From Playful Exhibits to LOM: Lessons from Building an Exploratorium Digital Library

Holly Fait, Sherry Hsi
The Exploratorium
3601 Lyon Street
San Francisco, CA 94123 USA
+1 415 674 2862 / +1 415 674 2809
hollyf@exploratorium.edu; sherryh@exploratorium.edu

ABSTRACT
The Exploratorium, an interactive hand-on science museum, is developing an online collection of science learning and teaching resources to better serve educators' needs for pedagogically-rich instructional resources via the Web. Several challenges arise when designing a digital library for formal K12 education audiences using the Learning Object Metadata standard. These problems are multiplied when attempting to catalog the wide variety of informal learning digital resources from the Exploratorium's ever growing website and exhibit-based resource collections. This paper shares key challenges and early solutions for the creation of an educational metadata scheme, new vocabularies, and strategies for retrofitting existing informal learning science resources into learning objects.

Categories and Subject Descriptors
H.3.7 [**Information Storage and Retrieval**]: Digital Libraries--Collection, Dissemination, Standards Systems

General Terms
Design, Experimentation, Standardization

Keywords
Science museum, metadata, informal learning, exhibits

1. INTRODUCTION
The Exploratorium is developing a digital library collection of educator resources based upon interactive media and instructional materials derived from its exhibits and exhibit-based teaching programs. The Exploratorium is an interactive, hands-on science museum in San Francisco; several hundred exhibits about science, art and perception are on display at a time, out of more than 650 designed on-site. In addition, new exhibits are continually designed at the Exploratorium and customized for other science centers as rotating exhibitions. In the spirit of the Exploratorium's vision of informal science learning, exhibits are designed to promote playful exploration, inquiry, and discovery of scientific phenomena. Since the opening of the museum in 1969 and the launching of the first web pages in 1993, the Exploratorium has a long history of developing educational materials, curricular resources and handcrafted web pages derived from its unique collection of exhibits and exhibit-based examples of science phenomena, as well as from its inquiry-oriented teacher institutes and professional development programs. As of 2004, the Exploratorium has 15,500 web pages, 600 streaming media webcasts, 10,000 digital images, and hundreds of electronic publications on topics ranging from everyday science topics to issues concerning scientific research.

It is the goal of the Exploratorium digital library project to make these rich assets available to both local museum web visitors, and to broad range of learners and K12 educators beyond the San Francisco Bay Area through interoperability with the National Science Digital Library (NSDL). This goal, however, is not easily achieved given the scope and nature of existing literature and practices on the development of learning objects for use in learning object metadata (LOM)-cataloged repositories. Specifically, practices for retrofitting existing digital assets to fit within a learning object framework are rarely addressed in current literature on learning objects and LOM. The discussion of developing libraries that use LOM as their main metadata standard revolves almost exclusively around developing learning objects from scratch—a practice that is too time consuming, expensive, and unsustainable as a design practice with the preservation and public dissemination of Exploratorium educational resources.

During the past year, the digital library team has developed guidelines for cataloging digital exhibit-based science resources, using established metadata standards, while also capturing the unique museum quality and pedagogical value of Exploratorium digital assets. We present this work first with an overview of existing work in the digital library field; next we present our project in detail in order to inform the types of problems we faced with respect to retrofitting learning objects and using LOM. Finally, we present early solutions to those problems and discuss how our solutions can inform the field.

2. USING A METADATA STANDARD
There currently exists a wealth of on-going digital library projects, based at various institutions and with different goals in mind. The importance of digital libraries and motivation to continue growing collections is simply found in how they provide increased reuse opportunities for digital assets, wider dissemination of institutional collections, and improved access for both local and distant learners. Equally important to the success of developing digital libraries are choosing and correctly using metadata standards such as LOM, Dublin Core, CanCORE, MARC and EAD to name just a few. Selection of one standard or variation

JCDL'05, June 7–11, 2005, Denver, Colorado, USA

on a standard comes with both advantages and disadvantages for catalogers (i.e., having a metadata standard format that matches the assets they want to catalog versus having to adapt assets to a standard that does not quite fit.)

Our story is no different. The Exploratorium digital library project chose to use IEEE LOM 1484 as our metadata standard for numerous reasons. Foremost, LOM offered the best fit for cataloguing the various informal learning asset types in our collections. Also, because LOM is widely used, it lends itself to interoperability with the NSDL through simple mapping to Dublin Core (DC). Finally, LOM offered a framework that was otherwise unavailable through other widely used metadata standards (Dublin Core) for describing learning objects.

