learners) through a Semantic Web-based middleware system known as the Massive User Modelling System (MUMS) (Brooks et al., 2004). In this system, learner interaction information is packaged up using the Resource Description Framework (RDF), and corresponds to one or more RDF schemas.² These packages, known as *events*, are then forwarded from e-learning applications to higher-level applications, in particular software agents, where they can be analysed and acted upon.

Our initial work in applying the ecological approach to learning object metadata is based primarily around our content management system, iHelp Courses. This system

delivers standard IMS Content Package (Smythe et al., 2004) formatted learning objects to learners and typically includes text, video and interactive exercises. The contents of these packages are very similar to the CISCO learning object model (Barritt and Lewis, 2002) and learners must complete a short quiz both before and after the learning object is delivered. These quizzes contain multiple choice questions, which are related to the content that is being taught. Each question/answer pair in a quiz is mapped to a particular domain concept expressed in our domain ontology, as well as an entry in an educational objectives ontology. Our domain ontology is a large (more than 1000 nodes) RDF graph that represents the relationships between concepts covering basic computer science for non-majors, focusing on web technologies (HTML, ECMAScript, etc.) and the history of computer science. Our educational objectives ontology is based on the work done by Anderson and Krathwohl (2001), which itself is built off of work done by Bloom and Krathwohl (1956). This ontology allows us to represent both the cognitive process and the level of knowledge that a student obtained in a given topic.

Consider an example in Figure 1 taken from a lesson on operators. If the learner answers true, it shows they can understand procedural knowledge of both the topics 'Realtional Operators' and 'Logical Operators'. If the learner answers false, they have not demonstrated any knowledge or ability in particular. Using the case of the former as an example, the results can then be expressed in RDF (shown graphically in Figure 2). It is worth noting that the content management system itself knows only about the learner and the question/answers he/she has submitted (the left-hand side of the figure) – Semantic Web rules can be used to add values to the RDF with a derived understanding of the competencies a learner has gained after the fact.

Figure 1 Sample question asked to students in iHelp Courses

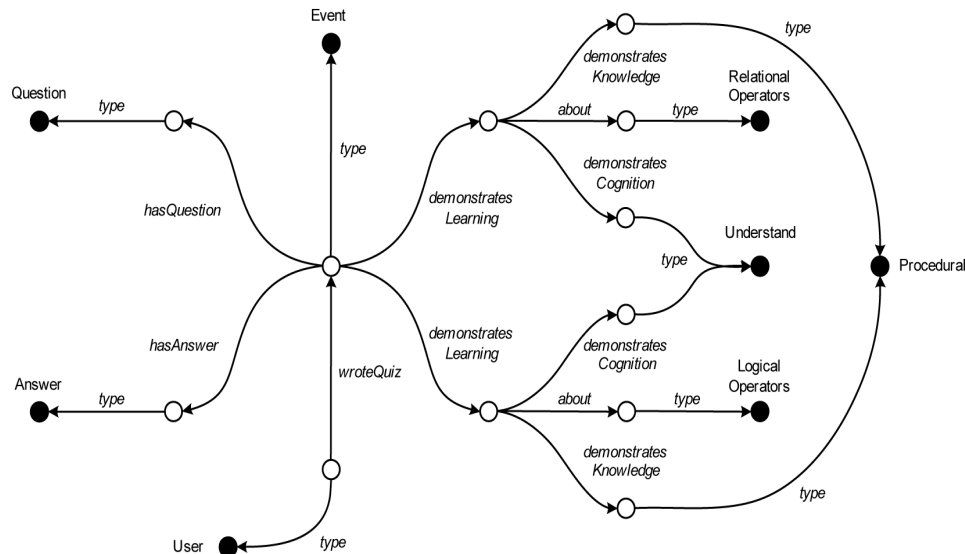
<p><i>Question:</i> What is the result of the operation $((2 < 9) \ \&\& \ (3 > 2))$?</p> <p>i) true</p> <p>ii) false</p>

This form of collecting learner competencies is relatively agile – there is no need to indicate 'correct' and 'incorrect' answers; instead, the instructional designer can indicate explicitly which answers demonstrate which competencies. This is especially useful in multiple choice tests where there is often one best answer, but several other options still demonstrate some smaller set of this knowledge. Further, the open world model of RDF allows for an arbitrary number and type of statements to be made about a learner's interaction with a quiz. This is useful in making multiple assertions about learner knowledge (as in the above).

