Functional Requirements of Metadata System: From User Needs Perspective

Ya-ning Chen, Shu-jiun Chen, Hon-chung Sum, and Simon C. Lin
Computing Centre, Academia Sinica, Taiwan, ROC
{arthur, sophy, htshen, sclin}@sinica.edu.tw

Abstract

This paper aims to construct a common set of functional requirements for metadata systems based on the research of metadata system specifications at the National Digital Archives Program（NDAP）in Taiwan. Eight projects of the NDAP are chosen as a case study to examine the functional requirements of metadata systems across museums, libraries, archives, and herbaria communities. To enable deeper analysis of functional requirements, interviewing and system prototyping are also employed. The study concludes with four findings for metadata system functions and proposes two recommendations, including（1）a set of recommendations for content experts, metadata professionals, and system designers to develop a clear and appropriate metadata system; and（2）the spectrum of functional requirements for metadata system design and testing.

1 Introduction

Since 2001, the National Digital Archives Program（NDAP）in Taiwan has been initiating many projects regarding digital library research and practice. Essentially, the NDAP is composed of a wide range of content-oriented digital library projects across the arts, humanities, social sciences, and biodiversity disciplines. Related research and development of information technology are frequently employed in order to support content-oriented projects in the creation and representation of digital contents. Among these advancements, metadata plays a chief role to help organize digital information. Metadata requirements, which involve both comprehensive description and structured relationship, are often very complicated beyond traditional relational database techniques, and thus become one of the grand challenges for system design. This paper aims to examine three metadata issues as follows. First, how can content experts properly express the needs for a metadata system? Next, how can metadata professionals act as a reliable bridge to properly transfer new metadata requirements to system designers? Finally, how do system designers quickly develop an adequate metadata system for content-oriented projects?

2 Background

To create a common set of functional requirements for metadata systems is no simple task. In practice, the requirements will very much depend on the projects involved. Since the aim is to construct a set of requirements common to all projects involved, guidelines for evaluating the needs of each project are necessary.

These guidelines can be drawn from published literatures related to metadata system development and evaluation. For example, Koch categorizes the components of a metadata system into eleven categories. These categories include: 1) creating metadata, 2) automatic extraction and production, 3) conversion between metadata formats, 4) subject description, 5) help systems, 6) encoding, structure, and syntax, 7) exchange/transfer of metadata, 8) harvesting, indexing, search, and browse of metadata databases, 9) tools for metadata repositories and metadata storage, 10) metadata display tools, and 11) integrated environments [1]. The MetaWeb Project (in Australia) uses two categories: editors and generators [2]. Currie (from the Education Network Australia (EdNA) Project) puts metadata tools into three categories: metadata manager, metadata creator, and thesaurus viewer [3]. The categories created by these three projects can serve as a tool for constructing functional requirements of metadata systems.

An Request for Information (RFI) document written by the States Services Commission (in New Zealand) uses a different approach by creating a set of requirements based on user needs. [4] These are summarized below:

- Links to externally managed thesauri, controlled lists, and directories.
- A built-in process for managing the workflow quality assurance of metadata.
- A role-based security system controlling access to all features of the system.

System evaluation is yet another perspective from which requirements can be created. In one example, the Federal Geographic Data Committee, FDGC (in USA) provides a formal checklist for the evaluation of metadata tools. This checklist includes metadata exchange, usability, administration, and tool reliability as major evaluative points [5]. A second example, a case study on Global Information Locator Service (GILS) metadata,

focuses on four considerations: completeness, profile, accuracy, and serviceability [6].

To sum up, these literatures offer various criteria for needs evaluation and development of metadata systems; overall, the different viewpoints can be generalized into software design, user requirements, and system evaluation. However, two research issues remain unexplored, namely:

- What functions should be achieved by a metadata system?
- How does one illustrate the functions, roles, and relationships between metadata systems and external systems or resources, such as geographic information systems or digital library systems?

3 Methodology

3.1. Techniques

Three data collection techniques are used in this study: case study, interviewing, and experimentation. For the case study, subjects are drawn from NDAP projects, based on these two criteria, in order to accommodate the varying timelines of the projects:

- Content experts presented their metadata requirements clearly for the 2002 project plan.
- The metadata requirements were developed by a series of system analysis and normalization tasks.

