AN ONLINE PROCESS FOR POSTING AND META-TAGGING LEARNING OBJECTS AND INTEGRATING CONTENTS INTO DIGITAL LIBRARIES

Shankar Mahadevan[1], Saifur Rahman[2] and Peter Wiesner[3]

Abstract *– Digital libraries are fast expanding into the role of independent educational entities that aspire not only to complementing traditional classroom teaching, but also allow open electronic learning for distance and continued education. Many studies have defined a concept of "learning object" to address these issues. But in attempting to solve the problem, the definitions have emphasized some aspects of the digital library while leaving the other issues to be solved later. In this paper, we have proposed a comprehensive learning object (LO) model, along with the associated system model, that will allow complete and flexible integration of content into digital libraries. The process will be user-centric (both for knowledge developers and learners), scalable and interoperable with legacy and existing content databases and display systems. This paper covers how the LO model was integrated into the core of the library's content development, discovery, and delivery process. The results of the experiment in terms of ease-of-use, flow-control, and feasibility of the model are documented.*

Keywords *– Digital libraries, system profile/architecture, learning object model, metadata, content development, content discovery, content delivery.*

1.0 INTRODUCTION

Digital library is fast evolving into a multibillion-dollar enterprise, both in the academic and the commercial sectors [1]. Besides augmenting traditional forms of teaching, it, in itself, is becoming a self-contained institution that aims to provide education on demand. With increased proliferation of the Internet and the availability of related software and hardware technologies, the time-to-inception and maintenance has progressively decreased. But, this development and spread of digital libraries is asynchronous, and disjoint across various disciplines, educational institutions and commercial affiliates. Even standardizing bodies, that aim to bring about consistency in the implementation and the use of digital libraries are many in number, restrictive (in some cases proprietary) and non-interoperable [2]. Though, some common concepts have emerged that have an underlying unifying theme. One such concept is the "data model" or "object model" - a metadata-centric content model [3, 4, 5, 6, 7].

In this paper, we propose a unique object model for educational resources at the core of learning. Our model shows how learning (and resources associated with it) can be transformed from teacher-centric to learner-centric education, and how content can be transformed from multimedia resources (application or technology specific data) to metadata-based learning modules. This "learning object (LO) model" can be easily molded to use any known and popular metadata and/or content packaging standards. The features and complexity are functions of the owner's (digital library's owners/operators) choice of these standards. Thus making it versatile and inherently scalable. Its implementation is made simpler because of its object-oriented abstraction and "tree-based" file structure. In a digital library, a sound and efficient LO model is critical for pedagogy. But, before one attempts to define and use a LO model, it is essential to evaluated the nature of pedagogy. Figure 1 shows how pedagogy has evolved along with the digital age.

2.0 MODERN PROFILE OF A DIGITAL LIBRARY

The traditional model, Figure 1(a), has been the centerpiece of imparting education up until the last century. It is a closed model with little room for flexibility and enhancement. Schools and universities, which represent education managers here, structure curricula and standards for judging the level of learning. Here a teacher (the knowledge developer) creates relevant educational resources (the content) to provide step-by-step learning instructions in their respective field of expertise. The students (the learners) use these resources that fit the curriculum, of their desired field of study, to gain education. Thus, there are two distinct paths to pedagogy: the learning path (attending designated standardized in-class lectures etc) and the resource path (via recommended books, home-works, etc) from the educator to the learner.

An overall architecture, to represent this new dynamics to gain pedagogy, is proposed in Figure 1 (b). It represents the objective modularity of the underlying digital nature of relationship between various parties involved in education. At the core is Metadata, which is defined as data about data. It enables smooth flow of control among different entities, actions and functionalities provided by the digital library. Among others, the three main types of Metadata are: content-centric, user-centric, and pedagogy-centric metadata. Content-centric metadata deals with content description and packaging and are used while searching for the content.

