The Design of Metadata Interchange for Chinese Information and Implementation of Metadata Management System[*]

Chao-chen Chen
Department of Adult and Continuing Education, National Taiwan Normal University
Head, Department of Public Services, National Central Library
Taipei, 10764, Taiwan, ROC
cc4073@tpts1.seed.net.tw

Hsueh-hua Chen
Department of Library and Information Science, National Taiwan University
Taipei, 10764, Taiwan, ROC
sherry@ccms.ntu.edu.tw

Kuang-hua Chen
Department of Library and Information Science, National Taiwan University
Taipei, 10764, Taiwan, ROC
khchen@ccms.ntu.edu.tw

ABSTRACT

With the development of Internet, digital libraries/museums have received worldwide attention and many developed countries are doing extensive researches on digital libraries/museums. In Taiwan, many institutions have digitized their collections. In addition, major research projects such as Digital Museum Project, Digital Archive Program, Digital Local Document Program, and CMNet are either in progress or about to begin. To digitize resources and present them on the WWW is the first priority; nevertheless, it is more important to organize these resources based on their characteristics, so users may retrieve and use them effectively. In digital library/museum systems, metadata plays a crucial role: its formulation has to begin with understanding the user demands and object characteristics, and standardization is essential for

---

the interoperability among information systems. Currently, regardless of their metadata formats, most digital library/museum systems use XML or SGML as their importing/storing/exporting language. In particular, with the advantage of SGML to supplement the limitations of HTML, XML syntax is being promoted by the Internet communities with great efforts. This paper will discuss issues related to the development of metadata and introduce the features, structures, functions and use of Metalogy, a general-purpose XML/Metadata system, which is developed under the Digital Museum Project funded by National Science Council, Taiwan.

Keywords: digital library, digital museum, Digital Museum Project, XML, metadata, Metalogy

I. Introduction

Recently, with the rapid development of Internet, researches on digital library/museum have received worldwide attention; and all developed countries are supporting these researches with great enthusiasm. Our country has rich cultural heritage with a wide range of world-class treasures. In addition, many organizations and research institutions in Taiwan possess abundant collections of rare books, historical remains, artifacts and documents on Taiwanese culture. In the past, they were not open to the public due to preservation considerations. Now, through the powerful Internet, we will be able to present these valuable resources on the WWW. Besides increasing public exposures, it will preserve the physical resource that might be otherwise deteriorating.

In Taiwan, major institutions those have digitized their rare collections include National Taiwan University, Academia Sinica, National Central Library, National Palace Museum, National Museum of History, National Museum of Natural Science, and so on (Chang, 1999). To digitize

these valuable resources and present them on the web is their primary task. However, it is much more important to organize these resources based on their characteristics. Therefore, users may retrieve and use them effectively. It is obvious that metadata is vital to digital library/museum systems.

From the perspective of users, a digital system should contain basic functions of retrieval, browse, and link to other related web resources. Usually, a digital library system with a large volume of data will apply database management systems to manage its bibliographic records, digital objects, and www links. In order to avoid the possibility of linking to an invalid URL, people often use mechanism similar to Handle System to maintain links.

At present, retrieval technology can be classified into full-text search and field-based search. Full-text search does not require metadata description of the resource, but it yields lower precision rate. For non-textual images, sound or multi-media information, full-text search could not be used. Therefore, manual creation of metadata to establish field-based bibliography data for various types of digital objects is a pivotal step for digital libraries.

The first step in formulating a metadata is to understand the user demand and object characteristics. Next, the second step is to consider the interoperability among information systems, which depends on the use of standard. For example, users may apply and adapt an internationally-accepted metadata format, such as Dublin Core (Dublin Metadata Core Element Set, http://purl.oclc/dc), EAD (Encoding Archival Description, http://lcweb.loc.gov/ead/), FGDC (Federal Geographic Data committee, http://www.fgdc.gov/), GILS (Government Information Locator Service, http://www.access.gpo.gov/su_docs/gils/index.html), TEI (Text Encoding Initiative Headers, http://www.uic.edu/orgs/tei/)) and metadata syntax, such as SGML, XML, and HTML. Currently, regardless of the metadata format they choose, most digital

library/museum systems use XML or SGML as their metadata syntax. In particular, with the advantage of SGML to supplement the limitations of HTML, XML syntax is being promoted by the Internet communities with great efforts. This paper will discuss issues related to the development of metadata and introduce the features, structures, functions and use of Metalogy, an XML/Metadata general-purpose system, which is developed under the Digital Museum Project funded by National Science Council, Taiwan.

