Canadian Journal of Learning and Technology
Volume 28(3) Fall / automne, 2002
CanCore: Metadata for Learning Objects

Norm Friesen

Anthony Roberts

Sue Fisher

About the Authors

Norm Friesen is an Information Architect at Athabasca University and has been working in the area of Web-based learning since 1996. He is currently completing his PhD in educational technologies at the University of Alberta. He can be contacted at: norm.friesen@ualberta.ca

Anthony Roberts currently manages the TeleCampus repository for TeleEducation New Brunswick. Besides educational technologies and standards, his research interests include feminist ethics, poststructuralist thought and environmental studies. He can be contacted at: toni@telecampus.edu.

Sue Fisher is a children's librarian and also works at the Electronic Text Centre at the University of New Brunswick. She can be contacted at: fisher@unb.ca

Abstract

The vision of reusable digital learning resources or objects, made accessible through coordinated repository architectures and metadata technologies, has gained considerable attention within distance education and training communities. However, the pivotal role of metadata in this vision raises important and longstanding issues about classification, description and meaning. The purpose of this paper is to provide an overview of this vision, focusing specifically on issues of semantics. It will describe the CanCore Learning Object Metadata Application Profile as an important first step in addressing these issues in the context of the discovery, reuse and management of learning resources or objects.

Rsum

La perspective de ressources ou d'objets d'apprentissage numériques réutilisables, rendues accessibles grâce à des architectures de logithèque et des technologies de métadonnées coordonnées, a attiré une attention considérable dans les communautés de l'enseignement et de la formation à distance. Cependant, le rôle critique des métadonnées dans cette perspective soulève des questions importantes qui ne datent pas d'hier sur la classification, sur la description et sur le sens. L'objectif de cet article est de donner un aperçu général de cette

perspective, en se concentrant tout particulièrement sur les questions sémantiques. L'article décrit le Profil d'application des métadonnées des objets d'apprentissage CanCore (CanCore Learning Object Metadata Application Profile) comme une première étape importante dans l'examen de ces questions et ce, dans le contexte de la découverte, de la réutilisation et de la gestion des ressources ou des objets d'apprentissage.

Introduction

In Understanding Media, McLuhan (1964) suggests that the content of any new medium, at least initially, is provided by the medium that it is in the process of supplanting: the content of early writing, as in Homer's Odyssey, is the spoken word and the content of early cinema has been identified as theatre or vaudeville (Manovich, 2001; p. 107). Developments in Web technology and the use of this technology in education also seem to follow this pattern. Exclusive concern with document appearance and presentation —characteristics inherited from the print world— are gradually giving way on the Web to multimedia formats and distributed organizational mechanisms. Similarly, in distance education, the Web initially took as its content the lectures, overheads, discussions and other aspects of the traditional classroom. Many of these aspects - down to the closed classroom door, the obligatory teaching assistant and the classroom whiteboard - have been faithfully transferred onto the Web via password-protected course management systems like WebCT and Blackboard. However, attempts to replicate the face-to-face classroom seem to be giving way to distributed systems of "learning objects" that exploit the intrinsically decentralized and modular nature (Manovich, 2001) of Web-based content.

Access or discoverability of Web-based resources has typically been facilitated through the use of search services or engines such as AltaVista or Hotbot. In the simplest terms, these services make Webpages and Websites discoverable by finding matches between the character-combinations or "strings" entered by the searcher, and those occurring somewhere in the textual contents of Web documents themselves. The problems that this technology presents to users in general and educators in particular are both familiar and manifold: tens or hundreds of thousands of "matching documents" are retrieved in response to almost any search string, educationally appropriate resources are difficult to find and evaluate, and multimedia or interactive content is not directly searchable given. The inadequacy of this search technique springs, in part, from the fact that it works only with mere character combinations, matching those typed as searches against those occurring in Web pages. These search services have no real way of understanding, evaluating or registering the significance of these character combinations or the potential purpose or value of the resources. These services, in other words, only recognize the external properties or the appearance of words, seeing them simply as "formal squiggles" (Dreyfus, 2001). This rather artificial and rigid method reflects the overarching epistemology of western approaches to knowledge production. Knowledge is understood as simple classification, as in Kant's Critique of Pure Reason and Descartes' understanding of bodies and knowledge as mechanisms.