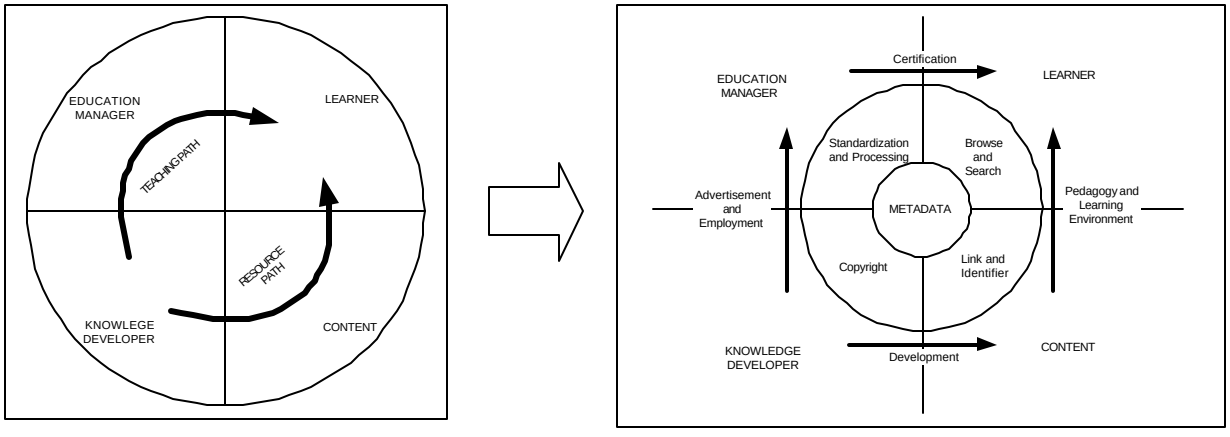
[1] Shankar Mahadevan, Virginia Tech, smahadev@vt.edu
[2] Saifur Rahman, Virginia Tech, srahman@vt.edu
[3] Peter Wiesner, IEEE, p.wiesner@ieee.org

User-centric metadata deals with the creation and maintenance of user profiles. Pedagogy-centric metadata deals with evaluation of learning and result tracking. The value of elements in the metadata is used to initiate actions among various entities within a digital library. The interaction and inter-use of metadata elements between these entities is complex and often overlapping, while the functionality provided by the metadata elements to each entity is unique [8, 9]. The knowledge developer quadrant encompasses professors, teachers, subject experts, and alike – those who generate content having educational value. Additionally, in the role of a reviewer, they determine the learning value and quality of the contents. Besides this traditional role, their participation in the digital library provides them with added value of copyright and recognition for the content generated by them. The author information and copyright is preserved via the metadata association of the resource. It is imperative that they understand and use metadata standards (especially those relating to content-centric development and use) developed by education managers. This allows easy identification and use of the content among various groups thus maintaining consistency and smoother flow-control across the digital library.

The content quadrant represents the physical resources that contain educational information. These resources can be of any size and format – as long as it can be electronically handled. The resource ownership, description, grouping and association are encoded in its identifying metadata file. Without the identifying metadata file, though the pedagogical value of the resource may not be lost, the mere fact that it has pedagogical value would be difficult to estimate. Thus, its use across the digital library is in jeopardy.

The last quadrant, the learner is the ultimate end-user for all pedagogical service activities. They use the metadata to seek content. It may involve many alternate methods including: basic keyword search, advanced Boolean search, browsing through taxonomy, etc. Without the metadata, any information about the resource or its pedagogy is hard to find. In digital medium, metadata also helps to identify learner-centric resources. Some advanced systems, when allowed, can filter/seek learner-centric modules based on the user-profile metadata.

It is clear that the interaction between various quadrants occurs via actions invoked by values in the metadata elements. As long as each quadrant can access the desired metadata, they can exist mutually independent of each other. For example, universities do not necessarily need to maintain content banks. It may be more efficient for other third-party entity (possibly a commercial database company) to maintain such content banks. What is required, is that the third party implement the metadata standard set by the university, when accepting content from authors (possibly pay royalty in return), following which it should share (possibly for a fee) with the affiliate university. The university should then, based on the quality of the metadata, set requirements of certification. The learners, may at anytime search the metadata repository and then download (possibly for a fee) the chosen type of content and gain education at the desired pace of study. The type and fashion of lessons studied by the learner is stored in his/her user-profile. If the user so desires, then, may share this profile with the university. Following which, the university, when satisfied of the fulfillment of the course requirement, may provide the desired certification (at a fee), to the learner, for the acquired knowledge. The following section defines an application independent definition of a LO that best fits the above system model of the digital library.



(a) TRADITIONAL PROFILE (b) MODERN PROFILE

FIGURE 1
EVOLUTION OF PEDAGOGY ENDOWMENT VIA DIGITAL LIBRARIES

3.0 LEARNING OBJECT MODEL

As mentioned earlier, content can be of any size and format and sometimes in mixed formats. Moreover the material can be presented as text, graphics, animated graphics (that is, graphics that move), audio, video, or a combination of these. The associated metadata represents a standardized set of rules and methods that allow collecting and processing this content. The concept of learning object encapsulates these requirements [10].

