2006

The Semantic and Syntactic Model of Metadata

Jian Qin
Syracuse University, jqin@syr.edu

Javier Calzada Prado
Universidad Carlos III de Madrid, fcalzada@bib.uc3m.es

The Semantic and Syntactic Model of Metadata

Jian Qin
School of Information Studies
Syracuse University
4-187 Center for Sci & Tech
Syracuse, NY 13244
USA
jqin@syr.edu

Javier Calzada Prado
Departamento de Biblioteconomía y Documentación
Universidad Carlos III de Madrid
Calle Madrid, 126
28903, Getafe (Madrid)
Spain
fcalzada@bib.uc3m.es

Abstract:

As more information becomes "born digital", metadata creation is increasingly becoming part of the information creation process. Current metadata schemes inherit much of the library cataloging tradition, which has shown limitations on representing "born digital" type of resources. Through analysis of issues of metadata schemes and review of metadata research and projects, the authors propose an ontology-based approach to building a modular metadata model in which semantics and syntax may be integrated to suit the needs for representing "born digital" resources. The authors use an learning object ontology as an example to demonstrate how the semantics and syntax may be built into a modular model for metadata.

Keywords:
Metadata standards; Digital libraries; Ontologies; Semantic modeling.

1 Introduction

Three areas of work are essential for metadata to perform its functions: semantics to define the meaning of data, syntax to specify the data binding structure, and vocabulary to control the language (Duval et al 2002). Duval et al maintain that because syntax language such as XML is still under development, it is necessary to keep metadata semantics separate from syntax, which has been witnessed during the first decade of metadata development. As more information is "born digital," metadata creation is increasingly becoming part of the information creation process. This fundamental change has a significant implication for metadata development. The "born-digital" trend has caught the attention of metadata and digital information developers. One of the strategies in addressing the challenge is expanding metadata standards by adding structural and/or content elements (Becker et al 2003; Dushay 2002; Kostur 2002). This raises questions in the paradigm of separating semantics and syntax in representing information that is created digital.

The first question is related to the current model of metadata. Early metadata experiments, including the one initiated at OCLC and contributed by librarians (Jul 1995), used the MAchine Readable Cataloging (MARC) format to encode the description data for web sites and pages, which is the data-binding format of the 2nd edition of Anglo-American Cataloging Rules (AACR2). The metadata schemes developed from these experiments are greatly influenced by library cataloging practice. Elements in metadata schemes have similar linear structures as those defined in AACR2. Syntactic structures for these elements are provided in separate specifications, which may deviate slightly from the element definition due to the need to adapt to the syntactic language. Gaps between metadata semantic and syntactic structures resulted in duplicate efforts in binding the same data elements with various languages and application programs, which leads to widely varied data binding models and implementations. If MARC format has successfully converted the card catalog into machine-readable form three decades ago, it is very unlikely that metadata standards will repeat the history again simply by following the footprint of MARC in the "born-digital" information environment.

The second question is the amount of semantics offered in current metadata standards. Due to the traditional cataloging influence, metadata standards generally contain limited semantics for machine processing. On one hand, common semantic elements in metadata schemas such as title, author, subject index terms, and description are often far from enough when finer metadata representation is needed. This forces developers to expand the metadata semantics with methods and technology suitable in their context, which results in widely varied practices and duplicate efforts. On the other hand, the fast growth of digital information is difficult enough for human catalogers to keep up with even for such limited metadata semantics. Much of the information about an object has to be left out of the metadata record. To enrich the semantics in metadata schemes while increase the amount of machine-processable data, a promising solution lies in a new metadata model that will standardize the metadata development and provides extensible and powerful semantics and syntax for utilizing the fullest potential of the "born-digital" information.

The limitations and future prospective of metadata standards call for a formal metadata model to address the issues related to metadata semantics and syntax. In this paper, we propose to build a semantic and syntactic model using an ontological approach as a solution to the problems mentioned above. The remainder of this paper contains the following sections: 1) review of metadata modeling and other related research literature, 2) methodology for data collection and processing about selected metadata schemas, including the rationale of using the ontological approach to building the model, 3) analysis of the semantic characteristics of metadata elements in schemas under examination, 4) discussion of the philosophy and principles for building the semantic and syntactic model, and 5) discussion of the implications and conclusions.

