CanCore Guidelines Version 2.0: Introduction

History of Introduction Document

Date	Version	Comment	Person
June 6, 2002	1.1	Based on IMS Learning Resource Meta-data 1.2.1	Sue Fisher
December 6, 2002	1.8	Based on 1484.12.1 LOM	Norm Friesen
August 13, 2003	1.9	Revisions incorporating feedback	Norm Friesen
November 20, 2003	2.0	Final revisions incorporating feedback	Norm Friesen

CanCore Guidelines Version 2.0: Introduction

By Dr. Norm Friesen, Sue Fisher, and Anthony Roberts.

CanCore Guidelines Overview

The Learning Object Metadata standard (IEEE 1484.12.1-2002 or LOM; also known as IMS Learning Resource Meta-data) is both complex and general in character, contains a broad range of elements, and leaves open many possibilities for interpretation. CanCore seeks to simplify and interpret this standard in order to help implementers and record creators with design, development, and indexing work. CanCore began this task by identifying a subset of the many elements in the LOM—selecting them on the basis of their simplicity and their utility for resource discovery and sharing. The current document, however, provides recommendations for the semantics and syntax of *all* LOM elements based on best practices in indexing and metadata communities. CanCore provides guidance on interpretation and implementation at a level of detail much greater than the normative information provided in the LOM—but in no cases does CanCore seek to diverge from the general, normative framework provided by the LOM. If CanCore's recommendations are used as a basis for LOM interpretations and implementations generally, the potential for the interoperability of LOM implementations will be greatly increased.

This introductory document addresses the following topics:

Definitions: Metadata, Learning Objects, and the LOM

Metadata, or data about data, acts similarly to a card or record in a library catalog, providing controlled and structured descriptions for books, recordings, and "resources" generally through searchable attributes such as title, author, date, location, description, and subject. And also like a library catalog record, a metadata record can be located separately from the resource it describes, or packaged with (or embedded within) that resource.

This document interprets and refines a particular metadata standard developed specifically for the description of educational objects, or reusable digital educational resources. In this document, the term Learning Resource is used. This is done to avoid the controversy and confusion surrounding the term Learning Object (Friesen, 2003), and to reflect the fact that CanCore can be used to describe things *other* than those that conform to object-oriented design principles.

As mentioned above, the metadata standard with which CanCore is primarily concerned is known as the Learning Object Metadata (LOM) standard, and is equivalent to the IMS Learning Resource Meta-data specification. The LOM standard comprises a multiplicity of parts: A datamodel, which defines the metadata elements and their general characteristics; and a number of bindings, which indicate how these elements are to be encoded in XML and other formats.

What Is CanCore?

CanCore's Approach to Metadata

CanCore's understanding of metadata is based on conventional and emerging practices and techniques in library and information sciences. CanCore believes that much can be gained by combining the practices and solutions developed by librarians, catalogers, and indexers for long-standing information management problems.

One metadata standard other than the LOM itself that has been especially significant for CanCore is the Dublin Core Metadata Element Set (DCMES). This element set "was created to provide a core set of elements that could be shared across disciplines or within any type of organization needing to organize and

classify information" (Dublin Core Metadata Initiative, 2002b). The CanCore guidelines make significant reference to normative statements, recommendations, and documents issued by the DCMI and members of the Dublin Core (DC) community in general.

CanCore's element subset has been defined in keeping with the DCMI's minimalist approach. According to this approach, minimizing the variety and complexity of metadata elements is seen as ultimately benefiting metadata creation and implementation, and as assisting in resource discovery. CanCore sees this as applying to learning object metadata as well, and understands a law of diminishing returns as governing metadata use and creation. Like any other data management processes (such as data normalization or the control of information quality), the creation of metadata requires an investment of resources. However, the relationship between investment in metadata creation and the resulting level of resource discoverability is not linear. The more elements from a metadata set that are implemented, the greater the investment of resources that is required. In addition, the more data elements used, the greater the chances for error and divergence among record creators and implementations. CanCore's work proceeds from the assumption that a few well-chosen and well-implemented metadata elements will enhance resource discovery in a cost-effective manner. Consequently, CanCore has defined a subset of 46 "active" LOM elements.

Semantic versus Syntactic Interoperability

Another important aspect of CanCore's understanding of metadata is captured in the word interoperability: "the ability of two or more systems or components to exchange information and to use the information that has been exchanged" (IEEE, 1990). For metadata to be shared and used, both within and between communities of practice, at least two aspects of interoperability need to be supported:

1. The syntax and the protocols used to encode and transmit metadata records have to be specified in an unambiguous and consistent manner across systems.

2. Within the resource description that is the metadata record itself, ambiguity needs to be minimized, and common ways of using and describing learning resources need to be established and formalized.

Much work has been done in the area of syntactic and protocol interoperability. Prominent examples of this include IMS work in developing and maintaining bindings for LOM, and the Metadata Harvesting Protocol of the Open Archives Initiative. Less has been done to develop a consensus and refine understandings regarding LOM semantics. There are precise ways of specifying the XML tags that make up a LOM record, but few precise ways of specifying the values that should go in between many of these tags.

Because semantic interoperability relies on human languages—which are by their very nature ambiguous, inexact, and incomplete—such interoperability can only be achieved incrementally, by degrees. The way it is achieved, moreover, is quite different from ways in which syntactic and protocol interoperability are attained. Common understandings of terms and meanings must be negotiated and made explicit within communities of practice. The *CanCore Guidelines* document is the product of a process of consultation with a loosely constituted community of practitioners. This process is ongoing as the guidelines, documents, and the interpretations they contain are updated and validated. These guidelines also provide a basis for the further development of this consultation. It is only by establishing shared meanings where possible (and by acknowledging the limits of common semantic ground where consensus is not possible) that semantic interoperability can be optimized.