II. Digital Museum Project

National Science Council (NSC) of Taiwan launched "Greeting a New Millennium--A Cross-Century Technology Development Program with Concern for the Humanities" in May 1998 with an intention to strengthen researches on humanity/social science and science education. Digital Museum Project was part of this Theme Project. Its main goals are to integrate and establish a digital museum with an emphasis on Taiwanese culture for Chinese people and to develop educational contents on the Internet (Huang, 1999). By establishing and promoting educational culture/art/science content on the web through the powerful Internet, the public may retrieve or browse information freely; consequently, users may experience its enrichment and enjoy the lifelong learning (Wang, 2000). Furthermore, by promoting digital collections, NSC hopes to stimulate the technology development of multimedia and the growth of content industry.

Since then, Digital Museum Project (DMP) has been progressed into the second phase. During the first year, NSC invited experts/scholars with experience on digital collections to form a collaborative mechanism to promote digital museum researches. Projects can be categorized into two types: topic-based projects and technical support projects (NSC, 1999a). In addition,

DMP Extension (DMPE) is responsible for training and promotion by serving as a bridge for library/museum communities, teachers, industries, and DMP staff.

In respect to contents, topic-based projects in the first phase include local spotlight and traditional culture. Specifically, there are two comprehensive projects on local culture: Discovery of Tamsui River and Taiwanese Aborigines--The Ping-pu Race. On natural science and environmental ecology, there are Butterfly Ecology and Native Plants and Fishes of Taiwan. On traditional culture, there are three projects: Traditional Thoughts and Literatures (The Four Books, Lou-Chuang, Poems of the Tang Dynasty), An Immortal Palace--Han Dynasty Culture and Burials, and Firearms and Ming-Ching Dynasty Warfare (NSC, 1999b).

System technical support projects in the first phase contain Electronic Cultural and Natural Resource Atlas (establishing a common coordination system for time, space and language symbol) and Understanding Ancient Texts--The Written Knowledge Network. In addition, it has three information technology projects: Resources Organization and Searching Specification (establishing metadata interchange format for Chinese information, thesaurus, and searching specification, so individual topic-prototype systems may have international interoperability), Digital Collection System Technology Development, and Research on System Evaluation Standard ("Discovery of Tamsui River" was used for empirical studies on developing evaluation standard and system evaluation methods in order to enhance higher quality and better dissemination effect) (NSC, 1999b).

DMPE was established in August 1998. Its goals are to train skillful staff on digital collection and to promote research results to various communities in our society. Through seminars, training courses for professionals, training sessions for elementary- and middle-school teachers, e-news and news articles on mass media, such as newspapers, magazines, and journals. DMPE

have increased collection institutions and industries' knowledge about digital library/museum, enhanced public interest about digital library/museum, improved the web resource utilization skills of elementary- and middle-school teachers, and trained professionals for digital library/museum (DMPE, 2000).

Currently, Digital Museum Project is in the second phase (Jan. 2000 - Dec. 2000).   During the second phase, it is open to all interested participants.   Among nearly 90 proposals, twelve were funded, including: 1. Treasurers of the National Palace Museum, 2. The World of Xuanzang and the Silk Route, 3. Discovery of Tamsui River (II), 4. Native Artist Digital Museum--Yu-Yu Yang Art Research Center, 5. Historical Photos of Taiwan, 6. Architectural History of Taiwan, 7. Mystery of Human Body, 8. Web Maintenance of "Taiwanese Aborigines--The Ping-pu Group", 9. Ancient Texts and Popular Songs of Tang and Sung Dynasties (II), 10. Native Freshwater Fishes of Taiwan (II), 11. Chinese Medicine and Acupuncture, and 12. Biology-Cultural Diversification of Orchid Island.   Among the twelve, four of them are carried on from the first year's project (#3, 8, 9, 10) (DMPO, 2000).

Technical support projects have been reduced to two: 1. The Implementation for Resources Organization and Searching Specification in Digital Museums, and 2. Technology Development of Digital Watermarking and Software Tools (DMPO, 2000).

