Developing Community-Oriented Metadata Vocabularies: Some Case Studies

Shigeo Sugimoto[1], Wonsook Lee[2], Mika Murotani[2], Mitsuharu Nagamori[1], Mitsuyoshi Moriyama[3]

[1] Institute of Library and Information Science, University of Tsukuba, and University of Library and Information Science, Tsukuba, Ibaraki, Japan
[2] Graduate School of Information and Media Studies, University of Library and Information Science, Tsukuba, Ibaraki, Japan
[3] Okayama Prefectural Culture Center, Okayama, Okayama, Japan

Abstract

The Internet has made possible a global information environment where we can access information resources regardless of distance. However, current Internet technologies have not sufficiently matured yet to provide a community-oriented information infrastructure where a user can access information resources using terms and metadata schemas designed for his/her community. Metadata plays an important role to help realize a community-oriented information environment.

This paper presents two case studies which include the development of domain-specific subject vocabularies – a core subject vocabulary for a subject gateway for library and library-and-information science (LIS) resources, and subject vocabularies of a portal service for a regional community. These case studies show that small subject vocabularies are useful for these community-oriented services, and that maintenance is a crucial issue for the development and use of the vocabularies.

In order to build a community-oriented information environment in the Internet, we have to solve two contradictory requirements for metadata schemas – specialization (or localization) in a community and interoperability among communities. This paper shows a conceptual model to understand crucial aspects to solve the contradictory requirements.

1. Introduction[1]

The Internet and WWW have made possible a global information environment. The progress of the Internet over the past ten years has significantly changed our information environment. Internet search engines and directory services such as Google and Yahoo! have been globally accepted as a basic service to help users access information resources on the Internet, and digital libraries have been recognized as a fundamental function in libraries and information-centric organizations for their users. On one hand, this progress has produced global accessibility to information resources for individual users, i.e., geographical distance is no longer a fundamental barrier to access information. On the other hand, the diversity of user communities and information resources often causes difficulties for a user to find appropriate information from vast amounts of information resources on the Internet. Community is an important factor to solve the difficulties because meaning of words and phrases often depends on communities. In other words, a community oriented vocabulary is an important resource to realize a community oriented information environment.

Metadata, which has been widely recognized as a key component for the Web and digital libraries in various aspects, is obviously important to create a community-oriented information environment. Local or domain-specific communities would need to define metadata schemas and controlled vocabularies in accordance with their requirements in the case that their requirements are difficult to be satisfied only by those defined for the global communities. On the other hand, community-oriented specialization of schemas and vocabularies would raise a bar for interoperability issues for cross-community use of metadata and information resources. In addition, long-term maintenance of the schemas and vocabularies is a crucial aspect for the communities. Thus, we need to satisfy the contradictory requirements to metadata in order to create a community-oriented information environment.

The central topic of this paper deals with vocabulary of terms included in metadata schemas. A

[1] The studies described in the paper are supported in part by JSPS Scientific Research Grant #15300078 and a collaborative research project grant between Infocom Corporation and the Research Center for Knowledge Communities, University of Tsukuba.

metadata vocabulary is a set of terms defined in a metadata schema. In general, there are two types of metadata terms – terms that define properties of resources and those that define schemes to encode property values. In this paper, we use "metadata vocabulary" as a generic term that means both of these two types. We use "property vocabulary" and "value vocabulary" to mean vocabulary of resource properties (or metadata elements) and encoding schemes, respectively. We use "subject vocabulary" to mean a value vocabulary to express subjects of resources. For example, Dewey Decimal Classification (DDC), Nippon Decimal Classification (NDC), and Library of Congress Subject Heading (LCSH) are subject vocabularies. This paper shows two case studies and a conceptual model of metadata schema. In both of the case studies, subject vocabularies for domain and community-oriented metadata have been investigated. The conceptual model proposes to split metadata vocabularies and structural (or syntactical) constraints of metadata schema in order to identify requirements for metadata interoperability.