2 Literature Review

Metadata models have been one of the research frontiers in recent years. Researchers from various backgrounds use different approaches to analyze the domain and seek the best and most effective ways to build the metadata model. These approaches can be divided into two broad categories: element-based expansion and ontological modeling.

Element-based expansion is essentially "customization" of metadata standards by either expanding the standard elements or adding new local elements. This approach is common in digital library projects where representing domain digital information requires specialized metadata elements but they are absent from the standard being adopted. Examples include the GREEN project (http://appling.kent.edu/NSDLGreen/GreenDLMetadata.htm), DLESE (http://www.dlese.org/Metadata/dlese-ims/index.htm), and GEM (http://www.geminfo.org/Workbench/gem2.html), among others. The expansion of standard elements may take domain specific markup languages and other relevant standards as the extended structure for the domain knowledge. The GREEN project, for instance, added elements from the Mathematic Markup Language (MathML) for the mathematic formulas and expressions in the metadata schema to create a customized version of the LOM scheme (Shreve and Zeng 2003). The customization of metadata schemas tailors the elements to fit local representation needs while the core elements comply with a metadata standard. However, element-based expansion still maintains the linear structure, i.e., hierarchical relationship among elements. Horizontal associations among elements can only be established at data binding (either in form of database tables or XML schemas).

Ontological modeling of metadata takes an object-oriented view of all elements in a metadata scheme and reorganizes them as concepts, concept properties, instances, and relations. General ontology modeling related to metadata includes the <indecs> metadata framework (Rust and Bide 2000) and the Functional Requirements of Bibliographic Records (FRBR) (Plassard 1998). There have been quite a few publications discussing the models, but implementation of such models is still in experimental stage (Hickey and Vizine-Goetz 2001). Lagoze and Hunter (2001) build a conceptual model to facilitate interoperability between metadata ontologies from different domains. Their model uses *Entity* as the root class and assigns three categories—*Temporality, Actuality*, and *Abstraction*—as its subclasses. The next level of subclasses includes *Artifact, Event, Situation, Action, Agent, Work, Manifestation, Item, Time*, and *Place*. The properties of these concept classes are defined as a set of relations such as "isPartOf," "inContext," "contains," phaseOf," and "hasRealization." As the authors state, this model is syntax-neutral and they suggest to use the Resource Description Framework (RDF)/XML as the data binding language.

While Logoze and Hunter try to create a metadata model without the influence of traditional cataloging practice, other ontology projects attempt to build metadata models based on existing metadata standards and controlled vocabulary. Kamel Boulos et al (2001) developed a Dublin Core (DC) metadata ontology for the health informatics domain, in which the *Subject* element in DC was populated with the Unified Medical Language System (UMLS) and clinical codes. Using controlled vocabulary to build ontology-based metadata schemas is another approach. Qin and Paling (2001) analyzed the controlled vocabulary from the Gateway to Educational Materials (GEM) and constructed an ontology to represent the facets of subject, pedagogy, relation, audience, educational level, format, and language in learning objects. Their metadata model uses *Resource* as the root concept which has *Resource Type* as subclass (e.g., lesson plan is a subclass of resource) and the above mentioned facets are global properties that may be inherited by the subclasses of *Resource*. No matter whether ontological modeling begins from scratch or is based on existing metadata schemas or controlled vocabulary, a common characteristic among the projects is that they all use an object-oriented approach to analyze the information objects and their content. This builds the technical condition necessary for modularized and reusable metadata schemas.

One application in ontological modeling is building domain ontologies for content representation and categorizing digital objects. Khan et al (2004) created a domain-dependent ontology to represent the context and meaning of audio objects' content. The most specific concepts in this ontology were considered as metadata. By using automatic context extraction techniques, the more general concepts in the ontology were used to categorize audio objects. Khan et al demonstrate how metadata may be generated and audio selection customized using the ontology model.

To summarize, element-based expansion is common in metadata creation and an easier way to adopt a metadata standard. One disadvantage, among other things, is the limitations in offering finer-grained semantics at conceptual level and in establishing relationships between related concepts, which can only be established at the implementation stage. Ontological modeling as a promising methodology is still being explored. Experiments with domain-dependent ontologies have been conducted in metadata extraction and information retrieval. However, questions remain on how to construct the metadata model to maximize the potential of born-digital information objects and to bring semantics and syntax together to minimize the implementation efforts.

