Citation: Qin, J. & N. Hernández. (2006). Building interoperable vocabulary and structures for learning objects. *Journal of the American Society for Information Science and Technology*, 57(2): 280-292.

Building Interoperable Vocabulary and Structures for Learning Objects

Jian Qin ♣
School of Information Studies, Syracuse University, Syracuse, NY 13244
Email: jqin@syr.edu

Naybell Hernández
School of Information Studies, Syracuse University, Syracuse, NY 13244
nhernand@syr.edu

Abstract

The structural, functional, and production views on learning objects influence metadata structure and vocabulary. We drew on these views and conducted a literature review and in-depth analysis of 14 learning objects and over 500 components in these learning objects to model the knowledge framework for a learning object ontology. The learning object ontology reported in this paper consists of 8 top-level classes, 28 classes at the second level, and 34 at the third level. Except class *Learning object*, all other classes have the three properties of preferred term, related term, and synonym. To validate the ontology, we conducted a query log analysis that focused on discovering what terms users have used at both conceptual and word levels. The findings show that the main classes in the ontology are either conceptually or linguistically similar to the top terms in the query log data. We built an Exercise Editor as an informal experiment to test its ability to be adopted in authoring tools. The main contribution of this project is in the framework for the learning object domain and methodology used to develop and validate an ontology.

1. Introduction

Representation of learning objects involves both content and metadata. Like many other digital objects, learning objects have structures filled with content components, such as learning objectives, procedures, concepts, practice, and assessment. They also need metadata to describe who the creators are, what the learning objects are about, and who has what right over the learning objects. The metadata practice is typically a distributed effort in today's network environment, which results in two contradictory forces in the creation and use of learning objects. On the one hand, creators of learning objects do not use a controlled vocabulary for labeling the content components and structures. As a result, learning objects come in a wide

♣ To whom all correspondence should be addressed.

variety of structures with various labels even for the same type of objects in the same subject area. This makes metadata representation extremely challenging. On the other hand, learning objects need metadata in order to be found and selected by users. Due to the unstructured content and inconsistent naming of content components, automatic metadata generation is difficult, if not impossible, especially for finer metadata representation.

The issues of vocabulary in learning objects have attracted researchers' attention in recent years. Developers of learning object authoring tools have incorporated structured components such as type of learning object, text area, and media component (Rice University, 2003; Trivantis Corporation, 2003). In the metadata community, educational metadata schemes such as IEEE Learning Object Metadata (LOM) and the Gateway to Educational Materials (GEM) metadata set have been widely adopted by educational digital library projects with local modifications. Although the Open Archive Initiative (OAI) provides a venue for the interoperability of metadata across digital libraries, there are few similar efforts in the learning object design and creation community. While the instructional design and digital library communities actively advocate for the creation of sharable, reusable, and interoperable learning objects, the vocabulary work has lagged behind.

The need for a controlled vocabulary for educational objects and digital libraries in general has caught the attention of researchers, including the National Science Digital Library (NSDL) Vocabulary Workshop (Hillman, 2004) among others. The consensus is that controlled vocabulary is fundamental to the discovery and interoperability of metadata and the objects it describes. Questions remain, however, on two fronts: What concepts should be included in controlled vocabulary for digital learning objects, and to what level of detail should we define a concept so that digital library personnel can use it as either an element name or value for an element? These two questions reflect the problems identified in the workshop's summary document (Sutton, 2004): many metadata creators do not use controlled vocabularies; when they do use such vocabularies, inappropriate ways of encoding them often lead to the loss of such enriched semantics. While many factors may contribute to the problems, the lack of a controlled vocabulary that users understand and that meets their representation and search needs should probably take most of the blame.

Library cataloging and indexing services have long used thesauri (e.g., ERIC Thesaurus) to represent the intellectual content of information objects. However, representing digital objects needs far more specific terms than the ones available in traditional thesauri (Qin & Godby, 2003). In the digital environment, terms in a controlled vocabulary form a knowledge model for a subject domain and are expected to function as labels for and relations between categories of data. This means that, ideally, the knowledge model will eventually be converted into a data model for the implementation stage. Thesauri do not have the mechanisms for shaping the data model as the knowledge model is being defined. On the contrary, ontologies as a new form of knowledge modeling fit well into this role, since they can model not only the metadata elements but also define the vocabulary for both elements and element values. What makes ontologies more advantageous in the digital environment is their ability to establish relationships between data/concept properties by using an object-oriented approach.

This paper proposes a guiding framework for representing the conceptual and application areas of learning objects. Through a review the facets of learning objects and by examining

current metadata standards related to education, the authors discuss the limitations of metadata standards in representing structural components in learning objects, which justifies for the need for an ontology. We will describe our approach in constructing the learning object ontology through query log mining. The following sections are divided into: 1) related research; 2) issues in learning object metadata; 3) a framework for the learning object domain; 4) methodology; 5) constructing a learning object ontology that includes four subsections: concept classes and properties, concept relationships, validation, and an example of ontology application; and 6) discussions and conclusion.

2. Related Research

Learning objects[1] in the context of this paper refer to digital materials created for learning or educational purposes. The creation and use of learning objects involves a broad base of participating communities. Each community defines the concept of learning objects in their own context and uses a set of terminology to define their view on learning objects. Studying these views will help us understand the differences and relations between them and gain insights into building an educational ontology. We summarize the research on learning objects from three different views in the following sub-sections.

2.1 The Structural View

The structural view reflects the way that educational institutions structure their academic programs. As shown in FIG. 1, a curriculum consists of courses, a course contains lessons, a lesson includes sections, and so forth. The IEEE LOM working group of Learning Technology Standards Committee (LTSC) maintains that a learning object may be a course, or one of its assignable units such as a lesson, section, and component object (LTSC, 2001). The structural view serves the need for academic programs to deliver systematic knowledge and training in a discipline or subject domain.

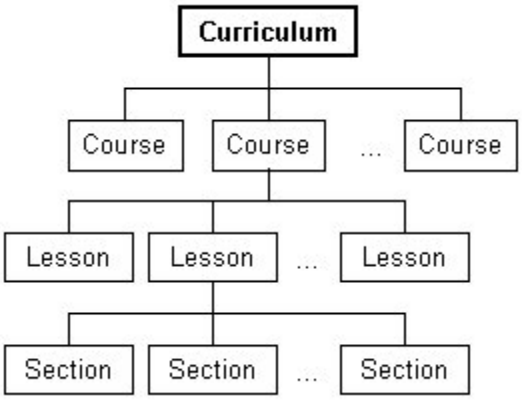


FIG. 1. Structural facet of learning objects

[1] The instructional design and training communities often use another term “learning objects” to refer to those specifically created for learning purposes. In the context of this paper, we use “learning objects” to include learning objects and other educational materials.

2.2 The Functional View

The functional view of learning objects is closely related to instructional design and technology. Rather than building learning objects as courses, the functional view treats learning objects in the context of "unit of study." Koper (2002) proposes an integrated model of learning object types as shown in FIG. 2. In this model, each unit of study plays the role of a framework and encapsulates various types of learning objects such as learning objective, prerequisite, role (learner and staff), activity, and environment. Each type may contain subtypes. For example, the Environment type has eight subtypes, each of which performs a different function.