The rest of this paper presents two case studies of metadata centered research and a conceptual model of metadata schema for interoperability. The first case study is a development of a subject vocabulary for the ULIS-DL metadata database that has about 27000 records of Simple Dublin Core metadata for Web resources published by and/or useful for libraries and library and information science (LIS) institutions. The study has created a reasonably small set of subject terms, which we call core subject vocabulary that includes about 1000 terms and covering approximately 90% of the resources collected. We have developed an XML-based software to create a subject directory from the metadata database using the core vocabulary which is encoded in Web Ontology Language (OWL). Section 2 describes ULIS-DL and the core vocabulary.

The second case study is a development of a set of vocabularies for an information navigation service named Digital Okayama Dai-Hyakka (Digital Encyclopedia of Okayama). This navigation service collects Web contents published by and/or for the local communities and provides links to the resources with simple and descriptive metadata. DODH assumes both librarians and non-professionals as its catalogers. Its metadata schema is defined based on Simple Dublin Core and it uses a subject vocabulary designed for the local community. Section 3 describes DODH and its basic design issues.

In section 4, this paper describes a conceptual model of metadata which is proposed to understand requirements for interoperability of metadata. This paper presents a layered model of metadata schema in order to separate application specific and neutral components and to separate semantic and syntactic components. This explicit separation is useful to identify interoperability between schemas and functional requirements to use metadata across multiple schemas. A metadata schema registry designed for Dublin Core is discussed based on the model.

Section 5 discusses the vocabulary issue from the aspects of maintainability, interoperability and re-usability of subject vocabularies and metadata schemas.

2. Development of a Core Subject Vocabulary for ULIS-DL

2.1 Overview of ULIS-DL

The Digital Library service at the ULIS Library (ULIS-DL) started in February 1999[7]. The principal purpose of ULIS-DL is to build a subject gateway to resources useful for libraries and LIS institutions. We have collected the resources published by libraries and LIS institutions mainly in Japan, and created metadata for the resources. The metadata element set, called ULIS Core, is defined based on the 15 Simple Dublin Core elements with a few ULIS-DL specific elements. Pronunciation information is included as an optional sub-element of every element. As of summer 2003, ULIS-DL had created more than 40,000 metadata records, and about 26,000 of those records are the metadata for the Internet resources published by libraries and LIS institutions. A single metadata record is basically created for a single page in a Web site, so that one or more metadata records are created for a single Web site. A single metadata record includes one or more *Subject* elements of Simple Dublin Core, each of which contains a single subject term. These metadata are provided for users through a text retrieval interface, but no directory style interface has been built.

A major issue to enhance the usability of ULIS-DL has been (semi-)automatic creation of a directory style interface for navigating users to appropriate resources in addition to the text-based retrieval function. A subject vocabulary is required to create the directory interface. However, since the resources of ULIS-DL are collected from the limited subject domains, we could not find a well-established subject vocabulary or subject classification scheme appropriate as a subject vocabulary when we started the service. Therefore, ULIS-DL defined a guideline for catalogers which very weakly controls description of *Subject* elements but it does not give a well-controlled subject vocabulary for the *Subject* elements.

2.2 Development of Core Subject Vocabulary for ULIS-DL

Based on these experiences, we have developed a small subject vocabulary in order to build a directory style interface for ULIS-DL which shows subject terms sorted in hierarchical order, and a list of resources associated to every subject term. A preliminary evaluation of the *Subject* element values showed that there are more than 15,000 distinct text strings in the raw metadata, which includes typographical errors, inappropriate use of upper/lower case letters and so on. We also found that a set of subject terms assigned to a page in a Web site significantly overlaps to that of other pages in the same site and that the divergence of the number of metadata records per site is significantly large.

After having the raw metadata normalized, we followed the following steps to create a core subject vocabulary.

(1) Removal of overlapped subject terms in a single site: In order to avoid over-counting of a term that frequently appears but only in one or a few sites, we merged metadata records of each single Web site into one record and got a merged set of subject records.

(2) Removal of *Creator/Contributor/Publisher* element values: We created a set of subject terms by extracting distinct text strings from *Subject* elements of the set of merged metadata records. At this phase, *Subject* element values of a site which are the same as a value of *Publisher, Creator, or Contributor* elements of the site were excluded from the set because values of these elements are inappropriate as a subject term. This set created through this process is called the primary subject term set.