Some of the disadvantages associated with any generalized metadata scheme, and in our case LOM, have become more publicized in recent years. These issues revolve around describing fully the different types of learning objects housed at various institutions, and around the production of learning objects.

Numerous groups, including [2], [6], and [9] cite the tendency for LOM users to extend or alter strict LOM in order to better describe their collections and make them more accessible. For instance, [2] describes the phenomena of not fully using LOM (i.e., only using a small subset of LOM fields that tend to line up with other standards like DC) and thus only describing the information aspects of assets instead of the information content. Similarly, many DLs must extend LOM to fully describe their collections. The tendency of collections to develop specialized or extended vocabularies for different fields in LOM, or additional fields entirely, in order to best catalog their assets is discussed in [3]. While LOM was designed to be extensible, this practice of extending LOM, or using only small subset of LOM fields can hinder development and interoperability. Further, developing additional vocabularies for assets that don't fit a standard requires skills that are both complex and time consuming to master.

Another interesting disadvantage of using LOM and learning objects is less often directly cited in current literature, but apparent from current discussions: Learning objects and use of LOM tends to rely on the assumption that learning objects are only developed from scratch, and does not include discussion on how to retrofit existing learning resources to fit a learning object format. For instance, [7], [4], [10] and many others take the perspective of having created, or expecting creation of new learning objects addressing issues of accessibility, reusability, and interoperability. This approach, while helpful to new collections for digital libraries, is less useful if there is a legacy of digital content to be cataloged. Additionally, the approach of developing learning objects from scratch is difficult, expensive and increases development time for even simple learning resources [5].

The Exploratorium has struggled with taking our existing, high quality informal learning and teaching resources (which are both educator reviewed, selected, and validated in educator communities) and separating them into appropriate learning objects in order to most successfully apply the LOM standard. Unfortunately, LOM fell short in allowing us to accurately and reliably describe informal science learning assets, and thus, we faced a new challenge of developing our own vocabularies for existing and extending LOM fields.

3. EXPLORATORIUM DIGITAL LIBRARY

3.1 History & Objectives

Given our extensive collection of digital resources created over the years and growing pains associated with managing these educational resources, the museum began an institution-wide digital assets management initiative. This initiative began in 2000 with digitization and cataloging of resources for museum and educator audiences into a digital library archive (funded by IMLS) [1]. More recently, we began the development of K12 educator metadata and metadata schemes for a learning resources collection, and its interoperability and inclusion into the National Science Digital Library (NSDL) [14].

Objectives and on-going challenges for our library included three main issues. First, identifying, selecting and cataloguing digital assets form our collection of hands-on museum exhibits. Second, establishing interoperability with NSDL, and finally, creating guidelines for cataloging digital exhibit-based science resources using established metadata standards while at the same time, capturing the unique cultural, pedagogical, and scientific value of individual assets.

3.2 The Team

Our digital library team consists of staff drawn from different departments across the museum, and in some cases, taking on multiple roles: a project manager, librarian, content developer, education researcher, science teacher, computer scientist, cataloger, science writer/editor and digital rights specialist. Domain specialists were brought in as advisors to the project to provide formative feedback and guidance to the Exploratorium team. We also partnered with expert metadata and database developers at SMETE [15] to build backend components of the library.

3.2 Implementation

The Exploratorium digital library consists of our metadata and asset repositories, a web portal for educators to search our collections and a backend system (i.e., search engine and tools, metadata harvesting and additional services like OAI interoperability with NSDL and software for component communication).

The backend architecture was developed in collaboration with SMETE. Functionality includes off the shelf solutions as well as customized architecture and code solutions:

- Tomcat, an application server [13]
- Velocity, a template engine for developing dynamic WWW pages [13]
- Lucene as a search indexing and query tool [13]
- Struts, a programming framework for servers [17]
- OAI harvesting protocol (provided by SMETE)
- Software "pipes" to exchange data between our asset store, Lucene and Velocity (provided by SMETE)

3.3 Assets Management and Metadata

3.3.1 Asset Management

The asset management system, (Canto™ Cumulus) [11], was initially put in place to catalog digital images prior to the start of the learning resources collection of the Exploratorium Digital Library project. The Exploratorium had selected this off-the-shelf asset management system with the creation of the first image library experiment [1], given the ease of installation and use by museum staff, and the limited availability of low-cost, institutionally-supported technical programmers available to the museum. We continue to use this software today, as it provides ease of scalability and development, advanced APIs, and accessible administration tools.