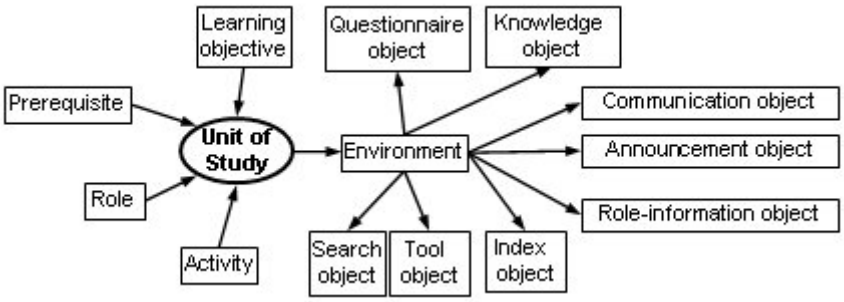


FIG. 2. The functional facet of learning objects (source: Koper, 2002)

The concept "unit of study" is more prevalent in industrial e-learning than in academic studies. Some of the Fortune 500 companies, e.g., Cisco and Honeywell, started developing learning objects for e-learning in the early 1990s (Barron, 2000). Barritt (2002) argues that a learning object is based on a single learning or performance objective that is presented through content, practice, and assessment items. Text and media elements contained in these items form the building blocks of a learning object. While these elements may be reused to develop or assemble new learning objects, a learning object may also be reused in a lesson, module, unit, course, and then curriculum. Cisco differentiates learning objects as "Reusable Learning Objects (RLOs)" and "Reusable Information Objects (RIOs)." RIOs include template content types such as concept, fact, procedure, process, or principle that respond to a single learning objective. A lesson or RLO combines five to nine RIOs with an overview and summary (Barritt, 2002; Cisco Systems, 2003).

Another functional view divides learning objects into instruction, collaboration, practice, and assessment objects (ASTD & SmartForce, 2002). Lessons, workshops, seminars, articles, white papers, and case studies are examples of instruction objects. Collaboration objects include mentored exercises, chats, discussion boards, and online meetings. Practice objects include all kinds of simulations such as role-play, software/hardware, coding, and conceptual simulations. Assessment objects consist of various tests such as pre-assessments proficiency assessments, performance tests, and certification prep tests. Similar to this classification, the instructional design community holds that a learning object has to have concept, practice, and assessment to form its entirety in order to achieve a learning goal. Lack of any of these three components would make a learning object incomplete (Acker, 2002).

2.3 The Production View

The production view covers the form or format aspect of learning objects, including whether or not there are any component objects in a learning object, how they are produced (individual or aggregated), and in what form they will be delivered and used. Wiley (2000) offers his taxonomy of learning objects based on the characteristics summarized from how learning objects are physically produced—dynamically assembled from multiple smaller media objects or otherwise static objects (FIG. 3). The column on the left side of FIG. 3 is a list of production attributes summarized by Wiley (2000), characterizing learning objects from fundamental to generative-instructional. The column on the right contains attributes of use and reuse that are applicable to make further categorization of learning objects by each of the production attributes.

The production view also includes those by media type and format. Media types include, for example, simulation applet, interactive illustration, animation, streaming audio/video, and interactive map. Media formats are the ones defined in template lists of object types as seen in metadata standards, e.g., Dublin Core's type element (DCMI, 2002) and a similar element in the IEEE Learning Object Metadata (LOM) (LTSC, 2002). Learning objects can also be classified by product form, e.g., lecture notes, tutorial, and bibliography.

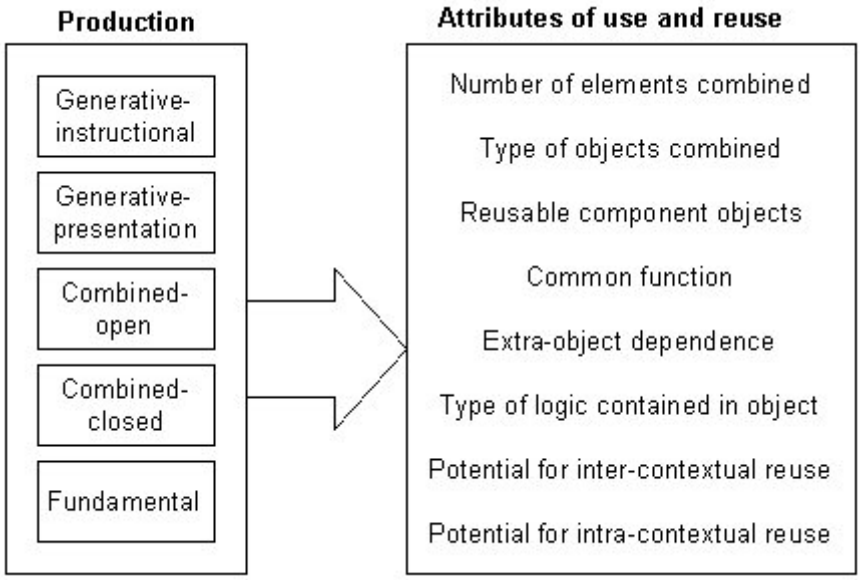


FIG. 3. The production view of learning objects (based on Wiley, 2000)

2.4 Metadata for Learning Objects

The views on learning objects summarized above influence the metadata representation in different ways and not all the views receive equal attention. The standards activities led by the IEEE Learning Technology Standards Committee (LTSC) demonstrate the mainstream in learning object metadata, which cover areas of learning technology, digital rights, metadata, and structured definitions related to instruction. The Learning Object Metadata (LOM) from LTSC prescribes the metadata elements in nine areas to represent a learning object: general, life cycle, meta-metadata, technical, educational, rights, relation, annotation, and classification (LTSC,

2004a). This standard bears a strong functional view as evidenced by the purpose statement of LTSC LOM Working Group:

- "To enable learners or instructors to search, evaluate, acquire, and utilize Learning Objects.
- To enable the sharing and exchange of Learning Objects across any technology supported learning systems.
- To enable the development of learning objects in units that can be combined and decomposed in meaningful ways.
- To enable computer agents to automatically and dynamically compose personalized lessons for an individual learner.
- To compliment [*sic*] the direct work on standards that are focused on enabling multiple Learning Objects to work together within a open distributed learning environment. …" (LTSC, 2004b)

The functional view of learning objects is also reflected in another line of work that deals with the issues of pedagogy. Metadata groups have been exploring ways of representing pedagogical aspects in metadata standards. The Gateway to Educational Materials (GEM) is one of the early metadata standards that include the element Pedagogy (GEM, 2002). Dublin Core Metadata Element Set (DC) Education Working Group recently released a proposal of a new element "Instructional method" (DC-ED, 2004). The metadata community is debating what vocabularies should go into the pedagogy/instructional method element while input from the instructional community raises questions on the choice of element name and values and how these terms can accommodate various learning theories (Mason, 2004).

There is a general sentiment across metadata and instructional communities that a vocabulary is needed to achieve the objectives as stated in the purpose document from LTSC. Researchers have explored various approaches in developing metadata vocabulary, including building ontologies (Forte, et al, 1999; Qin & Paling, 2001; Qin & Godby, 2003; Greenberg et al, 2003) and data collection techniques (Tennis, 2003). However, how will we obtain the vocabulary and validate it? To address this question, let us first take a closer examination of metadata standards.

3. Issues in Learning Object Metadata Standards

Researchers often refer to vocabularies used in metadata standards as ontologies (Greenberg et al, 2003), because they define not only metadata element names but also provide value space for the elements. In the educational metadata field, LOM is a standard that many metadata application profiles follow, which in turn is compatible with DC. Table 1 summarizes the total number of elements and educational elements in five metadata standards and application profiles. It is worth noting that, while LOM has 90 elements, other application profiles, based on either DC or LOM, have many fewer elements. The educational elements for each metadata scheme also vary according to which base scheme they use. Our further study of these schemes raised several issues.

