On the Importance of Metadata and Learner Interaction Data in SCORM 2.0

Sarah Riley, Leen-Kiat Soh, and Ashok Samal

Computer Science and Engineering Department

University of Nebraska-Lincoln

256 Avery Hall, Lincoln, NE 68588-0115 USA

{sriley, lksoh, samal}@cse.unl.edu

TEL: 402-472-6738   FAX: 402-472-7767

Abstract

This paper examines three key data and metadata components which are critical for providing interoperable adaptive functionality in e-Learning systems: **learning object metadata, learner information profiles, and learner interaction data** (log files). We give an example of one possible future e-Learning application, the learning object repository 2.0 (LOR 2.0), which would require these types of strict metadata and data standards. We conclude with recommendations for SCORM 2.0 which we feel would help bring recent research progress into mainstream e-Learning.

1. Introduction

Examination of the goals stated in previous versions of the Sharable Content Object Reference Model (SCORM) standard [11, 12], makes it clear that the ultimate goal of the SCORM standard is not simply to provide a format for interoperable learning objects (LOs), but to facilitate **adaptive, real time delivery of learning content based on individual learning contexts**.

"In contrast to classroom learning, these approaches enable the pace, sequence, content and method of instruction to better fit each student's learning style, objectives, and goals." [11]

"once sharable learning objects exist and are commonly available, they can be assembled in real time, on demand and then delivered to learners as needed."[11]

"The vision of the ADL {Advanced Distributed Learning} Initiative is to provide access to the highest quality learning and performance aiding, that can be tailored to individual needs, delivered cost-effectively anytime and anywhere." [12]

In order for the SCORM community to meet these goals, we must develop SCORM 2.0 such that it will facilitate the current research and development of these types of technologies:

- Automated LO metadata generation
- Automated categorization of LOs

- Empirical evaluation of LOs
- Algorithms to match students and LOs
- Algorithms to diagnose learner behavior patterns

Researchers who are working on these problems have made exciting progress in recent years. Examples include a student motivation diagnoser which achieves significant results after analysis of only 10 minutes of navigation and assessment log file data[4], learning object metadata generators such as in [1] which can extract partial metadata from a learning object automatically, teacher advice generators which create models of individuals, groups, and classes, automatically detect potential problems, and generate feedback for the instructor to pass on to students [3], and many more [2,5,6,7,8]. Unfortunately, much of this research is not portable to mainstream e-Learning applications because most of the tools and methods developed require some or all of the following elements, in machine-readable and machine-processable form:

1. ***Learning object metadata***
2. ***Learner information profiles***
3. Detailed ***learner interaction data*** (log files)

While SCORM 2004 3rd edition [12] has begun to address many of these concerns, most notably by the addition of the user interaction data model in the Run Time Environment (RTE) [14], all three of these areas need to be extended or revised to allow for the development of the tools and methods which will bring the goals of SCORM out of the research arena and into mainstream education and training.

2. Limitations of the current SCORM standard

This section will attempt to quickly cover some of the challenges researchers face when attempting to develop advanced e-learning applications while conforming to the current SCORM standards. The main areas we will focus on are: 1) the ***limitations of the current learning object metadata*** standards, 2) the ***absence of a learner information profile***, and 3) the ***limitations of the current learner interaction tracking*** capabilities.

2.1 Limitations of IEEE Learning Object Metadata (LOM)

The Institute of Electrical and Electronics Engineers, Inc. (IEEE) 1484.12.1 – 2002 Standard for Learning Object Metadata (LOM) [10] is currently a component of the SCORM 2004 Content Aggregation Model (CAM) [13]. The LOM is intended to facilitate the interoperable categorization and discovery of a learning object (LO) in a learning object repository (LOR). However, a recent study which was summarized in [9] on the LOM showed that the actual usage of the standard falls short of facilitating this goal.

Many of the metadata fields in the LOM standard are only loosely defined and many learning object authoring tools do not provide a simple interface for populating the large metadata set in the LOM. Additionally, the LOM can be extended to include fields and vocabularies which are not part of the standard. Unfortunately, these problems compound and lead to ***incomplete, inaccurate metadata, which may or may not contain extensions***. These limitations make

automatic classification and discovery of arbitrary learning objects by use of the LOM nearly, if not entirely, impossible.