> *A learning object is defined as a structured electronic resource that encapsulates high quality information in a manner that facilitates pedagogy. It has a stated objective and a designated audience. It has ownership and associated intellectual property rights.*

The definition highlights two aspects namely, "learning" and "object" with the underlying theme being "ownership" and "quality". It is immutable in the sense that it may not be changed or altered by the user. It must function independently (self-contained and standalone). Also note LO may have other LOs embedded in it.

The "learning" aspect of the LO definition is very important. Digital libraries are more than a simple content repository or a hosting platform, promoting continuous education and life-long learning. It will facilitate the acquisition, assessment and conversion of contents into LOs while fostering the assimilation of those LOs into learning modules and instruction. The "object" aspect of the LO definition relates to digital/electronic format of the resources. As digital constructs that encapsulate both information and behavior, the "objects" in LOs are analogous to objects used in object-oriented modeling (OOM). The analogy helps visualize how LOs will be packaged, processed and transported across the digital library as well as utilized in course building. Digital libraries have to assist in streamlining the processing of these objects. OOM concepts such as encapsulation, classification, polymorphism, inheritance and reuse can be borrowed to describe the operations on LOs in the digital library. Ownership and quality are the requisite attributes of any resource within a library. Ownership relates to the intellectual property of the resource. Quality relates to the following facets:

- Subject matter accuracy and authenticity,
- Pedagogical effectiveness or the educational value,
- Relevance of the information in a resource relative to its objective,
- LO features that represent post-publication usefulness such as easy discovery and use/re-use by learners and teachers,
- Technical "soundness" of LOs.

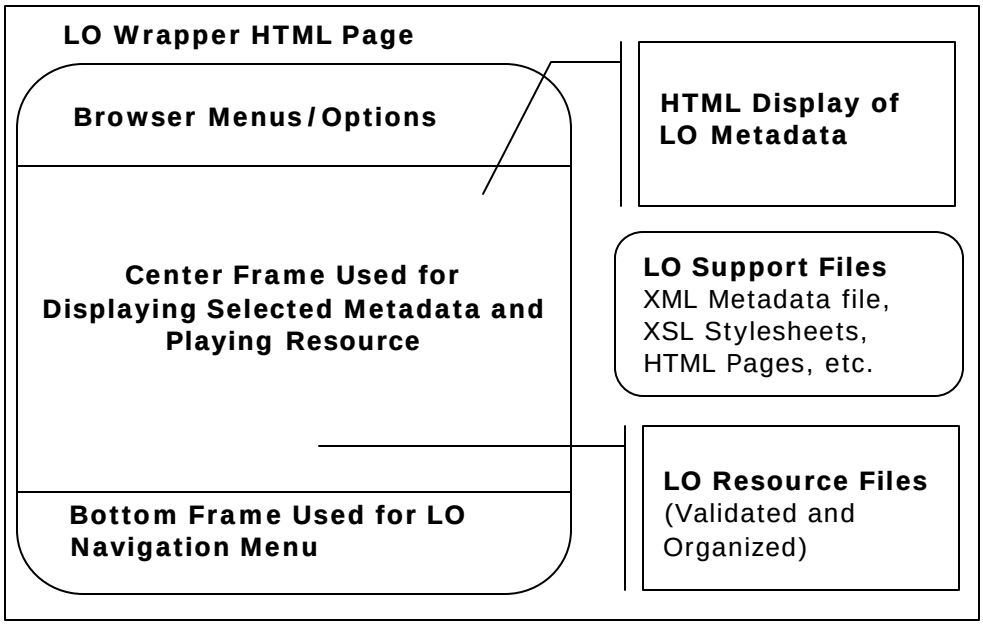


FIGURE 2
LEARNING OBJECT MODEL

Unlike in a physical book, where one knows that the appropriate side of the cover is the starting point which is then followed by a table of content and then the text, digital content as such inherently don't have any structured starting or continuation point. Asking authors to do this job is tedious and automatic chronology extraction tools have so far been not quite successful. We have proposed to implement a simple wrapper as a part of the LO which allows for its flexible use. Figure 2 shows the conceptual model. The LO will always play within a web browser. The front of the LO will be a designated frame-based HTML page that will list the essential set of data describing the author and usage in the top frame. Upon initiating the appropriate link, the contents of the LO play in this frame. The bottom frame is used to navigate through the LO. Options here might be to re-start and/or to view more complete set of LO related information. Due to modular format of the LO and need for versatility of use, additional support files are provided, which give instructions to users for appropriate use of the LO.