The eight projects for this case study consist of:

- Digital Archives Project of Chinese Painting and Calligraphy at the National Palace Museum （museum） http://www.npm.gov.tw/dl/03/index03.htm
- Digital Archives Project of Chinese Antiquities at the National Palace Museum （museum） http://www.npm.gov.tw/dl/02/index02.htm
- Digital Archives of Rubbings and Archaic Texts （library） http://mip.iis.sinica.edu.tw/rubbing/
- Digital Archives for the Grand Secretariat Archives （library） http://www.ihp.sinica.edu.tw/~mct/
- Digital Library Project for Official Economic and Diplomatic Archives （archives） http://dipeco.sinica.edu.tw/
- Digital Archives Project of Academia Historica （archives） http://www.drnh.gov.tw/
- Zoological Research of Taiwan: Fish and Mollusks（herbarium） http://www.drnh.gov.tw/
- Digital Library of Taiwan Herbarium（herbarium） http://taiwanflora.sinica.edu.tw/eindex.html

The interviewing phase uses a questionnaire to ascertain the needs of each project. This questionnaire includes the following questions:

- Current status: including project's goals, predictive schedule, amount and type of collections, used standard, and legacy systems and records.
- Requirements for metadata elements: including element's name both in Chinese and English, definition, and guideline.
- A context figure for delineation of object's relationship and granularity.
- An associative figure for indication of metadata element relationship.
- Metadata element's name, qualifier, code value and related standards, such as name authority file of Library of Congress for author name.
- Metadata instance: including metadata element name, qualifier and example.
- Functionality for metadata element: including display priority, maximal length, mandatory or optional, multi-value, link, default, fix value, and uniqueness.
- Metadata retrieval and display: label that element should be included into the simple or the advanced mode.
- A figure to elucidate the workflow procedures.
- A figure to illustrate the relationship between metadata system and miscellaneous system.

The results of the questionnaire are then analyzed according to difference, distribution, and preference of metadata requirements. Generalizations are then made based on the analyses of the questionnaire results.

Lastly, a prototype based on system specification is developed to examine issues of uniqueness, relationship, availability, and degrees of difficulty for metadata requirements.

3.2. Standards

The standards used in these examples follow a parallel-standard paradigm consisting of both a basic set of widely recognized standards and a domain-specific set of standards corresponding to each domain. Namely, the Dublin Core Element Set acts as the default, and domain-specific metadata standards are used in parallel for each domain. Examples of such standards include CDWA, CSDGM, EAD, and HISPID.

4 Discussion and Result

Based on case study, interviewing, and system prototyping for the eight projects in the NDAP, several

findings have been made. These findings are discussed in the following sections.

4.1. Requirement Category

Based on the data collected from the eight projects used in this study, a total of thirty-two requirements have been compiled and categorized according to their relevance to metadata science. These categories, as shown in Table 1, are input and maintenance, retrieval, display, interoperability, management, and automation of operational activities. Of particular concern are the requirements listed under "interoperability", which, in our opinion, are quite few in number. Indeed, the interoperability of metadata systems has been a pressing concern for many digital library projects, and also has been a key focus for research and development. During interviews we have conducted as a part of this study, project members have expressed strong desire for their systems to be capable of efficient data retrieval and sharing, so that, one day, data could be shared in the international community with research teams from other sciences. The apparent lack of sufficient requirements in the "interoperability" category, despite strong desire to have more interoperable systems, indicates to us that, perhaps, the knowledge and expertise regarding data interoperability and data storage is relatively foreign to project members.

4.2. Core Requirements

In order to conserve effort toward developing common system components and functions—and thus achieve more cost-effective system development—a list of core requirements should be offered by system designers. Our criteria for core requirements are that 1) they are demanded by more than six out of eight projects, and 2) they require the metadata system to offer functionality immediately. In our case, thirteen items pass the criteria for core requirements. In sum, they are: record creation, modification, deletion, multi-value attributes, select-list menu, simple search, advanced search, simple display, import and export using XML document, links to other databases, cataloguing history, setting private attributes, and authentication management.