Table 1. Major learning object metadata standards and application profiles

Standard	Base scheme	Number of elements	Educational elements with value space
Education Network Australia (EdNA)	Dublin Core	15+8	Type, curriculum, document, event, audience, spatial
Gateway to Educational Materials (GEM)	Dublin Core	15+8	Audience, format, grade, language, pedagogy, object type, subject
IEEE Learning Object Metadata (LOM)	IEEE LOM	90	Interactivity type, learning object type, interactivity level, semantic density, intended end-user role, context, difficulty, relation kind, purpose
CanCore	IEEE LOM	30*	Interactivity type, learning object type, semantic density, intended end-user role, context
UK LOM Core	IEEE LOM	46*	Interactivity type, learning object type, interactivity level, semantic density, intended end-user role, context, difficulty, relation kind, purpose

*: Not including the 2nd level elements.

The first issue is that the design paradigm of metadata standards essentially remains the same as that of library cataloging. Traditionally, librarians create cataloging records manually describing, indexing, and classifying, because the physical materials are not directly processable by computer. In this process, the record creation is separate from the material content creation. Digital learning objects, on the contrary, are processable directly by computer. This creates a necessary condition for processing digital learning objects directly and generating metadata records with little or no manual cataloging. However, elements in metadata standards have inherited much of the structure and semantics used in traditional cataloging, which are more suited to a human cataloger entering data for the elements than to computer programs processing and generating metadata. Researchers have experimented with the natural language processing (NLP) approach to generating metadata automatically (Paik, et al, 2001; Liddy et al, 2002). This approach needs sophisticated programs to analyze documents and insert linguistic and semantic markups between words and phrases in documents for automatic metadata extraction. While the NLP approach achieved comparable performance to manually created metadata in both Liddy (2002) and Paik's (2001) experiments, it is uncertain if the same would be true in much larger collections and if the process would be economic.

Adding to the traditional cataloging paradigm, another related issue is the lack of suitable, specific vocabularies that automatic metadata generation needs. As studies have found, traditional vocabularies and knowledge structures such as the ERIC Thesaurus and Library of Congress Subject Headings (LCSH) use very broad terms to describe the content and physical attributes of library materials, which are unsuitable for digital object representation (Forte et al, 1999; Qin & Godby, 2003). The technology trend now is to use markup schemas to create structured content. This requires vocabularies as the underpinning semantic infrastructure in order to be successful. Although there have been vocabulary building efforts for educational

metadata (DC-ED working group and the NSDL metadata management group, for example), they are focused on the cataloging aspect rather than on a broader base such as for structured content in digital objects.

Finally, we know very little so far about the vocabularies that users use in searching educational digital libraries. There has also been little research in validating the vocabularies used in metadata. The lack of this knowledge is hindering the advances of digital object representation in breaking the bottleneck problem of vocabulary. As more and more digital objects in education and other domains bear structured content, the demand for vocabularies and conceptual structures in the form of ontologies will increase and become urgent.

The core of these issues falls in one key research question for this study: How should we build a learning object vocabulary and if we build one, how can we validate it? In addressing this question, we proposed a conceptual framework for the learning object domain and developed an ontology based on the framework. We then used the query log mining results from an educational digital library to validate the ontology. The following sections will 1) explain the conceptual framework, 2) describe the methodology we used to create the ontology, 3) present the structure and vocabulary in the ontology, and 4) discuss the validating result.

4. A Framework for the Learning Object Domain

The learning object domain traverses a number of relevant fields, including instructional design and learning theory, information science, and technology. Clancy (1997) proposed a conceptualization model in which he summarizes the relationships between situated cognition and human knowledge, practice, and representational artifacts. Applying Clancy's conceptualization model to a learning situation, the knowledge would include a learner's conception of his or her activities; his/her practice would be the ways he/she learns, reads, discusses with others, and writes; the description would include the papers and emails he/she writes, comments posted on the bulletin boards, etc. The representation of learning objects in this sense is a process of capturing the characteristics of knowledge, practice, and description through standard vocabularies, rules, and associations to facilitate learner's coordination, formalization, and interpretation activities. We modified Clancy's model to formulate a conceptualization model of the learning process (FIG. 4).

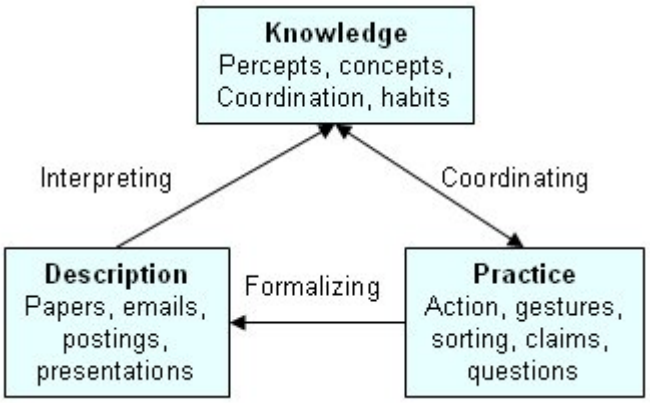


FIG. 4. Simplified view distinguishing between conceptualization (knowledge), action in the world (practice), and outcomes of learning (descriptions), modified based on Clancey, 1997.

The conceptualization model integrates the three views on learning objects discussed in the literature review section. It also raises the expectation for metadata to absorb and refine the learning practices in order to facilitate the interpretation of knowledge in learners' and instructors' activities. We developed a framework to operationalize the model (see Table 2). As discussed in Section 2, learning objects have structural, functional, and production facets and their creation and use involves learning theory, instructional design, disciplinary knowledge, and enabling technologies (e.g., computer science, linguistics, and information technology). Each of the use aspects in Table 2 interacts with those in content aspects in different ways. For example, a learning object may contain different types and levels of learning content in a discipline or subject; it may be in text or mixed with multimedia and used for reading or practice. On the content dimension, the learning object has a set of learning objectives to accomplish, is suitable for one or more learning models; all the functions – content, practice, and assessment – are enabled by an array of technologies, among which metadata and ontologies provide a semantic infrastructure for both content and use aspects. This framework has an emphasis on use—disciplinary knowledge vs. application, and learning outcomes—learners' competencies in analysis, comprehension, evaluation, synthesis, and application (Bloom, 1956).

Table 2. A framework of domain knowledge in learning objects

<table>
<tr><td colspan="2"></td><td colspan="3">Use aspects</td></tr>
<tr><td colspan="2"></td><td>Content</td><td>Presentation</td><td>Application</td></tr>
<tr><td rowspan="4">Content aspects</td><td>Disciplinary knowledge</td><td>Types, levels</td><td>Multimedia, text, mixed</td><td>Reading, playing, listening, practice</td></tr>
<tr><td>Learning theory & instructional design</td><td>Objectives, learning models, contexts</td><td colspan="2">Structure, naming, relationship, pedagogy</td><td>Comprehension, analysis, synthesis, evaluation, application</td></tr>
<tr><td rowspan="2">Enabling technologies</td><td colspan="2">Database, XML, authoring tools</td><td colspan="2">Graphic user interface, tools for annotation & recommendation</td></tr>
<tr><td colspan="4">Metadata, ontologies, repositories of learning objects</td></tr>
</table>

This framework suggests that the representation of learning objects cover content, presentation, and application not only from the disciplinary knowledge perspective, but also from the learning theory and instructional design perspective. To accomplish this multidimensional representation, the key is establishing cross-relationships between concepts involved in the framework. For instance, when a learner is reading a text about a subject, relevant practice and assessment materials are present in the context; or when an instructor is looking for a learning object, information is also provided about the learning models and pedagogical methods. While building cross-relationships between concepts is not new in traditional vocabulary construction, innovative ways to achieve it have been developed in the past decades due to information technology advances. The enabling technologies included in the bottom of Table 2 are critical in supporting both content creation and use aspects. An ontology as a conceptual modeling tool and vocabulary-tuned knowledge base provide the underpinning semantic infrastructure for representing learning objects. Under this framework, we created a learning object ontology.