(3) Creation of candidates of the core subject vocabulary: We created a set of subject terms from the primary set by extracting terms that appear N or more times (N>1) in the set of merged metadata records. This set is called a candidate term set, CTS-N.

(4) Evaluation of the candidate sets: We created CTS-2, 3, 4 and 5 and evaluated the ratio of coverage of the set over the whole metadata. We used the "uncoverage" ratio, which is defined as the ratio between a number of metadata records that do not have any of the core subject terms and the total number of metadata records, in order to evaluate cost-effectiveness of CTS-N. Table 1 shows CTS-N and its uncoverage ratio (N=2, 3, 4, 5).

We chose CTS-5 as the core subject vocabulary in this study because it covers approximately 90% of the total records, and also because we considered its size as the most reasonable to manually organize the subject vocabulary terms, i.e. classifying the terms and defining relationships between the terms.

We classified the CTS-5 terms into eight categories, which are (1) Web terms, e.g., links, (2) Library terms, e.g. OPAC, (3) Organization and facility information, e.g. floor guide and access, (4) Type of libraries, e.g. university library and public library, (5) Organization names and service names, (6) Place names, (7) General subject terms, and (8) Reference tools, e.g. dictionaries, thesauri. Then, we classified terms in these categories into detailed categories up to the third level to constitute a hierarchical structure of subject terms. We assigned a proper subject term to each node of the tree. Every subject term of CTS-5 was associated to a leaf node as an occurrence term in the metadata. We encoded the tree structure in OWL. A few example descriptions is presented in Figure 1.

Table 1. Coverage of Candidate Term Sets

Total number of metadata records = 26358

Total number of Subject element values in the primary term set = 28797

Total number of Subject element values excluding Publisher/Creator/Contributor values = 26107

	number of Subject terms	number of excluded records	Uncoverage ratio
CTS-2	3979	1519	6%
CTS-3	2045	2083	8%
CTS-4	1366	2590	10%
CTS-5	1025	2801	11%

3. Development of Community-Oriented Subject Vocabularies

3.1 Lessons Learned from Preceding Study: Internet Public Library – Asia

Internet Public Library – Asia (IPL-Asia) is a project to collect Internet resources useful from the viewpoint of public libraries published in Chinese, Japanese and Korean (CJK) languages and to provide information about the resources in all of the CJK languages[4][9]. As part of this project, we collected subject terms to classify the resources published for children. We found that a subject vocabulary designed for children's resources is indispensable to classify the resources and to build a directory style navigational interface in addition to the subject vocabularies widely used in the library communities, e.g., Universal Decimal Classification (UDC) and Nippon Decimal Classification (NDC). This is because the domain of the resources is narrow and the subjects are community-specific, the subject terms of those well-established vocabularies are difficult for children to understand, and appropriate terms should be chosen to express the subjects in accordance with the age levels of the children, e.g., first to third graders, fourth to sixth graders, seventh to ninth graders, and higher to general public. Moreover, we learned that maintenance of the subject vocabularies by a community (or communities), and that re-usability and interoperability of the vocabularies among neighboring communities are crucial since costs for creation and long-term use of the vocabularies are not negligible.

Based on the experiences in IPL-Asia, we have defined the following guidelines to build subject vocabularies for community-oriented metadata.

(1) Create a core subject vocabulary which should be a reasonably small set of subject terms.

(2) Create subject vocabularies by tailoring the core vocabulary and associating appropriate expressions to every subject term in order to present the subject terms in accordance with the properties of users, i.e., age range and language.

