

Designing and Implementation of Chinese Metadata Standards: A Case Study on Metadata Applications in Peking University Rare Book Digital Library

Xiao Long

Peking University Library
Beijing, P. R. China 100871
E-mail: lxiao@lib.pku.edu.cn

Chen Ling

CALIS Administrative Center
Peking University
Beijing, P. R. China 100871
E-mail: chenl@calis.edu.cn

Abstract: Based on the designing and implementation of metadata standard of Peking University Digital Library, this paper discusses the Chinese metadata standard and its related issues. It involves the metadata methodology – Metadata Standard Framework, and an analysis of an applied case – the generation of Rubbing Metadata Standard of Peking University Rare Book Digital Library.

1, Introduction

With the development of digital library, metadata is becoming a practical method from a theoretical research. Metadata standard, as a descriptive technique of digital resources, its design and definition establish the foundation of the construction of the digital library. At present, even though some metadata projects and standards have been generated, there hasn't produced any Chinese metadata standard based on Chinese resources.

Peking University Rare Book Digital Library (RBDL) is one of the most important components of Peking University Digital Library. It is built based on the large special collection of Peking University Library, total 180,000 titles and 1,600,000 items of Chinese rare books, rubbings, ancient atlases, Dunhuang Scrolls, and old journals published before 1949. For the better organization and use of these digital special resources, a series of Chinese metadata standards aiming at the characteristics of different objects are being developed.

In this paper, it will discuss the Chinese metadata standard and its related issues. Firstly, there is an introduction about the metadata methodology – Metadata Standard

Framework of Peking University Digital Library. Secondly, there is an analysis of an applied case – designing and implementation of Rubbing Metadata Standard of Peking University Rare Book Digital Library.

2, Methodology: Metadata Standard Framework of Peking University Digital Library

As we understand, metadata issue involves three aspects of concepts:

Metadata: It is the data to be used in describing the attributes of a certain object, in locating, administrating, and assisting retrieving the resources.

Metadata standard: A set of rules that are used in describing a certain type of object (resource), including: content structure and elements, semantic rules, syntax structure and rules, meta-language and DTD.

Metadata standard framework: Standards or rules that are followed in designing a metadata standard for a specific type of object (resource). In reality, it is nonobjective metadata or methodology, which defines the concepts, functions, structure, format, workflow standard, syntax structure and rules, and semantic rules of metadata standard at a higher level.

At present, Peking University Digital Library has completed the preliminary scheme of the General Framework of Metadata Standard, which is currently guiding the designs of metadata standards of Peking University Rare Book Digital Library.

As a methodology, the Metadata Standard Framework of Peking University Digital Library comprises of:

- Functions of metadata
- Fundamentals for designing metadata standard
- Workflow for designing metadata standard: step by step
- Content structure and the elements
- Semantic definition rules and related authorities
- Meta-language and Syntax Structure

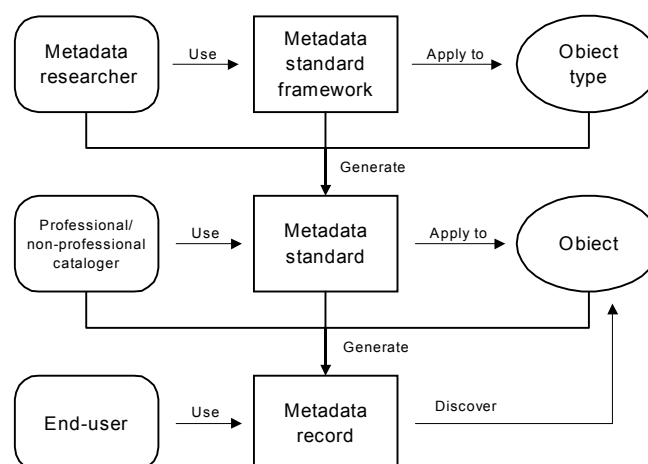


Fig. 1: Chinese metadata standard framework

Functions of metadata

In organizing and using digital resources, metadata plays the roles as following:

Description: The basic function of metadata, which describes the object's contents, attributes, location, etc.

Location: The information about location of the resources, like DOI, URL, URN, ISSN, and the others of the objects. It is very helpful for the discovery and retrieval of the resources.

Administration: Information about the rights management (copyright, ownership, accessing right, electronic watermark, and signature, etc.) and management of digital object, etc.

Retrieval: Indexing and organizing the information about the objects, establishing their relations, providing a retrieval system with multiple levels and entries.

Evaluation and selection: For users to know more about the object and to make their decision of selecting and downloading the information.

Interaction: Interactions between the system and metadata users. To a certain extent, the structure of metadata opens to both of the catalogers and the end-users of the digital library. In this way, the system can allow catalogers to create some elements and their relations, to set the fields for searching information, to construct authorities. Also, by creating a communication channel, the system can automatically absorb feedback or comments to metadata records or the contents of resource from the end-users.