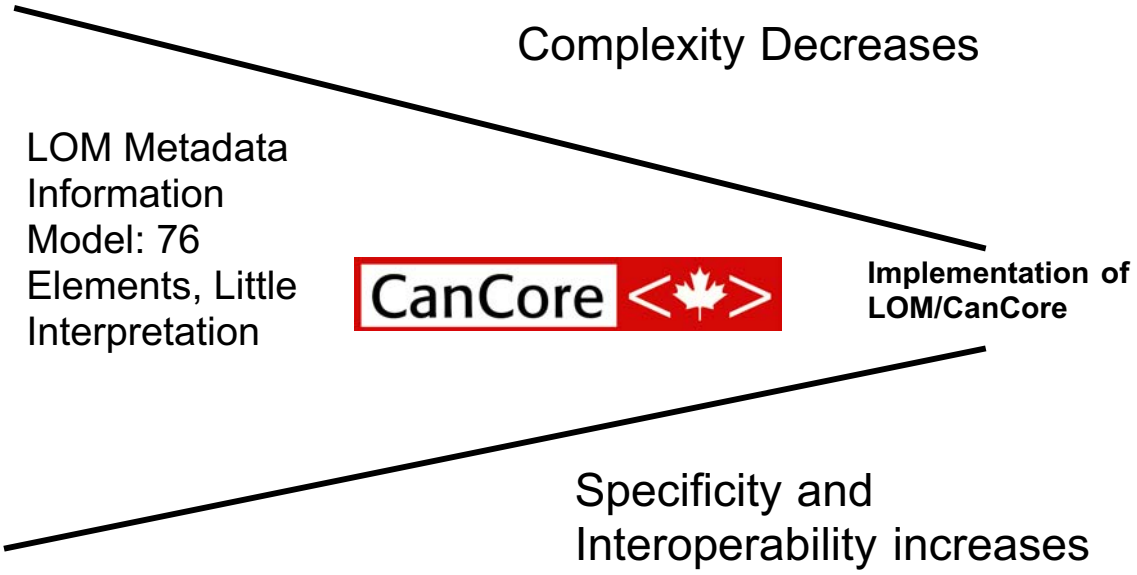
Besides serving as a guide for decision-making in the implementation process, the CanCore guidelines can serve as a straw man for teasing out differences and ambiguities in understanding, definition, and use between different communities and user/implementer groups.

For more about semantic interoperability and CanCore, see Friesen, 2002.

CanCore as an Application Profile

In choosing a subset of elements from the LOM standard, and in explicitly specifying the meaning of all of the LOM elements, CanCore is developing a Metadata Application Profile. In a document written jointly by representatives of the LOM and Dublin Core communities, an application profile is defined as "an assemblage of metadata elements selected from one or more metadata schemas and combined in a compound schema" (Duval, Hodgins, & Weibel, 2002). In the case of CanCore, these elements have been chosen from only one metadata schema.

However, CanCore has done much more than select elements. CanCore provides a great deal of fine-grained information about each element in the LOM—information that takes the form of recommendations, examples, and references to other interpretations. In this sense, CanCore represents an application profile that is perhaps more accurately captured in an earlier definition of this term provided by Clifford Lynch: “customizations of [a] standard to particular communities of implementers with common applications requirements" (Lynch, 1997). This diagram schematizes CanCore's approach:



The CanCore element subset, as well as its best practice guidelines, has been created through close consultation with participants in a variety of public projects and organizations. These have included, at various stages:

- Academic Technologies for Learning, University of Alberta
- Alberta Learning
- Athabasca University
- British Columbia Open University
- Centre de recherche LICEF
- CETIS
- Department of National Defence, Canada
- Eisenhower National Clearinghouse
- Electronic Text Centre, University of New Brunswick
- ARIADNE Foundation for the European Knowledge Pool
- Galbraith Media
- Learning and Teaching Scotland
- Library and Archives Canada
- Manitoba Education and Youth
- Memorial University of Newfoundland
- Ontario Ministry of Education
- Open Learning Agency of British Columbia
- TeleEducation NB
- University of Calgary

This predominantly public-sector participation does not imply that many of the indexing and data management best practices incorporated into the *CanCore Guidelines* are not relevant to other sectors and applications. Nor does it mean

that the subset of elements identified by CanCore might not also be useful for business and training applications of LOM. While users in these other sectors might decide to depart from some of CanCore's recommendations and choices, their decisions would likely benefit by referencing the rationale and interpretations explicated and disseminated in the *CanCore Guidelines*.

For more about the policies and projects motivating the development of CanCore, see Friesen, Mason, and Ward, 2002.

CanCore as an Element Subset

CanCore recommends the use of only certain LOM elements for the purposes of interoperation in distributed environments. CanCore recommends these elements on the basis of two criteria:

- their utility for resource discovery and description, as well as metadata record-sharing between independent projects and repositories that have implemented the LOM
- their use in other application profiles and by implementers participating in the development of the CanCore application profile

However, these recommended elements are to be understood as a subset only in a limited sense.

This subset is used in clearly different ways in the context of record-creation versus the context of metadata record storage and sharing:

1. In the context of creating or populating metadata records, the CanCore element subset does not represent a set of mandatory elements, which must be filled out or supplied for each metadata record created. Any number of elements can be provided or omitted in the metadata creation process. No elements are mandatory, either in the LOM itself or in the CanCore element subset. Also, the CanCore element subset has not been developed with the intention of limiting the kinds or number of elements that can be used in the metadata creation process. Record creators or implementers are welcome to include elements outside of the CanCore subset according to their local requirements. These elements could be other LOM elements, extensions to the LOM element set, or from other metadata specifications or requirements. As mentioned above, the CanCore element subset has been developed to simplify LOM implementation, with each element being selected on the basis of its utility for resource discovery and sharing.

2. For metadata storage and sharing, the CanCore subset is not intended as an acceptable minimal element set for systems that support the storage and sharing of metadata. These systems should be able to process, store, and share *all* of the elements in the LOM, not just the elements identified

in the CanCore subset. This will help to ensure that systems supporting CanCore will be able to interoperate with systems that support the LOM in general. It will also help to ensure that data elements not included in the CanCore subset will not be lost when LOM records are transmitted from one system to another. To maintain this type of interoperability, and to ensure that all the LOM data elements are included in implementations, CanCore has deliberately *not* made available any binding document or schema that *just* incorporates those elements in the CanCore subset.

Different metadata elements become important in different contexts. More obvious examples of some of these contexts are represented by the creation, search, and display of metadata records. This CanCore element subset represents a comparatively large group of LOM elements that are likely to be useful in a number of contexts. Smaller subsets will be useful in more specialized or narrowly defined contexts (for example, for searching, for search results display, for browsing according to resource interrelationships, etc.). This is illustrated in the table of elements below. This table shows the elements selected and recommended for the CanCore subset, as well as suggested, smaller element groups that *could* be derived from this subset:

CanCore Element Subset		Possible Search Subset	Possible Results Subset	Automatically Supplied Values	Values Supplied by Record Creator
No.	Element Name				
1	General				
1.1	Identifier				
1.1.1	Catalog				
1.1.2	Entry			Yes	
1.2	Title	Yes	Yes		Yes
1.3	Language	Yes	Yes		Yes
1.4	Description	Yes	Yes		Yes
1.5	Keyword	Yes	Yes		Yes
1.8	Aggregation Level		Yes		Yes
2	Life Cycle				
2.1	Version				Yes
2.3	Contribute				
2.3.1	Role		Yes		Yes
2.3.2	Entity	Yes	Yes		Yes
2.3.3	Date				Yes
3	Meta-Metadata				
3.1	Identifier				
3.1.1	Catalog			Yes	
3.1.2	Entry			Yes	
3.2	Contribute				
3.2.1	Role			Yes	
3.2.2	Entity			Yes	
3.2.3	Date			Yes	
3.3	Metadata Schema			Yes	