(3) Encode the vocabularies in an ontology description language such as XML Topic Maps and OWL. This encoding is essential not only for automatic creation of subject directories from metadata records but also for interoperability of the subject vocabularies and for long-term maintenance of the subject vocabularies.

```
<uo:Subject rdf:ID="uo_1"
            uo:print_name="Library Terms">
  <uo:sub_subject_id rdf:resource="#uo_2"/>
  <uo:sub_subject_id rdf:resource="#uo_9"/>
</uo:Subject>
<uo:Subject rdf:ID="uo_2"
            uo:print_name="Resource Search">
  <uo:super_subject_id rdf:resource="#uo_1"/>
  <uo:sub_subject_id rdf:resource="#uo_3"/>
  <uo:sub_subject_id rdf:resource="#uo_7"/>
</uo:Subject>
<uo:Subject rdf:ID="uo_3" uo:print_name="opac">
  <uo:super_subject_id rdf:resource="#uo_2"/>
  <uo:sub_subject_id rdf:resource="#uo_4"/>
  <uo:sub_subject_id rdf:resource="#uo_5"/>
  <uo:sub_subject_id rdf:resource="#uo_6"/>
</uo:Subject>
<uo:Subject rdf:ID="uo_4" uo:print_name="OPAC">
  <uo:super_subject_id rdf:resource="#uo_3"/>
</uo:Subject>
<uo:Subject rdf:ID="uo_5"
    uo:print_name="Online Public Access Catalog">
<uo:Subject rdf:ID="uo_9"
            uo:print_name="Digital Library">
  <uo:super_subject_id rdf:resource="#uo_1"/>
  <uo:sub_subject_id rdf:resource="#uo_10"/>
</uo:Subject>
```

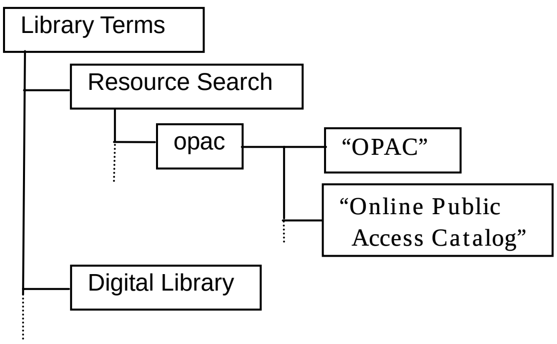

Figure 1. Subject terms described in OWL and Category Structure

3.2 Development of Community-Oriented Subject Vocabularies

Based on the above guidelines, we have built a subject vocabulary for the Internet resources collected by Digital Okayama Dai-Hyakka (DODH). The vocabulary, which is called Okayama Kids Vocabulary (OKV) in this paper, is designed primarily for children. In parallel to OKV, the prefecture of Okayama has built a subject vocabulary for classification of the Internet resources published by the government offices of the prefecture. Both of these subject vocabularies are designed to be simple because the subject terms will be used by the general public and children, and because metadata will be produced by non-professional catalogers. In addition to these vocabularies, NDC is used by librarians.

As shown in Figure 2, each of the subject terms of OKV has four presentation labels chosen in accordance with user ages, i.e. first to third graders (junior level of elementary school), fourth to six graders (senior level of elementary school), seventh to ninth graders (junior high school level), and eighth or higher graders (high school to general public). Presentation labels for the youngest age group has to be readable and understandable for children of that age, so that we can use only Hiragana, Katakana and a limited set of Kanji (Chinese characters)[2] and we have to re-phrase subject terms into plain words or phrases to make it easier for children to understand. For higher graders of elementary schools, we also use easier words and phrases for presentation labels, and we add pronunciation scripts[3] to the subject terms expressed. Presentation labels designed for junior high school students are almost the same as those for general public but some of them are re-phrased and

---

[2] Hiragana and Katakana are syllabic characters. First graders first learn Hiragana, then Katakana and a very limited set of Kanji. Students continue learning Kanji through elementary and junior high schools.

[3] In Japanese, every single Kanji character (Chinese character) has one or more pronunciations. Pronunciation of a Kanji basically depends on a word in which the Kanji is included but there are many exceptions. Pronunciation scripts, or Yomi in Japanese, helps children understand the subject terms.

there is additional pronunciation information.

OKV has eight major subject term groups. Each of the major groups has sub-groups of terms. As of January 2004, OKV has approximately 280 subject terms. Each subject term is included in one or more groups. The subject terms are collected from several major portal sites dedicated to resources for children and tailored them in accordance with the requirements of OKV. For example, some terms are specific to activities and events at schools and regional communities such as excursion, commencements, entrance examinations, and so on. The metadata schema of DODH is based on the Simple Dublin Core. OKV is to be used to create directories from DODH metadata and search aids.