Fundamentals for designing metadata standard

The fundamentals for designing metadata should be based upon the three kinds of requirements of: professional/non-professional catalogers, the characteristics of resources/objects, and the end-users of digital library. They are as following:

Simplicity and accuracy: Since most of the catalogers who do the descriptions are the experts, specialists, and the students in certain subject domains, the metadata standard should be simple enough for these non-professional catalogers to learn and to use. On the other hand, metadata should also be able to describe the object accurately and keep the retrieval effectively, avoiding the inaccuracy along with the simplicity. However, it is not easy to balance between the simplicity and accuracy.

Specialization and generalization: For the reason of differences of multiple objects and their characteristics, it is impossible to describe them by only one metadata standard. Varying metadata standards should be designed. At the same time, considering the commonness of the different objects, the generalization of metadata standard is also required. In other words, one metadata standard should apply to some different objects as many as possible.

Interoperability and interchangeability: Interoperability of metadata represents the supports for the interoperation between heterogeneous systems. In other words, the metadata based on the current Chinese metadata standards should not only be operated by each application software of Peking University Digital Library, but also

be implemented in the systems of other organization or institutions. In the practical applications, interoperability acts as interchangeability. It means that the current metadata can be converted easily as the common metadata in other operating systems, with the prerequisite of the least loss of information that the metadata contains. Therefore, both semantic definition of the elements and metadata structure ought to be considered carefully, especially its coherence with the semantic rules of some existing generalized metadata standards, such as Dublin Core.

Extensibility: Since the wide diversity of the digital resources and their applications in different fields, metadata standard can merely provide a general meaning description to digital objects. Some special contents aiming at the specific attributes of the objects are not included in the metadata standard. However, considering the further and more accurate descriptions in a certain practical application, metadata standard should be allowed to extend some elements or attribute values. This extension is generated under the instruction of the metadata standard.

User requirement: Designing the metadata standard is for the purpose of better discovery of the resources to users. Therefore, user requirements should be the final standard for framing the structure, selecting the elements, and making syntax and semantic rules. Additionally, it is also important to provide interaction channel between system and the end-users, for absorbing their comments or feedback to metadata record and the objects.

Content structure and the elements

A metadata standard of Peking University Digital Library comprises of three parts: descriptive metadata, administrative metadata, GIS metadata (especially applied for the Peking University Rare Book Digital Library). The Chinese Metadata Standard Framework defines the following generalized elements, which should be given priority in adoption while designing a certain metadata standard and especially be in accordance with their semantic rules.

Descriptive metadata: Using for the description of external and internal attributes of the objects, such as: title, creator, date, subject, language, etc. It includes:

- Core elements: a generalized component for all kinds of the objects. Currently, for the documents and document-like objects, it is designed in accordance with the Dublin Core, but just adopting the element names and definitions.
- Local core elements: a common part for the local collections of Peking University Digital Library.

- Unique elements: designing for a specific type of object based on its characteristics.

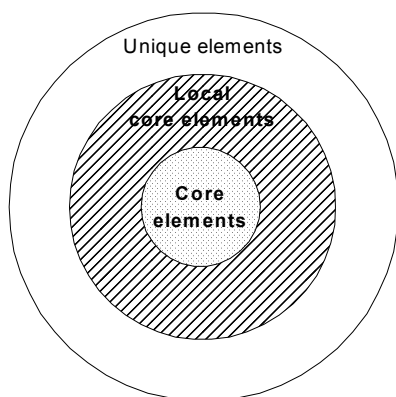


Fig.2-1: Descriptive Metadata

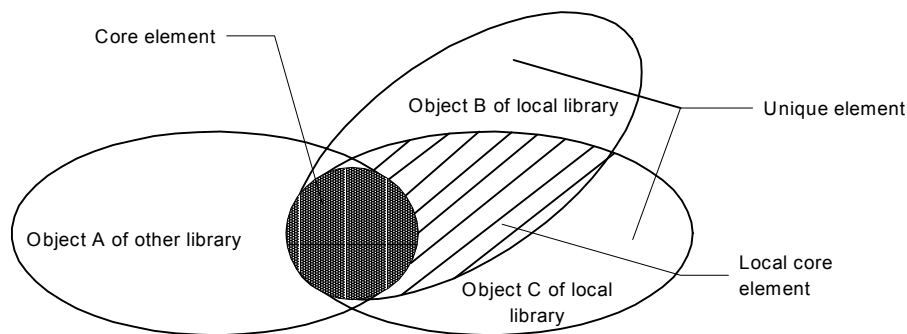


Fig.2-2: Structural relations between the metadata standards based on the different objects

Administrative metadata: For management information of the object. It consists of three elements:

- Object creation statement: digitizing method, condition, date, and digitizer, digital copyright, etc.
- Instance: universal digital resource identifier, like DOI, URI, URN or URL; image size, image format, resolution, modified date, other statements, accessing right.
- User remark/comments: for the feedback or comments of end-users.