Metadata: What is Data All "About"?

What has been widely suggested as a solution to these problems is to turn attention to the actual meanings of the words in Web documents, and to provide a textual meaning for non-text-based Web resources. Attempts to capture these meanings have become the raison d'être of Web-based descriptive metadata. "If there is a solution to the problem of resource discovery on the Web," as one metadata introduction explains, "it must

surely be based on a distributed metadata catalog model" (Gill, 2001; p.7).

In this sense, metadata functions in a manner similar to a card or record in a library catalogue, providing controlled and structured descriptions of resources through searchable "access points" such as title, author, date, location, description and subject. But unlike library catalogue records, a metadata record can either be located separately from the resource it describes, or be embedded or packaged with it. Also, many visualize this metadata as being distributed across the Web, rather than collected in a single catalog.

What is much less often mentioned in discussions of educational metadata, is that this approach to information management effectively inserts a layer of human intervention and interpretation into the Web-based search and retrieval processes. This is a layer where words are emphatically not just understood as "formal squiggles" that match other formal character strings, but as actual bearers of meaning and significance, a significance that extends beyond occurrence algorithms for relevance and other systematic means of ascribing search results with meaning. When searching metadata — whether it is distributed across the Web or collected in a conventional library catalogue — documents and other resources are seen as relevant to a given search not because of the letter or word combinations they contain. Instead, their value and purpose is assessed only according to the way they are represented and interpreted in the metadata that describes them. Documents in this new vision of the Web would not be determined as relevant to a specific subject or category not as a direct result of their contents, but because of how a metadata creator or indexer has understood their relevance.

This shift in emphasis implied in the application of metadata can be understood in terms of a shift from data manipulation and processing to the creation, interpretation and assessment of information or knowledge. Data, information and knowledge are often conceived of as forming a hierarchy, where each successive layer is differentiated from the last through a process of interpretation and mediation. Merriam-Webster (2001) defines data as "information in numerical form that can be digitally transmitted or processed", i.e., as pure, un-interpreted fact, perception, signal or message. Information, as characterized by management guru Peter Drucker, is data that is "endowed with relevance or purpose" (1988; p. 4). Information, in other words, forms the contents of the data signal or message. Knowledge, finally, is defined in terms that associate it even more closely with human understanding, intention and purpose: as 1) "the fact or condition of knowing something with familiarity gained through experience or association"; 2) "acquaintance with or understanding of a science, art, or technique"; or 3) "the fact or condition of being aware of something" (Merriam-Webster, 2001).

In this context, to characterize an interpretation of the meaning or purpose of a digital resource as "metadata" seems misleading. For in order to be "about" something — or to deserve the prefix "meta"— data needs to be endowed with purpose and relevance. On their own, the 1's and O's of a digital description (or any other digital resource) are not "about" anything in particular; to acquire relevance or "aboutness," this raw datum needs to be transformed into interpreted information or knowledge. In this sense, metadata as data that has significance or is "about something" is a contradiction in terms.

In addition, the resource or learning object being described would also most likely not simply be raw data. In order to have value for learners and for this value to be indicated in a metadata description, the resource itself also has to have the status of information. Consequently, a term like "metainformation" — interpreted information describing a resource that is similarly understood and interpreted according to its educational potential — might be more appropriate and less misleading. Only by clearly indicating and understanding how metadata is to function as a rich and complex description of resource meanings, purposes and contexts, will it be possible to realize the potential of specifications, profiles and technologies developed for metadata.