The LO related information forms the basis for content-centric metadata. The metadata can be stored in any of the widely available standards for data preservation and use [9, 10, 11]. Note this standard should be uniformly implemented for all functions within the digital library. Alternately, crosswalks may be developed for inter-operability. Having understood the various characteristics and features of an LO, let us now look at file structure that are required to implement such an LO. It is a "tree-based" implementation for scalability. The top package is split into four groups (see Figure 3 for illustration):

1) LO Preamble or the wrapper: LO preamble file initializes the LO and displays information regarding what is to follow. The LO resources are initialized through this file. It should always have the same and designated filename, so as to be scalable and compatible across the digital library. This file is located at the top level within the LO folder. It is an HTML file that has two frames - one on top of another. The top frame is used to "play" the resources and bottom frame is used to navigate the LO. The buttons, display and positioning may be optimized as required.
2) LO Metadata File: LO metadata file contains information about the LO. It is based on the desired metadata standard, a system parameter uniform across the digital library. Generally, it is an XML file [11], which contains the creator's name, organizational affiliation, LO's ranking, objective, pre-requisites, and many other types of information. The recommended location of the file is at the top level within the package.
3) Support Files: Support files are optional. They may be required in order to perform functions provided in the bottom frame of the wrapper. All support files must be located in a sub-folder labeled *SupportFiles* within the top LO folder.
4) LO resources: LO resource files are the collection of digital constructs (the objects) that convey educational information. These files are located in a sub-folder labeled *Resources* within the top LO folder. It may contain sub-folders as desired by the knowledge developer. The relative paths are accordingly adjusted, to allow proper functioning.

Figure 3 shows the complete file structure of the proposed LO file package. When an LO is embedded into another LO, the same structured is preserved, but it is placed within the *Resources* folder as it is "object" for the upper level LO.

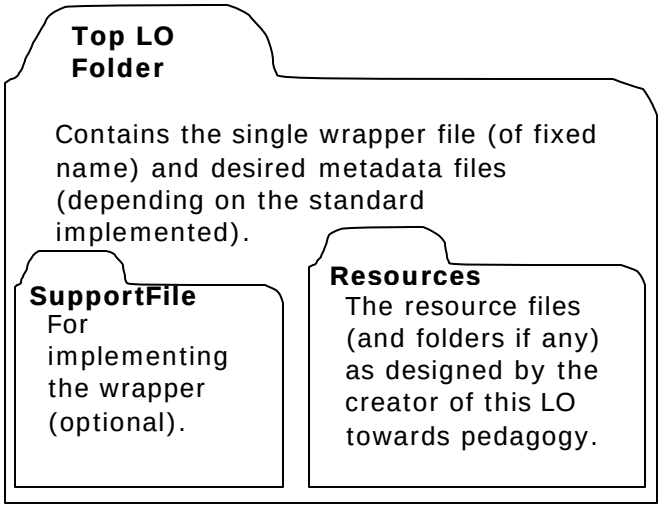


FIGURE 3
FILE STRUCTURE OF A DLNET LO

4.0 DLNET – A COMPLETE SYSTEM IMPLEMENTATION

The objective of the Digital Library Network for Engineering and Technology (DLNET) is to develop a specialized collection of engineering and technology-related content targeted at the practicing engineer and technologist [12]. We have implemented a limited model of the modern pedagogical endowment for the DLNET. The corresponding LO model was also implemented for flow-control and user/system interactions. The model basically allows content creators to "package" resources for the educator to the learners. The core metadata in this case is the content-centric metadata. It is based on the IMS-IEEE Metadata standard [9, 11]. Note these are the education managers (DLNET being a public domain project, its service are free, hence other groups of the education manager quadrant are not represented here). A simple in-house XML based metadata format was developed to track the user-action in the user-centric metadata. The LO is an organized

standalone resource that encapsulates quality information. This information must enable learning.