This method has at least two advantages compared to the traditional metadata approach. Firstly, the RDF data model differentiates itself from the LOM data model in that it is explicitly based upon an open world assumption – the idea that the metadata you have does not necessarily represent the total set of metadata that exists for this learning object (both in terms of instance data as well as ontological definitions for that data). While the LOM specification does not address this issue specifically, the majority of

implemented systems take a closed-world approach instead. The open-world approach described above can be easily extended to include misconceptions (a form of bug libraries) where question/answer pairs can be associated with typical wrong answers and suggestions about the conceptual issues a student may be having because of this.

Figure 2 Graphical representation of RDF model. Empty circles are instances while filled in circles are classes. Namespace prefixes and instance values have been omitted to aid in readability



Secondly, RDF allows for more complex statements to be made, where metadata can be further annotated with more metadata. Take again the instance data described in Figure 1. These data could be refined to indicate that the quiz question was first seen on some given date and that the learner saw the question on the screen for 5 seconds before choosing to answer. The LOM is incapable of expressing this kind of data, as it only supports extension through taxonomical *classifications*, which are learning object centred – each new classification applies to the whole of the learning object and not to a specific piece of metadata that represents the learning object.

The assessment data, once collected, can be attached to the learning object, as suggested in the ecological approach. Instead of just associating the raw data with the object, the data can be summarised by subtracting all of those competencies demonstrated in the pre-test from those demonstrated in the post-test. This then shows the net gain of knowledge the learner has achieved by interacting with the learning object.

During the 2004–2005 regular academic sessions, we began implementing this approach and collected the assessment data from approximately 50 learners using our online course and feedback from the instructor. During this time it became apparent that learners were frustrated with the pre-tests as they often lacked sufficient knowledge to understand the question being asked. We are addressing this by changing the pre-test from actual assessment to a declaration of self knowledge, where learners indicate what they feel is their level of knowledge in each topic using a Likert scale. The values on this scale are then mapped to approximate entries in the educational outcome taxonomy being

used. There is a delicate balance between asking too many questions, risking that some of them may be off topic, and not asking enough, thus missing potential useful metadata entries. This is a trade off we are trying to mitigate by working closely with the instructor and instructional designers for the course.

We are also actively exploring the augmentation of this self-assessment data with information extracted from usage data gathered during learners' interactions with the online course (e.g. which learning objects were read, how long they were read, what order they were read in, etc.) and with synchronous and asynchronous discussion forums associated with our iHelp peer help system (e.g. postings read, time dwelt on a posting, chats participated in and general availability online). The MUMS tool, discussed above, is crucial to this effort, since it allows ready customisation of data mining algorithms to analyse these data and naturally facilitates the integration of information discovered by these algorithms. We are also investigating the use of probabilistic models, such as those presented by Zapata-Rivera and Greer (2004), to help us to combine this disparate metadata and draw sound inferences from it.

These techniques allow for the collection of a wide variety of usage-based and user-based metadata about a given learning object, but do not predict how or even if that learning object is going to be useful. Since our aim is to reason over the metadata to adapt to a given learner or to support a particular pedagogical goal, strong inferencing capabilities are needed. Working specifications and toolkits for the Semantic Web (e.g. Hewlett-Packard Development Company, 2004; Horrocks et al., 2004) show this to be a promising area, more so than creating rules for the analysis of the LOM. We are devising special purpose-based algorithms to achieve goals such as recommending learning objects for particular learners (as in the research paper recommender system discussed by Tang and McCalla (2003)) and are also working on the more general problem of devising full instructional plans, with sequences of learning objects customised for a particular learner. We have also explored how to create entire taxonomies of such purpose-based algorithms (in an e-commerce recommender system, Niu et al. (2005)), demonstrating how notions of 'anytime' computation and procedure inheritance can enhance the power, the robustness and the scope of these purpose-based algorithms and the knowledge engineering effort required to create them. Finally, we are currently redesigning and augmenting our iHelp suite to take advantage of the wide range of data we can mine to achieve such goals as finding appropriate peer helpers or appropriate postings on the asynchronous discussion forum (as described more fully by Winter et al. (2005)).