GIS metadata: For the application of GIS. It matches the elements about the temporal and spatial information in descriptive metadata, and is operated by the data processing technicians. There are two elements in it: Spatial (coordinates) and Temporal.

Semantic definition rules and related authorities

There should be the corresponding semantic rules for the definition of the elements while designing the metadata standard, which involves:

Description rules: Different digital resources can develop their own specific description rules, with the prerequisite of following the element definitions in the Metadata Standard Framework. Description rules include: the contents that the element describes; the methods for handling the special cases; description examples; etc. Especially, the objective description is emphasized.

Element statement: It states if the element is repeatable and required, and if it has the sub-elements, qualifiers, and authority files.

Authority files: Including: name authorities (for personal and place names); chronology; subject authorities; classification/taxonomy. Name authorities and chronology will be used in the forms of mapping lists, which emphasizes providing accurate and entire results to end-users through the interrelated retrieval, rather than only controlling the description traditionally.

Retrieval statement: Standards for duplicate checking; arrangement principles; index extractive rules; Chinese Pinyin retrieval; retrieval entries; related technologies; etc.

Meta-language and Syntax Structure

For the purpose of interoperability and interchangeability of metadata, the Chinese Metadata Standard Framework stipulates the meta-language and related syntax applied in the implementation of metadata.

XML is the meta-language for describing the metadata. The method for metadata syntax definition is XML Schema or XML DTD.

Like Dublin Core, each element is defined using a set of ten attributes from the ISO/IEC 11179 [ISO11179] standard for the description of data elements, including:

Name: The label assigned to the data element.

Identifier: The unique identifier assigned to the data element.

Version: The version of the data element.

Registration Authority: The entity authorized to register the data element.

Language: The language in which the data element is specified.

Definition: A statement that clearly represents the concept and essential nature of the data element.

Obligation: Indicates if the data element is required to always or sometimes be presented (contain a value).

Datatype: Indicates the type of data that can be represented in the value of the data element.

Maximum Occurrence: Indicates any limit to the repeatability of the data element.

Comment: A remark concerning the application of the data element.

3, Case and Analysis: Rubbing Metadata Standard of Peking University Rare Book Digital Library

Peking University Library has the rubbing collection of 60,000 items. Digital Rubbing Collection will be the initial project of Peking University Rare Book Digital Library (RBDL). The rubbing metadata standard is designed under the instructions of the Metadata Standard Framework of Peking University Digital Library and defined as RBDL Rubbing Metadata Standard Version 1.0.

Workflow: step by step

Researches on existing metadata standards: Ensure if there is any existing metadata standard to be applied directly. Or, at least, some positive characteristics or ideas of these metadata standards could be absorbed. So far total eight metadata standards are studied and compared, like CDWA¹, VRA Core², GILS³, Dublin Core⁴, FGDC⁵, MOA II⁶, as well as the combinations of the metadata and content data, such as EAD⁷ and TEI⁸.

Analysis of the objects and their relations with rubbing experts: Three main aspects were analyzed: the relations among the objects; minimal description unit; description elements. For example, a serial rubbing, is the description unit for the

whole series or just sub-rubbing? For obverse side and reverse side of a certain rubbing, should they be described in a whole record or two separate records?

Investigation to resource users for the purpose of understanding their requirements: An investigation was conducted among the professors, researchers, graduates of the Department of History, the Department of Chinese Language and the Department of Archaeology. They provided a great deal of information about their requirements, especially about the contents that they are most interested and often been described, and the entries that they search most frequently. This is very beneficial for designing the metadata standard.

Preliminary scheme of the metadata standard was accomplished: Including its structure, format, elements and their definitions, related rules and authorities, etc.

Test description by professional/non-professional catalogers and **correction** of metadata standard.

Software development.

Test description in the system and **re-correction** of the metadata standard.

Final design of the metadata standard was completed.

Key points

Figuring out the relation of original object (tombstone or others), rubbing and digital rubbing (digital image): Rubbing is the reproduction of the original object. Digital rubbing is the reproduction of the rubbing. Three of them overlap all the internal attributes of the original object based on the contents. Therefore, the metadata records of above three can't be independent and self-exist, but rather should be integrated with different identifiers as a whole record.