Implementation of the LO model involves following three steps: harvesting metadata (information about the LO and its creator), resource validation and content packaging. A simple JAVA application was built to assist knowledge developers in this task. DLNET, with some added extension has adopted the IMS-IEEE LOM and content packaging standards. The metadata information helps catalog the LO and, later to identify it during search. As defined earlier, the DLNET LO model has four components: the preamble, the metadata file, the support files and the resource files.

Once the LO is created, the knowledge developer logs in to the DLNET website and uploads the LO. The LO is then separated into educational content and the metadata. Members of the peer group review the content and metadata, and upon passing, the metadata is added to the metadata repository and content is added to the content repository. Public can search or browse through the metadata at anytime. But, they are required to login to download the content. This way a user-profile database is developed. As of now no plans to certify are envisioned, but can be easily implemented to complete the digital library picture.

5.0 SUMMARY AND CONCLUSIONS

A digital library can be successful only if the underlying system architecture is robust. While the hardware and software aspects play a major part in implementation, it is the interaction and functionalities that are provided by the entities associated with digital library, that will allow efficient pedagogy. The modern profile of the digital library is such that these entities, while acting independently, have a unifying goal to sustain the paths to knowledge from the educator to the learner. It is modular and hence more flexible in meeting the changing landscape of education endowment. Furthermore, digital libraries must themselves be scalable to integrate various libraries representing different disciples together. All this is possible only because of the central and critical role played by metadata. The educational content is processed base on the values of elements in the metadata. Thus, metadata and content should inherently be grouped together by association. The learning object model defines such an association.

The LO model should in itself be easily accessible across various quadrants of the digital library architecture. The proposed concept of "learning object" helps authors and users alike in streamlining the learning program development process. Its physical representation is hence standardized. The system designer, based on the available underlying hardware and software capabilities, can decide the metadata and packaging formats. It, thus, integrates the digital library abstraction into the actual digital media.

A beta-version of these concepts has been successfully tested with volunteers and implemented as a part of the DLNET project. Above all digital libraries, like the DLNET are evolving to be more than just a repository of learning materials. Digital library patrons can judge and selectively choose what they want to learn about and then have a system that supports that chain of learning. Hopefully, in the longer run these ideas will promote other digital libraries to standardize not only the development of learning programs, but also the overall system architecture so as to allow easy across-the-board acceptance of digital libraries as a credible and qualitative source of education - anytime and anywhere.

6.0 REFERENCES

[1] Ubel R., "Engineers turn to e-learning", *IEEE Spectrum*, October 2000, pp. 59-63.

[2] Mallak L. A., "Challenges in Implementing e-Learning", *Portland International Conference on Management of Engineering and Technology*, 2001, pp. 298-299.

[3] Ozsoyoglu G., Balkir H., et al., "Electronic Books in Digital Libraries", *IEEE Proceedings on Advances in Digital Libraries*, 2000, pp. 5-14.

[4] Adam N. R., Atluri V., et al., "A Dynamic Manifestation Approach for Providing Universal Access to Digital Library Objects", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 13, No. 4, July/August 2001, pp. 705-716.

[5] Wiley D., "Peer-to-peer and Learning Object: The New Potential for Collaborative Constructivist Learning Online", *IEEE International Conference Advanced Learning Technologies*, 2001, pp. 494-495.

[6] Carchiolo V., Longhew A., et al., "Learning through Ad-hoc Formative Paths", *IEEE International Conference Advanced Learning Technologies*, 2001, pp. 96-99.

[7] Anido L., Llamas M., et al., "A Distributed Object Computing Approach to E-Learning", *3rd International Symposium on Distributed Objects and Applications*, 2001, pp. 260-269.

[8] Mukaiyama H., "Next Generation Digital Library – Architecture and Implementation", *International Conference on Digital Libraries*, Kyoto, 2001, pp. 105-112.

[9] Anido L., Llamas M., "A contribution to the e-learning standardization", *2nd IEEE Conference on Standardization and Innovation in Information Technology*, 2001, pp.295-308.

[10] DLNET Team, "Learning Object Definition", *White Paper*, October 2001, http://www.dlnet.vt.edu/.

[11] DLNET Team, "DLNET Schemas for Metadata and Content Packaging", *White Paper*, November 2001, http://www.dlnet.vt.edu/.

[12] "Pathways to Progress: Vision and Plans for Developing the NSDL", March 20, 2001.

7.0 ACKNOWLEDGEMENT

The work reported here was supported in part by a grant from the National Science Foundation, grant number: DUE-0085849