A Practical Metadata Approach towards Chinese Rubbings

Ya-ning Chen
arthur@sinica.edu.tw
Shu-jiun Chen
sophy@sinica.edu.tw

Computing Centre
Academia Sinica
Taipei, Taiwan, R.O.C.

Voice: +886+2+27899861
Fax: +886+2+27836444

Abstract

Chinese rubbings, rubbed from original objects of stone, wood and bronze, are a special rare collection and deemed precious historical and cultural heritage. The study reports the metadata planning for the Digital Rubbings Project initiated at Academia Sinica which houses one of the largest and most valued Chinese rubbings around the world. In response to Chinese rubbings across various domains, with very different from book and journal collection, the Metadata Working Group aims to develop a standard set of workflows for the metadata applications. Firstly, this paper adopts IFLA's FRBR model as an analysis framework to design metadata elements and distribution in a context way. Then a set of workflows for rubbings metadata analysis and implementation are highlighted. The workflows involve the acquisition of user needs and data requirements, metadata analysis, mapping, system implementation, guidelines and tutorials for Digital Chinese rubbings. The research findings include (1) metadata types and relations should not be limited to objects only; (2) re-examination of IFLA's FRBR's model; (3) a collaborative model for metadata applications. Finally, this paper suggests that "project management" and "a generic metadata framework" are essential for those organizations conducting more than two metadata projects simultaneously.

Introduction

Academia Sinica has dedicated the Digital Library and Museum Project which is

funded by the National Science Council (NSC) in Taiwan since 1998 as part of the "Greeting a New Millenium--A 21st Century Science Development Program with Concern for the Humanities as a Main Theme. The strong needs of metadata implementation was raised among 10 subject-oriented projects at the very beginning stage. The major purpose of this paper addresses a practical issue of metadata applications in Digital Library Project by use of a case study of Chinese Rubbings at Academia Sinica in Taiwan.

Metadata for Chinese rubbings

According to an informal survey of stakeholders around the world, currently Chinese rubbings are housed by different organizations across Japan, USA, France, Mainland China and Taiwan The Academia Sinica in Taiwan are the major stakeholders. In Academia Sinica, Chinese rubbings are categorized into four types: Buddhism Images, Stone Tablet's Text (inscriptions), Tomb Tablet's Texts, and others. The Stone Tablet's Text is the key collections among these four. Generally research of Chinese rubbings is across arts, history, and social sciences. Within this paper, the Chinese rubbings will be taken as a case study for metadata implementation and discussion.

The Metadata workflow

Metadata workflow is an essential means for metadata applications, and provides a fundamental framework to get consistent analysis and result of metadata. There are six phases of metadata workflow composed of : (1) acquisition of user needs and requirements, (2) metadata analysis, (3) mapping, (4) system implementation, (5) guidelines, and (6) tutorials.

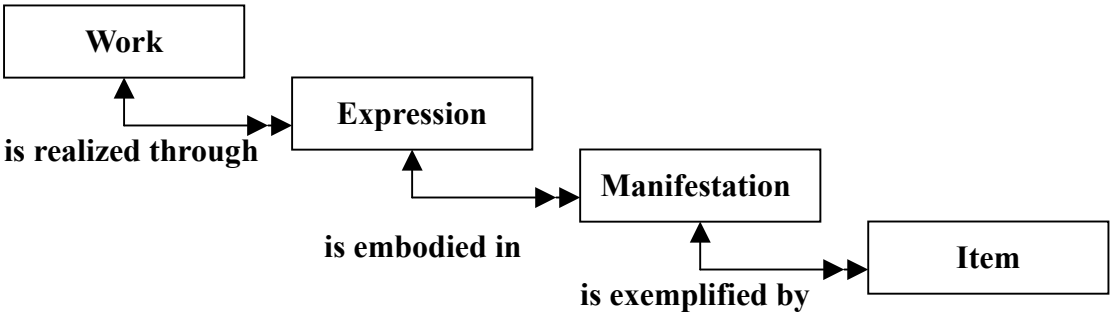

Figure 1: IFLA's FRBR model (IFLA Study on the Functional Requirements for Bibliographic Records, 1998, p. 13)

Acquisition of user needs and data requirements

The first step to do metadata is to make sure the user needs. After interviewing with stakeholders, user needs and data attributes can be identified and transformed into metadata. The extent of describing metadata for rubbings will dependent on user's need, especially in view of stakeholders or content providers. During the interviewing session, some key points must be ascertain by questions which include functions of description, indexing and retrieval, representation, access, rights management, etc..

Metadata analysis

In Academia Sinica, it is a critical issue to develop a systematical approach based on some theoretical principles for metadata analysis. Metadata analysis is based on SMART (Sinica Metadata Architecture and Research Task) model which is abridged from IFLA model. IFLA developed the 'Functional Requirements for Bibliographic Records' (FRBR) in 1998 when completion of a pilot project study of functional requirements for cataloging records. This IFLA's FRBR model creates four functional levels for bibliographic records as Figure 1 shows. This model has been adopted by INDECS/DOI community as a common model in support of interoperable metadata in e-commerce field (Bearman, Miller, Rust, Trant, and Weibel, 1999). However, the model currently has not been designed for metadata analysis around the world when one institution initiates to start implementation of metadata.