3 Methodology

To address the questions raised from literature review, we chose to study a number of representative metadata schemas to examine their structures and vocabularies, rather than conducting a formal survey with a scientific sampling method. Our main purpose is to gain insights into the extent to which metadata standards were adopted, where the expansions to these standards occurred in the adopting schemas, and what semantic and syntactic characteristics existed in the schemas and expansions.

3.1 Data Collection

We realize that it is impossible to examine all metadata schemas used by all digital libraries. The selection criteria were based on two considerations: whether the digital library has a strong presence of metadata development and a metadata team, and whether the metadata schema has its own controlled vocabulary and expansions. The six metadata schemas included in our study were chosen from six digital libraries that met the two considerations and had separate sites for metadata information: the Digital Library of Theses and Dissertations (NDLTD), the Digital Library for Earth System Education (DLESE), the Alexandria Geospatial Digital Library (ADL), the Gateway to Educational Materials (GEM), MERLOT (Multimedia Educational Resource for Learning and Online Teaching) and ARIADNE (Alliance of Remote Instructional Authoring and Distribution Networks for Europe).

A relational database in Microsoft Access was then created to collect data on schemas, elements and subelements. The following data fields were included in the database: name and URL of digital libraries, schemas used in digital libraries and their version, and elements and subelements that belonged to each of them. Detailed information was collected about elements and subelements: name given to each tag, type of element or subelement, description available and semantics embedded. Types of elements or subelements were defined according to the reference schema or standard chosen and declared as such by each digital library or, if necessary, defined as an expansion or locally developed element or subelement, as indicated by the following categories: "DC" (Dublin Core), "DC element expansion", "LOM" (IEEE Learning Object Metadata), "LOM element expansion", "IMS", "IMS element expansion" and "local element." In addition to element coding, we also followed the data entry rules below:

- When tag names were not identical to the ones in the adopted metadata standard, they were considered as element expansions, even if they referred to the same concept.
- Some subelements recurred in several elements or subelements during database binding. We entered these recurring subelements as "expansions."
- An element or a subelement was considered as local when it was developed by the adopting schema.
- If a subelement is identical to one of those in DC or LOM, it was marked as "DC" or "LOM" (even if they were found in a locally developed element), stating the fact that it had not been altered from the standard considered.

3.2 Data Processing

After all data had been collected (resulting in 95 elements and 311 subelements), we ran several queries to merge data fields for elements and subelements as well as categorized both elements and subelements for descriptive statistical processing.

One important step in data processing is categorizing the elements and subelements. Researchers have categorized metadata with more or less similar groups (Greenberg 2001; Gilliland-Swetland, 2000; Lagoze et al 1996). Common groupings include administrative data, descriptive/discovery data, intellectual content data, technical data, and rights data. Based on previous research on metadata groupings, we categorized all elements in our data set into four groups:

(1) **Administration**: This group includes elements that are mainly used for managing and tracking metadata. It includes time and cataloging agency related data.
(2) **Description**: Any element describing a digital object's discovery or access characteristics is categorized into this group. Included in this group are content, descriptive, time coverage, scientific data, and rights. They mainly perform user-oriented tasks.
(3) **Education**: Most schemas included in this study contain a number of educational elements, but little research has been done in examining the elements in this group. It is also a user-oriented group.
(4) **Technical data**: This group contains the elements dealing with the physical characteristics and system requirements for using the objects. Elements in this group have the greatest potential to be extracted automatically by computer programs.

Data categorization was conducted by both researchers in parallel. The categorization results were compared and differences were discussed and cleaned in order to ensure the accuracy and consistency.

3.3 Analysis

Once the data set was ready, we ran frequency and cross-tabulation analysis: elements and subelements by schema, type of element (DC, LOM, or local elements), and category. The findings are presented in Section 4 and discussion of results and modelling in Section 5.

4 Findings

The six metadata schemes we studied had 95 first level elements. The frequency analysis of these elements reveals that only a handful of elements were identical: title, description, rights, and technical. The rest of elements had one occurrence each, though many of them were semantically similar or identical. We further analyzed their corresponding 311 subelements by dividing them into four types as shown in Table 1. While each of the metadata schemes adopted at least one standard, they all created a large number of local elements, counting slightly over one-third of the total.