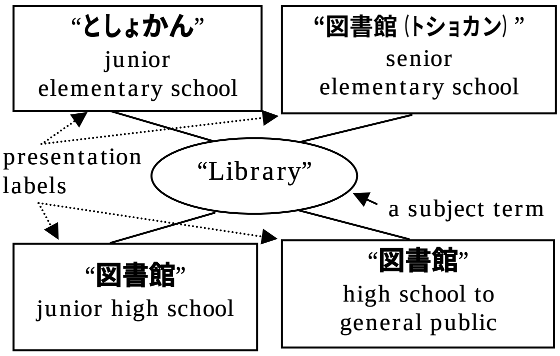

Figure 2. A subject term with multiple presentation labels

3.3. Discussion on Subject Vocabulary Maintenance

In our preliminary study, we have built the subject term vocabulary for IPL-Asia using XML Topic Maps in which each subject term is defined as a topic and associated with multiple presentation labels in the CJK languages. We applied the multi-lingual subject vocabulary to the IPL-Asia metadata and the DODH metadata in order to build subject-based directories of the resources. This experimental study, which is a straightforward approach, has shown the feasibility to build a user interface that has multiple presentation modes.

From this study, we have learned that ontology description languages such as XML Topic Maps and OWL are useful not only for encoding the vocabulary in a machine understandable form but also for maintaining the vocabulary for long term. Vocabulary maintenance is a crucial issue even if OKV is a small set of terms since it evolves over time, for example, evolution of subject terms and subject groups, and update of presentation labels. XML-based encoding is not a panacea but will help to decrease the cost of maintenance.

We consider that a metadata schema registry, which stores reference descriptions of metadata schema elements and provides them to users via the Internet, is a useful tool to store and maintain the subject vocabularies. The metadata schema registry has a function to register multiple labels in different languages associated with a single term.

4. A Conceptual Model of Metadata and Metadata Schema for Interoperability

4.1 Basic Concepts of Dublin Core and a Model of Metadata

4.1.1 Simple Dublin Core and Qualified Dublin Core

Dublin Core has been defined as a set of fifteen elements for cross-domain resource discovery. The use of these fifteen elements for metadata records, with no additional qualifiers, and with only plain-text strings as values, is known as "Simple Dublin Core". By design, any of the fifteen elements is optional and repeatable. This set has been approved as an international standard-ISO 15836. "Qualified Dublin Core", in contrast, uses the elements together with qualifiers that increase the richness and precision of description. Qualified DC has two types of qualifiers – element refinement and encoding schemes.

DCMES is a stable but not a closed set. DCMES evolves in accordance with requirements to express resource properties and value-types which are not expressible using existing ones. The following sections show underlying key concepts of Dublin Core[10]. The article reported by Andy Powell shows an abstract model from an architectural viewpoint[5].

4.1.2 Warwick Framework – Basic Framework for Extensibility

Since the Internet is a very diversified environment, it is useless to assume that a single metadata element set will meet the needs of all domains and purposes. It is also impractical to develop metadata sets application by application: the result would be expensive and chaotic, and interoperability would be non-existent. On the other hand, it is desirable for application developers to use established metadata schemas and adopt them in accordance with local requirements. The Warwick Framework, a conceptual model that resulted from the 2nd Dublin Core Workshop in 1996, gave an early expression to the notion of metadata as modular components that may come from more than one metadata schema[3]. In this model, a metadata instance is expressed as a container which contains one or more packages, each of which is expressed in a given metadata schema. The Resource Description Framework (RDF) provided a practical realization of many of the ideas of the Warwick Framework.

The Warwick Framework is important as a model for modular metadata on the Internet. No single metadata schema is sufficient to all applications. Rather, it is necessary to adopt appropriate elements from various schemas in accordance with the functional requirements of an application. Application Profiles described in section 4.1.4, which provide a framework to adopt one or more element sets in accordance with an application, could be also considered as a realization of the Warwick Framework.

4.1.3 The Dumb-Down Principle as a Basis for Interoperability

The Dumb-Down principle gives a guideline for qualification. The Dumb-Down principle suggests that a value of a qualified element has to be consistent as a value of the element without any qualification. For example, given the following qualified values:

(1) (Element Refinement) Date Accepted: "2002-12-11", and

(2) (Encoding Scheme) Language: "en" encoded in RFC 1766.