The SMART model is expanded into work, expression, manifestation/media, and item/rights management levels as Figure 2. Compared with FRBR, only two differences exist for these models. First, the four functional levels in FRBR are extended into include more matters in content, such as manifestation and media, as well as item and rights management. Second, the SMART model emphasizes on more detailed associations among four levels from top to bottom. Based on this model, the horizontal and vertical categories of relationship would be established. The horizontal relationship is to indicate the relation at same level, but each level is with different meanings and functions. For example, different works would be connected with specific relation, such as part and whole, reference, dependent and sequential relationship, so as for item/rights management level could be the transferred history of copyright, ownership, access control. Within the vertical one, the relationship focus will be more hierarchical. For instance, the 'Phantom of the Opera' is a work and can be expressed both in a novel and an opera that share identical but different expression levels and relations. The novel one can be manifested both in paper format of a book and in digital form of CDROM.

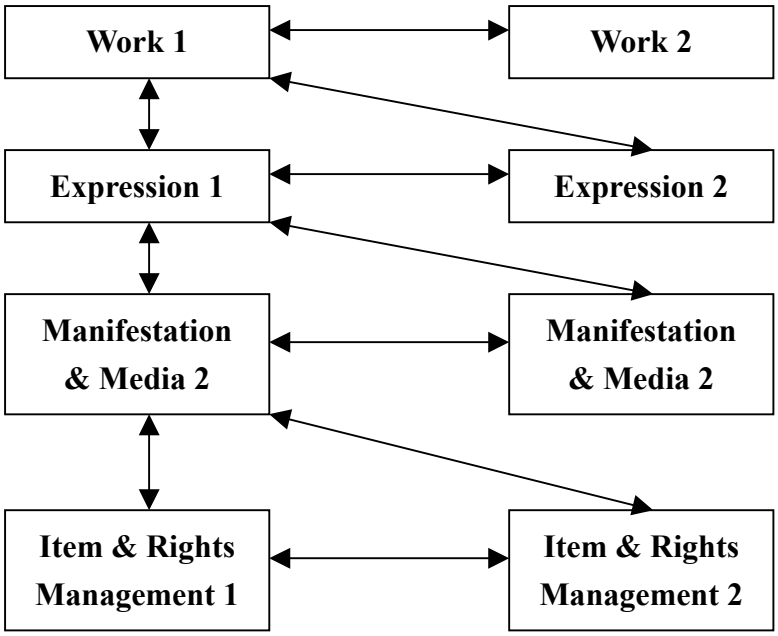

Figure 2: SMART model

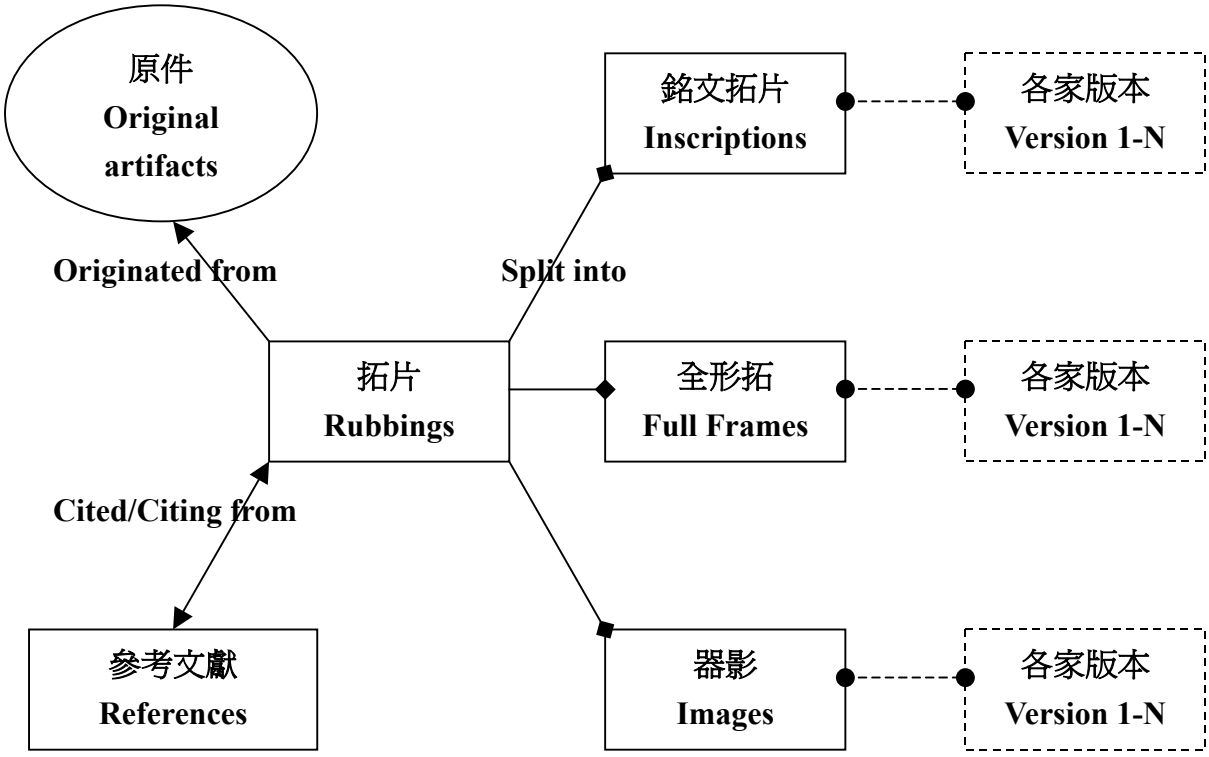

Figure 3: An analysis of Chinese Rubbings for various objects

The SMART model has been used as a metadata analysis model for many metadata projects at Academia Sinica. The application in the case of Chinese Rubbings has the following results.

Analysis 1: Structure and context of metadata objects

- The relationship of Chinese rubbings consists of different objects related to rubbings, such as original copy or artifacts (原件), cited and citing references, seals (印記), as well as different rubbings forms (full frames (全形拓), inscriptions (銘文拓片), and images (器影)) and versions (版本) for different researching purposes. This examination helps to clearly identify the distinction of elements which are distributed in various objects. *(please refer to Figure 3)*
- The Chinese rubbings can be separated into diverse components from the whole to the tiny one, according to the requirements from stakeholders or researchers' definitions. This shows the fundamental information unit in the metadata process, and gives requirement for system representation and design. *(please refer to Figure 4, 5)*

Analysis 2: Attributes and distribution of metadata elements

- Examine whether any metadata element for a specified object is skipped or lost by the categorized distribution of metadata elements. With the model the metadata team can find out the distribution pattern of metadata elements and provide users the diagnosed result. This help to ensure whether this skipped gap is accordant with requirements or not, and to re-arrange or add new metadata elements for Chinese rubbings. (Table 1)
- Examine whether new elements or existing elements should be created or deleted. This model aggregates all elements by levels to enable the decision of any integrating the related elements into a structured one, splitting from one element into more, or constructing a deeper level by use of qualifiers.
- Define descriptive indication and annotation sections. In descriptive indication section any attributed information of an object could be described as metadata elements without distinction, because these information are objective based on agreements for all viewpoints. These elements include rubbing title (拓片題名), calligrapher (書畫人), rubbing maker (拓工), decoration (圖像), dimension, and so on. In annotation section some metadata elements could be explained or annotated into deeper dimension in terms of various perspectives for research, and these different annotations are juxtaposed.