5 Conclusions

Our research is beginning to show us how to merge Semantic Web techniques with the ecological approach to allow the capture of at least some aspects of learners' end use of learning objects and then use this information to inform a variety of e-learning applications. In this way we hope to overcome the rigidities of standard approaches that require humans to explicitly attach metadata to learning objects. We believe that in any reasonably complex environment universal agreement on the semantics of the learning objects will effectively be impossible, as will anticipating the metadata required for all purposes for which the metadata might be used. Instead of the standard 'prescriptive' view, we propose a 'descriptive' view that takes advantage of the huge amount of

end-use data and information about learners that will typically be available in an e-learning context. We also take an explicitly procedural perspective, emphasising the creation of processes that can use this information adaptively and appropriately to the purpose and end-use context.

The move to a process-based view of metadata is not without difficulties. If assertions about a learning object increase over time as more learners interact with that learning object, then the time required to make sense of these assertions for a particular purpose will also increase. Further, using Semantic Web technologies to capture these assertions essentially makes metadata machine readable only. Thus, we see a shift of the consumers of learning object metadata to computational agents acting on the behalf of instructional designers, instructors, students and content management systems instead of those entities directly. Similarly, the production of metadata also shifts away from human producers to more automated methods. We see a deeper integration of metadata creation within learning content management systems in particular as a promising method of generating experience-based annotations for learning objects. While standards for capturing the interaction between a learning object and its user are currently quite limited, we believe that the Shareable Content Object Reference Model (SCORM) run-time environment specification (Advanced Distributed Learning, 2004) is a start in the right direction. This specification allows for explicit tracking of the user within a content management system. While there is as yet no standard way of saving or sharing these tracking models, we see this as a natural extension to the specification.

While we believe that the benefits of having many specific local ontologies in contrast to a single global taxonomy like the LOM are numerous, there are some detriments. In particular, the number of available elements now increases from 76 to hundreds or thousands as more specific ontologies are developed and more end-use-oriented metadata is collected or inferred. The key to dealing with this explosion of elements is that in our approach these elements are only suitable for certain purposes – that is, vendors and applications will still only support those pieces of metadata that fit with their ability to capture data as well as meet their needs and goals. The rest can be ignored as any given purpose is fulfilled. We see this as being similar to the principles of a free market – those ontologies and metadata that are useful to publishers and consumers (human or software) will continue to be used and gain in popularity and acceptance, while those ontologies and metadata that are not useful will die out and be replaced by those that are. As with free market systems a number of issues exist: barriers to entry (how to get your ontology or metadata marketed and used by others), duplication of resources in competition, etc. Identifying and mitigating these issues are of interest to us in our future research.

Finally, we are concerned about the implications of our approach for the interoperability of e-learning artefacts. This challenge is the primary reason globally applicable standards such as the IEEE LOM exist. The problem is that once deviation is allowed from a standard (as in our approach), interoperability begins to get severely hampered, both within a given learning object repository and between repositories. To combat this, we anticipate that agents native to our metadata repositories will be needed to reason over and potentially summarise metadata to convert it to a more standardised form for export. Unfortunately, this is a lossy process, in that in many situations the summarised data cannot contain multiple records and the new semantics we would like to introduce (through deeper learner modelling) are meaningless to external

repositories. We are thus also investigating the notion of negotiated understanding between agents who contain some intersection of understanding – essentially an automation of ontology mapping done without the interaction of a human author. With this, external repositories may be able to learn (potentially over time when presented with examples), the appropriate usage for classes of learning objects and be able to enrich their ability to answer queries for a material more effectively. Standardisation efforts could then be aimed at creating clear semantics for negotiation, ontology expression (e.g. RDF/RDFS/OWL) and mapping protocols instead of just on explicitly defining a set of metadata fields.