CanCore Element Subset		Possible Search Subset	Possible Results Subset	Automatically Supplied Values	Values Supplied by Record Creator
No.	Element Name				
3.4	Language			Yes	
4	Technical				
4.1	Format	Yes	Yes	Yes	
4.2	Size		Yes	Yes	
4.3	Location		Yes	Yes	
4.6	Other Platform Requirements		Yes		Yes
4.7	Duration				Yes
5	Educational				
5.2	Learning Resource Type	Yes	Yes	Yes	
5.3	Interactivity Level		Yes		Yes
5.5	Intended End User Role	Yes	Yes		Yes
5.6	Context	Yes	Yes		Yes
5.7	Typical Age Range	Yes	Yes		Yes
5.9	Typical Learning Time		Yes		Yes
5.11	Language				Yes
6	Rights				
6.1	Cost	Yes	Yes		Yes
6.2	Copyright and Other Restrictions		Yes		Yes
6.3	Description				Yes
7	Relation				
7.1	Kind			Yes	
7.2	Resource				
7.2.1	Identifier				
7.2.1.1	Catalog			Yes	
7.2.1.2	Entry			Yes	
8	Annotation				
8.1	Entity				Yes
8.2	Date			Yes	
8.3	Description				Yes
9	Classification				
9.1	Purpose			Yes	
9.2	Taxon Path				
9.2.1	Source			Yes	
9.2.2	Taxon			Yes	
9.2.2.1	Id			Yes	
9.2.2.2	Entry	Yes	Yes		Yes
9.4	Keyword	Yes	Yes		Yes

Conceiving of element subsets in this way helps to address one of the problems that the LOM initially presents to implementers and record creators alike: How to

deal with all 76 elements of the LOM or the 61 total (46 "active") elements in the CanCore subset in creating records and presenting them to users. The search and results subsets indicate that 13 and 21elements (respectively) may be used for these purposes. Only 21active elements may need to be supplied by human record creators, and 20 active elements may be supplied automatically.

In the CanCore subset, as in the LOM itself, every element is optional. CanCore does not otherwise identify elements as Optional or Mandatory. Given the diverse needs of the e-learning community, it seems impractical to stipulate an element set that would be satisfactory for every context. Some have argued, for example, that at a minimum the Title element should be required. While this is logical for repositories that deal with autonomous, named resources, it does not make sense for projects that deal with streaming video or other dynamically created, interdependent learning resources.

While every element is optional, certain principles should be brought to bear on the use of elements. If an element is used, at least one value or data-bearing sub-element should be included within it. If a pair or grouping of elements makes sense only when they appear together (as is the case with the sub-elements in Classification), then the elements should be selected and used as a group.

CanCore and Metadata Creation Tools

CanCore sees the development of tools to support the creation of high-quality LOM as indispensable to the successful implementation of the LOM standard. However, the creation of such tools must *not* be focused on the needs of LOM users as a single homogenous group. Instead, these tools must be developed in such a way that they directly address (or can be adapted to address) the requirements of particular user groups and local implementations. An example of a metadata creation tool that addresses the needs of a particular project and set of users is provided by CUBER (personalized Curriculum Builder for a Federated Virtual University of the Europe of Regions; see http://www.cuber.net/Test-Prototyp/ai/guide_start.jsp). Examples of tools that present the complexity, flexibility, and specialization of the LOM structures and encodings directly to end users are numerous (see http://www.cancore.org/lomsurvey.html for examples of these).

As a result of this need for adaptation and simplification, the CanCore guidelines are themselves not intended as a generic set of rules and recommendations for end users generally. They are instead intended to inform the development of metadata creation and management tools, and to serve as the basis for documents and supports that would be used in their implementation. (For more about the audience for which these guidelines are intended, see the Who Is This Document For? section below.)

As this introduction is being written, new technologies are being developed or implemented that may change the way metadata is created in practice. For

example, artificial intelligence technologies may soon be able to populate a number of the fields of a LOM record. Additionally, semi-automated metadata components already integrated into content creation tools may also be able to supply values for certain LOM elements (for example, it may be possible to derive some LOM elements from the document properties information that is already automatically generated by some popular word-processing programs).

However, this guidelines document is written at a time when these types of solutions are either not readily available in educational communities, or are not designed or used specifically for the purposes of LOM record-creation. This guidelines document seeks to address the present needs and requirements of public institutions, projects, and educators. Consequently, the understanding of metadata that informs it is one that emphasizes human agency and interpretation, as well as existing classification practices, rather than mechanisms of automation. Others have advocated a similar understanding or approach in the metadata world in general (for example, see Milstead & Feldman, 1999; Lucas, 2000).

Avoiding Common Metadata Misconceptions

LOM has some important characteristics that it does *not* share with some other metadata specifications.

Perhaps most importantly, many aspects of LOM are less suited for direct exposure to end users than those of the Dublin Core and some other metadata specifications. These aspects include metadata element titles and the hierarchical structure in which these elements are embedded, as well as the encoding required for a number of element values. Tools for creating LOM often seem to suggest that presenting users with an expandable, hierarchical listing of LOM element titles with text boxes and other form components is acceptable in practice. However, a brief consideration of one or two LOM elements reveals this to be unadvisable—except for the most highly trained end users. For example, to enter something as simple as an author or creator of a learning resource, while referring only to the terms and structures associated with the LOM, requires the following:

1. 2.3.2:LifeCycle.Contribute.Entity must be recognized as the correct element for this information.

2. Author must be ensured to be the most appropriate value for the element 2.3.1:LifeCycle.Contribute.Role, from a list that includes values such as Script Writer, Subject Matter Expert, and Publisher—none of which are unambiguously defined or differentiated from Author.

3. The specialized vCard (virtual business card) markup required for this element must be entered. This may involve: 1) locating an existing vCard record; 2) using a separate vCard creation tool (whose output would then

have to be coordinated with the LOM tool); or 3) entering vCard markup and values manually. In the latter two cases, the user would be required to decide which of the dozens of vCard elements are appropriate for the record.

LOM also does *not* provide a means for the management or enforcement of digital rights associated with a learning resource or for version tracking and control. Although LOM does provide elements that address some of these concerns very generally, these elements do not do so in a way that is sufficient for the requirements of many projects. This is most conspicuous in the case of rights management, for which LOM provides only four elements, none of which provide the possibility of inserting a machine-readable linkage to a rights-management service or expression document. The emerging vision for managing these types of information is one involving distributed, specialized, modular metadata records and lookup services.

Finally, LOM also does *not* encapsulate or otherwise imply a specific model for learning resource content. It does not provide a detailed account, for example, of the way that this content is aggregated (into lessons, units, courses, etc.) or of the particular ways that learners can interact with this content. As in the case of rights management, LOM provides elements that address such concerns on a very general level. However, these elements do not address these issues on the level of specificity that is available through other specifications and forms of expression. (Examples of content models that provide this level of detail include Educational Modelling Language and IMS's Learning Design specification).