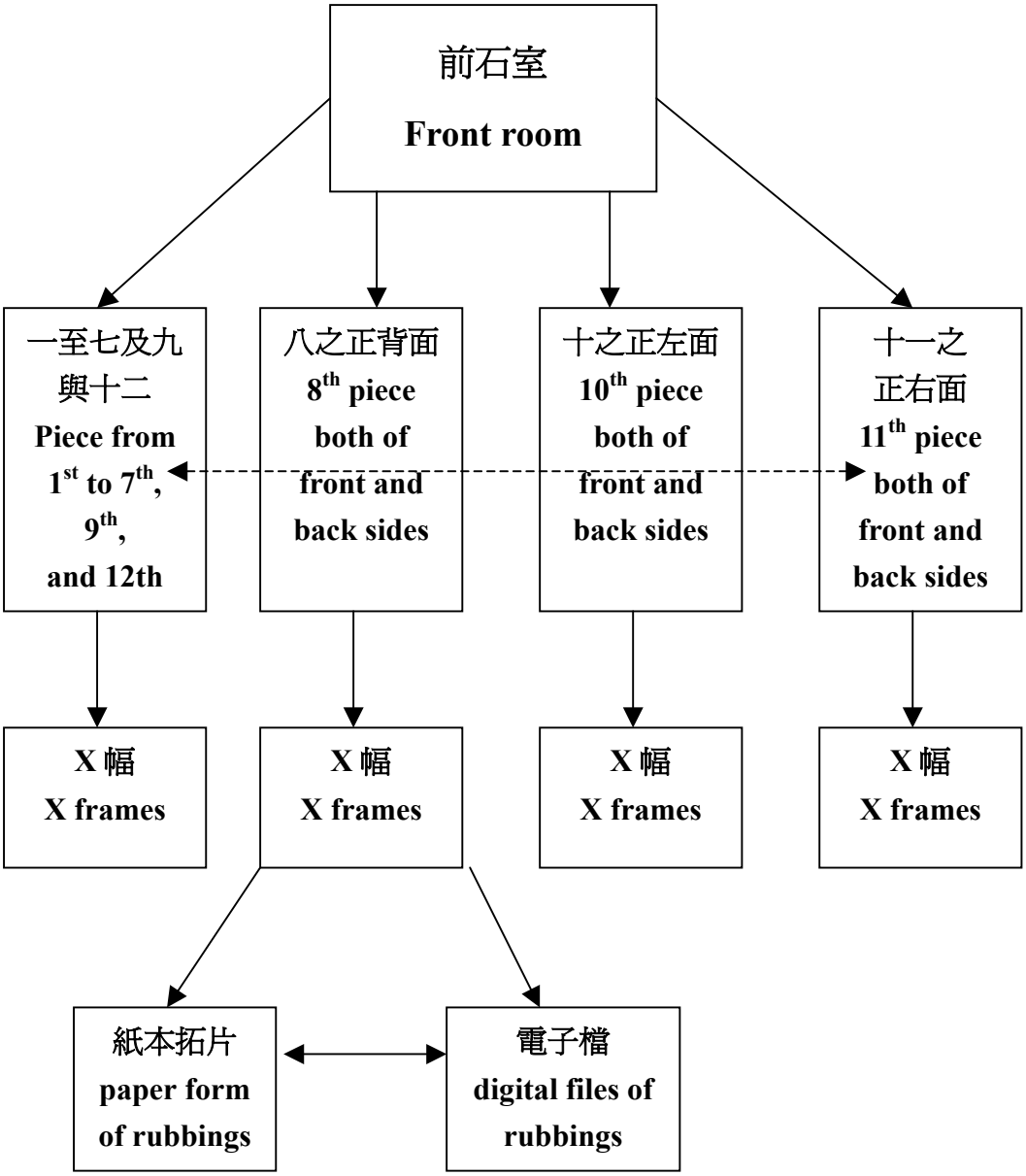

Figure 4: An analysis of context and information unit for Wu Liang Tomb (武氏祠) Chinese rubbings