Acknowledgements

This paper represents ongoing work within our laboratory and is an elaboration and extension of ideas that were originally published by Brooks and McCalla (2005) at the International Workshop on Applications of Semantic Web technologies for E-Learning. This project was funded in part by the Government of Canada through the LORNET NSERC research network.

References

- Advanced Distributed Learning (2004) 'Shareable Content Object Reference Model (SCORM) Content Aggregation Model (CAM)', Version 1.3.1.
- Anderson, L. and Krathwohl, D. (2001) *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, Addison Wesley Longman, Inc.
- Anderson, T. and Wason, T. (2003) 'IMS Learning Resource Meta-data Best Practices and Implementation Guide', Version 1.1.
- Bannan-Ritland, B., Dabbagh, N. and Murphy, K. (2000) 'Learning object systems as constructivist learning environments: related assumptions, theories and applications', in D. Wiley (Ed). *The Instructional Use of Learning Objects (On line Version)*, AIT/AECT.
- Barritt, C. and Lewis, D. (2002) 'Reusable learning object strategy', Available at: http://www.cisco.com/warp/public/10/wwtraining/elearning/implement/rlo_strategy_v3-1.pdf.
- Bloom, B.S. and Krathwohl, D.R. (1956) *Taxonomy of Educational Objectives, the Classification of Educational Goals, Handbook I, Cognitive Domain*, David McKay Company Inc.
- Brooks, C. (2005) 'Versioning of learning objects', MSc Thesis, University of Saskatchewan, Saskatoon, SK, Canada.
- Brooks, C., Kettel, L. and Hansen, C. (2005) 'Building a learning object content management system', *World Conference on E-Learning in Corporate, Government, Healthcare and Higher Education (E-Learn 2005)*, Association for the Advancement of Computing in Education (AACE).
- Brooks, C. and McCalla, G. (2005) 'Flexible learning object metadata', *International Workshop on Applications of Semantic Web Technologies for E-Learning (SW-EL '05)*, Available at: <http://www.win.tue.nl/SW-EL/2005/swel05-aied05/proceedings/2-Brooks-final-full.pdf>.
- Brooks, C., Winter, M., Grees, J. and McCalla, G. (2004) 'The massive user modelling system', *Seventh International Conference on Intelligent Tutoring Systems (ITS04)*, Springer-Verlag, pp.635–645.