3.3.2 Metadata

We reviewed and selected several fields from LOM, adapting them to Exploratorium assets, as well as creating additional vocabularies. Some of our assets were best defined using terms not typically found in suggested LOM vocabularies, including "Web Exhibitions", "Museum Exhibits" and "Teacher Professional Development". In order to make these important asset types distinct, we adapted the field titled *Educational Resource Type*. These fields and the motivation for using them are shared further in Section 5: SOLUTIONS. The following is a subset of the final metadata scheme developed for our assets.

Fields included from LOM (not all listed)

- Record Name
- Resource Title
- Brief Caption
- Learning Resource Type
- URL
- Type of Resource
- Grade Level

Additional & Adapted Fields

- CA Standard
- Curricular Area (3 levels, e.g.: Physical Science→Light and Optics→Color)
- Learning Resource Type (vocabulary)
- Teacher Description
- Teaching Topic
- Resource Description
- Phenomena

3.4 User Interface

Our web interface (Figure 1) was developed in-house, based on usability studies with middle and high school teachers and iterative design techniques. Our first basic implementation (not yet available to the public) allows users to search, view resource information, and download digital assets. Current functionality includes simple and advanced search, a browsing interface based on mappings to curricular areas, teaching tips, and resource descriptions. Additional functionality is slated for development after a full evaluation of the current implementation is completed. This includes indexes to science standards, grade levels, featured resources, a user tutorial, and links to related resources.

4. THE PROBLEMS

Because the Exploratorium is an institution with a long tradition of exhibits and exhibit-based resources designed for curiosity and play, and a wealth of digital assets developed for informal learning, we face a 'happy problem'. On one had, most of our digital assets are high quality, developed for a wide range of users and have stood the test of time in the field. On the other hand, the creative and collaborative nature of the Exploratorium has led to a rich collection of digital resources with little consistent infrastructure and no unified voice to describe its contents. While our resources were developed in a cultural context of museums and informal learning, our target audience was K12 teachers in formal settings who teach within a grade level and value materials indexed to state and national content standards. For this project, our focus was not on developing new learning objects, but instead on retrofitting existing digital assets into an existing LOM standard. The problem was threefold and applied, in some cases, to a subset of our asset types.

Figure 1: Resource Record view of the web interface for the Learning Resources Collection in the Exploratorium Digital Library.

4.1 Cataloguing Legacy Assets

Development of highly reusable learning objects is time consuming and expensive, as shown by a handful of digital library projects and learning object experts [5], [10]. During the project, we learned quickly that simply cataloguing existing assets was extremely time intensive and required great expertise, along with additional infrastructure to help catalogers use a similar voice and vocabularies. We sought a solution to assist our catalogers with their task and to maximize the quality of newly created metadata.

4.2 Maintaining an "Exploratorium-ness"

We also needed a way to maintain the cultural, instructional, and scientific value of Exploratorium assets in the metadata authored. For instance, the term "snack" is commonly used within the Exploratorium community to describe a specific type of hands-on learning activity and tool, but it is formally defined in the LOM Education Learning Resource Type field as an activity. Due to the unique nature of our assets, placing them in a uniform repository of learning objects--accessible to users inside and outside the Exploratorium community--was critical to their continued use. Further, the creation of new learning objects to take the place of existing resources was not a goal of the project and would contradict the goal of recording existing museum

culture. The unique combination of cultural, instructional and scientific value imbued in all Exploratorium collections is explained by many in our community as an "Exploratorium-ness".

Finally, having found a way to organize the team catalogers in the production of high quality metadata, and methods for maintaining the "Exploratorium-ness" of our assets, our third problem rested on the nature of the initial creation of our assets.

4.3 Asset Format

Despite beginning our digital library project with previously digitized assets, we still faced challenges in retrofitting certain digital formats and presentation schemes to best fit a reusable learning object structure. For example, a number of our assets were web sites, online collections, or online exhibitions about a given topic, place, or floor exhibit (e.g. Science of Skateboarding, Microscope Imaging Station, or Saturn.) Each of these assets might contain zero or more learning objects within it, or could be entire sites that are best described as only one learning object. Many of our library development and cataloguing problems were not faced when cataloguing simpler items like science publications and hands-on activities, but instead, when cataloging web exhibitions that contained multiple media types and teacher professional development resources.