5. Methodology

The goal of the learning object ontology is to provide a conceptual model for capturing vocabularies related to the concepts in the domain as specified in the framework. A major difference between this ontology and traditional thesauri is the greater specificity, because the terms will serve as element names and values in document type definitions and metadata schemas. The ontology was developed in four phases: data collection, concept modeling, ontology validation, and an example application. In the data collection process, we studied literature on learning and instructional design and vocabulary used in educational metadata standards. We also conducted an in-depth study of 14 learning objects and over 500 components within these objects. The component types included interactive illustrations, java applets, tables, data sets, text blocks, and keywords used in these objects. All the data were entered into a database and then output into a statistical program for analysis. The sources and survey provided useful information for the expectations and requirements for learning objects as well as first-hand knowledge of existing learning objects.

In the conceptual modeling phase, we drew concepts and properties from data analysis and used Protégé, an ontology editor from Stanford University (http://protege.stanford.org), to construct the ontology. Two principles were followed whenever possible or applicable:

1) use simple and explicit terms to represent concepts and properties because a term may be used in schemas as element names or values and such elements may be used for structuring content or description metadata; and
2) focus on "representing" rather than describing learning objects, i.e., including structural, pedagogical, and functional concepts that are traditionally excluded by educational metadata standards.

Ontology validation implies either syntax or semantic validation, a process which usually verifies whether the encoding is well-formed or a value is legitimate for a given element. While both syntax and semantic validations are important, we focused only on the semantic validation at this stage of the ontology construction. More specifically, we were mainly concerned with the validity of the ontology—the vocabulary, concept structure, properties, and relationships. Much of the literature on ontology evaluation and validation discusses technical validation (Bench-Capon et al, 1998; Damjanoviæ et al, 2003) by using validation programs, which cannot satisfy our need for evaluating the validity of the developed ontology. Another way to validate an ontology is direct validation that involves using human subjects, usually users of the ontology, to conduct experiments for obtaining their opinions about the appropriateness of an ontology. This type of experiment is often difficult to perform because a large-scale experiment would be prohibitively time-consuming and costly. Even though a small-scale experiment may be feasible, its representativeness and reliability would be questionable.

To avoid these pitfalls, we adopted an indirect validation by mining query logs of an educational digital library—the Gateway to Educational Materials (GEM). The query log data used for validation cover a four-month period in 2003 (February, March, April, and August). This period generated 411,898 queries (the query log mining result will be reported in a separate paper). We wrote SQL programs to dissect the queries in order to obtain a master list of query components. The master list was then cleaned and coded for counting frequencies of terms and

query types. We also conducted an in-depth analysis of the terms covered in our ontology to compare conceptual similarity and term similarity. The result is reported in Section 6.3.

The last phase of this project was to create an exercise authoring tool that embedded the concepts and properties applicable to the tool. This Exercise Editor is only an informal experiment to test whether the concepts and vocabulary defined in the ontology can be used in tool development for learning objects. Further evaluation of the ontology needs to be performed in several areas, which would involve participation by instructors, learners, and developers in different stages of learning objects—creation, use, and tool development. This study was focused on ontology construction only.

6. Constructing a Learning Object Ontology

Ontologies provide semantics for content, presentation, and applications by defining concepts and their relationships in a domain. At various stages of learning object production and use, ontologies can contribute to:

- Modeling the structure of a learning object through classes and class properties;
- Normalizing structural element names through a controlled vocabulary;
- Establishing concept relationships through the hierarchical structure and cross references; and
- Providing consistent semantics and structures for database schemas, search and browsing interfaces, and presentation of content.

6.1 Concept Classes, Properties, and Instances

There are eight top classes in the ontology, including learning object, learning objective, learning content, learning practice, learning context, assessment, pedagogy, and technical attribute. Each class may have subclasses. A class may have its own properties (local properties) or inherit properties from an upper class. The properties of a class serve as a schema for capturing instances of that class.

Learning object: It acts as the container for content, practice, and assessment functions. Hence, the structural view of learning objects fits well into this concept class. Learning objects may be in forms of course, lesson, module, or unit of learning. They also have common attributes such as title, creator, owner, and date of creation. For example, a tutorial is a type of learning object (a subclass of the *Learning object* class); an online tutorial teaching how to catalog with the Connexion system at OCLC Online Computer Library Center would be an instance of a tutorial as well as an instance of *Learning object* since it is a subclass of *Learning object.* The left column in FIG. 5 shows the class structure and the main attributes (slots) of learning objects.

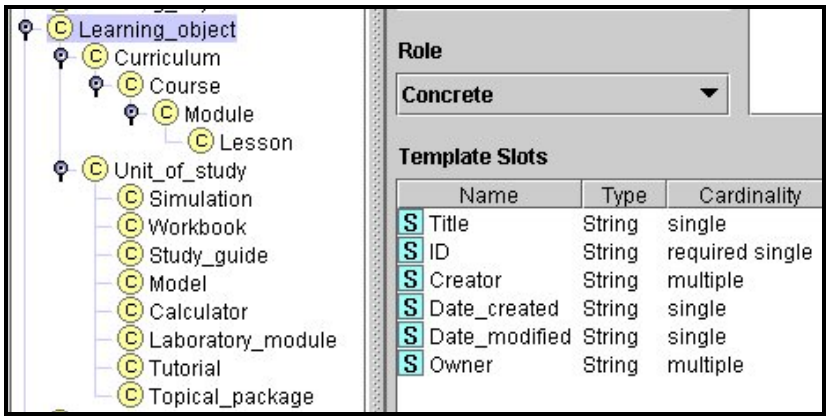


FIG. 5. Subclasses and attributes of the *Learning object* class

Learning objectives: Among the different opinions about how a learning objective statement should be formulated, Bloom's taxonomy of learning objectives has won the broadest acceptance in the instructional design community. It contains six levels of cognitive learning: knowledge (information), comprehension, application, analysis, synthesis, and evaluation (Bloom, 1956). We adopted these terms as subclasses of the *Learning objectives* class.

Learning content: This class includes two aspects: one is the type of content and the other the disciplinary knowledge. We used fact, concept, principle, procedure, and process to distinguish between types of learning content (Cisco Systems, 2003). As for disciplinary knowledge, traditional thesauri have established vocabularies from which we borrow.

Learning practice and assessment: Practice includes problems and exercises that learners can apply the knowledge to solve. Assessment measures the outcomes of learners by using various assessment methods and tools.

Pedagogy: This class has two subclasses: learning model and teaching method. The former stresses learners and their activities, and the latter focuses on the instruction side of learning.

Learning context: Learning contexts may be related to broad environments such as on-the-job training or formal education, and may be labeled with much more specific purposes, e.g. programming skill training and multicultural education.