- Cardinaels, K., Meire, M. and Duval, E. (2005) 'Automating metadata generation: the simple indexing interface', *The 14th International World Wide Web Conference 2005 (WWW 2005)*, *International World Wide Web Conference Committee (IW3C2)*.
- Cooke, J., et al. (2003) 'Computer science tutorial page', Available at: <http://www.cs.usask.ca/resources/tutorials/csconcepts/index.html>.
- Dublin Core Metadata Initiative (2001) 'Dublin Core Metadata Initiative (DCMI)', Available at: <http://dublincore.org/>.
- Friesen, N. (2004) 'Final Report on the "International LOM Survey"', *Technical Report Document 36C087*, Canadian Advisory Committee for ISO/IEC JTC1/SC36.
- Greer, J., McCalla, G., Cooke, J., Collins, J., Kumar, V., Bishop, A. and Vassileva, J. (1998) 'The intelligent helpdesk: supporting peer-help in a university course', *Fourth International Conference on Intelligent Tutoring Systems (ITS 1998)*, Springer-Verlag, pp.494–503.
- Hammouda, K. and Kamel, M. (2005) 'Data mining in e-learning', in S. Pierre (Ed). *E-Learning Networked Environments and Architectures: A Knowledge Processing Perspective*, Springer.
- Hewlett-Packard Development Company (2004) 'Jena 2 – A Semantic Web Framework', Available at: <http://www.hpl.hp.com/semweb/jena.htm>.
- Horrocks, I., Patel-Schneider, P., Boley, H., Tabet, S., Grosz, B., Dean, M. (2004) 'SWRL: a Semantic Web rule language combining OWL and ruleML', Version 0.7.
- Learning Object Metadata Working Group of the IEEE Learning Technology Standards Committee IEEE, Inc. (2002) *IEEE P1484.12.1-2002*, Draft Standard for Learning Object Metadata.
- McCalla, G. (2004) 'The ecological approach to the design of e-learning environments: purpose-based capture and use of information about learners', T. Anderson and D. Whitlock (Guest Eds). In *Journal of Interactive Media in Education (JIME)*, Special Issue on The Educational Semantic Web, Vol. 1, p.18, Available at: <http://www-jime.open.ac.uk/2004/1>.
- McCalla, G., Vassileva, J., Greer, J. and Bull, S. (2000) 'Active learner modelling', *Intelligent Tutoring Systems 2000 (ITS2000)*, Springer Verlag, pp.53–62.
- Niu, X., McCalla, G. and Vassileva, J. (2005) 'Purpose-based user modelling in a multi-agent portfolio management system', *Proceedings of the Workshop on Decentralized, Agent-Based and Social Approaches to User Modeling (DASUM-05)*, held in conjunction with the Tenth International Conference on User Modeling, Edinburgh, pp.66–69.
- Palmér, M., Naeve, A. and Paulsson, F. (2004) 'The SCAM framework: helping Semantic Web applications to store and access metadata', *European Semantic Web Symposium (ESWS 2004)*.
- Recker, M. and Wiley, D. (2001) 'A non-authoritative educational metadata ontology for filtering and recommending learning objects', *Interactive Learning Environments Journal*, Special Issue on Metadata, pp.1–17.
- Simon, B., Massart, D., van Assche, F., Ternier, S., Duval, E., Brantner, S., Olmedilla, D., Miklos, Z. (2004) 'Simple Query Interface (SQI) for Learning Repositories', Version 0.8.
- Smythe, C. and Jackl, A. (2004) 'IMS Content Packaging Information Model', Version 1.1.4.
- Tang, T. and McCalla, G. (2003) 'Smart recommendation for an evolving e-learning system', *Workshop on Technologies for Electronic Documents for Supporting Learning, International Conference on Artificial Intelligence in Education (AIED 2003)*.
- Vassileva, J., McCalla, G. and Greer, J. (2002) 'Multi-agent multi-user modeling in I-help', *User Modeling and User-Adapted Interaction*, Special Issue on User Modelling and Intelligent Agents, Vol. 13, No. 1, pp.1–31.
- Winter, M. and McCalla, G. (2003a) 'An analysis of group performance in terms of the functional knowledge and teamwork skills of group members', *Workshop on User and Group Models for Web-based Collaborative Environments, Ninth International Conference on User Modeling (UM 2003)*, pp.35–45.

- Winter, M. and McCalla, G. (2003b) 'An analysis of group performance in terms of the functional knowledge and teamwork skills of group members', *Proceedings of the International Conference on Artificial Intelligence in Education (AIED 2003)*, Sydney, Australia, July, pp.261–268.
- Winter, M., Daniel, B. and Brooks, C. (2005) 'Towards automatic discovery of peer helpers from a large message board system', *Workshop on Usage Analysis in Learning Systems, the 12th International Conference on Artificial Intelligence in Education (AIED 2005)*.
- Zapata-Rivera, J-D. and Greer, J. (2004) 'Inspectable Bayesian student modelling servers in multi-agent tutoring systems', *International Journal of Human-Computer Studies*, Vol. 61, No. 4, pp.535–563.

Notes

¹More information on these e-learning applications can be found at <http://ihelp.usask.ca>.

²Examples of these schemas are available at: http://ai.usask.ca/mums/best_practices.