4.3. Function Refinement and Appropriateness

In general, any system function can be implemented by a variety of approaches; however, one may find that adopting one of different methods can develop the same function. To understand the spectrum of system functions is useful toward attaining smooth communication among content experts, metadata professionals, and system designers, so that they may refine the requirements of system functions in a suitable way. These refinements are grouped into five categories:

a. Multi-value Attributes

There are four archetypes proposed for multi-value attributes: a single multi-value element, multiple multi-value elements, a whole set of multi-value elements, and a hierarchy set of multi-value elements （as shown in Figure 1). This function is often utilized to achieve two objectives in a web-based user interface environment. First, it is convenient for the creation and maintenance of specific metadata elements, which are required for repeated data input. Second, this function is useful to group homogeneous or similar elements together based on considerations for metadata creation, display, and retrieval.

b. Select-list Menu

Three generalized formats of select-list menu exist: single column, hierarchy menu, and one-to-one menu. A set of values can be grouped into a select-list menu of a single column, convenient for inputting data by selecting or clicking with a mouse. A hierarchical menu, based on the hierarchical relationship between values, is the second format. In our projects, we encountered two variants for this function: a two-level hierarchy, and a three-level hierarchy. The third format, based on the parallel relationship between two independent values, is a one-to-one menu describing such a relationship.

c. Search and Display

Metadata search and display functions are separated into two kinds: simple and advanced. The simple approach is centered on general users, and therefore selective metadata information is provided. Advanced search and display focuses on research and management for content experts, and information is much more comprehensive than in the simple approach. More importantly, a seamless connection between the simple and advanced approaches is constructed in order to guide the user toward the appropriate information.

d. Cataloguing History

In order to attain quality assurance, a history log, which covers metadata creation, maintenance, and verification, is strongly requested by content experts. Three general options are given: latest log, detailed log, and verification log. The latest log keeps a record of whoever last created or updated the metadata record. The detailed log, on the other hand, keeps a complete history of transactions for the metadata record. The verification log, which employs a verifier to justify record accurateness, is maintained in order to indicate verification of records.

Table 1. Requirements for a metadata system

Functional Item	Description
1. Input and maintenance	❙ User can create, delete, modify, and cut and paste metadata records through a web-based interface. ❙ Multi-value attributes. ❙ Select-list menu: including single column menu, hierarchy menu, and one-to-one menu. ❙ Maintenance of values and codes for select-list menu content. ❙ Preview of images.
2. Retrieval	❙ Including the simple and advanced search options. ❙ Keyword, Boolean operator, and limitation.
3. Display	❙ Including the simple and advanced display options. ❙ Content experts can select display format (including labels and information).
4. Interoperability	❙ To provide an XML-based mechanism for metadata import and export. ❙ In addition to XML-based approach, other text files can also be imported and exported. ❙ The metadata system can link to external systems, such as an authority file system for person, a gazetteer, and a management system.
5. Management	❙ To track transaction of record activities, such as creator, modifier, verifier, and related date and time. ❙ To verify the accurateness of metadata records based on standard requirements, and to offer related messages for record modification. ❙ To offer the function to de-duplicate a record based on specified criteria, for the process of record creation and modification. ❙ To set up attributes in private and public groups, in order to keep specified information for internal use only. ❙ To offer different levels of authorization, such as criteria based on task or job responsibility.
6. Automation of Operational Activities	❙ To offer inventory function in order to perform routine tasks related to decoration, repair, checking, and so on. ❙ To provide a circulation function to conduct tasks related to check-out/in, claim, and so on. ❙ To provide the function of a report generator in order to generate related reports or statistic data based on specified criteria from metadata elements.

e. Setting Private Attributes

This function is used to separate some information from end users for metadata search and display, and the information can be attributed into two types: management-based and research-based. A metadata system, in addition to describing cultural objects, also acts as a base for management information systems to record internal information, such as exhibition, inventory, and so forth. Additionally, a metadata system can be employed as a research tool for content experts to record annotated research information for objects. When information has not yet been justified, researchers are not inclined to release it out without a series of examinations according to research methodology. Therefore, the function of setting private attributes is necessary to label information as separate from the public.

4.4. Common Functions

In addition to core requirements, seven functions are regarded as candidates for formulation of common functions by all projects. Although these functions are not reflected in specifications of metadata systems directly, they have often been raised during interview sessions. The seven common functions are summarized as follows:

- To provide different levels of information granularity: such as fonds, sub-fonds, series, files, and items for archive description.
- To offer a mechanism to integrate external resources seamlessly, including sharing of metadata creation and maintenance: such as whether databases of person profiles can be shared.
- To convert legacy systems and records into new ones: can legacy systems and records be integrated or converted into new ones smoothly? What are the methods and procedures?
- To equip customized options for report generation: can specific reports be generated? When is the right time to raise these requirements during system analysis and development?
- To incorporate miscellaneous tools, in terms of metadata creation, retrieval, display, and so forth; e.g., a cross-walking mechanism between the Western and Chinese calendar systems.
- To implement structured relations for existing metadata standards: such as offering a variety of Encoded Archival Description (EAD) structures for archives, ranging from fonds, series, files, to items levels.
- To enable multi-lingual processing. How does one record the Japanese and Chinese characters that have not been defined by the BIG-5 or Unicode character sets? Is it necessary to create a namespace mechanism for rarely used characters?