III.   Resources Organization and Searching Specification

Before NSC launched Digital Museum Project, the authors and their colleagues initialized a metadata research team, ROSS (http://ross.lis.ntu.edu.tw), under National Taiwan University Digital Library/Museum (NTUDL/M) Project to study Metadata Interchange for Chinese

Information (MICI) in March 1997. Its research scope contains the following: to understand the history and features of collections, to study various metadata formats both domestically and internationally, to understand relations among metadata, database and the system framework, and to understand information demand and retrieval behavior of potential users. ROSS held that, our metadata should be able to describe attributes of the collections, to provide users with the mandatory access points, to enhance interoperability among different digital libraries to exchange information, and to take consideration of the quality of cataloging. Most digital collections of NTUDL/M were historical documents. After studying the characteristics of historical documents, ROSS made in-depth studies of the metadata of similar types of collections, including CIMI (Computer Interchange of Museum Information, http://www.cimi.org, describing museum art collections) and EAD (describing archival information). Nevertheless, due to cultural and characteristic differences, these metadata formats are not sufficient to describe Chinese special collections. Hence, it is necessary to focus on the research on Chinese metadata, which is the main goal of ROSS (Chen, Chen, & Chen, 1999).

Resources Organization and Searching Specification Project was carried out by the authors in 1998, and it is continued in 2000 entitled "The Implementation for Resources Organization and Searching Specification in Digital Museums" as a technical support sub-project of Digital Museum Project, NSC. Research of the second year project is focused on related issues on information organization and retrieval in Chinese digital libraries/museums, which include data storage and management system design, user-demand and information retrieval behaviors, and integration among different systems.

Besides historical documents, ROSS began to work on metadata for other resource types (objects, ancient maps, photos/pictures and butterfly specimen) since November 1998. During the process of metadata development, in addition to the frequent discussions with experts/scholars,

we studied how similar digital museums record their collections. In the first year, ROSS was responsible for the metadata development for two topic-based prototypes, "Discovery of Tamsui River" and "Butterfly Ecology". In the second year, the main task of ROSS is to develop a management system capable of handling various types of metadata, Metalogy, for all topic-based prototype projects.

IV. Metadata Interchange for Chinese Information

Metadata Interchange for Chinese Information (MICI) adopt the Dublin Core's 15-element as its basic structure. However, in order to describe the attributes of our rich cultural heritage and be more precise on the semantics of the collection descriptions, element qualifiers were added to the appropriate elements based on the attributes of collections. As a result, while extending the scope of its application, it is compatible with international standards as well. This set of Dublin Core-based MICI with self-defined qualifiers is named as MICI-DC.

MICI-DC has been used to catalogue various types of resources: historical documents, maps, photos/pictures, calligraphies, objects, and Buddhism scriptures/paintings. In addition to DC's official qualifiers, individual institutions may define their own qualifiers based on the attributes of its collections. Users may choose DC's 15-element and qualifiers and adjust the orders of these elements according to their needs. This will be compatible with international standards, and meanwhile, allow users with great flexibilities in meeting the local needs. In order to make it easier for users to catalogue resources using MICI-DC, a tagging guide was complied with explanations and examples on the 15-element and their qualifiers. Thus, users may DIY their MICI-DC without further assistance. For details on MICI-DC, please see Appendix 1.

V. Metalogy, an XML/Metadata System

With the advantages of SGML and free from the complications of SGML, XML is widely applied on the web.   In addition, it may supplement the flexibility and preciseness which HTML lacks of.   XML syntax has been promoted by the Internet communities and the Database communities with great enthusiasm.   Thus, when ROSS was about to design a metadata management system, we decided to use XML syntax as a basis for information interchange among databases.   However, in addition to syntax, we need to consider the semantics. Currently, there are various types of metadata formats, and many communities are developing their own metadata format to fulfill their domain-specific need; thus, flexibility is essential for a metadata management system.   That is, one cannot develop a system based on one particular type of metadata; it should provide users the freedom to choose their own metadata types. Therefore, developing a general-purpose XML/Metadata system is the main concern.   The designing concept and structure is described below:

1. The features and structures of Metalogy (version 1.0) System

Metalogy, an XML/Metadata management system, is developed under the Digital Museum Project funded by National Science Council over a one-year period.   This system may be used to develop databases for any digital museum, digital library, and digital archive in various subjects.   Functions include database set-up by the DTD, metadata edit, authority file (thesaurus) edit, retrieval (including both Window and Web interfaces), and import/export of XML files. Features of this system include:

a.   System schema is mainly based on the input DTD

b.   System allows co-existence of different types of DTD

c. System is capable to retrieve different formats of data at the same time

d. System allows users to adjust the element format designated by DTD and access restrictions based on the schema

e. System allows users to define their hyperlink, index, retrieval and display of elements with a user-friendly interface.

f. Data import/export conforms with its DTD format

g. System is capable to determine whether the imported data conforms to the designated DTD format and to check for duplication of input data

h. System is capable to process structured element, multimedia, and texts.

i. System contains management functions of access control, transaction log, and so on

j. System has web search function, which allows end-users to retrieve information from the database via WWW interface.

Structure of Metalogy system can be shown in Figure 1.

2. System developing tools and its functions

The developing environment of Metalogy is Delphi 5.0, and the programming language of Web Searching is ASP.   The back-end database systems could be Oracle or SQL sever.

Currently, Metalogy has developed the following functions:

a. Input DTD to set up a database

Simply import any type of XML DTD, users may set up a corresponding database and access the cataloging display screen.   Please refer to Figure 2.

b. Define the system schema

DTD declaration does not define the data format, conversion specification, maximum length

of input character, authority control, index file, and so on.   As a result, while importing

DTD, although the schema will be automatically generated by the system, it is necessary to

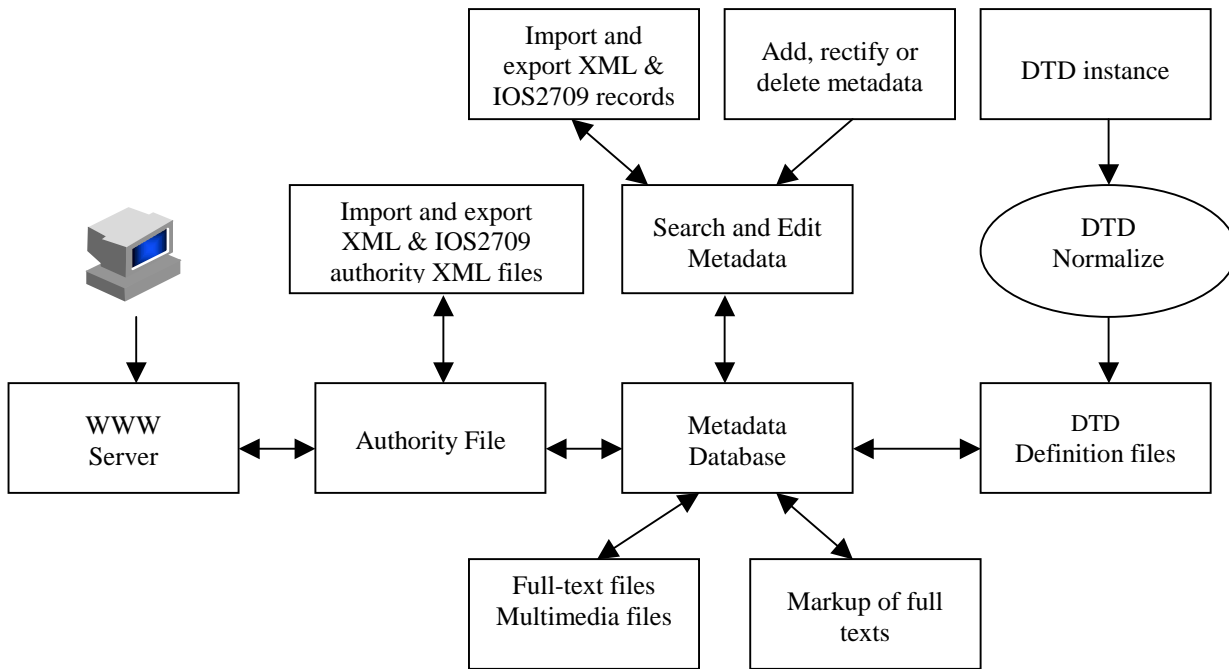check and modify the schema manually.
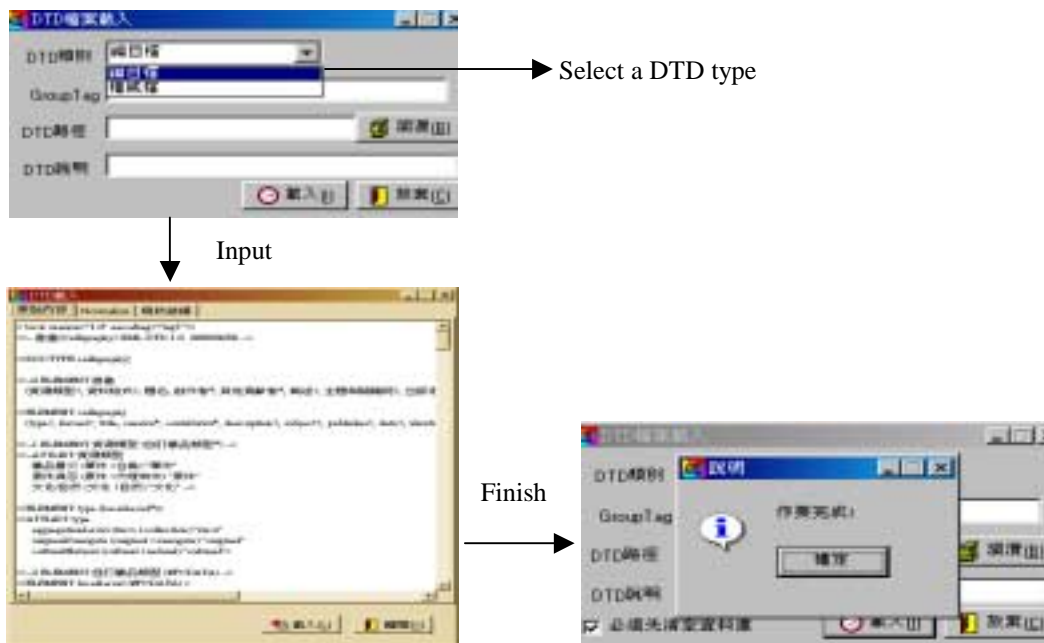