Then, assuming that the qualifications in the above examples, *Accepted*, *RFC 1766* and the component names of the value structure (i.e., *name*, *affiliation* and *contact*) are removed. The values of example 1 and 2, "2002-12-11" and "en" are still consistent with their elements after the removal.

Dumbing-down is a crucial function for metadata interoperability in the global community since local communities can extend their schemas in accordance with their requirements, and at the same time they can also keep their metadata interoperable with other metadata communities.

4.1.4 Application Profiles

Dublin Core Metadata defines the vocabulary of metadata, i.e., terms and their meanings, but in general does not specify the encoding or syntactic characteristics. An exception is the feature included in Simple DC that is "Any of the 15 elements is optional and repeatable." Local applications, however, may have domain specific requirements appropriate to a given domain or application:

- Title, Creator and Description might be mandated but others are optional,
- Use only Title, Creator, Description, Date and Language elements,
- Use the 15 elements of Simple DC and some elements from other metadata sets such as the IEEE Learning Object Metadata (IEEE LOM), and so forth.

These requirements can be defined independently of the vocabulary definitions. Description of this application-specific syntactic feature is called an *application profile*. Any application can have its own application profile, which specifies a set of metadata vocabulary terms used in the application as well as syntactic or structural features of the particular application. Figure 3 shows a model of application profiles. The vocabulary terms could be borrowed from one or more source schemas. More importantly, the application profile could be used to define a mapping between the application's scheme to a global scheme(s), which is crucial for interoperability.

4.2 A Conceptual Model of Metadata Schema for Interoperability

A metadata schema defined for an application is composed of three layers:

(1) Layer 1 - Semantic Definition Layer: Definition of terms used in the schema. In other words, definition of metadata vocabulary, i.e. metadata element set. In general, two types of metadata terms are included in the metadata vocabulary – property vocabulary and value vocabulary[1]. A property vocabulary, or in other words element vocabulary, is a set of property terms, for example, elements and element refinement qualifiers of DCMES. A value vocabulary is a set of value terms, for example, encoding schemes. Definition of each term should primarily include a primary name and

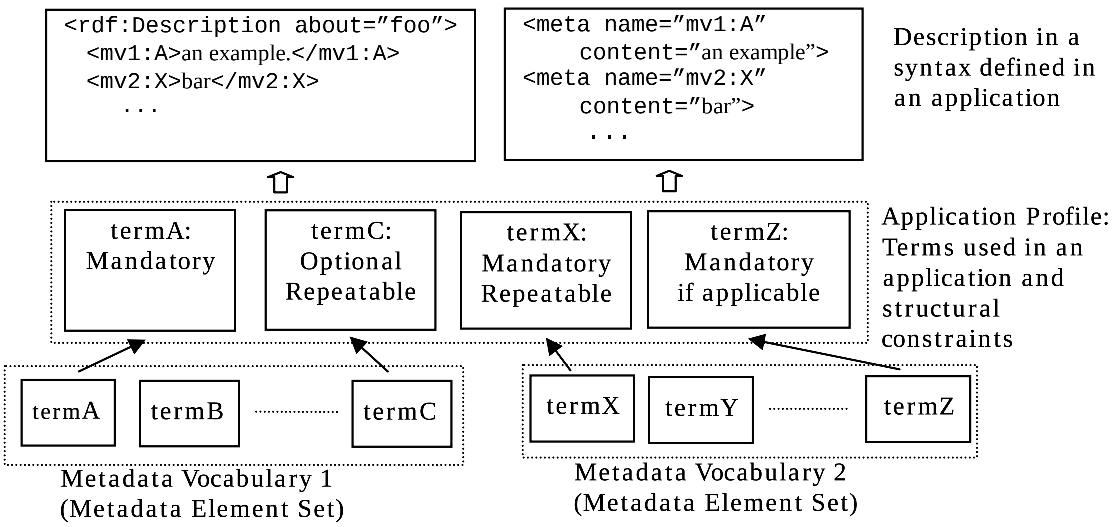

Figure 3. Concept of Application Profile

its meaning. Thus, a vocabulary definition gives the semantic basis of a metadata schema.