Technical attributes: System and application types and file attributes are the two main subclasses in this class.

FIG. 6 presents details for each class in the ontology. In designing the class properties, we followed a rule from our experience in developing ontologies, that is, whenever applicable, a class should have the three properties of preferred term, synonym, and related terms. These properties serve as the schema for capturing and mapping vocabularies for that class.

Using the *Pedagogy* class as an example, we demonstrate how a class and its properties support vocabulary capture. FIG. 7 (a) displays the properties associated with *Pedagogy*. Each property includes a name, a type, and constraints, i.e., the number of occurrences of the property value. In FIG. 7 (b), all the instances belong to the *Teaching method* class. Since it is a subclass of *Pedagogy*, it inherits *Pedagogy*'s three properties as the data capture schema (FIG. 7 (c)).

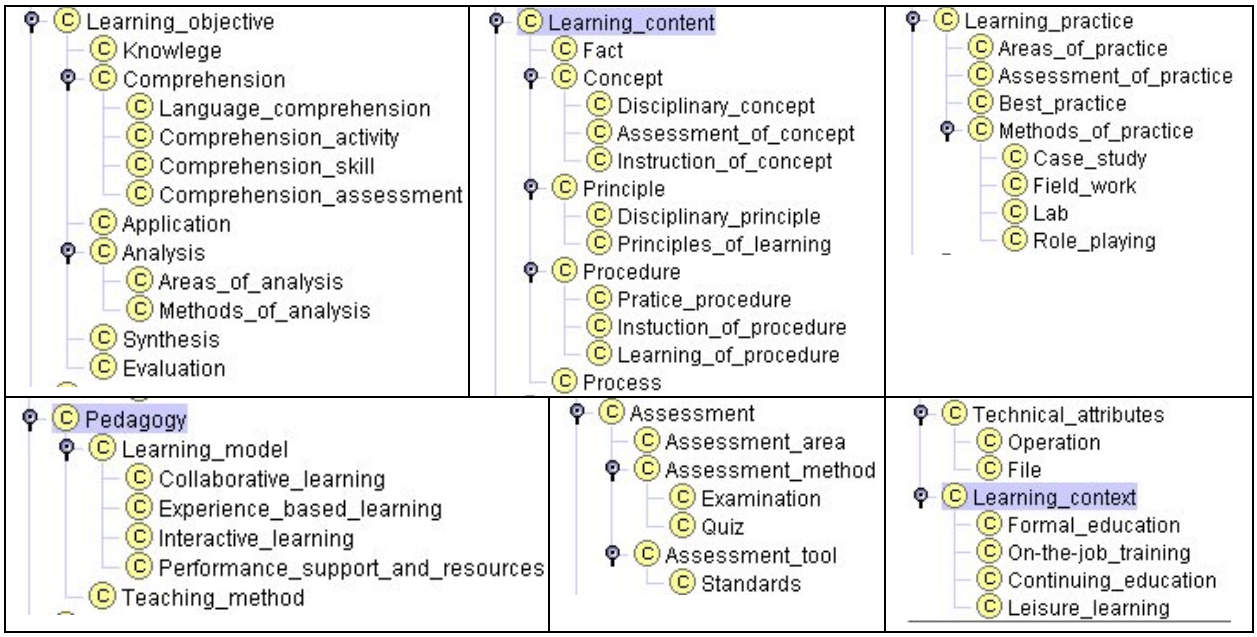
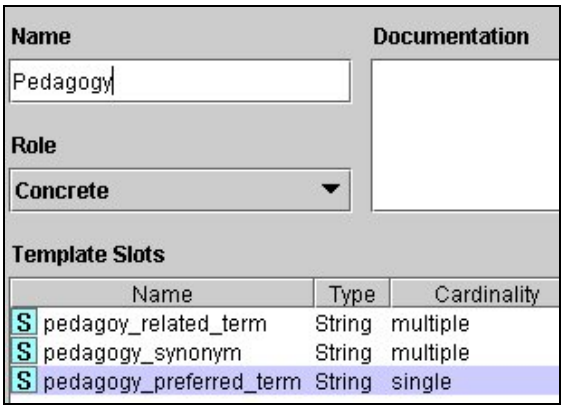
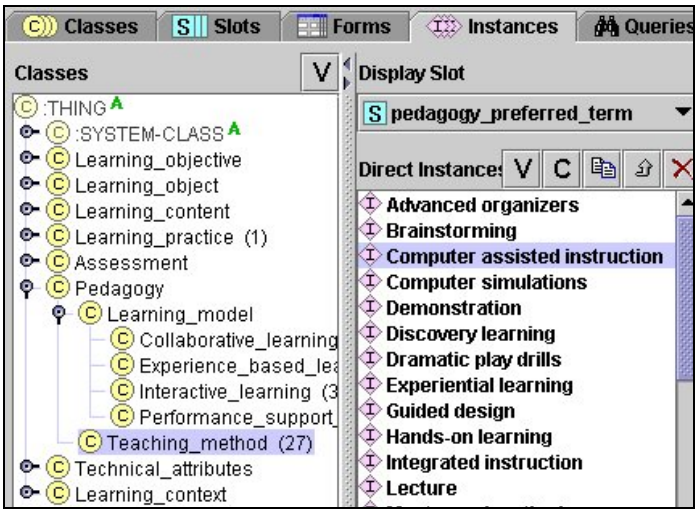


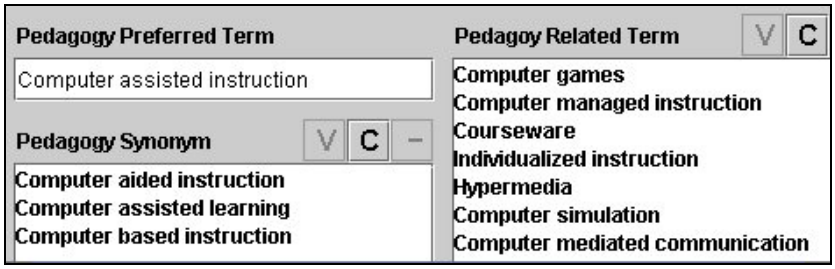
FIG. 6. Concept classes and subclasses in the learning object ontology



(a) The *Pedagogy* class has three properties (slots)



(b) Instances shown on the right column belong to *Pedagogy→Teaching method*



(c) *Teaching method* inherits three properties from its parent *Pedagogy* and uses them as the schema to capture vocabulary for the instance

FIG. 7. Demonstration of using concept classes and properties as an instance-capturing tool

6.2 Concept Relationships

Defining classes in an ontology sets the stage for mapping concept relationships. We used two methods to define relationships between concepts—through a lower-upper class relationship and by referencing a class through a property type. The latter method is available in Protégé (Noy et al, 2001), the ontology editor we used for this project. The properties of a concept are similar to fields in a database table—they have a name, type, cardinality, and facets (value space or referenced classes). Property types that are frequently used include Integer, String, Symbol, Class, Instance, and Boolean. Proper use of property types "Class" and "Instance" can give ontologies a great advantage to reusing and associating concept classes that have been defined elsewhere in the ontology. FIG. 8 displays the properties associated with *Learning object.* All the subclasses under *Learning object* automatically inherit all the properties listed in the right side of FIG. 8. Some properties have instance or class as the type. This means that the value for those properties is restricted to the classes provided in the "Other facets" column. In other words, the classes work as the "value space" or "domain" for the property with which they are associated. For example, subclass *Tutorial* is a kind of learning object and a tutorial teaching students how to use OCLC Connexion cataloging system is an instance of class *Tutorial*, which is also an instance of the *Learning object* class. As a result, it bears all the properties for both *Tutorial* and *Learning object*. When a property type is "class," such as "Content type" in FIG. 8, the allowable value for this property will be any or combination of any of Process, Fact, Principle, Procedure, Concept classes (including their subclasses) in the "Other facets" list. If a property type is "instance," the allowable value for the property would be the instances of the classes given in the list.