The History of CanCore

The CanCore Initiative was established in November 2000 to address common concerns regarding information management and resource discovery within a number of Canadian public-sector e-learning projects. These include the Broadband Enabled Lifelong Learning Environment (BELLE) and Portal for Online Objects in Learning (POOL) projects, sponsored by Industry Canada, a department of the Canadian federal government. These projects also include the Campus Alberta Repository of Educational Objects (CAREO), the LearnAlberta.ca portal, and TeleCampus, sponsored by provincial educational ministries. All of these entities, together with the Electronic Text Centre at the University of New Brunswick, were the founding partners of CanCore. Their key concern in forming CanCore was to synthesize efforts with respect to metadata creation and sharing.

Since its inception, CanCore has:

- devised a workable subset of the LOM, known as the CanCore Element Set (http://www.cancore.ca/elementset1.1.html);
- become a participant in the IMS Global Learning Consortium, Inc. through the sponsorship of Industry Canada;

- become a participant in the IEEE LTSC (Institute of Electrical and Electronics Engineers, Inc. Learning Technology Standards Committee) through the sponsorship of Athabasca University;
- become a part of the Canadian delegation to the International Standards Organization (ISO) Committee for Information Technology for Learning, Education, and Training (ISO/IEC JTC1/SC36, the committee that is reviewing the possibility of making the LOM an ISO standard);
- developed informal ties with the Dublin Core Metadata Initiative; and
- written and presented numerous papers in the field of LOM (see the Bibliography below).

CanCore's Guidelines Document

Document History

This guidelines document itself has developed through two phases. In the first phase, a draft was produced in June 2002 based on the IMS version 1.2.1 Learning Resource Meta-data specification. In the second phase, this draft was updated during 2003 to reflect the version of LOM that has been approved as a standard by the IEEE (note that the IMS specification will be revised and made identical with the IEEE standard).

During both phases of its development, a guidelines working group contributed to the creation of the guidelines document. This group comprised the following individuals and organizations:

Pierre Bernard, Centre de recherche LICEF
Markiana Eliuk, Slavic Studies Program, University of Alberta
Sue Fisher, Electronic Text Centre, University of New Brunswick
Deb Fralick, Alberta Learning
Norm Friesen, CAREO and CanCore
Louis Guerette, Centre de recherche LICEF
Carolyn Guinchard, Alberta Learning
Brian Lamb, University of British Columbia
Karin Lundgre, Centre de recherche LICEF
Mike Magee, University of Calgary
Gerry Paille, Open Learning Agency
Anthony Roberts, TeleEducation NB
Lori Tozer, Electronic Text Centre, University of New Brunswick

Who Is This Document For?

The intended audience for this document is diverse: Systems administrators, metadata managers, and individual record creators will all benefit by making use of this document. On its own, this document is *not* intended as a reference for end users who may be creating or searching metadata. Although it provides recommendations that can be used as tips or cues for populating or searching metadata fields, any such recommendations should be checked for their

appropriateness in the context of a particular project, site, or implementation. The administrators, managers, implementers, and record creators who use this documentation will hopefully do so as a part of a decision process of implementation and application. These guidelines are intended to support these processes, and to provide a solid basis for creating end user documentation and other supports.

Possible scenarios in which these guidelines are put to effective use—each derived from real-life implementation and support requests directed to CanCore—include the following:

1. A group of French language or immersion schools in a particular jurisdiction are developing a learning resource collection for internal use. As a part of this process, they are surveying standards-based tools and support available for creating such a collection. They are particularly interested in integrating a complex, hierarchical set of learning outcomes set by regional governance authority, and want to ensure that these learning outcomes are appropriately accommodated in the LOM. General documentation available on the CanCore Website can assist them in their survey of tools and other supports. The CanCore guidelines provide them with recommendations and guidelines for creating standards-based educational metadata for their collection. These guidelines also provide specific recommendations and examples derived from work in other jurisdictions related to the integration of hierarchically ordered learning objectives into the Classification element group in the LOM.

2. An existing, national database wishes to align itself with e-learning standards. However, in doing so, those responsible for its maintenance are wondering about the data normalization, element title translation, and other crosswalking issues that may be involved but are not addressed in detail by standards documents. The CanCore guidelines provide these implementers with interpretations of element titles and meanings, and with detailed definitions of vocabulary terms used in the LOM. These guidelines also provide recommendations regarding the formatting and other characteristics of element values. In this way, reference to CanCore maximizes the potential for the creation of interoperable metadata through the crosswalking process.

3. A private firm with existing e-learning systems and content wishes to become compliant with the Sharable Content Object Reference Model (SCORM). The technical management at the firm has acquired a SCORM test suite, but is looking to provide in-depth support for the creation and maintenance of metadata. Because the CanCore guidelines provide recommendations for all of the metadata elements defined as mandatory for SCORM content, CanCore's guidelines document can provide substantial assistance in this. The firm decides to develop and integrate a

> well-documented, semi-automated interface for metadata creation with their company's systems. The firm is able to adapt CanCore's descriptions of the purpose and application of various metadata elements for use in its documentation, and refers to CanCore's recommendations to provide appropriate defaults and data-checking features for the interface.

Why Are Best Practice Guidelines Important?

Guidelines such as the ones provided in this document are indispensable for effective implementation, and have been created for a wide variety of metadata specifications and implementations. Examples include the Dublin Core usage guide (entitled "Using Dublin Core"; http://www.dublincore.org/documents/usageguide/), the CIMI "Guide to Best Practice: Dublin Core" (http://www.cimi.org/public_docs/meta_bestprac_v1_1_210400.pdf), and the Metadata Object Description Schema user guidelines ("MODS User Guidelines Version 2.0"; http://www.loc.gov/standards/mods/). In each of these documents, an explication of the data model at hand is followed by descriptions, interpretations, and examples provided on an element-by-element basis. The first of these, the DC usage guide, explains its purpose in terms of consistent metadata creation:

> [An] important goal of this document is to promote "best practices" for describing resources using the Dublin Core element set. The Dublin Core community recognizes that consistency in creating metadata is an important key to achieving complete retrieval and intelligible display across disparate sources of descriptive records. Inconsistent metadata effectively hides desired records, resulting in uneven, unpredictable or incomplete search results. (DCMI, 2003)

Especially in the case of the CIMI documentation, significant reference is made to best practices as they have emerged in the field of cataloging and indexing, and as they are encoded in cataloging rules, such as the Anglo-American Cataloguing Rules, 2nd edition (AACR2).

The complexity of the LOM, as well as its widespread adoption, would seem to underscore the need for similar guidelines in the e-learning community. The apparent lack of publicly available, normative interpretation and explication of LOM elements represents a conspicuous gap that this document hopes to address.