2.2 The need for a SCORM 2.0 learner information standard

There are several existing learner information standards which might serve the learner information profile needs of SCORM 2.0, including IMS Learner Information Profile (LIP) [15], IMS ePortfolio [16],and IEEE Public and Private Information for Learners (PAPI). The main arguments for the inclusion of a learner information profile in SCORM 2.0 are:

- These learner information profiles can *migrate with a learner* from one LMS to another.
- It would *encourage the development of interoperable applications and plug-ins* which leverage learner information to provide adaptive capabilities

Although the inclusion of a learner information profile would remove some flexibility from SCORM-conformant LMS systems, the benefits would be substantial. Not only would it bring us one step closer to the potential for lifetime learner profiles, it would also encourage the development of applications and plug-ins which would be also be interoperable with any SCORM-conformant learning environment.

Suggestion of an appropriate learner information standard for SCORM 2.0 is outside the scope of this paper. However, we will make some recommendations as to the *types of information categories* which should be included and the *general properties* which would be desirable for a SCORM 2.0 learner information standard.

At this point in time, there appears to be little consensus on what categories should be included in a student information standard. This has led to the inclusion of a large number of categories, some of which are potentially ambiguous and/or redundant. In fact, the current standards appear to suffer from many of the same problems as the LOM: ambiguity of categories and fields, a large data model, and the ability to create extensions again leads to extremely low levels of real life interoperability. Whichever learner information standard is incorporated into SCORM 2.0 should attempt to cover at least the following information categories:

- *Identification Information*, probably including a Universal Identifier of some sort.
- *Demographic Information,* such as gender, ethnicity, nationality, age, etc.
- *Preferences*, such as learning styles, interests, etc.
- *Competencies*, which should include a history of learning objects used, including the universal identifiers, a time stamp, and the pass/fail status. There may be additional measures which could be included in this category.
- *Goals*, which would outline the learning or training goals for the student along with time frames for these goals. These may be set either by the student or by an instructor.

There remains a lot of work to be done in the Competencies and Goals categories. Progress is currently being made in the areas of Ontological Engineering and the Semantic Web which may lead to a more comprehensive solution to these categories. For the time being, it may be

preferable to implement only those components which can be unambiguously defined for the sake of interoperability.

Additionally, whichever standard is ultimately chosen should have these general properties:

- It should be machine-readable and machine-processable.
- It should include an XML binding for the import and export of both full and partial student profiles.
- It should allow for the merging of partial records into a comprehensive record.
- It should be designed with privacy concerns in mind.
- The implementation of a database for storage, import, and export of these student models should be mandatory for SCORM 2.0-conformant LMSs.

There are many benefits to the incorporation of a learner information standard in SCORM 2.0. Most notably, it will ***allow for the development of interoperability along another dimension: the creation of context-aware applications which could be plugged into any SCORM 2.0-conformant e-Learning system***. The possibilities are nearly endless, but might include adaptive learning object sequencers, small group formation tools, and student motivation or disengagement detectors.

2.3 The importance of student interaction tracking mechanisms

Learner interaction tracking capabilities are arguably the most critical aspect of SCORM 2.0 in terms of facilitating adaptation, because they constitute a log of the student's actual interactions with the learning object and give us a personal glimpse into each student's learning experience. These interactions can be utilized by a large variety of algorithms and applications of interest, such as:

- A *learning style diagnoser* which analyzes a student's interactions with several learning objects to diagnose the student's problems or difficulty in learning.
- A *student and group modeler* that can be used to reduce the complexity of adaptive, just-in-time content discovery or provide class information to distance learning instructors, through identifying and characterizing student learning styles or preferences.
- A *learning material analyzer* which mines all student log files for a particular learning object in order to identify the elements of a learning object which may need improvement, or how the learning object relates to other learning objects for sequencing purposes.

At a minimum, SCORM 2.0 should facilitate tracking:

- ***Detailed assessment interactions***: We should be able to track the complete time-stamped history of student assessment interactions. For example, if a student navigates to an assessment page, clicks true, then false, then clicks the help button, then true again, we should track each of these interactions with time-stamps.

- *Navigation Interactions*: We should be able to reconstruct the student's navigation through the various pages of the learning object by keeping time-stamped logs of this behavior.

While the SCORM 2004 RTE interactions data model are a move in the right direction, here are additional points to consider:

a. An interaction log file should be a log of an individual student's session with an individual learning object. Thus, if a student interacts with a learning object on more than one occasion, there should be a *log file for each session*.
b. The *LMS implementation requirements* for the mechanism for storing, importing, and exporting student log files should be part of the SCORM 2.0 standard.
c. Rather than requiring each interaction to be logged by an individual call to an API function, the *API should be extended* to include high level function calls which would turn on a subset of logging behavior for the entire learning object. This should greatly simplify the implementation of tracking functionality. For example, a single function call might turn on navigation tracking or detailed assessment tracking.