Figure 1: Structure of Metalogy System

Figure 2: Input DTD

c. Metadata cataloging

After choosing the cataloging Meta type, users may add, rectify or delete any of these records. While editing a record, users may duplicate, delete, or bring out the sub-element based on the mapping.   If this element is input by code or authority control, besides direct input, users may browse or retrieve through the display of code or authority screen.   In addition, users may call for the retrieval function directly on the cataloging screen to check for the needed records immediately.

d. Establish thesaurus and authority file

Users may construct the thesaurus and authority file by the same process as the metadata cataloging records.

e. Management and descriptions of digital objects

Users may carry out the cataloging with brief descriptions by a single multimedia file or a batch of multimedia files. If the description is completed prior to metadata cataloging, users may build the multimedia link while cataloging.   If it is done afterwards, users may rectify the metadata to proceed with the link.   For importing large volume of multimedia files, batch processing is recommended.

f. Search function

Search function may be carried out by a single Meta field or all the Meta fields in the database.   It may be conducted by exact search or fuzzy search.   Please refer to Figure 3.

g. Search function on authority file

Same as described in the search function.

Figure 3: Search Function

h. Import XML files

Metalogy is capable to export designated format of XML files and import XML files from other systems.   After XML DTD is imported to Metalogy, Metalogy will accept XML files with one single record or more records.   Users may set decision rules beforehand, so during the import of XML files, system will be able to determine whether a particular record is in the system database.   Please refer to Figure 4.

i. Export XML files

Through XML, Metalogy may exchange metadata with other systems and export well-formed XML files for user access.   Users may export certain number of records or a batch of records through search, and they may set up variables to define the field for export. Please refer to Figure 5.