5 Findings

After thorough consideration of the results derived from the study, four findings remain to be achieved:

5.1. Coverage of Functions

If a cross-examination is conducted on projects' requirements based on the preceding discussion, we may find that requirements of the eight projects fall nearly into three areas that are generalized in literature review: user requirements, software design, and system evaluation. Consequently, one requirement is not covered by the literature's generalization but still can be categorized into user requirements, namely, a web-based charging mechanism for future e-commerce application. On the other hand, six requirements are not represented so clearly, and can be generalized into concealed and unrepresented items. Concealed items are either those items for which content experts have presented their requirement unclearly, or those items which have been performed with a background approach; these items include help guidance, management, and reliability. Unrepresented items are the responsibility of metadata professionals acting on behalf of content experts and include transferring of multiple metadata formats, thesaurus control, and the decreased duplication of data input.

5.2. Gap between System Implementation and Requirements

How to describe objects of original and digital counterparts and their relations is a challenge for all projects. If system designers apply techniques of relational databases to developing metadata systems, some points deserve to be examined:

- Is system performance slow and cost-effectiveness poor from employing too many tables for a database schema's indexing and linking?
- If the one-to-many principle of metadata description for original and digital objects is adopted, instead of the one-to-one principle for each object, how does one describe metadata information focused on rights-management, access control, and long-term preservation for digital objects?
- If the one-to-one principle is employed, common metadata elements are forced to be input again and again. This issue can also be proven true by a series of analyses based on International Federation of Library Associations and Institutions (IFLA) Functional Requirements for

Bibliographic Records (FRBR) model or Entity-Relationship (ER) model. How do we build up a bi-directional linking to metadata records for various original and digital objects consistently?

- No matter whether the one-to-many or one-to-one principle is deployed, how does one indicate relations between original and digital objects, and various digital objects for the same original object, such as master file and derivative files?

As a result, metadata requirements for comprehensive description and rich relations have become a challenge for information techniques used in the digital library domain.

5.3. Relationship between Metadata Systems and External Resources

In general, "thing" objects play an important role in metadata information systems, and other objects, such as "person name" and "place name", are regarded as supportive roles for "thing". Therefore, one may find that other supportive objects are included into a tiny part of most metadata formats, and are treated as subordinate objects to "thing". According to our requirements for metadata systems, other objects are emphasized as an independent and separate external resource. As a result, two independent metadata systems, based on "person name" and "place name", are developed.

From the perspective of system integration, external resources have two meanings in our cases. First, another two metadata systems can be considered as external resources for one of three systems. Our requirement for integration of external resources is centered on how to construct a bi-directional linking and synchronous reflection of data updated across various metadata systems. For instance, the same element in a "thing" metadata system will be correspondingly updated when a metadata element completes a data update in a "person name" system. External resources existing on Internet Web formats are also integrated to allow discovery of additional resources related to research themes.

5.4. A Mechanism for Metadata Exchange and Transfer

Our requirements for a metadata exchange and transfer mechanism are very different from those of other digital library projects around the world, in that most such projects adopt only one metadata standard. In our case, two requirements for metadata exchange and transfer seem to strive toward opposite goals. From the cultural heritage standpoint, a metadata system has to adopt a domain-specific, comprehensive metadata standard. On the other hand, each project is demanded to offer a federated search service over a variety of systems for the public. Our approach is composed of two major components as follows: database schema, and a cross-walking mechanism. First, an internal database schema is comprehensively designed based on a domain-specific metadata format. Next, a mapping service is provided for on-demand conversion among a wide range of metadata formats. Therefore, a parallel-format principle to achieve the above two requirements for exchange and transfer is adopted. Namely, the Dublin Core Element Set is default as a basis for federated search, while another domain-specific metadata standard is used as a reference for system development, such as Categories for the Description of Works of Art (CDWA), Content Standard for Digital Geospatial Metadata (CSDGM), EAD, Herbarium Information Standards and Protocols for Interchange of Data (HISPID), and so forth.