3   An e-Learning Dream Scenario: Learning Object Repository 2.0

One potential e-Learning application which might emerge from the research arena into mainstream e-Learning in the near future is the learning object web crawler, or Learning Object Repository (LOR) 2.0.

The Learning Object Repository of the future will operate like a web crawler. The automated discovery of learning content, a common discovery portal, and the ability to rank learning material based on the *actual content* and the *actual usage* will cause this LOR 2.0 to eventually replace the current learning object repository model. This LOR 2.0 will have a human search portal as well as a program interface which will allow for the development of tools such as a personal learning application which will automatically search and retrieve learning objects from LOR 2.0 on behalf of its learner.

3.1 A lesson from the not-so-distant past…

In the early days of the World Wide Web, new web pages were manually listed under categories within a simple taxonomy and metadata on the topics in the pages were added using the HTML META tag by the content authors. These metadata and the classifications were generally trusted by search engines, and used to deliver relevant content to the end user.

The manual classification of web pages was error prone which certainly was not ideal, but a more insidious problem soon arose: once people began to see the web was a good source of revenue, web page authors began to use false keywords in the META tags to increase their web search hits.

New search engines entered the scene, now crawling for new web pages using spiders and using the actual content of the web page along with empirical usage data to list and rank the search results.

3.2 The not-so-distant future

The LOR 2.0 will use web spiders to search for both SCORM conformant learning objects and empirical usage metadata instances. The LOR 2.0 will use state of the art text classification algorithms to automatically categorize learning objects according to topic based on the actual content of the learning object.

It will utilize semantic web technologies to manage the various versions of a learning object and ontological engineering to situate the learning object within a knowledge domain. Empirical Usage Metadata instances (essentially, user interaction log files) will be associated with the learning object, and automated reasoners within the LOR 2.0 will use this to mine learning context rules from the various sources of metadata. This will allow for fast and seamless access to the learning object records which most closely match not only the desired topic, but also the *desired learner context*.

The LOR 2.0 will have both a basic keyword search interface and an advanced search interface which in additional to allowing you to select a learning topic will allow you to enter learner parameters such as age, gender, nationality, learning styles preferences, etc. in order to obtain search results such as the one in Figure 1. Alternatively, programs such as the proposed personal learning application can *upload a student model to automatically retrieve appropriate learning objects to provide for seamless adaptive just-in-time content delivery.*

Elementary Sorting Algorithms
This learning object introduces and compares the basic quadratic sorting algorithms. It covers bubble sort, insertion sort, and selection sort and includes interactive…read reviews
Search Relevance: 92% more details
Ontology: ComputingCurricula2001
Topic: AlgorithmsAndComplexity.FundamentalComputingAlgorithms.
        QuadraticSortingAlgorithms

Figure 1: A potential human-readable search result from LOR 2.0

If a person were to click on the 'more details' link from the above record, they would have the ability to review a large collection of empirical and qualitative statistics on the learning object. Some statistics which might prove useful are: the total recorded students who have used the learning object, the average age of the student who has used the learning object, the percentage of females / males who have used the learning object successfully, the minimum, maximum, and average time to successfully complete the learning object, the reading level of the tutorial text (vocabulary difficulty), and so forth. These metadata details will constitute a significant improvement over the current search metadata because it will be derived solely from the empirical usage evidence rather than opinion.

3. 3 LOR 2.0 Concluding Remarks

LOR 2.0 will most likely become the future LOR model simply because the current learning object repository model, is unsustainable and un-scalable. In addition, as the learning object economy begins to flourish, we may find that manually entered metadata instances become just as suspect as the HTML META tag. The future of the learning object economy depends on the ability to automatically categorize learning objects according to their content and usage history with actual students. SCORM 2.0 should implement learning object, student information, and interaction tracking data in order to establish a foothold in upcoming e-Learning technologies such as LOR 2.0.