The ontology has 8 top-level classes, 28 at the second level, and 34 at the third level. Except class *Learning object*, all other classes have three properties of preferred term, related term, and synonym.

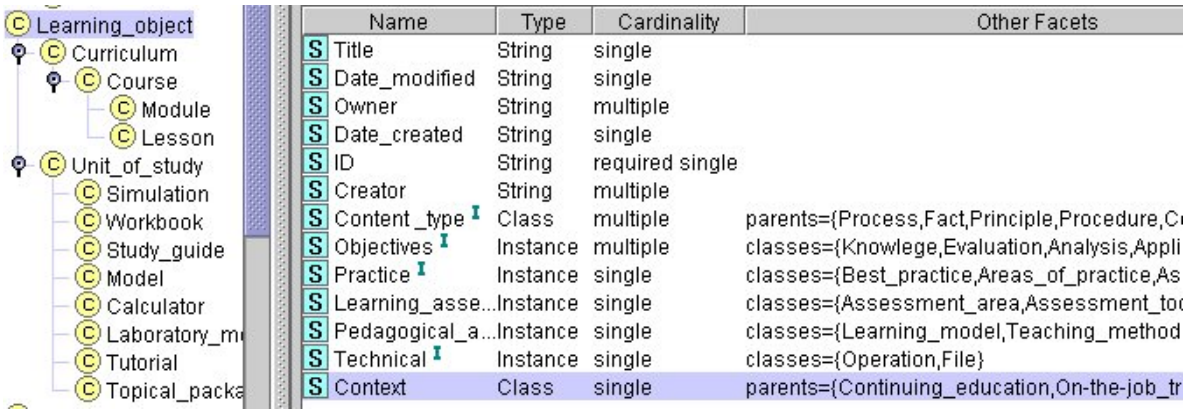


FIG. 8. Property type "Instance" is used in the class *Learning object*. Each of the properties that have Instance as the type is associated with subclasses in a top class

6.3 Validation

To test the appropriateness of the ontology, we conducted a query log mining using data taken from the GEM system. The intention was to discover to what extent the vocabulary in the ontology was similar to the query terms at both the conceptual and term levels. By conceptual level, we mean that even though a class is not an exact match of query terms or vice versa, the two may be conceptually similar.

The overall range of keyword occurrences in the whole data set distributed from the highest—12,440 occurrences for language arts, to the lowest—those that occurred only once. Approximately 1% of query terms counted for over 99% of the total occurrences. The keywords included in Table 3 fell into this 1% group and occurred in queries either as single words or in a phrase. Compared to the classes in the ontology, these keywords are either the same as or semantically similar to the classes in the ontology.

Table 3. Keywords used in the GEM queries

Pedagogical Keywords	Number of Occurrences	Pedagogical Keywords	Number of Occurrences
Process	3135	Application	180
Comprehension	1666	Exercises	140
Assessment	625	Principle	140
Project	491	Content	116
Creative	286	Facts	114
Analysis	282	Objectives	112
Concept	278	Context	108
Practice	231	Procedure	92

In an in-depth analysis of query terms, we wrote SQL queries to extract all the queries that contain the words in Table 3 to form separate files. Each subset of data was subsequently examined by the researchers so that meanings and patterns in query keywords and phrases could

be identified and normalized to represent the ontology concepts. Table 4 presents the categories for the main concepts in the ontology that were identified from the query log mining process.

Table 4. Categories of learning-related keywords in GEM queries

Analysis Areas of analysis Methods of analysis	Application Areas of application Application for jobs	Assessment Assessment areas Assessment methods Assessment tools	Comprehension Language Activities Skills Assessment
Concept Disciplinary Assessment Instruction	Content Disciplinary	Context Clues Vocabulary	Creativity Action Thinking Teaching
Facts Discipline	Practice Assessment Subject areas Best practice	Principle Disciplinary Learning	Procedure Practice Learning Instruction
Project Community Academic Methods	Process Operation Application		

The evidence from query log mining supports the main concept classes and most of the terms used as class names in the ontology. Further examination of the query data is needed to discover and generate more specific terms suitable for representing learning objects.

6.4 An Example of Ontology Application

We developed an Exercise Editor to explore how the ontology may be used in learning object authoring and in representing learning objects. The Exercise Editor is not a formal experiment, but rather, an exploration test, to find procedures, tools, and methodologies appropriate for applying the ontology. FIG. 9 is a screenshot from one of the Exercise Editor's interfaces—Add Objectives. The idea was to furnish instructors with structured learning objectives, procedures that students need to understand and follow, and assessment instruments and criteria. By using the Editor, learning object components were marked up with the vocabulary in the ontology (structured content). The Editor user may predefine learning objectives by entering them through the Add Objectives interface. Procedures and Assessment components may also be entered from separate tabs. The instructor can then assemble the exercise by choosing appropriate components from the component base (FIG. 10). It is also possible that an instructor finds an existing learning objective or procedure reusable for her or his exercise, so s/he does not need to repeat the work.

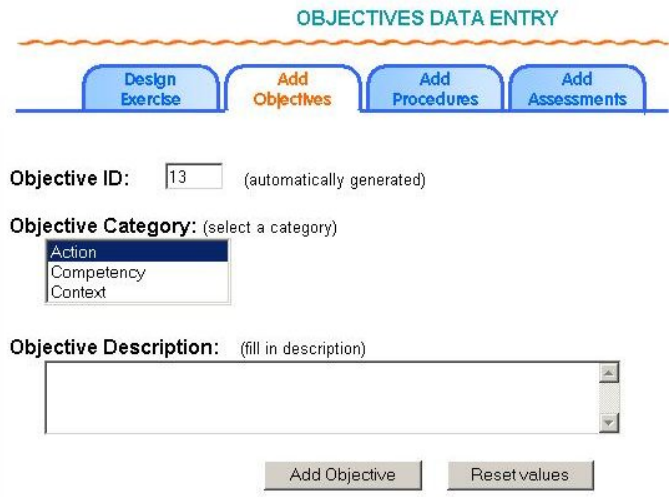


FIG. 9. A template for defining learning objectives

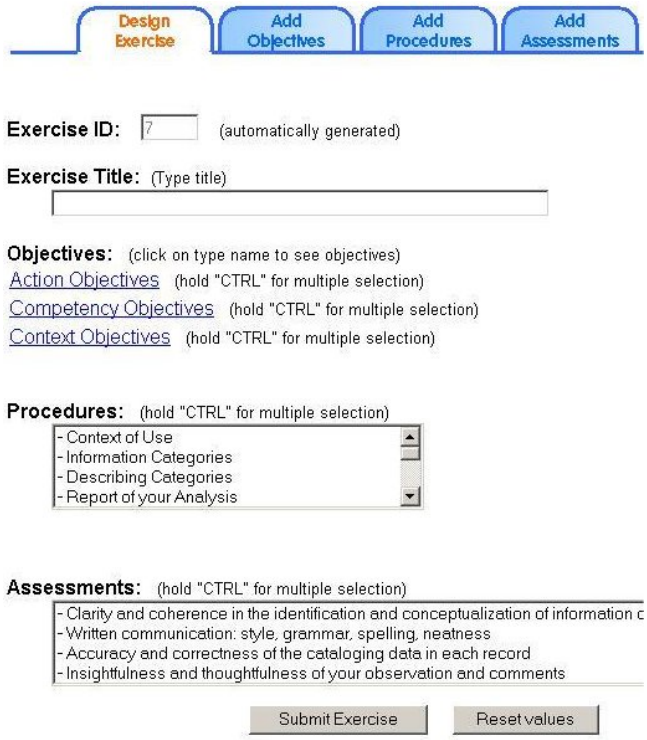


FIG. 10. An ontology-based Exercise Editor

FIG. 11 is a screenshot of an exercise generated by the Exercise Editor. It contains well-structured content marked up by learning-related vocabulary such as objectives and competency and can be dynamically displayed and manipulated.