4.1 Categories of Metadata Elements

To find out the distribution of metadata element categories, we divided all subelements into four main groups: administration, description, educational, and technical (Figure 1).

Within each group, elements were further categorized based on their semantics. As we can see in Figure 1, the largest group was Description. Figure 2 shows the *Relation* elements counted for 42% of all descriptive subelements; *Title* for 11%; URL or web location of the resource 11%; *Description* of the resource 8%; *Version* of the resource 7%; and the rest counted for only small fractions.

Table 1. Number of elements by element type and schema

Metadata scheme	Number of elements by element type				Total
	DC	IMS	LOM	Local	
ADL		5		56	61
ADN		71		19	90
ARIADNE			29	21	50
ETDMS	18			4	22
GEM	60			9	69
MERLOT			8	11	19
Total	78	76	37	120	311

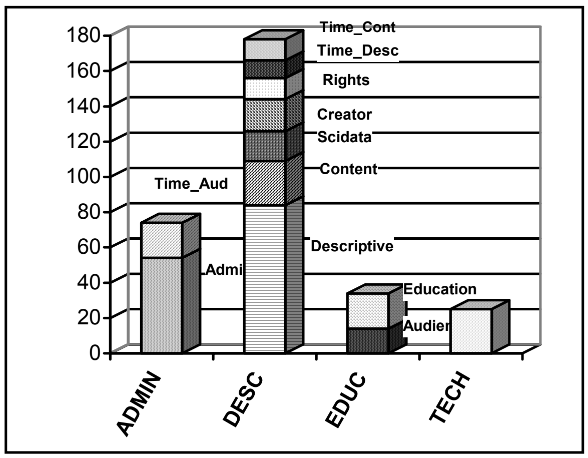

Figure 1. Subelements by category

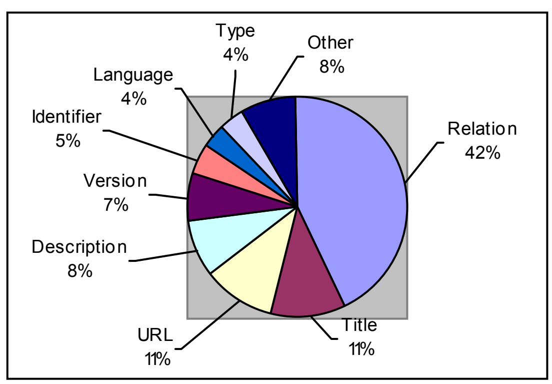

Figure 2. **Details of description subelements.**

The description group contained a large number of elements for content, creator and scientific data.

Content elements had to do primarily with four aspects, by frequency order: the subject of the resource, the abstract of its content, its discipline and keywords (Table 2). Subelements related to the creator of the digital object are straightforward—they deal with their names and roles. Those subelements that dealt with scientific data belonged to ADL and DLESE, containing specific data to attend the specific needs for representing coordinates, elevation, projection, etc. in the digital library collections.

Table 2. Content elements by semantic category

Semantic category	Number of elements
Subject	8
Abstract	4
Discipline	4
Keywords	3
Other (1 occurrence each)	6
Total	25

Table 3. Subelement types by semantic category

Semantic category	No subelement	Number of Subelements by Subelement type							Total
		DC	DC exp	IMS	IMS exp	LOM	LOM exp	Local	
ADMIN	0	1	0	4	16	0	2	20	**43**
AUDIENCE	1	0	3	2	5	0	0	0	**12**
CONTENT	3	2	7	1	3	0	0	15	**31**
CREATOR	1	3	3	1	2	0	2	5	**17**
DESC	10	18	21	2	10	3	8	17	**89**
EDUCATION	0	0	0	3	3	7	5	3	**20**
RIGHTS	2	1	0	1	2	0	1	3	**10**
SCIDATA	0	0	4	3	1	3	0	22	**33**
TECHNICAL	2	2	5	6	8	1	3	1	**28**
TIME	0	2	6	1	2	2	4	11	**28**
Total	19	29	49	24	52	16	25	97	**311**

Table 3 shows the distribution of subelements by subelement type and semantic category. An obvious pattern is that expanding on standard metadata elements is common practice and local elements were used in almost all categories. Most of the expansions in DC concentrated on descriptive elements, while spread across categories relatively evenly in IMS and LOM. It is worth mentioning that local elements counted for almost one-third in the total, among which administration and scientific data categories were at the top. It is surprising that only 41 LOM elements and subelements were adopted by the six metadata schemes while the standard has over 80 elements and subelements in total.