Educational Metadata: The Ieee Lom Standard

The importance of differentiating the management of raw data from the complexities of interpretive knowledge is well illustrated by some of the challenges presented by educational metadata implementation. The only officially approved standard for learning object metadata, "IEEE Learning Object Metadata" (or the "LOM"), provides a case in point. The IEEE (the Institute of Electrical and Electronics Engineers) provides a wide variety of e-learning and other specifications, almost all of them very particularly focused on issues of technical interoperability, such as data interchange, computer managed instruction, and platform/media profiles. The IEEE LOM standard has received widespread support from major players in the educational technology industry. The metadata specification in particular is being used or referenced in international repository efforts like MERLOT (merlot.org) and ARIADNE (ariadne.unil.ch), as well as in the U.S. Department of Defense SCORM initiative (www.adlnet.org).

The LOM standard ambitiously defines approximately 80 separate aspects or "elements" for the description and management of learning resources. These elements include generic informational items such as title, author, description, and keywords, technical aspects such as file size and type, and also include educational and interpretive aspects like "typical learning time" or "educational context".

However, the sheer number and variety of elements in this metadata specification has created widely recognized difficulties for its implementers. A consortium of stakeholders in e-learning standards development, the IMS, describes the situation as follows:

Many vendors [have] expressed little or no interest in developing products that [are] required to support a set of meta-data with over 80 elements. æMost have existing products that they hope could support a minimum baseline of elements that the learning resource community would agree to be essential. They also want to be able to make marketing statements such as "IEEE/IMS meta-data conforming document." (IMS, 2000)

Complicating matters, the IEEE LOM documentation provides only very brief and sometimes confusing descriptions of the purpose and character of its numerous metadata elements. For example, the element labeled "5.2 Learning Resource Type" is described in the LOM documentation only as the "specific kind of learning resource", and is associated with the following recommended vocabulary terms: exercise, simulation, questionnaire, diagram, figure, graph, index, slide, table, narrative text, exam, experiment, problem statement, self assessment, and lecture (IEEE, 2002). Even a casual glance at this listing reveals terms whose meanings overlap (e.g., diagram, figure, graph), and terms that describe content (table, slide) as opposed to the education application for which this same content can be used (exercise, lecture, exam). For the purposes of clarifying these and other ambiguities related to vocabulary terms, the LOM simply refers implementers to existing practice in general, and to the 1989 edition of the Oxford English Dictionary in particular. The latter source is a historical and etymological reference, and provides dozens of definitions for a term such as "figure," including "a person dressed in character."

The matter of deciding whether to use such elements or vocabulary values and deciphering what their intended purpose might be is no small task. The metadata that many of the LOM elements specifies or represents is not simply a unity of un-interpreted data, that can be simply passed, parsed and processed, as is the case with many of the variables outlined in more technically-oriented IEEE e-learning specifications. Instead, many of the elements in the LOM standard effectively require the intervention of human intelligence, of human

interpretation and understanding. Unfortunately, this is not explicitly recognized in LOM documentation — either in the form of clear and detailed element descriptions and term definitions, or in the form of general guidelines for interoperable interpretation and implementation.

As a result, the tasks of using the LOM metadata element set in a project or indexing task is a complex, resource-intensive undertaking, requiring elements to be chosen, interpreted, used, and then possibly reinterpreted by each group or individual collecting or developing resources. Varying implementations of this element set, moreover, threaten to create problems for the effective searching and exchange of metadata records between projects and jurisdictions.

The Cancore Metadata Profile

The CanCore Learning Object Metadata Application Profile takes as its starting point the explicit recognition of the human intervention and interpretation that separates raw data management from the information or knowledge that can actually be "about" something. The Canadian Core Metadata Application Profile, in short, is a streamlined and thoroughly explicated version of a sub-set of the LOM metadata elements. The CanCore element set is explicitly based on the elements and the hierarchical structure of the LOM standard, but it greatly reduces the complexity and ambiguity of this specification. As an application profile, CanCore represents a "customization" of a "standard" for the specific needs of "particular communities of implementers with common applications requirements" (Lynch, C. A., 1997). In the case of the CanCore application profile, these communities are constituted by public education (both distance and conventional forms) in Canada, including the primary, secondary and tertiary educational sectors. In interpreting and refining the LOM standard for the needs of this community, it is important to note that the CanCore initiative is not creating a new specification or standard. Instead, it is bridging the gap between the generalities and choices left open by the LOM standard and the very specific needs of implementers, projects and indexers.