Characteristics of the LOM/CanCore Datamodel

General Characteristics

As mentioned earlier, the LOM datamodel defines 76 data elements, covering a wide variety of characteristics attributable to learning resources and digital resources. The CanCore subset of recommended elements reduces this total to

61 elements, 46 of which are "active" or can be assigned values by record creators and systems.

Hierarchical: The elements in the LOM/CanCore datamodel are structured in relationships that are hierarchical. This means that there are a number of elements that are not directly assigned any values by record creators or indexing systems, but that simply contain other elements located on a different level in the hierarchy. Prominent examples of these container elements are provided by the nine main data elements in the LOM (General, Life Cycle, Meta-Metadata, Technical, Educational, Rights, Relation, Annotation, and Classification). Together, these nine elements or categories contain all of the other elements in the LOM. Such elements are more generally known as aggregate elements or element aggregations. The place of an element in a particular hierarchical structure is indicated by the decimal places in its number and in its full name (for example, the element 1.1.2:General.Identifier.Entry is located on the third level of the hierarchical structure).

Element Repetitions: Many of the elements in the LOM can be repeated to accommodate multiple values. If these elements contain further, hierarchically ordered elements, this means that these additional elements are repeated as well. More than one repeatable element in an aggregate hierarchical structure of elements can result in combined repetitions that can be complex in their structure and implications. CanCore's guidelines address these complexities and implications.

Size refers to the number of times that an element should be capable of being repeated.

Smallest Permitted Maximum (spm) relates to both the number of repetitions of an element (or element group) and to the number of characters (as specified in the **datatype**) to be accommodated by an element. It has different implications for record creators and systems developers:

- Record creators can use the number of characters and element iterations identified in the spm to describe a resource, with the assurance that distributed systems will preserve this information.

- Systems developers should ensure that their systems can store, transmit, and process at least the number of characters and iterations specified as the spm.

The spm does *not* refer to the maximum limit on the number of element repetitions or characters that the element accommodates. Instead, it specifies the smallest number of repetitions or characters that systems must be able to accommodate for each element. These systems must be capable of storing, transmitting, and processing at least the number of repetitions or characters

specified in the spm. The capabilities of these systems should not be limited to anything *less* than these numbers.

In the context of record creation processes for distributed systems, this number implies that systems and persons creating records can rely on at least that number of repetitions and characters being available to them. The spm in no way requires record creators to supply the number of characters or element iterations specified. For example, the element 1.4:General.Description has a size spm of ten with a datatype spm of 2000; this means that a record can be created with a single description of, say, 1998 characters, and up to ten separate descriptions can be provided with confidence in the fact that distributed systems will accommodate all of these characters and repetitions. More repetitions or characters could be added, but not with the assurance that they would be preserved by different, distributed systems.

Ordered versus Unordered: For ordered repetitions of elements, or groups of elements, the sequence of these element iterations is significant. Iterations are typically sequenced in terms of descending importance or priority.

Unordered implies that a list of repetitions is *not* significant, or is *not* in any way indicative of the importance or priority of the values listed. Where an element repetition is *ordered* and the order is *not* significant, CanCore recommends that repetitions or values simply be provided in alphabetical or numeric order.

LOM Datatypes

According to the LOM, a datatype is "a property of distinct values, indicating common features of those values and operations on those values" (IEEE, 2002). Simply put, the term datatype typically indicates the structure, quantity, and/or the general kind of content that is associated with a given element. For example, some elements require that a date be indicated; for these, a DateTime datatype is used. With the exception of the CharacterString datatype, each datatype in the LOM is compound in character: Each allows for the inclusion of two or more discrete values or pieces of information in a single metadata element. As an example, the LangString datatype allows both for an indication of the human language used and for the string of characters making up the human language expression itself.

The five datatypes in the LOM are as follows:

CharacterString
Typically, a value that is (at least to some degree) linguistically neutral, and that serves as a formal identifier or name (for example, ARIADNE, URL, ftp://cogito.ergo.su). In most cases, this datatype is to be used with the Unicode character set, or more specifically, with "the repertoire of ISO/IEC 10646-1:2000" (IEEE, 2002).

LangString
A combination of a language indicator and a string of characters formulated using the human language indicated. According to the LOM, LangString "represents one or more character strings. A LangString value may include multiple semantically equivalent character strings, such as translations or alternative descriptions" (IEEE, 2002).

The LangString datatype structure allows for repetitions for translated versions of its contents or values within the datatype itself, without any repetition of the element with which it is associated.

This datatype is to be used with the Unicode character set, or with "the repertoire of ISO/IEC 10646-1:2000, excluding the NUL-character (UCS character U00000000)" (IEEE, 2002). Either a two- or three-letter language code (ISO 639.1 or ISO 639.2) can be used to indicate the human language used in LangString. See the CanCore recommendations for 1.3:General.Language for more information.

DateTime
This datatype is a combination of two parts:

1. One expresses date and time in a prescribed, formatted manner. This part is based on ISO 8601:2000; a standard which is provided in summary form at http://www.w3.org/TR/xmlschema-2/#isoformats.

2. The second is a non-formatted, textual description of the date and time if these cannot readily be expressed or if textual information is needed to supplement the ISO 8601 value.

According to ISO 8601, this datatype should contain date information expressed in the YYYY-MM-DD format. For general or partial dates, values can be eliminated as necessary from right to left: YYYY (a year date only), YYYY-MM (a year and month date only).

Duration
Like DateTime, Duration is a combination of two parts: One for formalized, formatted expression of duration; and another for a non-formatted, textual description of the duration (if it cannot be otherwise expressed or if textual information is needed to supplement the formatted value).

According to the LOM datamodel, Duration (like DateTime) is based on the ISO 8601:2000 standard. The significant portions of this standard for expressing duration are summarized at http://www.w3.org/TR/xmlschema-2/#duration).

According to the LOM datamodel, the formatted portion of this datatype should contain duration information expressed, for example, as "P1M2DT1H20M25.55S." As an example that uses most of the datatype elements, the duration information expressed is equivalent to 2 days, 1 hour, 20 minutes and 25.55 seconds. In this example, P "is used as the time duration designator, preceding a data element representing a given duration of time" (W3C, 2001; see http://www.w3.org/TR/xmlschema-2/#isoformats). D stands for day, T for time, H for hour, M for minute, and S for second (or fraction thereof).

Vocabulary
This datatype is used for controlled vocabularies, or lists of terms or numbers that are specified in advance for a particular element. In a Web-form or other human interface, these might be presented as a pull-down menu or pick-list.