Category wise, expansions by way of local elements concentrated mainly in administration, content, description, and scientific data.

4.2 Vocabulary Use in Local Elements

A further examination was conducted to analyze the vocabulary use in subelements in the four largest categories. We found an interesting phenomenon across all the local expansions, i.e., an element "Type" used in XML data binding was mixed among the semantic elements, which has completely different semantics from the *Type* element in Dublin Core. This semantic- and syntactic-neutral element often hints a user-defined data type in the XML schema. The administrative subelements beginning with a Type element include following:

Type-Email-Address
Type-Larger Organization
Type-Notes
Type-Operators
Type-Metadata-Mapping
Type-Postal-Address
Type-Postal-Address-PO Box
Type-Postal-Address-City

Element names in the content category incurred wide variations in terms of both semantics and linguistic forms (Table 5). The content subelements fell into two categories: those for topical terms, which used the thesaurus construction approach, and those for time covered by the digital object content, which were ambiguous in their meaning and use.

Local subelements in the description category added more details to the common ones. For example, "hierarchy" was used in two elements to describe the item type and "Event name tied to coordinates" to link events and geographic areas. Scientific data categories contained many subelements particular to geospatial data.

Table 5. Local subelements in the content category

Body or Planet	Simple Time Period

Discipline	Spatial-Coverage
Level	Subdiscipline
Main Concept	Temporal-Coverage
Main Concept Synonyms	Time AD
Name	Time BC
Named Time and Period	Time Relative
Other Concepts	Type

4.3 Educational Elements

Since most of the metadata schemes included in this study are created for educational digital resources, we conducted an analysis of the details in educational elements. Figure 3 shows the distribution of elements in both educational and audience categories. In 32 subelements, the audience elements counted for the largest number. "Type" here belongs to data binding as discussed before, hence is not a real semantic element. Compared to other categories that had large number of local expansions, the educational elements were relatively poorly developed. The vocabulary used to label some of the elements was not immediately clear (e.g., Didactical context, Semantic density, Granularity), while others (Interactivity level, Interactivity type, Grouping) require intensive human judgment.

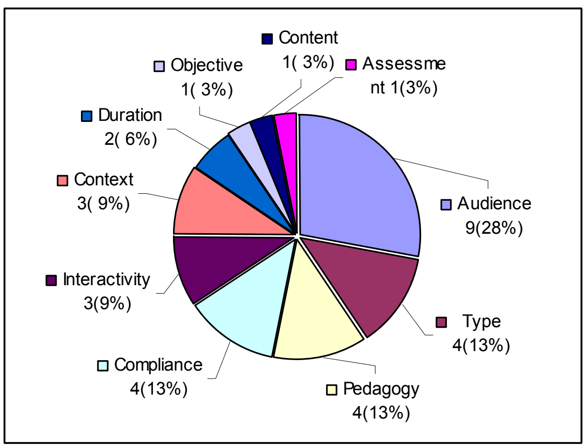

Figure 3. Detail of educational subelements

4.4 Summary of Findings

The findings from our survey data reveal at least three important facts:

1) Metadata standards provide limited semantics and have to be expanded to meet local needs;
2) Problems exist in local expansions in both semantic consistency and explicitness;
3) Metadata binding with XML brings in semantic and syntactic neutral elements as a method for bridging reusable or user-defined data types.

One implication from the data analysis is that, as technology evolves and digital information grows in both volume and complexity, we need to reexamine the principles and methods for metadata development.

5 Discussion

In previous sections, we discussed the reasons why we need to reexamine the metadata principles and what issues need to be addressed. The focal point of discussion falls onto what metadata models would be more extensible, scalable, and effective and more fundamentally, what underlying philosophy supports such metadata models. Based on the literature review and the analysis of data as well as previous research (Qin 2003; Qin 2004b), we propose an ontology-based metadata model that specifies concepts, properties, and relationships involved in metadata schemas by using an object-oriented approach.