Additionally, some of our most valuable assets were available only as large streaming media files, and often included many different learning objects in one file. Our team needed a systematic method for disambiguating these previously compiled learning assets into meaningful learning objects.

Figure 2: Top page of the Iron Science Teacher website showing an example of one web page with multiple links to live and archived streaming media webcasts.

5. SOLUTIONS

5.1 Two for One

The solutions to the problem of uniform voice and consistent, quality metadata authoring, discussed in sections 4.1 and 4.2, hinged on implementing a rigorous editorial policy, and by extending our metadata fields beyond LOM.

5.1.1 Editorial Policy

Our asset selection, editorial and cataloguing policy included development of selection criteria for existing assets, fill-in metadata sheets for our domain experts to use along with a content guide written by our expert teacher/cataloger to outline voice. A style guide was also included for use by our editorial department. This policy outlines explicit stages of metadata development that define when and by whom assets are cataloged and the responsibilities of the cataloger at any given stage. Specifically, the editorial policy defines that assets are first selected (from on-line repositories, existing digital libraries or from individuals in the institution) and curated by the library staff. Next, a science teacher with access to domain experts, using the voice defined in our content guide, populates the fields that we found were most highly valued by educators (grade level, standards and teacher description). Then, that content is edited and fitted to our style guides by editorial staff. Finally, a staff librarian completes additional metadata fields integral to our collections management efforts.

This approach allowed us to institute rules for voice and thus reduce the variability among completed metadata fields. Further, it allowed us to formally maintain the way terms and vocabularies are used in the Exploratorium, and make those terms available to our catalogers and end-users. Recall the term "snack". Formally, Exploratorium snacks are defined in the LOM Education Learning Resource Type field as activities, but are defined by our community simply as snacks. Our cataloguing rules and practices covered this difference to increase end-user ease-of-use by including both descriptors in our Teacher Description field (e.g. "This Exploratorium snack is a hands-on activity for….").

By setting rigorous editorial standards and setting style guidelines, we were able to both preserve the style of our most important metadata fields, as well as maintain the special nature and culture surrounding Exploratorium resources. With this framework in place, we have the flexibility to use multiple expert educators as catalogers and then implement simple peer review for adherence to styles before passing those assets on to our editorial staff or librarians for final cataloging. This effectively decreases the total time and effort one single cataloger without domain expertise might spend cataloging our assets.

5.1.2 Extending LOM

Some of our assets were best defined using terms not typically found or suggested in LOM, including "Web Exhibitions", "Museum Exhibits" and "Teacher Professional Development". In order to make these important asset types distinct, we included a separate vocabulary for use in the LOM Learning Resource Type field.

We also found that the fields available in LOM did not allow us to fully express the different phenomena or curricular topics covered by our assets. For our metadata field "Curricular Area", a set of controlled vocabularies were defined which included first, second, and third level categories. The first level includes large granularity educational topics such as Life Sciences, Physical Sciences and Earth Sciences; the second level, under Life Sciences, for instance, includes "Cells, Heredity & Genetics, Evolution..." The third level describes even smaller-grained subsets of level two

topics. These controlled vocabularies were derived from a synthesis drawn from contributions by staff teachers, review of existing educational digital libraries, and indices to school science textbooks from major publishers. A complementary field called "Phenomena" provided a way to use another set of controlled vocabularies used by museum staff, exhibit developers, and staff scientists to describe the scientific and/or visual phenomena being demonstrated in the digital asset.

5.2 Categorizing Web Assets

Our third problem, discussed in section 4.3, forced us to develop a categorization scheme to disambiguate our assets in a methodologically sound and repeatable manner. Each of the questionable assets (i.e., those that in their current incarnation were quite difficult to catalog as a single learning object) fell into one of three categories.