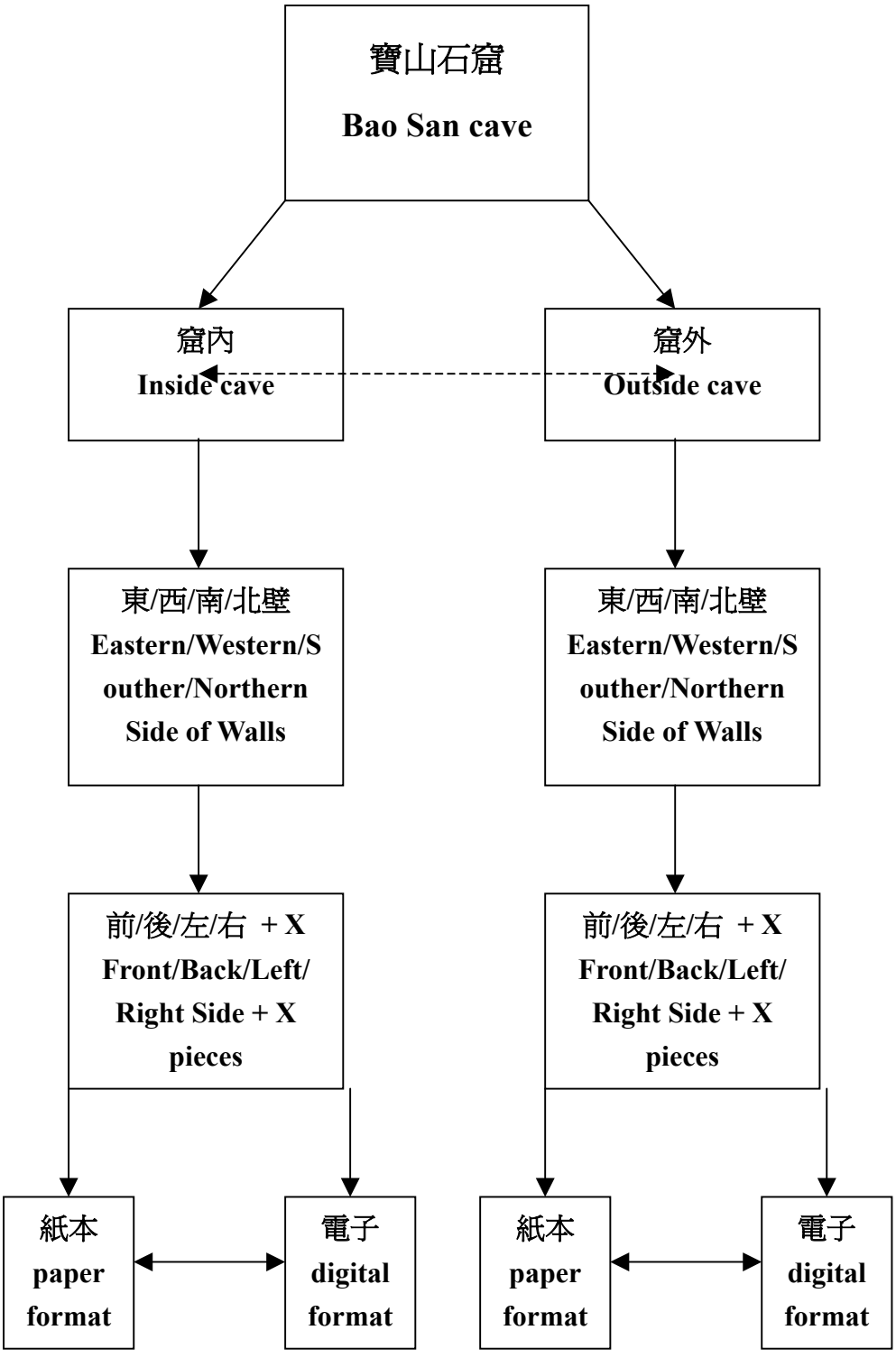

Figure 5: An analysis of context and information unit for Bao San Cave (寶山石窟) Chinese rubbings in Sui Dynasty