j. Access control

System provides management functions of establishing basic information and access control restrictions on users.   While initializing the system, users need to input the user name and password, and system will allow users to use Metalogy accordingly after verifying.
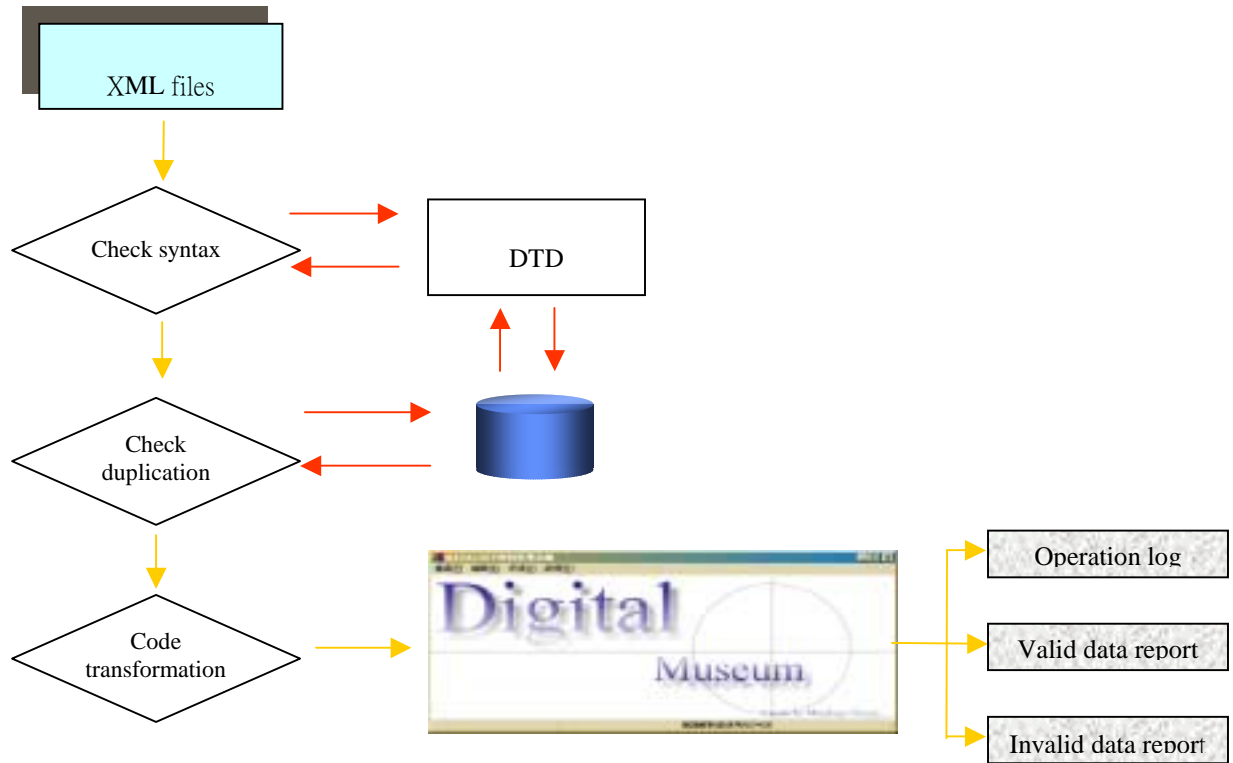
Figure 4: Import XML files

Select the path for storage

Key in the ID number

Select the export
XML files format
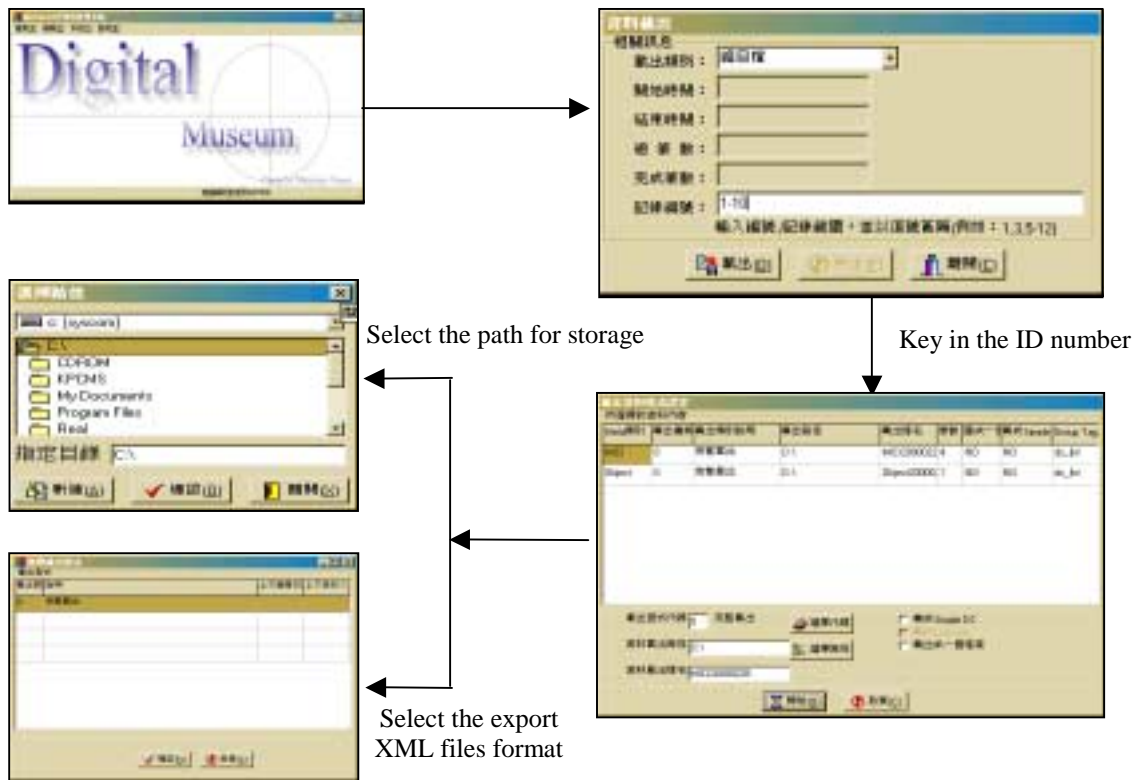
Figure 5: Export XML files