6 Conclusion and Suggestion

To develop an appropriate metadata system for content-oriented projects in the NDAP is a difficult task for content experts, metadata professionals, and system designers alike. Effective communication of the requirements is essential to ensure that they are understood properly, thus allowing the overall aim of the projects to be attained. In this paper, we offer several points on developing metadata systems for all parties involved with system requirements assessment and design. More importantly, we further extend the spectrum of functional requirements (e.g., by adding a functionality checklist for system design) for metadata system development as a basis for future extended use.

6.1. Recommendations for Communication

a. Content Experts

- The content experts should first ensure the project scope and objectives themselves, and requirements should be expressed systematically.
- Any revision on requirements has to be confirmed by both metadata professionals and system designers to avoid duplicated effort, and thus to decrease costs for system development.
- Content experts have to understand knowledge related to information technologies to a sufficient degree in order to present reasonable requirements.

b. Metadata Professionals

- Metadata project members should relay those requirements that are beyond metadata, but related to digital library technologies, to other members who engage in developing technology for digital library projects.

Figure 1. A multi-value element and instance

- In terms of a digital library workflow encompassing capture, organization, dissemination, rights-management, resources discovery, query, access, and archiving, there are various sets of metadata elements corresponding to this workflow and its tasks. Based on our requirements for metadata, one may find that the focus of state-of-the-art project requirements is still centered on description and structure. In the near future, metadata professionals should be proactive to guide content experts toward other areas of metadata requirements, in order to build a more integrated metadata system.

c. System Designers

- Let content experts be aware of current status and future development of information technology related to digital libraries.
- Offer a systematic service to examine the feasibility of metadata requirements.

6.2. Spectrum of Functional Requirements for Metadata System

By comparing metadata system requirements found both in literature and in our study, one may find that our requirements are both principled and practical, and that main categories are also offered (as shown in Table 1).

According to our research, the spectrum of functional requirements for metadata systems can be composed of these items, in order:

- The web is used as a user interface for metadata creation, maintenance, representation, and query, in order to manage metadata records in an efficient way.
- The web-based metadata representation has to offer different options for browsing, indexing, and searching for various targeted audiences.
- External resources, such as thesauri for "person name" and "place name", should be regarded as independent metadata systems. Synchronous data update and bi-directional linking are also required. Furthermore, projects should be able to share both metadata systems and external resources.
- Flexible customization is necessary to offer various options to accommodate a wide range of user needs.
- The ability of multilingual processing is required to handle Chinese and Japanese materials.
- A mechanism composed of more than two different metadata standards for import, export, and harvest, is an essential component of a metadata system, including conversion of legacy systems and records.
- The ability to control quality assurance for workflow and management.

- A management set of metadata elements for system and administrative management is also needed.
- Provide offline and online modes to promote the use of metadata systems widely.
- The facility of security functions for authentication and metadata exchange.
- The help facility should be equipped to guide users toward adoption of a metadata system.
- The metadata system can be planted across various platforms in order to interoperate with technology evolution.
- Dedicated personnel are required to keep system operations stable.
- The metadata system should be designed by a component-based approach to incorporate new software and hardware, to avoid obsolescence of information technology.
- Recovery and storage devices should be used to prevent data loss.

Acknowledgements

We'd like to thanks Mr. Shi Lin for help in developing the function of multi-value for metadata system, and figure for this paper.

References

[1] Koch, T. (2000). Metadata related tools. Retrieved June 19, 2002, from http://www.lub.lu.se/tk/metadata/dctoollist.html

[2] Campbell, D., & Zhan, Z. (1997). MetaWeb project: Analysis of metadata creation tools. Retrieved June 19, 2002, from http://www.dstc.edu.au/Research/Projects/metaweb/toolpost.html

[3] Currie, M., & Moss, M., Ip, A., & Morrison, I. (2000). The EdNA metadata toolsets: A case study. Retrieved June 19, 2002, from http://ausweb.scu.edu.au/aw2k/papers/currie/paper.html

[4] State Services Commission. (2001). Request for information: Metadata management and search tool. Retrieved June 19, 2002, from http://www.go.vic.gov.au/pdfs/rfi-pdf.pdf

[5] Federal Geographic Data Committee. (n.d.). Evaluation criteria. Retrieved June 19, 2002, from: http://www.fgdc.gov/clearinghouse/mitre/task2/evalcriteria.html

[6] Moen, W.E., & Stewart, E.L. (1998). Assessing metadata quality: Findings and methodological considerations from an evaluation of the U.S. government information locator service. In: IEEE international forum on research and technology advances in digital libraries: ADL 98 proceedings: 246-255, 1998. California, CA: IEEE computer society.