Metadata Elements	Description	Annotation	Work/Expression /Manifestion/Item	Rubbings	Original Artifacts	Digital	Others
撰文人Author	*		W	V	V		
書體Script	*		W	V	V		
文體Literary style	*		W	V	V		
刻法Relief	*		W		V		
圖像Decoration	*		W	V	V		
材質Material	*		W		V		
造形Shape	*		W		V		
墨色Ink	*		W	V			
釋文Transcription	*		W	V	V		
類型Type	*		W	V			
印記Seal	*		W	V			V
題跋Colophon	*		W	V			V
類目Class		*	W				V
出土狀況Provenance	*		W		V		
題名Title	*		W	V	V		
年代Date	*		W		V		
作品語言	*		W	V	V		
拓工Rubbing maker	*		W	V			
版本Edition	*		M	V			
數量Quantity	*		M	V			
裝潢Mount	*		M	V			
高廣Dimensions(高Heig	*		M	V	V		
登錄號Accession#	*		I				V
電腦檔編號	*		M			V	
收藏者Past Collector	*		I	V	V		V
入藏Acquisition	*		I				V
版權Rights(Owner/Credi	*		I			V	V
拓片狀況Conditions	*		M	V			
來源Source	*		M				V
著錄Reference	*		W				V
附註Notes		*	W	V	V		V
關連Relation	*		W				V

Table 1: Distribution of metadata elements for Chinese rubbings

Eventually, metadata elements of Chinese rubbings can be analyzed in a context way to be ascertain of different objects for rubbings, and be distributed into different levels to check the right location of metadata against with each metadata element.

Metadata mapping

The purpose of mapping project's elements into existing metadata formats is to achieve the semantic interoperability between different resources. Though many organizations have developed mapping tables and software for different metadata standards, some enhancing tasks are still needed to do for much broader applications and interoperability to metadata mappings. The SMART Projects develops the metadata record structure composed of CORE and INDIVIDUAL parts (see the Figure 6). The structure is designed for a communication bridge of diverse metadata formats and developed on an OO model basis. The CORE elements share the common characteristics of four types of rubbings. The specific elements for deeper or special requirements beyond CORE elements will be located into the INDIVIDUAL elements. CORE represents basic elements for all kinds of rubbings, but INDIVIDUAL is only required when the special needs arisen from the attributes of rubbings. Dublin Core currently is used as a core element set for the experimental trial in SMART Projects. The strength of CORE-INDIVIDUAL metadata record structure is that existing metadata formats and elements can be integrated into this unified structure smoothly, furthermore only one set of metadata elements schema will be maintained in system as Figure 7 & 8.