k. Message management

Without re-compiling the system, system managers could edit the error message based on their needs with the customized description, icon, and button in order to avoid misunderstanding of users due to the ambiguity of the message.

l. Web search function

It provides the same search function as the internal search menu. Please refer to Figure 6.

Figure 6: Web Search

3. Metalogy User's Manuel

Metalogy User's Manuel was compiled to guide users on how to use this system. It includes step-by-step diagrams and descriptions on how to install, setup, use and operate the system. Through this manual, users may easily manage its system by themselves without further assistance.

4. Metadata and DTD instance

While designing a digital library/museum/archive, one must design the metadata based on its data. In addition, while using Metalogy, one must use XML/DTD to express metadata. Since the formulation of Metadata and DTD require an in-depth understanding of the resource characteristics (which is rather time consuming) and need to consider the interoperability among systems, it is better to use some kind of pre-defined DTD. This year, the authors design several types of metadata and DTD for the National Palace Museum (including the metadata and DTD for calligraphies, objects, scriptures, exhibition, reference, name authority files, title authority files, geographical names thesaurus, time thesaurus) with a tagging guide and examples.

VI. Conclusion

Metadata technology is the core of digital library systems, and XML is the most popular syntax for metadata. Major types of metadata format under user include EAD, GILS, FGDC, MARC, CIMI, TEI, DC, and so on. In addition, many formats are extensions of the above metadata formats. Furthermore, one institution may hold different types of resource and use different types of metadata format. Those are the differences between digital libraries and traditional libraries. Thus, when designing a metadata management system, one should not base on a particular format. Rather, it is more appropriate for the system designers to use XML as a core that would be capable of handling various metadata formats and this is the concept behind the development of Metalogy. For the time being, Metalogy is available upon request and it is free of charge. User feedback and comments are welcomed for further improvement of Metalogy.

Reference

Chang, San-Cheng (1999). Digitalize national collection (In Chinese). In Science and Technology Advisory Group of the Executive Yuan of the Republic of China (Ed.), The 9th Executive Yuan of the Republic of China（SRB）meeting (In Chinese) (153-182). Taipei: Editor.

Hwang, Jenn-Tai (1999). Greeting a New Millennium -- A Cross-Century Technology Development Program with Concern for the Humanities (In Chinese), National science council monthly 27(7), 715-718.

Wang, Mei-Yu (2000). Achievement of National Science Council Digital Museum Project (In Chinese), National science council monthly 28(4), 249-253.