The CanCore application profile consists of 8 main categories, 15 "placeholder" elements that designate sub-categories, and 36 "active" elements for which data are actively supplied in the process of creating a metadata record. The CanCore profile includes eight of the nine main categories in the LOM standard: General, Lifecycle, Metametadata, Technical, Educational, Rights, Relation and Classification. These categories and the elements contained in each can be briefly described as follows: The first, "General", describes "context-independent features of the learning object", and in CanCore, this category includes seven active elements including title, language, coverage, and an element for full-text description of the resource's content. The second category, "Lifecycle", uses four active elements to describe the circumstances of the object's development, including its developers' (and other contributors') names, the date of its creation, as well as publication and version information. Elements in the "Metametadata" category describe the metadata record itself, identifying those who developed or validated the record, the natural language of the record, and the date it was created or validated. The "Technical" and "Educational" categories use 5 elements each to designate (among other things) the object's technical format, size, location and requirements, as well as its educational type, context, and age range. (A simplified vocabulary for the educational context suitable for the Canadian context is also provided by CanCore.) The "Rights" and the "Relations" categories employ three active elements each to describe terms and conditions for the use of the learning object, and its relation to other resources. "Classification", the last category, consists of four active elements that can be adapted to the use of almost any classification purpose or vocabulary, regardless of the type or the aspect of the object it might

describe. As one suggested application of this "catch-all" category, the CanCore profile provides a classification and vocabulary for granularity (or pedagogic type) to designate the object as a "program", "course", "unit", "lesson" or "component". (See www.cancore.ca/schema.html for more information about the metadata elements included in this protocol.)

Table 1. Overview of CanCore Elements

1 general	3 metametadata	4.6 otherplatform-requirements	6.3 description
1.1 identifier	3.1 identifier	4.7 duration	7 relation
1.2 title	3.2 catalogentry	5 educational	7.1 kind
1.3 catalogentry	3.2.1 catalog	5.2 learning-resourcetype	7.2 resource
1.3.1 catalog	3.2.2 entry	5.5 intendedenduser-role	7.2.1 identifier
1.3.2 entry	3.3 contribute	5.6 context	7.2.3 catalogentry
1.4 language	3.3.1 role	5.7 typicalagerange	7.2.3.1 catalog
1.5 description	3.3.2 entity	5.11 language	7.2.3.2 entry
1.7 coverage	3.3.3 date	6 rights	9 classification
2 lifecycle	3.4 metadatascheme	6.1 cost	9.1 purpose
2.1 version	3.5 language	6.2 copyrightand otherrestrictions	9.2 taxonpath
2.3 contribute	4 technical		9.2.1 source
2.3.1 role	4.1 format		9.2.2 taxon
2.3.2 entity	4.2 size		9.2.2.2 entry
2.3.3 date	4.3 location		9.4 keyword

By providing these types of simplifications and recommendations, the CanCore profile has already realized considerable economies of scale for its users. It has worked to prevent redundant or inconsistent interpretations, and to ensure that educational metadata and resources can be shared among their users as effectively as possible with others across the country and with LOM implementations internationally. As of the writing of this paper, the CanCore profile is serving these ends in the context of a number of national projects. These include the CANARIE-supported "Portal for Online Objects for Learning" led by TeleLearning (POOL; www.newmic.com/pool/) and the "Broadband Enabled Lifelong Learning Environment" (BELLE; www.netera.ca/belle/) projects, as well as the Learn Alberta project (http://www.learnalberta.ca/). CanCore also supports the Campus Alberta Repository of Education Objects (www.careo.org), a repository initiative with a focus on post-secondary education in Alberta. Funding and support for the development of the CanCore Protocol has been provided through these projects, as well as by the Netera Alliance, TeleCampus.edu, the Electronic Text Centre at the University of New Brunswick, the University of Alberta and Athabasca University.