(2) Layer 2 - Structural Constraints Definition Layer: Definition of syntactic features which does not depend on any particular implementation scheme. A set of terms used in the schema and structural constraints applied to each term should be included in a definition. Application profiles are given in this layer.

(3) Layer 3 - Implementation Dependent Syntax Definition Layer: Definition of syntax of metadata in an implementation, for example, metadata description syntax in HTML, XML, and RDF.

In addition to these definitions, each application schema developer would provide guidelines for creating metadata. Figure 4 illustrates relationships among schemas based on the conceptual framework.

The layer 1 of Figure 4 provides definition of metadata elements and qualifiers of a metadata schema. On the other hand, there are controlled value vocabularies and value encoding schemes which are defined neutrally to any specific metadata schema, e.g., DDC for expressing subjects and ISO 8601 for encoding dates. Every single term of a controlled value vocabulary is given its definition as well, e.g. a decimal number and its associated label(s).

4.3 Metadata Schema Registry

A metadata schema registry stores and provides metadata schemas not only for human but also machines. A metadata schema registry is a key software tool to enhance interoperability of metadata schemas expressed in all layers. For example, the DCMI metadata schema registry[2][6] which provides reference descriptions of DCMI terms is designed primarily for layer 1. Every DCMI term is expressed in RDF Schema. The DCMI registry provides the reference descriptions translated into more than 20 languages. Metadata schema registries are useful to store and provide all types of metadata vocabularies, i.e., application profiles, subject terms and other vocabularies.

There are some issues that are recognized important for future. For example, long term maintenance of metadata vocabularies is an important issue for metadata interoperability over time[8]. Interoperability of metadata schemas are obviously important for metadata interoperability. Re-usability of metadata schemas is important to decrease the cost of schema development and interoperability. Metadata registries have large potential to solve these issues.

5. Discussion – Creation, Maintenance and Interoperability of Community-Oriented Metadata Vocabularies

This paper has presented two case studies of metadata-centric services, both of which are based on Simple Dublin Core. Some lessons we learned from the case studies are as follows:

(1) Subjects of Web resources in a specific domain can be covered by a reasonably small set of subject terms even though the subject terms are very weakly controlled within the metadata creation process. On the other hand, the set of subject terms includes domain-specific subject terms which are not covered by large general-purpose subject vocabularies.

(2) A scheme to associate different presentation labels to a single term in a subject vocabulary is crucial not only to enhance usability of user interfaces in accordance with users but also to enhance maintainability of the vocabulary. XML-based vocabulary description is useful in this respect.

(3) Communities need subject vocabularies which are built in accordance with their domain, goal and languages. However, maintainability of the subject vocabularies is a crucial issue for the communities to build and maintain metadata for long term. From this viewpoint, a reasonably small vocabulary designed based on community requirements is advantageous, and software tools designed for

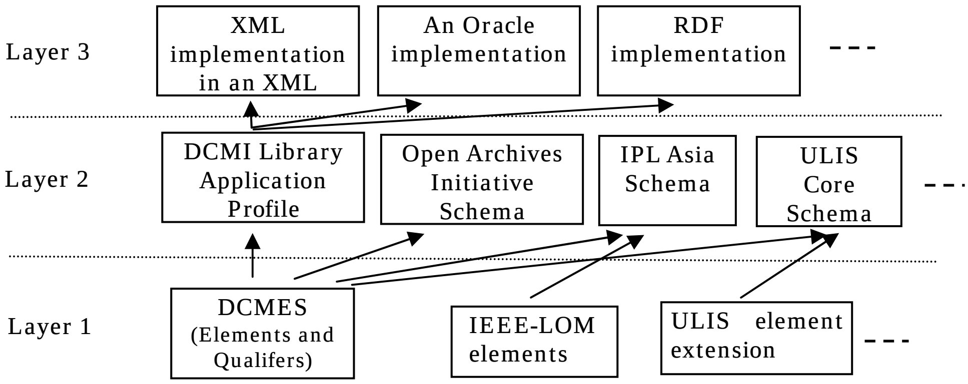

Figure 4. Layered Model of Metadata

vocabulary maintenance are crucial to enhance the usability and reliability of the vocabulary.