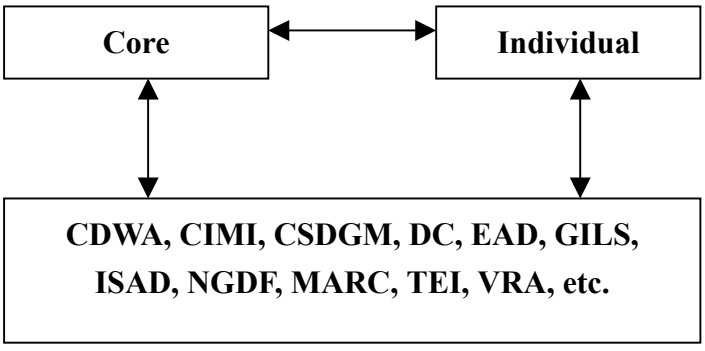

Figure 6: A metadata record structure

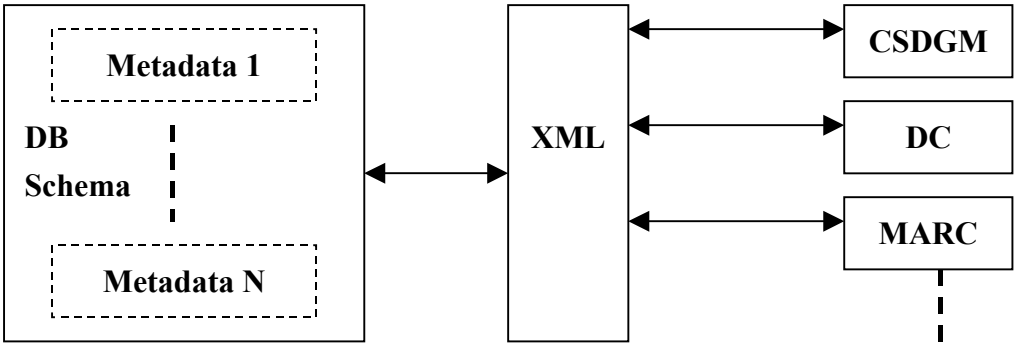

Figure 7: A relation of metadta structure and metadata formats

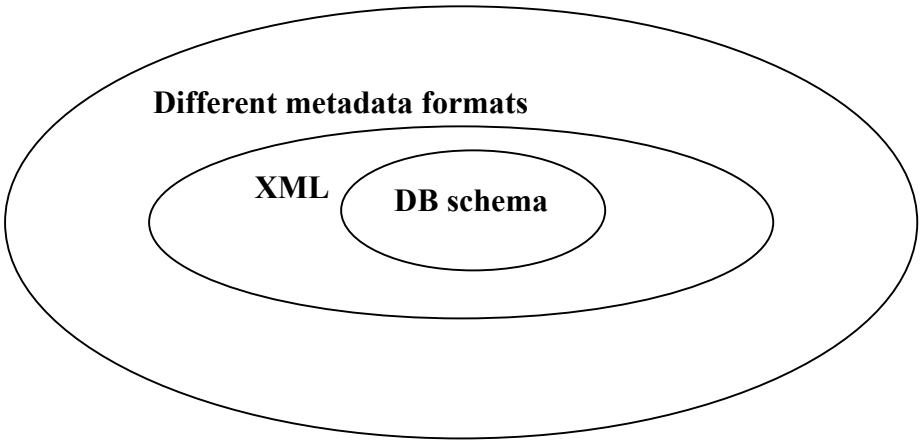

Figure 8: A relation of metadta structure and metadata formats

The import and export mechanism for interchange among different metadata formats in SMART will require XML DTD. Certainly resource description framework (RDF) will be the next task as part of interchange tool in the near future to achieve deeper semantic interoperability.

System implementation

Metadata systems allow the input of metadata descriptions which is based on the metadata analysis and standard, as well as a metadata reporting and search engine for use on the Internet. Software designers of metadata system will participate in the process of metadata analysis and interview with users to ensure the metadata requirements. In the process of designing a metadata system, some principles are persistent as follows (see Figure 9 -12):

- An OO approach is adopted to design for metadata elements and system.
- The capability of processing language characters should be multi-lingual and UNICODE standard.
- Only one set schema of metadata elements will exist in the system, but any metadata elements can be customized to the display labels according to Chinese rubbings scholars' needs.
- The system enables diversified metadata formats to import and export among different metadata systems by means of XML DTD mechanism. At current stage the Dublin Core Element Set has been completed for the purpose. Soon other metadata standards will be developed and integrated, such as MARC (e.g. USMARC, CMARC), content standard for digital geospatial metadata (CSDGM), and so on.

At the moment the metadata system developed by SMART project is available for trial and confirm whether the metadata system is appropriate for users. Certainly some revisions are required to do when the feedbacks are collected.