To ensure that further coordination and economies of scale can be realized, the CanCore initiative has more recently developed a comprehensive guidelines document. This extensive, 170-page document builds on two existing implementation documents or application profiles, and provides explanations of the meaning and use of each element included in the CanCore profile. In addition, it provides refinements of vocabulary definitions, discussions of best practice, multiple XML-encoded and plain language examples, and detailed technical implementation notes. The guidelines document is available for download at no charge from the CanCore Website (www.cancore.org/documents.html). Finally, the CanCore initiative is planning to provide training and further vocabulary recommendations, as well as other support and coordination services. Together, these products and services promise to realize further economies of scale, and to make possible an even higher level of interoperability among vendors, developers and repository efforts. In this way, the CanCore initiative has added and will continue to add considerable value to the IEEE Learning Object Metadata standard.

Repository Architecture & Metadata

To understand how the CanCore initiative and profile fits into the larger picture of learning objects and repositories, it is important to envision how metadata records developed in accordance with CanCore guidelines will be integrated into a number of repository architectures and distribution models. These architectures and models fall into two general classes: centralized architecture and commercial or for profit distribution models on the one hand, and distributed architecture, and public or open distribution, on the other.

Typically, a centralized repository architecture combines both learning resources and the metadata describing them in the same location - often on the same server, but at least subject to the same, central administration. This centralization provides the potential for exercising significant control over the distribution and availability of the learning objects housed in such a repository. As a result, a centralized architecture would likely be most suitable for vendors of commercial content, such as textbook publishers who wish to control access to their learning objects. It is important to note, however, that this control and management would take place in a layer that is separate from the metadata. Such control and management "layers" or systems are available from third party providers. However, the CanCore Application Profile _ like the LOM specification — provides informational elements that indicate the presence of such rights controls, and that describe commercial status and the pricing for the object under consideration.

For the CanCore profile to facilitate effective search and retrieval in such a scenario, it is essential that the metadata records describing the learning objects be themselves free from any access or management control systems. These metadata records — unlike the assets themselves — must be made openly available in an interchangeable format to other repositories or collections of similar metadata records. Such a format for interchange is provided by XML (eXtensible Markup Language). The CanCore intiative will be providing LOM-compliant specifications for creating and verifying such interchangeable XML records. These interchangeable descriptions can then be seen to serve as digital "calling cards" for the commercial assets they describe, exposing these products to wider markets and the users of any number of learning object repositories.

A distributed architecture provides an alternate method of organizing learning objects that seems especially well suited to the interests of public sector repository projects, and to larger collaborative efforts. Just as commercial content providers have been deploying distribution control mechanisms for their products, public sector educators, taking their cue from the academic research and open source communities, have been implementing practices for openly sharing and developing high-quality educational resources. For example, the MERLOT project (www.merlot.org) has adapted the practices of peer review of faculty research as the basis for encouraging the development, evaluation and reuse of learning objects for postsecondary teaching and learning. At the same time, in its highly publicized OpenCourseWare initiative, the Massachusetts Institute of Technology has taken the open source software development model as the basis for making many of its own course materials freely available on the Web (MIT, 2001).

Germaine to these types of open, collaborative approaches would be a repository architecture that is itself distributed and cross-institutional. Just as the efforts invested in the development, adaptation and review of the objects would be scattered across many institutions, so the objects themselves would be distributed across the Web, in locations provided most likely by their developers or supporters, or their sponsoring institutions. The metadata describing these resources, however, would remain centralized for fast and effective searching, sharing and control. This would allow the resources to be updated and otherwise maintained by their owners

while at the same time, permitting the metadata describing them to be shared and searched. This type of mechanism, perhaps more accurately described as a web gateway than repository (see the DESIRE Information Gateways Handbook; http://www.desire.org/handbook/), is being developed by the BELLE and POOL projects using the CanCore profile.