The first category (one web page with one asset) is included for completeness, but caused little to no difficulty in cataloging. The second category of assets included a single asset (or web page) that has more than one learning object within it. Some examples of these are a single online exhibit, a snack, a science article, and a web interactive such as a Flash animation. We chose two approaches: (1) Make two separate records with two different sets of metadata, where both records point to the same asset—forcing users to potentially scroll through the asset to find what they are looking for; or (2) for assets that are small enough, make only one record and describe all of the objects within the asset. In practice, the first approach is more effective for disambiguating different learning objects from a development point of view, but is possibly confusing to users. The second approach, while technically cumbersome for catalogers to describe each item in limited space, might be more understandable for end-users. We will investigate these possibilities in our upcoming user testing.

Table 1: Categorization of Web Assets

Description of Pre-cataloged asset	Learning Objects	Records
1. Single Webpage, One Asset	Single Learning Object	Single Record
2. Single Webpage, One phenomena, Multiple Assets	Single to Multiple Learning Objects	Single or Multiple Records
3. Single Webpage	Multiple Learning Objects	Multiple Records— pointing to the same asset
4. Large Media Files	Multiple Learning Objects	Multiple Records— pointing to different portions of the asset

The third asset category includes multiple combined assets (imagine a web site) that primarily describes one phenomena, but can cover more than one learning object. In this case, we were torn between trying to describe the richness included in a complete web site or collection of assets in a single record, multiple records, or one record per page or asset. This last solution created assets with a learning object granularity too fine to be effective. Similarly, just assigning one record to such a rich asset made the learning object granularity too large for real reuse. While the issue of appropriate learning object granularity is common [8], [16], the Exploratorium challenge of *retrofitting existing resources* into appropriate learning object granularities is not easily adapted or exemplified in current literature. We decided, for this specific problem, to create a record for the main, top page of a site and to create records for individual assets (possibly multiple page) within the site. Top pages, like a jump-page or table of contents, were treated the same way. We used our previously developed Educational Resource Type field to describe these sorts of mini-collections as "Web Exhibitions".

The fourth category includes only large or streaming media files. For example, Iron Science Teacher (Figure 2) is a website which contains archived and live webcasts of school teachers engaged in a friendly competition to build and present their best instructional lesson to a live audience. These types of websites posed interesting problems with respect to learning object mapping.

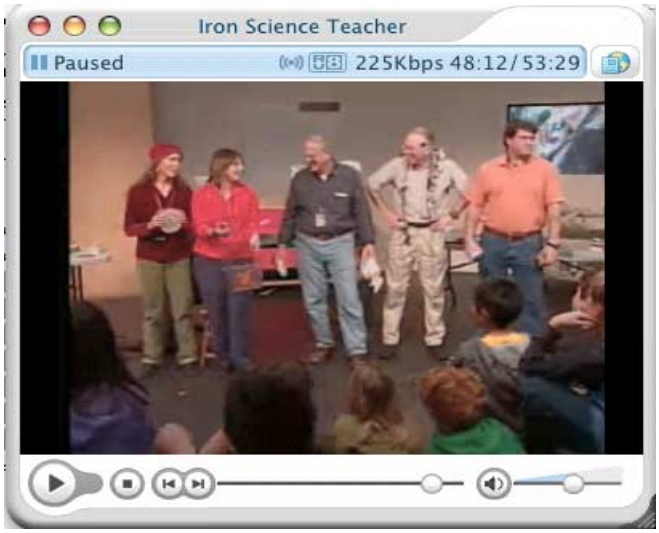

Figure 3: A frame from the Iron Science Teacher webcast show at the Exploratorium.

Each episode of an Iron Science Teacher is a one-hour live webcast, which is then archived on the website. Featured in the webcast, five teachers are given a short time to develop a lesson or demonstration to illustrate a certain concept using a mystery ingredient announced moments after the show begins. During the competition, each teacher presents his or her experiment to the audience to vote on their favorite. These webcasts, while extremely popular among teachers and visitors who want to review the show after a visit or share with colleagues and family, are extremely difficult to navigate. This is compounded when a user desires just one or two instructional ideas from within the entire one-hour video. From a cataloguing perspective, they have many of the same issues as elements in our second and third categories combined, but the format of the data introduces additional programming issues. For instance, if one webcast is split into four learning objects, each representing an experiment shown, there is still the issue of how to allow users to access the master file at the correct point. Similarly, because we are not

developing new content, the solution of slicing such webcasts into multiple coherent video files is too labor intensive.