4   Summary of SCORM 2.0 Recommendations

Researchers in Artificial Intelligence in Education, Intelligent Tutoring Systems, User Modeling, and Information Retrieval are making exciting progress in developing algorithms and applications which facilitate adaptive functionality in e-Learning systems. However, most of these researchers have not chosen to work within the SCORM standard, likely due to the difficulty in obtaining the needed student interaction data. In order to meet the original goals for SCORM by providing adaptive, just-in-time delivery of interoperable learning content, we should a close look at three key types of data which are often used to provide adaptive functionality in e-Learning systems:

1. ***Learning Object Metadata:*** Learning object metadata must be machine readable, and extensions should be avoided to maintain as much interoperability as possible.
2. ***Learner Information Profiles:*** SCORM 2.0 should include a standard learner information profile, such as IMS LIP or IEEE PAPI to allow for the creation of SCORM conformant algorithms and plug-ins for adaptability.
3. ***Learner Interaction Data:*** SCORM 2.0 should provide the capability to collect a log file of detailed assessment interactions and navigation behavior (with time stamps) for each student's session with a learning object. Additionally,
    a. SCORM 2.0 should include a LMS implementation standard for the storage, import, and export of student interaction log files.
    b. SCORM 2.0 should define higher level API calls to manage the tracking of student interactions instead of individual calls to the set function.

5   Conclusion

Incorporating and expanding on existing ***learning object metadata, learner information profiles, and learner interaction data standards*** for SCORM 2.0 will provide researchers with the raw materials needed to create SCORM 2.0 conformant e-Learning applications.   This should provide the impetus to bridge the gap between the state-of-the-art in e-Learning research and mainstream e-Learning systems.

Acknowledgements

References:

[1] Roy, D., Sarkar, S. and Ghose, S., "Automatic Extraction of Pedagogic Metadata from Learning Content", *International Journal of Artificial Intelligence in Education*, Vol. 18, 2008, pp. 97-118.

[2] Baker, R., Corbett, A. Roll, I., Koedinger, K., "Developing a generalizable detector of when students game the system", *User Modeling and User-Adapted Interaction*, Vol. 18, 2008, pp. 287-314.

[3] Kobsa, E., Dimitrova V., Boyle, R., "Adaptive feedback generation to support teachers in web-based distance education", *User Modeling and User-Adapted Interaction*, Vol. 17, 2007, pp. 379-413.

[4] Cocea, M., Weibelzahl, S., "Eliciting Motivation Knowledge from Log Files Towards Motivation Diagnosis for Adaptive Systems", *Proceeding of the 11th International Conference on User Modeling (UM2007)*, Springer-Berlin, 2007, pp.197-206.

[5] Devedzic, V., Jovanovic, J., and Gasevic, D. 2007. "The Pragmatics of Current E-Learning Standards", *IEEE Internet Computing* 11, 3 (May. 2007), 19-27.

[6] Kay, J., and Lum, A., "Exploiting Readily Available Web Data for Scrutable Student Models", *Proceedings of the 12th International Conference on Artificial Intelligence in Education*, IOS Press, 2005, pp. 338-345.

[7] Brooks, C. Greer, J., Melis, E., and Ullrich, C., "Combining ITS and eLearning Technologies: Opportunities and Challenges", Proceedings of the 8th International Conference on Intelligent Tutoring Systems, Springer-Berlin, 2006, pp. 278-287.

[8] Kiared, S., Razek, M., and Frasson, C., "The Pyramid Collaborative Filtering Method: Toward an Efficient E-Course", Proceedings of the 8th International Conference on Intelligent Tutoring Systems, Springer-Berlin, 2006, pp. 248-257.

[9] Friesen, N., *The International Learning Object Metadata Survey*. Retrieved August 7, 2008, from http://www.irrodl.org/index.php/irrodl/article/view/195/277/

[10] IEEE Learning Technology Standards Committee: 1484.12.1-2002 IEEE standard for Learning Object Metadata.

[11] SCORM 1.2 Overview, retrieved August 7, 2008, from http://www.adlnet.org.

[12] SCORM 2004 3rd Edition Overview version 1.0, retrieved August 7, 2008, from http://www.adlnet.org.

[13] SCORM 2004 3rd Edition Content Aggregation Model (CAM) version 1.0, retrieved August 7, 2008, from http://www.adlnet.org.

[14] SCORM 2004 3rd Edition Run Time Environment (RTE) version 1.0, retrieved August 7, 2008, from http://www.adlnet.org.

[15] IMS Global Learning Consortium Learner Information Profile (LIP) standard, http://www.imsglobal.org/profiles/

[16] IMS Global Learning Consortium Learner Information Profile (LIP) standard, http://www.imsglobal.org/ep/