As in the case of a centralized architecture, in a distributed repository model, metadata records describing the resources are in every instance freely available for interchange. Whereas the records created in a centralized, commercial model would typically be generated by the agency owning the assets, the records in a distributed repository would likely be created by those individuals who develop or contribute objects. Because these individuals would not likely be trained in classification or indexing, support documents and metadata quality control procedures would have to be provided. Both the processes of quality control and the creation of records by non-specialists would be facilitated by the relatively small number and simplicity of elements available in the CanCore profile, and by the explication and interpretation that CanCore provides. Again, these metadata records would serve as a type of "calling card" for the resource, providing the original developer with favourable exposure to a community of peers, and allowing any resource to be used, reused and potentially improved.

The free interchange of these records among commercial and public repositories opens the way for a third type of repository approach — one that is derivative of the distributed model outlined above. This type of repository can be characterized broadly as "a repository of repositories" where metadata is aggregated from other repositories and repository systems. In this case, these records would refer to resources available in any number of other centralized or distributed collections, and would provide the user with a single place for searching and accessing these resources, regardless of their point of origin. When a user searches such a central metadata store, she would be able to get simplified access to resources that are otherwise scattered across the Web, and in proprietary and public databases that may have a wide variety of access protocols and paths. When combined with customizable community-building features and other services, the aggregation of material presented in such a context would truly earn it the title of "portal": an effective starting point for educational users of the Web.
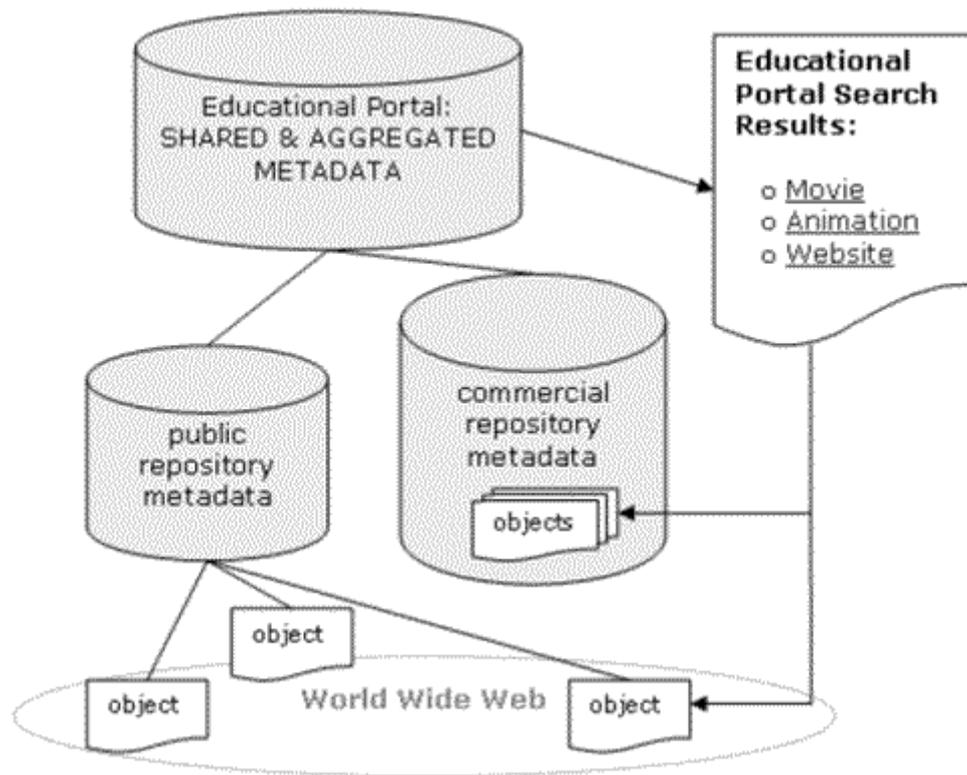
Figure 1. Repository Types and their interrelation

A diagram schematizing these repository types and their possible interrelation is provided above. An effective way of understanding the operation and efficacy of these types of repositories is provided by practices that are commonplace in the library world, namely, in the form of a "union" catalogue. A union or consortium library catalogue represents an aggregation of interchangeable records describing resources that are available across an entire library system or consortium. Such a catalogue would be the functional equivalent of a "portal" or "repository of repositories" described above sharing metadata records across repositories, with the resources they describe being seamlessly available to the user.