This datatype is a combination of two parts: The first identifies the source of the vocabulary or list of values, and the second is for the controlled vocabulary value listed. The vocabulary values are typically terms or short strings (using lowercase characters), or integers. A Vocabulary datatype labelled in the LOM datamodel as Enumerated refers to integers or other values that can be ordered (for example, from least to greatest; from 1 to 5); a Vocabulary datatype labelled as State in the LOM cannot be ordered in these ways. The source of the vocabulary is generally a namespace, URI, or URL. A common human-readable name of the vocabulary source can also be provided in addition to the source namespace, URI or URL. Ideally, this source namespace, URI, or URL should resolve to a list of acceptable values and their definitions. Elements that are subject to the Vocabulary datatype should reference openly referenced and maintained value sets.

In the case of certain elements, and for certain implementation purposes, the vocabulary values recommended for a particular element in the LOM may be suboptimal or unsuitable. In these cases, CanCore recommends that a local or custom vocabulary be used in conjunction with the LOM vocabulary as indicated below.

1. Use one of the vocabulary values recommended by the LOM. If no clearly suitable value is available, use the LOM value that, by virtue of its generality, connotative meanings, or other attributes, is at least not *un*suitable or misleading.

2. Repeat the metadata element, and provide both the source location and appropriate value(s) for the locally developed or custom controlled vocabulary.

(Note: This technique is based on the LOM recommendations for 5.6:Educational.Context: "Suggested good practice is to use one of the values of the value space and to use an additional instance of this data element for further refinement"; IEEE, 2002.) Where this technique is not possible (for example, where only one instance of the element is permitted), CanCore strongly recommends that the LOM vocabulary is used, but also recognizes that the use of other vocabularies may sometimes be unavoidable.

Value Spaces

Value Spaces is used in the LOM to indicate the specific kind of values or content (for example, a type of character set) that can be associated with a particular element. These often correspond to the datatype specified:

CharacterString In some cases, the datatype of CharacterString corresponds to a Value Space of "repertoire of ISO/IEC 10646-1:2000" (essentially equivalent to the Unicode character set).

Vocabulary In most cases, the datatype of Vocabulary corresponds to a Value Space that lists a series of words or numbers as recommended vocabulary values.

Langstring, DateTime, Duration The Value Space is unspecified.

LOM and Translation

There are a number of considerations important to keep in mind when implementing the LOM in non-English or multilingual contexts:

1. Translations of the LOM datamodel (element titles and vocabulary values) have been undertaken by the CEN/ISSS WS-LT (European Committee for Standardization/Information Society Standardization System WorkShop on Learning Technologies; http://www.cenorm.be/isss/Workshop/lt/). At the time of writing, these translations are at a preliminary stage, and have not been made publicly available in finalized form.

2. In the context of the XML binding for the LOM, LOM element names and the vocabulary values should not be translated. In an XML schema document or XML-formatted record, the original English versions of these names and values should be regarded as linguistically neutral strings or tokens. Equivalents in any language should be provided through the interface or any other mechanisms through which these element titles and vocabulary values are provided to end users. (French equivalents to some LOM vocabulary values are provided in the CanCore guidelines documentation as an indication of how this translation can be undertaken,

and for the purposes of avoiding semantic ambiguity that occurs when exclusively unilingual values are provided and defined.)

3. Identifying equivalents for elements and vocabulary values can be challenging, and may present barriers to semantic interoperability. For example, the common French translation for Publisher is *éditeur*, a word that, along with the French *rédacteur*, can also mean Editor. This difference in emphasis and association may create problems specifically for the translation and use of French values for the vocabulary values associated with 2.3.1:LifeCycle.Contribute.Role, and is illustrative of some of the kinds of challenges that translating values and titles can present. (In the French vocabulary values provided with these guidelines, this particular issue is addressed by equating both *éditeur* and *rédacteur* with the value Editor, and translating Publisher as *maison d'édition*.) Element titles and values provided in any language should be accompanied by definitions that clarify their meaning.

4. The CEN/ISSS WS-LT has released documents providing further guidance for the use of the LOM in multilingual contexts:

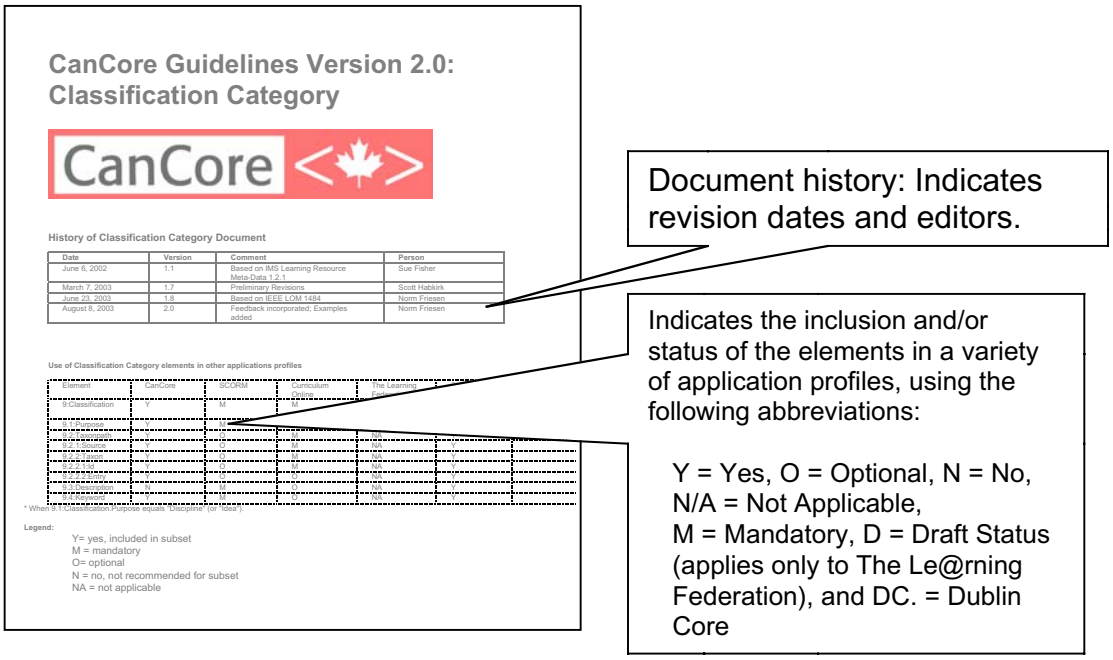
    - "The Internationalization of the IEEE Learning Object Metadata" (CWA 14643; http://www-gist.det.uvigo.es/~lanido/LOMInt/documentation/CWA-LOMINT-version-0-1-lanido.doc)
    - "Availability of Alternative Language Versions of a Learning Resource in the IEEE LOM" (CWA 14645; http://www.softlab.ntua.gr/~csgouro/cen-wslt/ALV-pre-final7.doc)

Guidelines Document Organization

The Nine Category Elements

The CanCore guidelines have been developed and structured around the nine element categories in the LOM (General, Life Cycle, Meta-Metadata, Technical, Educational, Rights, Relation, Annotation, and Classification). The guidelines for the elements in each category were originally developed as separate documents, and each can be downloaded individually, in PDF format from the CanCore Website (http://www.cancore.org).