NSC, National Science Council (1999a). National Science Council Digital Museum Project asks for subject project 2000 (In Chinese) [Announce]. Taipei. Retrieved Dec. 25, 1999 from the World Wide Web: http://www.nsc.gov.tw/announce/89digi_museum.html

NSC, National Science Council (1999b). The general situation of National Science Council Digital Museum Project (In Chinese). Retrieved Dec. 25, 1999 from the World Wide Web: http://www.nsc.gov.tw/y2k/dml/880209DATA2.html

DMPE, Digital Museum Project Extension, National Science Council (1999). National Science Council Digital Museum Project summary: Digital Museum Project Extension, National Science Council (In Chinese). Retrieved July. 11, 2000 from the World Wide Web: http://mars.csie.ntu.edu.tw/~dlm/plan/1st/intro12.htm

DMPO, Digital Museum Project Office, National Science Council (2000). An Introduction of National Science Council Digital Museum Project (In Chinese). Retrieved July. 11, 2000 from the World Wide Web: http://dm.ee.ntu.edu.tw/projects.htm

Chen, Hsueh-hua, Chen, Chao-chen, & Chen, Kuang-hua (1999). Metadata Interchange for Chinese Information. In Ching-chih Chen (Ed.), IT and Global Digital Library Development (pp.65-74). West Newton: MicroUse Information.

Appendix 1 : MICI-DC

Element	Qualifier			
Type	Aggregation Level	Item/ Collection		
	Original Surrogate	Original/ Surrogate		
	Cultural Natural	Cultural / Natural		
	DC Type	interactive resource		
		dataset		
		event		
		image		
		sound		
		service		
		software		
		collection		
		text		
	LocalLevel			
Format	Medium			
	Extent (Size, duration)	Quantity		
		Dimension	Name	
			MeasurementsUnit	
			Position	
Title	Main			
	Subtitle			
	Alternative			
Description	Acquisition	Method		
		Source		
		Price		
	Physical Description	Illustration		
		Color		
		Material		
		Attachments		
		FormWholeObject		
		Part Of Object		
		Scale		
	Abstract / Synopsis			
	Place			
	Collection Or Site Information	Locality		
		Name		
		Date Gathered		
		Field Number		
		Method of Collection		
		Type Of Site		
		Coordinates		
		Coordinates of Object		
		Phonomena		
		Accompanying Object		
		Cultural Layer		
		Geological Period		
		Age		
		Environmental Details		
	Seal Type	From Artist	Artist Inscription Seal Locality	

Element	Qualifier			
			Artist Inscription	
			Artist Seal	
		About Colophon	Colophon Locality	
			Colophon Writer	
			Colophon Seal	
			Colophon Full Text	
		About Label	Label Locality	
			Label Writer	
			Label Seal	
			Label Full Text	
		About Loose Leaf	Loose Leaf Writer	
			Loose Leaf Seal	
			Loose Leaf Full Text	
		From Collector	Collector Seal Locality	
			Collector Seal Inscription	
	Inscription	Series Number		
		Position		
		Category		
		Style		
		Content	InscriptionContent Full Text	
			InscriptionContent Image	
	Decoration	Series Number		
		Position		
		Category		
	Transcription			
	Mount	Volume Cover		
		Protective Covering Case		
		Book Case		
	Release	Edition Name		
		Binding		
		Style Form	Border/ Column	
			Center Boundary/ Row	
			Block Heart	
			Frame Mark	
		Lines Per Page		
		Font		
	Exhibition	Exhibition Name		
		Exhibition Size		
		Object Description	Exhibition Description	
			Recommendation	
			Web Description	
	Condition			
	Grade			
	Notes			
Subject	Subject Descriptor	Primary Subject		
		Secondary Subject		
		Other Subject		
		Situation		
		Function		
		Technique	Series Number	
			Position	

Element	Qualifier	
		Category
		Style And Movement
		Personal Name
		Corporate Body
	Keywords	
Creator	Personal Name	
	Dynasty	
	Birth Place	
	Corporate Body	
	Role	
Contributor	Personal Name	
	Dynasty	
	Birth Place	
	Corporate Body	
	Role	
Publisher		
Date	Cataloging Date	
	Created	
	Issued	
	Acquired	
	Modified	
Identifier	CallNumber	
	AccessionNumber	
	URI	
Source		
Relation	References	Reference Work
		Collection Catalogue
		Research Material
	Has Part	Title
		Creator
		Contributor
		Pagination
	Is Part Of	
	Citation	
Language	Cataloging Language	
	Item Language	
Coverage	Place	Place Of Use
		Scope Of Coverage
		Place Of Event
	Time	Period Of Use
		Date Of Event
	Lat Long	
Rights	Owner Name	
	Owner Country	