Conclusion

A shift in how we understand knowledge production promises also to change how we can search, find, exchange, describe and use digital resources. As rigid knowledge categories are deconstructed, the way we understand digital assets and data are giving way to more dynamic, fluid and more useful way of engaging and regarding these assets. Interoperable metadata, consistently and systematically implemented, is one of the lynchpins in achieving a vision of easy access to shared and reusable learning objects. The CanCore profile promises to provide this crucial functional element for those following it in their implementations.

As has hopefully become apparent in this paper, the interest of CanCore is not to compete with or supplant other standards efforts. Instead, its goal is to add value strategically to the widely accepted but difficult metadata model standardized by the IEEE. The CanCore initiative adds value to this model by simplifying and refining it, by developing vocabularies suitable to Canadian educational sectors, and also by providing interpretation and support in the form of guidelines documents and other services. In this way, the CanCore initiative is attempting to position itself not as an "authoritative" metadata solution, but rather one that is the easiest, and the most strategically attractive to implementers, administrators and educators alike.

References

Dreyfus, H. (2001). On the Internet. London: Routledge.

Drucker, P. F. (1988, January-February). The Coming of the New Organization. Harvard Business Review, 4-11.

Gill, T. (2001). Metadata and the World Wide Web. Retrieved on November 14, 2001 from http://www.getty.edu/research/institute/standards/intrometadata/pdf/gill.pdf

IEEE (2002). Draft Standard for Learning Object Metadata. Retrieved March 20, 2002 from http://ltsc.ieee.org/doc/wg12/LOM_WD6_4.pdf

IMS Global Learning Consortium. (2000). IMS Learning Resource Meta-data Best Practices and Implementation Guide. Retrieved [Month, Day, Year] from http://www.imsproject.com/metadata/mdbestv1p1.html.

IMS Global Learning Consortium (2001a). IMS Learning Resource Meta-data Best Practice and Implementation Guide. Retrieved March 20, 2002 from http://www.imsproject.org/metadata/ims_md_bestv1p2.html.

IMS Global Learning Consortium. (2001b). IMS Learning Resource Meta-data Information Model. Retrieved March 20, 2002 from http://www.imsproject.org/metadata/ims_md_infov1p2.html.

Kant, I. (1969). Critique of Pure Reason. Smith, N.K. (trans.). Bedford: St.Martin's

Lynch, C. A. (1997, April). The Z39.50 Information Retrieval Standard. Part I: A Strategic View of Its Past, Present and Future. D-Lib Magazine . Retrieved March 20, 2002 from http://www.dlib.org/dlib/april97/04lynch.html

McLuhan, M. (1964). Undestanding Media: The Extensions of Man . New York: McGraw-Hill.

Manovich, L. (2001). The language of new media. Cambridge, MA:MIT Press.

Merriam-Webster's Collegiate Dictionary (2001). Retrieved on November 14, 2001 from http://www.m-w.com/

MIT News Office (2001). MIT to make nearly all course materials available free on the World Wide Web. Retrieved March 20, 2002 from http://web.mit.edu/newsoffice/nr/2001/ocw.html.

Roschelle, J. K. J. (1996). Educational Software Architecture and Systemic Impact: the Promise of Component Software. Journal of Educational Computing Research, 14 (3), 217-228.

Wieseler, W, J. Katzman, J. Larsen, & J. Caton (1999). RIO: A Standards-based Approach for Reusable Information Objects. Retrieved March 20, 2002 from http://www.cisco.com/warp/public/10/wwtraining/elearning/learn/whitepaper_docs/rlo_strategy_v3-1.pdf.

——————————————————