In the CanCore documentation, the following information is provided for each element group as a whole:



Use in Element Subsets; Element Subsets Covered

SCORM: The Sharable Content Object Reference Model (at http://www.adlnet.org) profiles the LOM in addition to a number of other specifications that it integrates into an overarching model or architecture. It designates LOM elements as Mandatory or Optional for different types or levels of content aggregation. CanCore uses those designations provided by SCORM for what it calls Sharable Content Objects (SCOs), the most prominent type of content aggregation in the SCORM. SCORM documentation defines the SCO as "a collection of one or more [atomic] Assets that represents a single launchable learning resource that utilizes the SCORM Run-Time Environment to communicate with Learning Management Systems (LMSs) " (ADL, 2003).

Curriculum Online: This national, school-level intiative in the United Kingdom (http://www.curriculumonline.gov.uk/Default.htm) uses the terms Mandatory, Optional, Conditional (i.e., mandatory in some circumstances but not others), and Recommended. For some elements labelled Mandatory, the Curriculum Online documentation describes more precise circumstances under which this condition applies.

The Le@rning Federation: The Le@rning Federation [TLF; http://www.thelearningfederation.edu.au/tlf/] is an "initiative designed to create online curriculum content and infrastructure." This initiative takes a slightly different approach to metadata application profiling than the other initiatives covered here. It combines a significant number of elements from the LOM with elements from the Dublin Core and other metadata schemas. However, the elements included and excluded from the LOM

are still broadly comparable with those labelled as Mandatory, Optional, or otherwise in other LOM application profiles. In The Le@rning Federation application profile, those elements identified as Mandatory "must appear in metadata instances conforming to the profile." All other LOM elements included in The Le@rning Federation application profile are Optional. LOM elements that are simply not listed in this application profile are labelled in the CanCore documentation as Not Applicable (N/A).

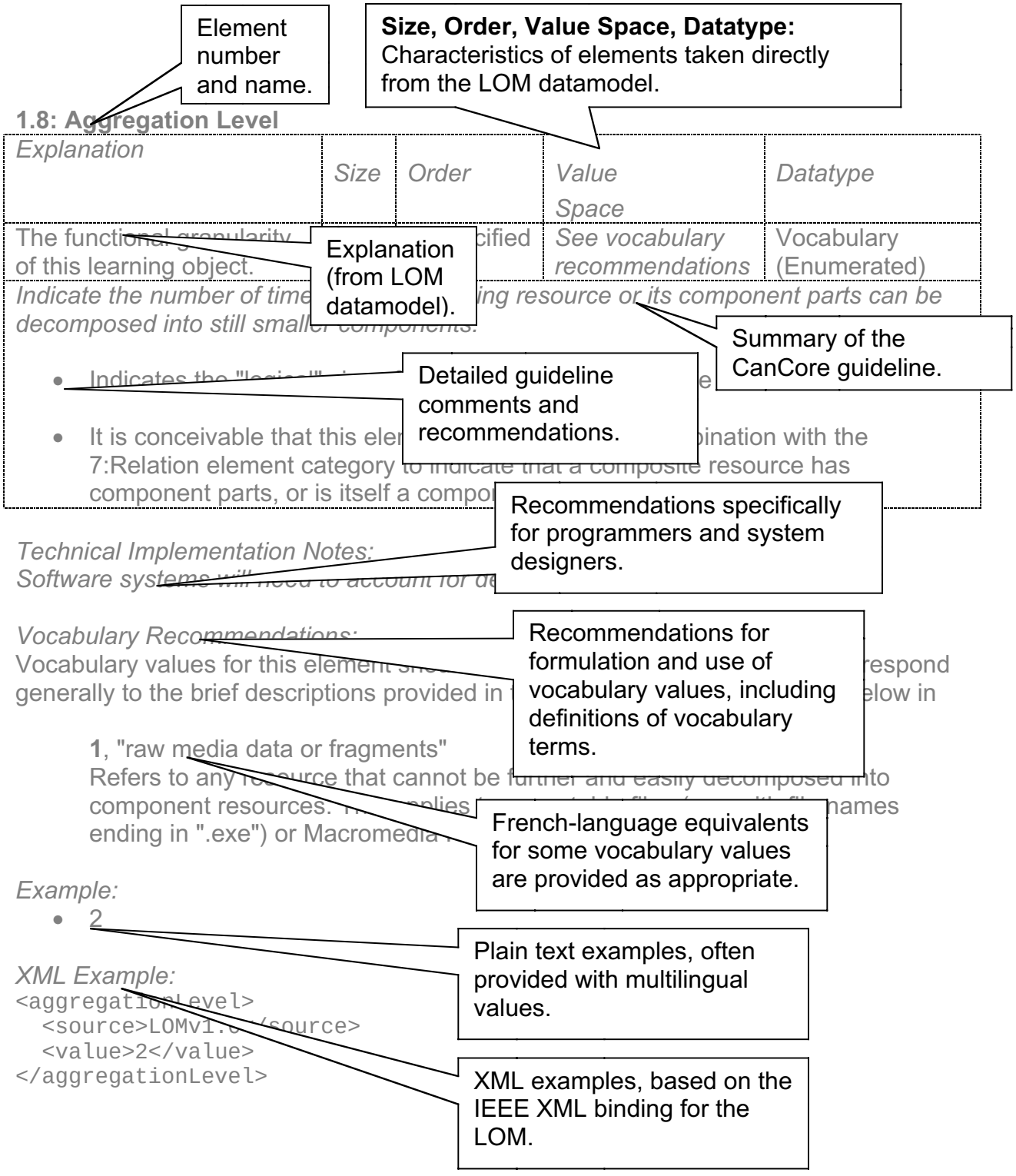
SingCORE The SingCORE initiative (http://www.ecc.org.sg/cocoon/ecc/website/singcore-17-jan-03.pdf) is described as "Singapore's Meta-data Schema for Labeling Digital Learning Resources." Its purpose is to customize the LOM standard for local Singapore needs.

UK LOM Core Like CanCore itself, the purpose of the uklomcore (United Kingdom Learning Object Metadata Core; http://www.cetis.ac.uk/profiles/uklomcore) is to "identify common practice and provide guidelines for metadata implementers, creators and users." Its use of the terms Mandatory and Optional is explained as follows:

> The UK LOM Core [UKCMF] identifies elements as being mandatory, optional and optional (recommended). Values must be supplied for all mandatory elements. Container elements are regarded as mandatory if one or more of their sub elements are also mandatory. Values for optional elements may be supplied if they are deemed to meet the requirements of the user community. It is recommended that optional (recommended*)* elements should be used where possible….
>
> It is recognised that many users will develop application profiles based on the [UKCMF] tailored to meet their own specific requirements. In order to maximise interoperability, all application profiles based on the [UKCMF] must contain all mandatory elements. Developers should publish their application profiles, including vocabularies, on a project website, for example, to ensure they are accessible to other [UKCMF] implementors (CETIS, 2003).