5.1 Underlying Philosophy of Modeling

Metadata is used for three main purposes: reuse, retrieval, and tracking (Rockley 2003). Reuse has two meanings—the element definition reuse (e.g., address elements can be defined once and reused in publisher, creator, and contributor elements) and the data reuse (e.g., address for the same author who created several digital objects). Metadata modeling must facilitate reuse in both senses.

Retrieval metadata is perhaps the oldest arena in metadata in its broadest sense (thus including bibliographic data in the traditional sense). Conventional retrieval elements such as author, title, and keywords still play a vital role in resource discovery, but the way they are constructed should enable local expansions in a consistent manner to avoid wild variations in semantics and syntax. This is the basis for enabling multiple-database searching and reducing duplicate implementation efforts.

Digital objects often need information for tracking who created or submitted the object and/or metadata and when it was created or submitted. The large number of time-related elements in our survey demonstrates the importance of such metadata elements. Tracking digital objects in large repositories may require use of tracking elements combined with other types of elements to narrow the search.

One thing that becomes clearer in the past decade of metadata activities is that developing access to digital objects can not simply copy the model from AACR2 and MARC. A more flexible, powerful model must be developed to accommodate the characteristics of digital objects and the needs for using these digital resources in non-traditional ways. As the World Wide Web Consortium (W3C) phases out the metadata activities into Web Ontology, the metadata modeling discussion can not come at a more appropriate time. The ontological, object-oriented approach to modeling metadata would also be in line with the Web Ontology development at W3C.

5.2 The Model

Metadata elements need to have an abstract model to consistently represent the semantics and syntax. Following the paradigm of RDF, we propose an abstract model that is simple and conforms with the RDF formal model while maintains scalability and extensibility for metadata schemes. The diagram in Figure 4 suggests that elements in a metadata scheme are concepts and have properties and instances (properties also have instances). Concepts (or classes) form the structure of a domain in which semantics, syntax, and properties are specified. While semantics refers to the meaning of an element and syntax to the encoding format, properties serve as a data model to capture instances and define constraints of concepts. Instances contain vocabulary, both controlled and free-text, for elements and element values with a consistent syntax.

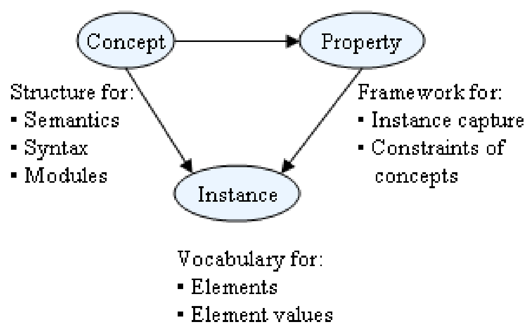

Figure 4. An abstract model for semantic and syntactic metadata

The main advantage of this model is that no matter how the domain concepts are structured, they will always be represented by a tuple of concept, property, and instances. This model may be used for any metadata scheme to define metadata structure and vocabulary. We will use a learning object ontology (Qin 2004a) as an example to demonstrate what this model means.

The learning object ontology created by Qin (2004a) contains a number of main concepts: learning objective, learning object, learning content, learning context, learning model, learning practice, and assessment. These concepts form the knowledge structure for the learning related content in learning objects. Figure 5 presents a portion of the concept classes in the ontology and direct instances for the *Learning object* concept. Each class in the ontology has properties of term, synonym, and related term, which are used as the data model to capture instances for the classes. The concept *Learning object*, for example, has direct instances as shown in the second column on the left. The instance *Figure* uses the word "Figure" as the preferred term, which has synonyms such as Illustration, interactive illustration, diagram illustration, photo illustration, chart, etc. and related terms such as figure title, figure type, and figure content. Their relationships may be expressed as:

Learning object is Concept
Which has property of {
Structure Term,
Structure Synonym,
Structure Term Related};
Figure is Instance of Learning object
Which has {
Structure Term {Figure},
Structure Synonym {
Illustration,
Interactive illustration,
...},
Structure Term Related {
Figure title,
...};

5.3 Modules

One of the main drawbacks in most metadata schemas is a lack of modular structure for the elements. It is common that dozens of elements are stuffed in a metadata schema as a very long list. Such a single list style of metadata elements makes metadata schema maintenance and implementation inconvenient and complicated.