Figure 9: Metadata system screens for set up elements

Figure 10: Metadata system screen for database schema

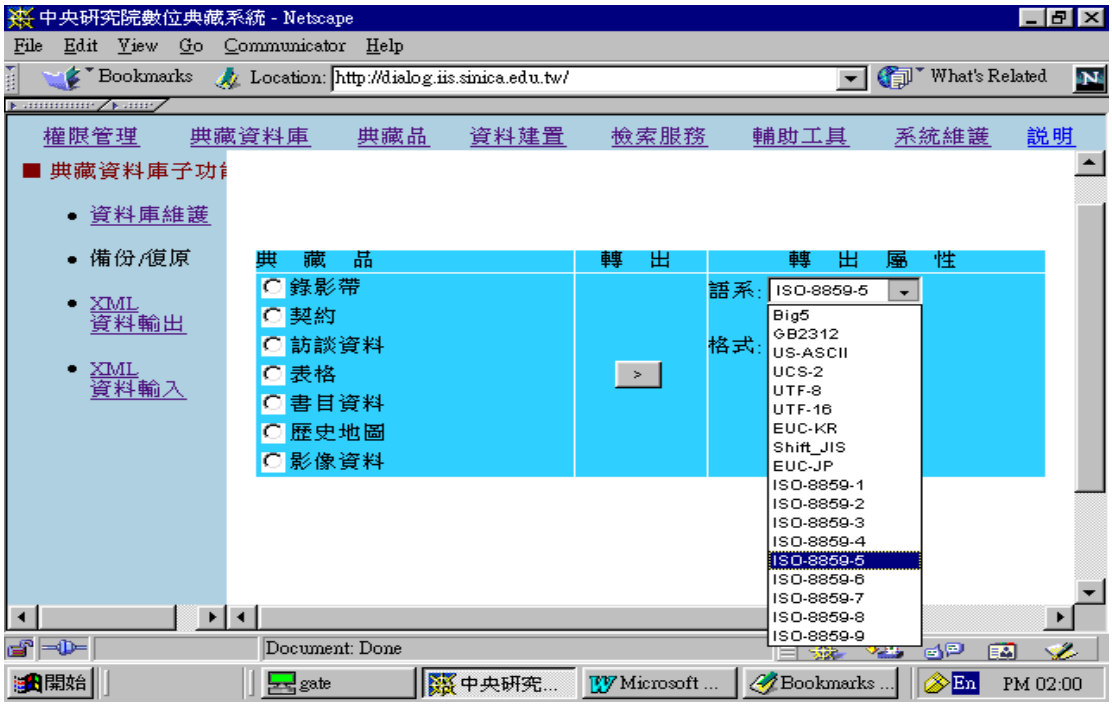

Figure 11: Metadata system screen for export

Figure 12: Metadata system screen for import

Guidelines and tutorials

The final procedure is to complete the metadata guidelines on the Chinese rubbings case to get a consistent result of metadata records from members of the Chinese rubbings project. The tutorials are designed to deliver the metadata task, for example, how to use the metadata, system and guidelines in order to achieve a well quality management for metadata system.

Findings

With the implementation of Chinese Rubbings case, the SMART members achieve several research findings that will be useful references for any implementation of metadata projects and tasks.

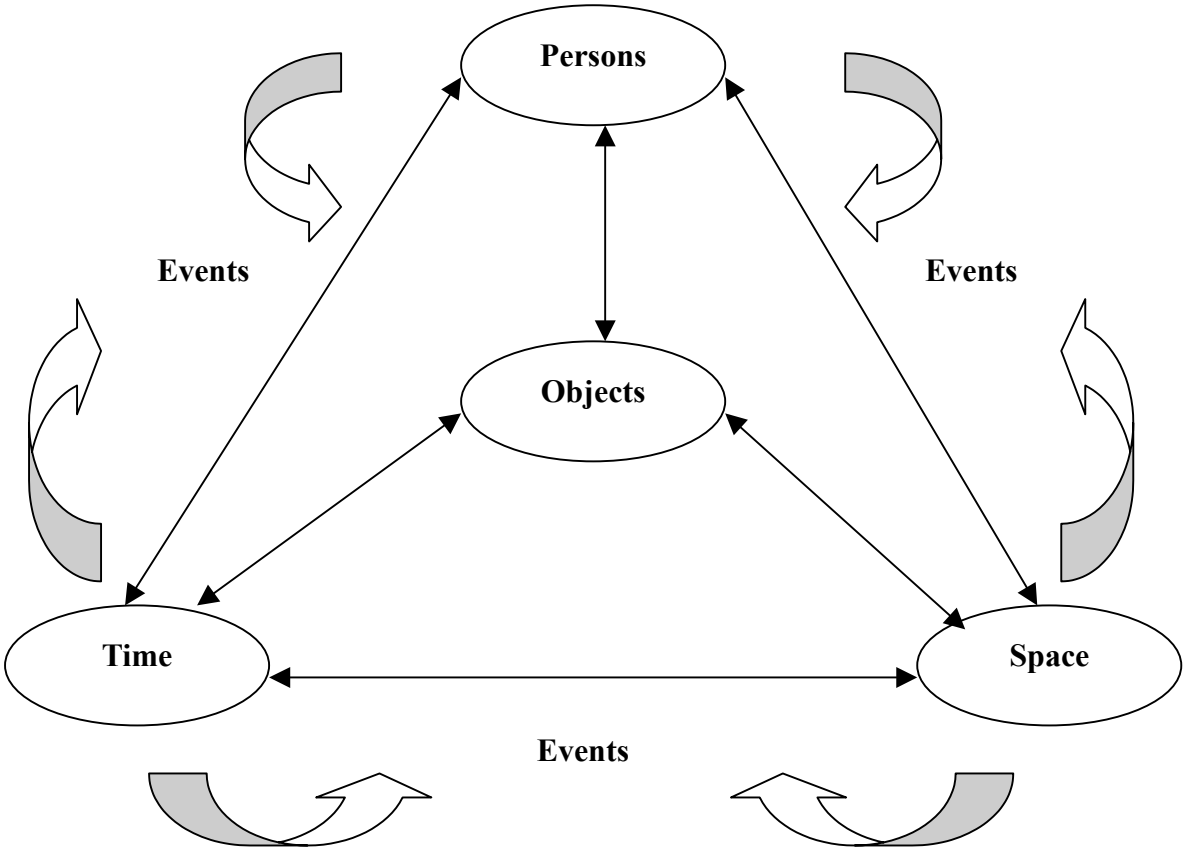

Figure 13: Types and relations of metadata

Types and relationships of metadata

- From current development of metadata standards, metadata formats can be seen in four directions. *The first* is originated from computer science for network protocols, such as IAFA/WHOIS++ Templates, LDIF, SOIF, and URC. *The second* is oriented for a diverse community of stakeholders including museums, archives, libraries and governments, such as CIMI, CDWA, ISAD, EAD, Dublin Core, and GILS. *The third* focuses on disciplinary domain. For example, TEI is designed by a group of the humanities and linguistics fields, so as CSDGM is for geospatial information. *The fourth* emphasizes on establishing one common framework across domains, such as Dublin Core and RDF.
- From the practices of the Chinese Rubbings in SMART Projects, we find that metadata types are a very diverse composition of OBJECTS, PERSIONS, TIME, SPACE and EVENTS, as well as ASSOCIATIOMS among these five types (please refer to Figure 13). It seems that most of metadata contents are nearly emphasized on OBJECTS, though some other cases have special concerns of SPACE, such as CSDGM in USA and national geospatial data framework (NGDF) in UK. In the Chinese Rubbings case metadata is proposed to integrate metadata, image of rubbings and the geographic information system (GIS) in order to discover new research meanings and themes. In the near future, gazetteer will be included to bring in unexpected associations and ideas ever before for rubbings.