Dublin Core The scope and character of the DCMES (http://www.dublincore.org) is explained elsewhere in this document (see especially What Is CanCore? CanCore's Approach to Metadata).

Elements and Element Groups within the Nine Category Elements



General Indexing Best Practices

When creating a metadata record, always keep the following indexing principles in mind.

Resource Discovery

The ultimate function of a metadata record is to enable users to find relevant resources. When creating a metadata record, always keep the user in mind. It is easy to become preoccupied with the properties of the resource being described at the expense of the knowledge base, search habits, and search vocabulary of the end users. These latter concerns should always be foremost in a record creator's mind.

Specificity of Terms

Use the most specific term possible to describe the learning resource. For example, an article discussing treatment for acute mylogenous leukemia should be described using this specific term, rather that the more general leukemia or cancer. Specific terms derived from a controlled vocabulary can be traced upwards through a series of broader terms; they cannot necessarily be mined hierarchically downwards.

Relevance versus Recall

The use of the most specific term will help ensure high-relevance search results for the end user. Less specific terms will result in greater search recall (for example, multiple search hits). Often searchers will want high recall searchers, but it is far easier to move from more specific, relevant searches to more general high-recall searches.

Derivation of Keywords

When assigning free text keywords, try to balance out the following two principles:

1. Literary Warrant: Terms assigned should be inherent to the resource being described, rather than being assigned externally by the record creator.

2. The Language of the End User: If the metadata is being created for a specific user community, the record creator should anticipate the terms the user community would use to locate the resource.

Spelling

CanCore is not prescriptive about spelling conventions. Vocabulary values specified in the LOM that are not integers use exclusively lowercase characters. The choice of language used in the element 3.4:Meta-Metadata.Language

governs the entire metadata record unless otherwise specified. Keep this in mind when deciding upon and adopting spelling conventions.

Formatting

Do not include any formatting in a metadata record, as it frequently cannot be translated to the XML environment. Issues regarding base character sets, accented characters, or other characters non-standard to Unicode should be discussed with technical implementation staff, and best practices should be put in place at the local implementation level.

Referenced Documents and Specifications

As mentioned above, this document references the e-learning standard currently known as the IEEE LOM 1.0, or as IEEE 1484.12.1-2002. This document references the "LOM XML Schema" posted on 5 February 2003, and submitted to the IEEE LTSC for standardization. These schema documents, as well as supporting materials, are available at http://www.cs.kuleuven.ac.be/%7Eerikd/LOM/.

These guidelines were also developed with consideration of the following:

- the Dublin Core Metadata Initiative (http://www.dublincore.org/), particularly its element descriptions (http://www.dublincore.org/documents/dces/), qualifier descriptions (http://www.dublincore.org/documents/dcmes-qualifiers/), and usage guide (http://www.dublincore.org/documents/usageguide/)
- the CIMI “Guide to Best Practice: Dublin Core,” a comprehensive, community-specific (museums) set of guidelines for using Dublin Core metadata: available from http://www.cimi.org/public_docs/meta_bestprac_v1_1_210400.pdf

Bibliography

ADL. (2003). ADL SCORM version 1.3 application profile: Working draft 1.0. Retrieved from http://www.digitale-medien.ufg.ac.at/cms/scorm_app_profile.pdf.

CETIS. (2003). UK learning object metadata core: Draft 1.0. Retrieved from http://www.cetis.ac.uk/profiles/uklomcore/uklomcore_v0p1.doc.

DCMI. (2002a). DCMI frequently asked questions (faq). Retrieved from http://www.dublincore.org/resources/faq.

DCMI. (2002b). Dublin Core metadata element set, version 1.1: Reference description. Retrieved from http://www.dublincore.org/documents/dces/.

DCMI. (2003). Using Dublin Core. Retrieved from http://www.dublincore.org/documents/usageguide.

Duval, E., Hodgins, W., Sutton, S., & Weibel, S. L. (2002). Metadata principles and practicalities. *D-Lib Magazine, 8* (4). Retrieved from http://www.dlib.org/dlib/april02/weibel/04weibel.html.

Friesen, N. (2003). Three objections to learning objects. In R. McGreal (Ed., in press), *Learning objects and metadata.* London: Kogan Page.

Howe, D. (Ed.). (1993). Free on-line dictionary of computing. Retrieved from http://foldoc.doc.ic.ac.uk.

Institute of Electrical and Electronics Engineers. (1990). *IEEE standard computer dictionary: A compilation of IEEE standard computer glossaries.* New York: Author.

Institute of Electrical and Electronics Engineers. (2002). *1484.12.1 IEEE standard for learning object metadata.* New York: Author.

Lucas, M. (2000). Demystifying metadata. *Mappa Mundi*. Retrieved from http://mappa.mundi.net/trip-m/metadata/.

Lynch, C. A. (1997). The Z39.50 information retrieval standard. Part I: A strategic view of its past, present and future. *D-Lib Magazine, April.* Retrieved from http://www.dlib.org/dlib/april97/04lynch.html.

Milstead, J., & Feldman, S. (1999). Metadata: Cataloging by any other name. *ONLINE, 23* (1) Retrieved from http://www.onlinemag.net/OL1999/milstead1.html.

Pearsal, J. (Ed.). (2002). *Concise Oxford English dictionary.* (2nd Ed.). Oxford: Oxford University Press.

Simpson, J. A., & Weiner, E. S. C. (Eds.). (1989). *Oxford English dictionary.* (10th Ed.). Oxford: Oxford University Press.

W3C. (2001). XML schema part 2: Datatypes. Retrieved from http://www.w3.org/TR/xmlschema-2/#isoformats**.**

Referenced CanCore Publications

Friesen, N. (2002). Semantic interoperability and communities of practice. In J. Mason (Ed.), *Global summit of online learning networks: Papers* (pp. 104-107). Adelaide: Educationau. Retrieved from http://www.educationau.edu.au/globalsummit/papers/nfriesen.htm.

Friesen, N. (2001). CanCore: Learning object metadata. In *Computers and Advanced Technology in Education: Proceedings of the Fifth IASTED International Conference, June 27-29* (pp. 526-529). Anaheim: ACTA Press.

Friesen, N., Mason, J., & Ward, N. (2002). Building educational metadata application profiles. *Dublin Core - 2002 Proceedings: Metadata for e-Communities: Supporting Diversity and Convergence* (pp. 63-69). Retrieved from http://www.bncf.net/dc2002/program/ft/paper7.pdf.