A modular data model is usually considered as more extensible and flexible because it can be managed separately and tested independently or combined as an integrated whole (Luna 1992). The abstract model we proposed allows metadata elements to be built in a modular style while still maintains structural and syntactic consistency. In this model, a concept or several concepts can be created as a simple module. Several modules may also be combined to form a new, complex module while the properties remain the same for elements in these modules. In the implementation stage, an adopter may choose to maintain a shallow metadata model in which individual modules are jointed together by an overarching schema at run time, or the adopter may choose to joint the modules before applications are developed.

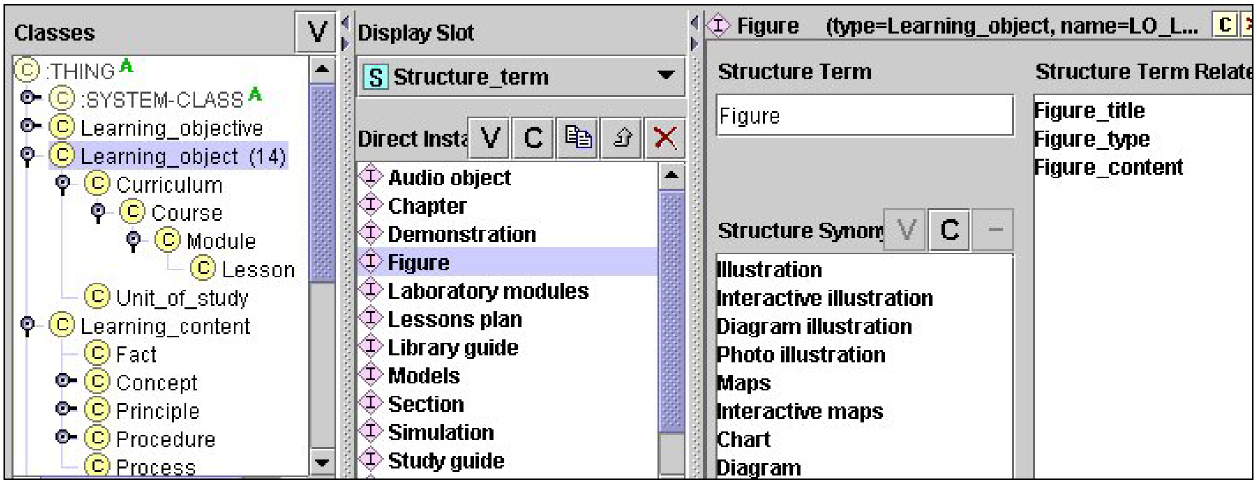

Figure 5. Concept classes and the direct instances for the *Learning object* concept in the learning object ontology (http://web.syr.edu/~jqin/LO/LOV2/)

Metadata modules include reusable modules and functional modules.

Reusable modules: The findings from our survey show that reusable elements occurred mostly in role and syntactic elements, such as name, address, email, and the elements in data binding. Another category that can be defined as reusable module is content elements. Regardless of element names, all elements in the content category may use a model of preferred name, synonym, and related terms. Reusable modules are similar to "user-defined" data types in object-oriented data modeling.

Functional modules: These modules will perform retrieval, tracking, and administration functions. They may overlap with reusable modules.

6 Conclusion

In this paper, we analyzed 311 metadata elements in six metadata schemas based on their type and category. Our findings show that large numbers of local expansions were made based on metadata standards but semantic inconsistencies and ambiguities existed across local expansions among the schemas. As we pointed out in the Introduction section, much of these problems are related to the underlying philosophy of metadata development, which is influenced primarily by traditional library cataloging. We proposed an abstract model of the concept, property, and instance tuple and explained the underlying philosophy of the model. Using the learning object ontology as an example, we also demonstrated what the model means and how it works for building a modular, extensible, and ontology-based metadata model.