Metadata differs from library catalogs

- In essence metadata is different from library catalog in terms of process matters, characteristic of objects, and approaches. Originally library catalogs are a kind of inventory list. With advancement of information technology and growth of publications, catalogs have evolved from card catalogs to OPACs and have attempted to extend into Internet space. In the age of digital/hybrid library, metadata has been introduced to organize born-digital information in a context way, but it is positioned to be different from library catalog.
- From the process perspective, objects of library catalogs mainly focus on books and serials, but metadata deals with a various combination of print on paper within an organization and born-digital information on Internet.
- The scope of library catalogs is traditionally for the collection of one organization and the characteristic of library catalogs tends to be static. Metadata is much more in a dynamic of associated digital information under network circumstances.

- The library catalogs emphasize on the information analysis of "aboutness."(Burnett, Ng., and Park, 1999) The metadata has extended to knowledge context of the resources with the associations among objects, persons, space, time, and events.

Re-examination of IFLA's FRBR model

The IFLA's FRBR model needs more case studies to verify its rightness for analysis of metadata elements.

- According to the concept of IFLA's FRBR model, WORK, EXPRESSION, MANIFESTATION, and ITEM are four independent entities for any objects. In the practical experiences of Chinese Rubbings, it is difficult to draw a clear line between WORK and EXPRESSION entities by virtue of the characteristics of rubbings. These example elements are type (類型), transcription (釋文), script (書體), literary style (文體), direction (文向), decoration (圖像), shape (造形), ink (墨色), line and character (行格), relief (刻法).
- However, this analytical consequence is not as identical as the case of the National Palace Museum (NPM) which contains lots of refine artistic objects. In the light of these artifacts at the National Palace Museum Project, we find that some object characteristics of EXPRESSION can be included at WORK entity, and some may be located into MANIFESTATION entity. These elements include pattern (款識), material (質材), color of enamel (釉色), quality of embryo (胎質), feature (特徵), luster of color (色澤), style of skill (技法), inlay (鑲嵌), shape (形制), and exhibition explanation (展覽說明). The question of whether the IFLA's FRBR model is appropriate for metadata analysis, or need to be abridged into lesser levels, still needs further case studies to verify.

A collaborative model

- The SMART team tends to build up a collaborative model for the diverse communities who involve the metadata task. In the Chinese Rubbings case, the collaborative model composed of rubbings stakeholders, metadata mangagers, and system designers. By means of metadata workflow, various voices will achieve to a common agreement.

Suggestions and conclusion

In the light of the Chinese Rubbings case study, we find that the IFLA's FRBR model is useful for metadata analysis, though some verification is needed to further clarify. It is also critical to develop an essential set of workflow to analyze metadata in a consistent way. Furthermore, the approach of CORE-INDIVIDUAL record structure

is initially proven as an effective means of metadata design and implementation. The Chinese Rubbings is only one part of the SMART Projects of the Digital Archive Project at Academia Sinica. There are more 10 projects waiting for metadata implementation. To cope with the emerging requirements of metadata for digital library projects, two critical issues should be addressed. These are "project management" for conducting metadata projects efficiently and "a generic metadata framework" for sharing and building up a unified system for all projects at next phase.

Acknowledgements

We would like to thank Malee Huang and Ming-shin Lin for their contribution of the whole metadata process, and Chi-wen Fann and Tzu-fan Chung for their help in developing the metadata system prototype, as well as thank Cheng-shang Wu for his assistance in providing and confirming, and I-ting Chang for her cooperative analysis of metadata elements for this case.

Reference

Bearman, D., Miller, E., Rust, G., Trant, J., and Weibel, S. (1999). A common model to support interoperable metadata: Progress report on reconciling metadata requirements from the Dublin Core and INDECS/DOI Communities. *D-Lib Magazine,* 5(1). (Online, Access Date: 18 Jan., 1999)
Available：http://www.dlib.org/dlib/january99/bearman/01bearman.html

Burnett, K., Ng, D.B., and Park, S. (1999). A comparison of the two traditions of metadata development. *Journal of American Society for Information Science*, 50(13), 1209-1217.

IFLA Study Group on the Functional Requirements for Bibliographic Records. (1998). *Functional requirements for bibliographic records: Final report.* München, Saur: 1998. (Online, Access Date: 24 Dec., 1998)
Available：http://www.ifla.org/VII/S13/frbr/frbf.pdf