Library catalogs, which are a typical metadata, have been created by professionals for high quality description of resources. Library community has developed a rich set of metadata vocabularies which are utilized well not only by libraries but also by library users, publishers and related communities. However, the conventional business model does not always work well in the Internet environment where huge number of diverse resources are published and used by divergent communities. Any community, which may be regional or domain specific, would be able to develop metadata vocabularies in accordance with their requirements, but maintenance of the vocabularies for long period of time seems to be a major and practical barrier for the community to develop the vocabularies. From this viewpoint, we consider that it is important to keep the vocabularies simple and small, and that vocabulary maintenance software has an important role.

The Internet has realized a global infrastructure to access resources regardless of geographical distance. However, community oriented information environment has not been realized well. Metadata schema designed for communities is a crucial component to realize the community oriented information environment. On one hand, communities need metadata schemas designed for themselves, but on the other hand, interoperability and reusability of metadata and metadata schema with other communities are crucial. Thus, it is required to satisfy two contradictory requirements, i.e. specialization vs. generalization.

6. Concluding Remarks

From our experiences in IPL-Asia and ULIS-DL, we have learned that these systems need subject vocabularies designed for them in addition to well established subject vocabularies such as UDC and NDC. The subject vocabularies created for DODH, which is an ongoing project hosted by the Okayama Prefectural Culture Center (OPCC), is being used by participating catalogers at OPCC and related organizations. We will continue our study to evaluate the usability and validity of the vocabularies and to build technologies to support maintenance of the vocabularies.

Community-oriented information environments need to satisfy the contradictory requirements. We believe that metadata schema registries and metadata schemas encoded in an XML-based ontology description scheme are key technologies to satisfy the requirements. We also consider that the model presented in this paper provides some basic guidelines to find solutions for the contradictory requirements.

Acknowledgments

We would like to express our thanks to Drs. Stuart Weibel, Thomas Baker, Koichi Tabata, Tetsuo Sakaguchi, Atsuyuki Morishima for their support. We would like to thank Huishan Jiang and Akiko Shingai for their contribution to the development of the software tools. We would like to express our special thanks to Infocom Corporation for their support to our collaborative research and development activities between University of Tsukuba and the Okayama Prefectural Cultural Center.

References

[1] Baker, T., et al. “Principles of Metadata Registries, A White Paper of the DELOS Working Group on Registries”http://delos-noe.iei.pi.cnr.it/activities/standardizationforum/Registries.pdf

[2] DCMI, “The Dublin Core Metadata Registry”, http://dublincore.org/dcregistry/index.html

[3] Lagoze, C. “The Warwick Framework – A Container Architecture for Diverse Sets of Metadata”, D-Lib Magazine, July/August 1996, http://www.dlib.org/dlib/july96/lagoze/07lagoze.html

[4] Lee W., wt al., “A Subject gateway in Multiple Languages: a Prototype Development and Lessons Learned” Proceedings of DC-2003, pp.59-66, Seattle, 2003 (http://www.siderean.com/dc2003/203_Paper76.pdf, accessed January 2004)

[5] Powell A., et al., “DCMI Abstract Model”, http: //www.ukoln.ac.uk/metadata/dcmi/abstract-model/

[6] Nagamori, M., et al., A Multilingual Metadata Schema Registry Based on RDF Schema, Proceedings of DC-2001, pp.209-212, Tokyo,2001

[7] Sugimoto, S., “Digital Libraries in Japan and at ULIS”, Proceedings of ISDL’99, pp.66-73, Tsukuba, 1999

[8]Sugimoto, S., et al., , “Versioning the Dublin Core across Multiple Languages and over Time”, Proceedings of SAINT 2001, San Diego, 2001

[9] Sugimoto, S., et al., “An Experimental Study on a Collaborative Development of a Multilingual Subject Gateway for Public Library Users”, Proceedings of Digital Libraries: IT Opportunities and Challenges in the New Millennium, Beijing, pp.118-130, 2002

[10] Sugimoto, S., Baker, T., and Weibel, S.L., "Dublin Core: Process and Principles", Proceedings of ICADL 2002, Singapore, 2002, (Lim, EP, et al. (eds), "Digital Libraries: People, Knowledge and Technology", LNCS 2555, pp.25-35, Springer)