The main contribution of this paper is that it raises questions on the metadata development direction and proposes an ontology-based approach that is simple yet allows for extensibility and consistency in developing metadata schemas. As more and digital information objects are created with structural elements, metadata schemes will need to be extended to include such elements in addition to metadata. With this vision of future metadata development, it becomes critical to have a simple abstract model for dealing with the complexity, scalability, and interoperability of metadata schemes.

Based on the ontology we created, our future research will continue the work on metadata modeling. This will include developing the modular semantic model with various functions and reusable data types, as well as the syntactic model that will provide effective and standard data binding formats.

References

Aslan, G. and D. McLeod. (1999). Semantic heterogeneity resolution in federated databases by metadata implantation and stepwise evolution. The VLDB Journal, 8: 120-132.

Becker, J., C. Brelage, K. Klose, and M. Thygs. (2003). Conceptual modeling of semantic navigation structures: the MoSeNa-approach. In: WIDM'03, 118-125. New York: ACM Press.

Duval, E., W. Hodgins, S. Sutton, and S. L. (2002). Weibel. Metadata principles and practicalities. 8(4): http://www.dlib.org/dlib/april02/weibel/04weibel.html

Dushay, N. (2002). Localizing experience of digital content via structural metadata. In: JCDL'02, 244-252. New York: ACM Press.

Gilliland-Swetland, A. J. (2000). Defining metadata. In: M. Baca (Ed.), Introduction to metadata: Pathways to digital information. Los Angeles, CA: Gettry Information Institute. Also available: http://www.getty.edut/gri/standard/intrometadata/2_articles/index.htm.

Greenberg, J. (2001). A quantitative categorical analysis of metadata elements in image-applicable metadata schemas. Journal of the American Society for Information Science and Technology, 52: 917-914.

Hickey, T. and D. Vizine-Goetz. (2002). Implementing FRBR on large databases. Fall 2002 CNI Task Force Project Briefing presentation. http://staff.oclc.org/~vizine/CNI/OCLCFRBR_files/frame.htm

Jul, E. (1995). OCLC Internet Cataloging Project. D-Lib Magazine, December 1995: http://www.dlib.org/dlib/december95/briefings/12oclc.html

Kamel Boulos, M.N., A.V. Roudsari, and E.R. Carson. (2001). Towards a semantic medical web: HealthCyberMap Dublin Core Ontology in Protégé-2000. In: Proceedings of the Fifth International Protégé Workshop, Sowerby Centre for Health Informatics at Newcastle, England, July 18, 2001, http://protege.stanford.edu/ontologies/dublincore/hcm_dc_in_protege_newcastle.pdf

Khan, L., D. McLeod, and E. Hovy. (2004). Retrieval effectiveness of an ontology-based model for information selection. The VLDB Journal, 13: 71-85.

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

Lagoze, C., C. A. Lynch, and R. Daniel. (1996). The Warwick Framework: container architecture for aggregating sets of metadata. http://cs-tr.cs.cornell.edu:80/Dients/Repository/2.0/Body/ncstrl.cornell%2fTR96-1593/html.

Lagoze, C. and J. Hunter. (2001). The ABC ontology and model. Journal of Digital Information, 2(2): http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Lagoze/

Luna, J. J. (1992). Hierarchical, modular concepts applied to an object-oriented simulation model development environment. In: Proceedings of the 1992 Winter Simulation Conference, 694-699. New York: ACM Press.

Plassard, M.-F. (ed.) (1998). Functional Requirements for Bibliographic Records: Final Report. UBCIM Publications - New Series Vol. 19. K . München: G. Saur.

Qin, J. and J. Godby. (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. (2004a). Learning Object Ontology. http://web.syr.edu/~jqin/LO/LOV2/.

Qin, J. and N. Naybell. (2004b). Ontological representation of learning objects: building interoperable vocabulary and structures. In: WWW2004, May 17-22, 2004. New York: ACM Press.

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

Rockley, A. (2003). Managing Enterprise Content: A unified Content Strategy. CMS Watch, http://www.cmswatch.com/Features/TopicWatch/FeaturedTopic/?feature_id=85.

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

Shreve, G. M. and M. L. Zeng. (2003). Integrating resource metadata and domain markup in an NSDL collection. Paper presented at the 2003 Dublin Core Conference, Sept. 28 -- Oct. 2, Seattle, WA.