One solution, similar to a hybrid approach for our second and third categories, combines creating a master record for the entire webcast, and then one record for each additional learning object within the webcast. We developed a solution for our user trials that uses multiple records, and allows users to "jump-in" to a particular point in the file, given an appropriate URL. For instance, one record from a webcast might describe an in-depth experiment running from timestamp 00:34:12 to 00:41:58. The URL associated with that asset record would point to the video at time 00:34:12. Admittedly, this solution might not be extensible because of the staff time needed to index digital videos. Beyond accessing the file at certain points is the issue of downloading and data storage. If educators want to download just one section of the video, they still must download the entire media file—a prospect that is daunting to people with slow or limited Internet connectivity. With respect to storage, it is not reasonable to keep multiple copies of large streaming media files within our asset store, and thus we are left to relying on URLs to remain somewhat static. Despite these issues, this is the most feasible solution available to us, given our goal of not developing new media, and time and staffing constraints.

6. CONCLUSIONS

In this paper, we have provided a small window into the issues and process by which a legacy of exhibit-based informal educational content and resources has been cataloged into a LOM-based scheme with some adaptations and additions.

Our discussion and working solutions suggest an initial set of guidelines and cautionary points for other informal science learning institutions who wish to build digital library collections using LOM with existing digital assets.

With additional feedback from other designers of digital libraries and formal testing with teachers in the next steps of the project, we will have empirical support for the utility of our metadata scheme and cataloging approach.

From our work-to-date, it is clear that a uniform standard is needed to assist science centers and other informal science institutions in retrofitting education resources to fit theoretical learning object structure and LOM, while also ensuring that, as digital libraries scale, newly developed learning objects can be added. Such a uniform standard is possible when science education digital libraries work synergistically to share and critique metadata schemes to arrive at better educational metadata standards.

7. ACKNOWLEDGMENTS

This work is based on work supported by the National Science Foundation National Science Digital Library Grant No. DUE-0333722. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the view of the NSF.

8. REFERENCES

[1] Falanga, R., Hidden treasures: managing digital assets at the Exploratorium. In *Computers in Libraries*, 22:4, April 2002, pp 14-20.

[2] Friesen, N. The International Learning Object Metadata Survey. *International Review of Research in Open and Distance Learning* (Nov. 2004), #40.

[3] Friesen, N and Nirhamo, L. Survey of LOM Implementations: Preliminary Report. In *ISO/IEC JTC1 SC36, Information Technology for Learning, Education and Training* 2003-09-03.

[4] Hong, J., Chen, H. and Hsiang, J. A Digital Museum of Taiwanese Butterflies. In *Digital Libraries* (San Antonio, Texas, 2000, 260-261). ACM, 260-261.

[5] Laleuf, J and Spalter, A. A component Repository for Learning Objects: A Progress Report. In *Proceedings of the JCDL* (Roanoke, Virginia, June 24-28, 2001). ACM Press, New York, NY, 2001, 33-40.

[6] Neven, F. and Duval, E. Reusable Learning Objects: a Survey of LOM-Based Repositories. In *Multimedia '02* (Juan-les-Pins, France December 1-6, 2002). ACM Press, New York, NY, 2002, 291-294.

[7] Polsani, P. Use and Abuse of Reusable Learning Objects, *Journal of Digital Information*, *3, 4.* Article No. 164, 2003-02-19.

[8] Quinn, C. & Hobbs, S. Learning Objects and Instructional Components. In *Educational Technology and Society*, 2002, 3(2). Retreived March 14, 2005 from the World Wide Web: http://ifets.ieee.org/periodical/vol_2_2000/discuss_summary_0200.html

[9] Saddik, A., Ghavam, A., Fischer, S. and Steinmetz, R. Metadata for Smart Multimedia Learning Objects. In *ACE 2000* (Melbourne, Australia, December, 2000). ACM, 87-94.

[10] Wiley, D. A. (2000). Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. In D. A. Wiley (Ed.), The Instructional Use of Learning Objects: Online Version. Retrieved January 22, 2005, from the World Wide Web: http://reusability.org/read/chapters/wiley.doc

[11] http://www.canto.com/

[12] IEEE Learning Technology Standards Committee (LTSC) (2001) *Draft Standard for Learning Object Metadata Version 6.1.* http://ltsc.ieee.org/doc/

[13] http://jakarta.apache.org/tomcat/

[14] National Science Digital Library http://www.nsdl.org

[15] http://www.smete.org/smete/

[16] http://www.reusablelearning.org/index.asp?id=28

[17] http://struts.apache.org/