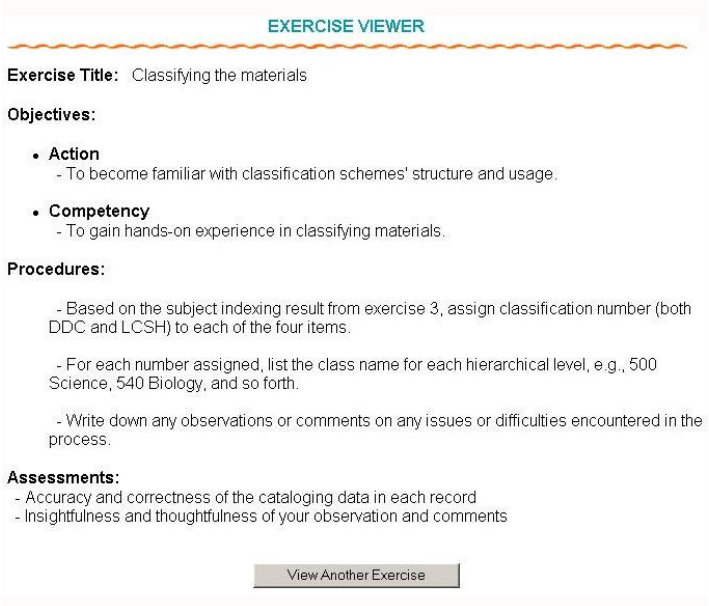


FIG. 11. The exercise generated by the Exercise Editor

7. Discussion and Conclusions

From ontological modeling and construction to the Exercise Editor, it becomes apparent that, while the ontology acts as a knowledge model for the content, presentation, and application of learning objects, its concept classes and properties should also be able to function as labels, values, or tags in database and/or encoding schemas. This is an important distinction between ontologies and traditional library classification schemes and thesauri, in which the classes and descriptors are usually too broad to be used as schema element labels and values.

In addressing the question of “how will we obtain the vocabulary and validate it,” our research demonstrated the process of building the ontology. As a knowledge structure and modeling tool, ontologies have more flexibility and functionality than traditional thesauri because the methods and technology allow for an integrated representation of both the content and metadata in digital objects. This promises to be a way to extend the traditional cataloging paradigm and take full advantage of digital objects in providing more effective methods and tools for digital object representation and use.

The Exercise Editor demonstrates a different design approach from other tools currently on the market, i.e., structural elements in learning objects should not be limited only to building blocks of text and media components. Adding richer semantics to the structural elements, as in the Exercise Editor, is what ontologies can offer to creators, vendors, educators, learners, and the like to fulfill the goals of content, presentation, and application of learning objects. One larger issue from this informal experiment is the lack of tools for implementing the ontology in applications. This translates into a gap between system development tools and knowledge modeling tools. On the one hand, collection building systems such as DSpace at MIT and FEDORA at Carnegie Mellon provide nice tools for incorporating metadata with digital objects, and ontology editors such as Protégé at Stanford offer powerful knowledge modeling capabilities. On the other hand, systems like DSpace do not offer mechanisms to incorporate

controlled vocabulary into metadata, and the ontology editors lack the tools for implementing the knowledge model in a system. To move further, either controlled vocabulary must be added to a collection development system or an ontology must be implemented in the system, both requiring the writing of programs to fill the gap, which is not only time-consuming and challenging, but also a repetitious waste of resources.

In summary, we discussed various views on learning objects and analyzed how these views affect the representation of learning objects. We proposed a framework based on Clancy's model to connect different areas of knowledge and technologies with the content, presentation, and application of learning objects. The ontology we constructed resulted from analysis of literature and was validated through query log mining. The main contribution of our work is the framework for the learning object domain and the ontology that reflects this framework. Our validation analysis is a unique approach in terms of the methodology, which provides evidence in supporting the main classes and structure of the ontology and the method itself as well. The next phase of research is to further analyze the query log data, and deduce more concept classes and properties as well as instance terms to enhance the ontology. The ontology also needs a larger scale and more formal evaluation and validation while implementing it in a prototype system.

REFERENCES

Acker, S. (2002). Personal communication.

ASTD & SmartForce. (2002). A field guide to learning objects. http://www.learningcircuits.org/2002/jul2002/smartforce.pdf

Barritt, C. (2002). Using learning objects in four instructional architectures. Networker Newsletter, 18(7). http://www.svispi.org/networker/2002/0702a1.htm

Barron, T. (2000). Learning object pioneers. Learning Circuits: ASTD's Online Magazine All about E-Learning, March. http://www.learningcircuits.org/mar2000/barron.html

Bench-Capon, T., Castelli, D., Coenen, F., Devendeville-Brisoux, L., Eagleston, B., Fiddian, N., Gray, A., Ligeza, A., and Vermesan A., (1998). Report on the 1st International Workshop on Validation, Verification and Integrity Issues of Expert and Database Systems, Information Research: an international electronic journal, 4(2): http://informationr.net/ir/4-3/paper55.html (electronic edition)

Bloom, B.S. (Ed.). (1956). Taxonomy of educational objectives: The classification of educational goals: Handbook I, cognitive domain. New York: Longmans, Green.

Caracciolo, C. (2003). Toward modular access to electronic books. Journal of Digital Information, 3(4): http://jodi.ecs.soton.ac.uk/Articles/v03/i04/Caracciolo/

Cisco Systems. (2003). Reusable learning object authoring guidelines: how to build modules, lessons, and topics. http://business.cisco.com/servletwl3/FileDownloader/iqprd/104119/104119_kbns.pdf

Clancey, W.J. (1997). The conceptual nature of knowledge, situations, and activity. In: P. Feltovich, R. Hoffman & K. Ford. (eds). Human and Machine Expertise in Context, 247–291. Menlo Park, CA: The AAAI Press.

Damjanoviæ, V., Gaševiæ, D., and Devedžiæ, V. B. (2003). Analysis of ontology validation possibilities. In: 4th Workshop on Computational Intelligence and Information Technologies, October 13, 2003, Faculty of Electronics, Niš, Serbia. http://cs.elfak.ni.ac.yu/ciit/w4/papers/09.pdf

DCMI. (2002). DCMI Type Vocabulary. http://dublincore.org/usage/terms/dcmitype/

Dodani, M. H.: (2002). The dark side of object learning: learning objects. Journal of Object Technology, 1 (Nov-Dec.): 37-42.

Forte, E., Haenni, F., Warkentyne, K., Duval, E., Cardinaels, K., Vervaet, E., Hendrikx, K., Forte, M. W., F. Simillion. (1999). Semantic and pedagogic interoperability mechanisms in the ARIADNE educational repository. SIGMOD Record, 28(1) (March): 20-25.

GEM. (2002). GEM project site: GEM 2.0. http://www.geminfo.org/Workbench/gem2.html

Greenberg, J., Sutton, S., & Campbell, D.G. (2003). Metadata: a fundamental component of the Semantic Web. Bulletin of the American Society for Information Science and Technology, 29(4), 16-18.

Grützner, I., Angkasaputra, N., & Pfahl, D. (2002). A systematic approach to produce small courseware modules for combined learning and knowledge management environments. In: SEKE ’02, 533-539. New York: ACM Press.

Hickey, T., O’Neill, E. T., & Toves, J. (2002). Experiments with the IFLA Functional Requirements for Bibliographic Records. D-Lib Magazine, 8(9), http://www.dlib.org/dlib/september02/hickey/09hickey.html

Hillmann, D. NSDL vocabulary workshop. http://metamanagement.comm.nsdl.org/cgi-bin/wiki.pl?VocabWorkshop

IFLA. (1998). Functional Requirements for Bibliographic Records: Final Report. http://www.ifla.org/VII/s13/frbr/frbr.htm

IMS Global Learning Consortium. (2003). Reusable Definition of Competency or Educational Objective (RDCEO). http://www.imsglobal.org/competencies/index.cfm

Kamel Boulos, M. N., Roudsari, A. V., & Carson, E. R. (2001). Towards a semantic medical web: healthCyberMap’s Dublin Core ontology in Protégé-2000. http://protege.stanford.edu/ontologies/dublincore/hcm_dc_in_protege_newcastle.pdf

Koper, R. (2001). Modeling units of study from a pedagogical perspective: the pedagogical meta-model behind EML. http://eml.ou.nl/introduction/articles.htm

Kostur, P. (2002). Connecting learners with content: a unified content strategy for learning materials. In: SIGDOC’02, 100-103. New York: ACM Press.

Laleuf, J. R. & Spalter, A. M. (2001). A component repository for learning objects: a progress report. In: Proceedings of the first ACM/IEEE-CS joint conference on Digital libraries (Roanoke, VA, June 2001), 33-40. New York: ACM Press.

Liddy, E. D. et al. (2002). Automatic metadata generation & evaluation. In: Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (Tampere, Finland, August 2002), 401-402. New York: ACM Press.

Liddy, E., Allen, E., Harwell, S., Corieri, S., Yilmazel, O., Ozgencil, N.E., Diekema, A., McCracken, N., Silverstein, J., and Sutton, S. (2002). Automated metadata generation & evaluation. In: Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval, 401-402. New York: ACM Press.

Longmire, W. (2000). A primer on learning objects. Learning Circuits, March. http://www.learningcircuits.org/2000/mar2000/primer.html

LTSC. (2001). IEEE P1484.3/D3 Draft Standard for Information Technology Learning Technology -- Glossary. http://ltsc.ieee.org/doc/wg3/Glossary-200103 09.doc

LTSC. (2002). Draft Standard for Learning Object Metadata. IEEE 1484.12.1-2002. http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf

LTSC. (2004a). IEEE P1484.12.3/D2 Draft for Extensible Markup Language (XML) Schema Definition Language Binding for Learning Object Metadata. http://ltsc.ieee.org/wg12/files/IEEE_1484_12_03_d2.pdf

LTSC. (2004b). IEEE WG12: Learning Object Metadata: Purpose of Proposed Project. http://ltsc.ieee.org/wg12/index.html

Martin, M., Agnew, G., Boyle, J., McNair, J., Page, M., & Rhodes, W. (2002). DRM Requirements for Research and Education. Discussion Paper. September 4. http://xml.coverpages.org/RE-DRMRequirements.pdf

Mason, J. (2004). [Comments in the discussion thread on] Proposal for an instructional methods element. http://www.jiscmail.ac.uk/cgi-bin/webadmin?A1=ind0408&L=dc-education

Merrill, M. D. (1998). Knowledge Objects. CBT Solutions, March/April. http://www.id2.usu.edu//Papers/KnowledgeObjects.PDF

Naber, L. & Kohle, M. (2003). E-nhance lectures. Journal of Digital Information, 3(4): http://jodi.ecs.soton.ac.uk/Articles/v03/i04/Naber/

Noy., N.F., Sintek, M., Decker, S., Crubezy, M., Fergerson, R.W., & Musen, M.A.. (2001). Creating Semantic Web contents with Protege-2000. IEEE Intelligent Systems 16(2):60-71.

Ochoa, S., Ormeno, E. G., & Pino, J. A. (2002). Reusing courseware components. In: SEKE ’02, 549-556. New York: ACM Press.

Paik, W., Yilmazel, S., Brown, E., Poulin, M., Dubon, S., Amice, C. (2001). Applying natural language processing (NLP) based metadata extraction to automatically acquire user preferences. In: Proceedings of the international conference on Knowledge capture 2001, 116-122. New York: ACM Press.

Polsani, P. (2003). Use and abuse of reusable learning objects. Journal of Digital Information, 3(4): http://jodi.ecs.soton.ac.uk/Articles/v03/i04/Polsani/

Qin, J. & Chen, J. (2001). A multi-layered, multidimensional representation of digital educational resource objects. In Subject Retrieval in the Networked Environment: Proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001, 90-96. München: K. G. Saur (published in 2003).

Qin, J. & Finneran, C. (2002). Ontological representation of learning objects. In Proceedings of the Workshop on Document Search Interface Design and Intelligent Access in Large-Scale Collections, July 18, Portland, OR. http://xtasy.slis.indiana.edu/jcdlui/uiws.html

Qin, J. & Godby, J. (2003). Incorporating educational vocabulary in learning object Metadata Schemes. In: Research and Advanced Technology for Digital Libraries: 7th European Conference, ECDL 2003, Trondheim, Norway, August 17-22, 2003, Proceedings, 52-57. Berlin: Springer.

Qin, J. & Paling, S. (2001). Converting a controlled vocabulary into an ontology: the case of GEM. Information Research: An International Electronic Journal, 6(2): http://InformationR.net/ir/6-2/paper94.html (January).

Rice University. (2003). Connexions: education for a networked world. http://cnx.rice.edu/aboutus/publications/ConnexionsWhitePaper.pdf

Rust, G. & Bide, M. The <indecs> metadata framework: principles, model and data dictionary. 2000. http://www.indecs.org/pdf/framework.pdf

Sutton, S. (2004). [NSDL Vocabulary] Workshop notes. http://metamanagement.comm.nsdl.org/Vocabulary-Workshop-Notes.html DC-ED. (2004). DC-Ed Element Proposal: Instructional Method. http://www.ischool.washington.edu/sasutton/8-21-04/

Tennis, J. (2003). Data collection for controlled vocabulary interoperability—Dublin Core audience element. Bulletin of the American Society for Information Science and Technology, 29(2), 20-23.

Trivantis Corporation. (2003). Standards for Lectora Publisher. http://www.lectora.com/product_info_standards.html

Tuso, G. & Longmire, W. (2000). Competency-based systems and the delivery of learning content. In W. Longmire et al (Eds.), Learning without Limits, v. 3, Infomania, San Francisco, 31-38.

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. 2000. http://reusability.org/read/chapters/wiley.doc