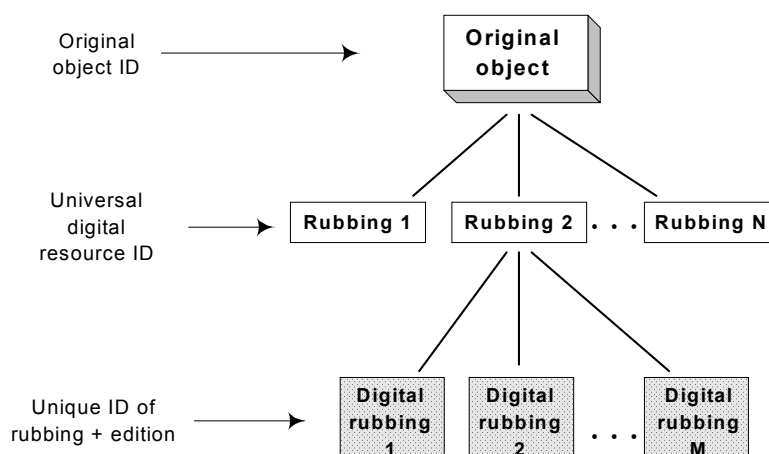


Fig.3: Relations between the original object, rubbings and digital rubbings

Making sure the description unit and relations among the rubbings: it is based on the analysis of:

- Relations between the original objects, different editions and different copies of rubbings (see fig. 4)
- Relations between the different rubbings (see fig. 5)

At last, the minimal description unit is fixed on the each copy of rubbing,

identified by the call no.

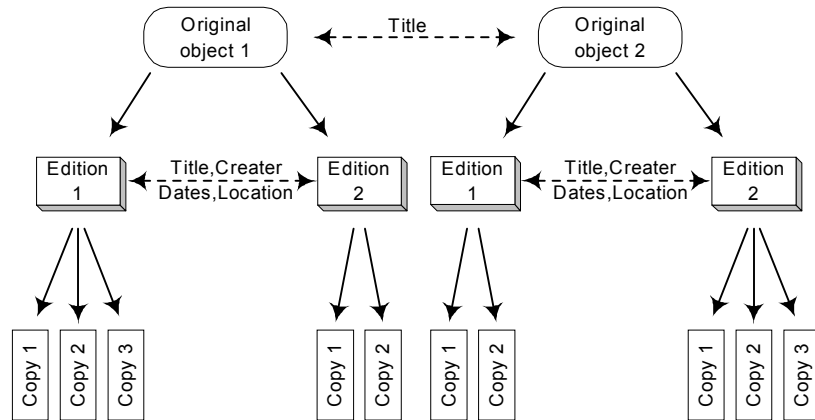


Fig.4: Relations between the original object, different editions and different copies of rubbings

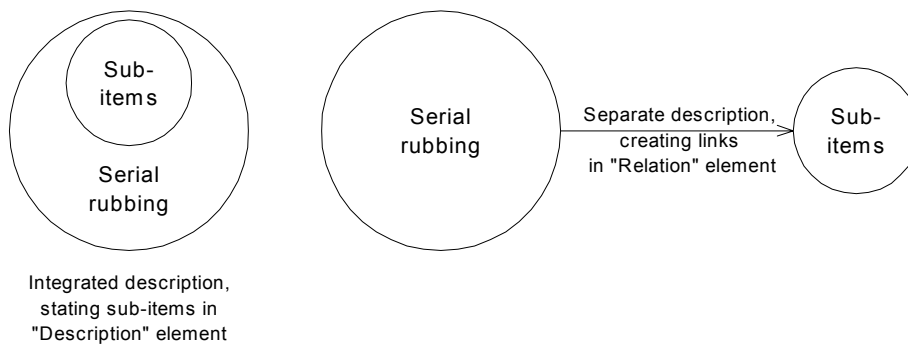


Fig.5-1: Contained relation between different rubbings: serial rubbing and sub-items

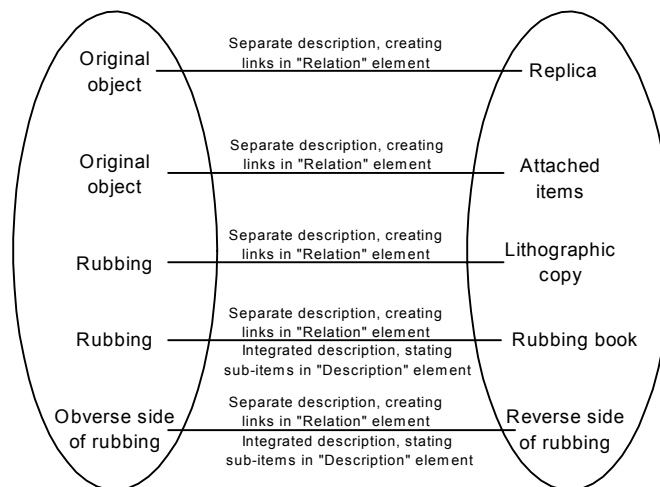


Fig.5-2: Parallel relations between different rubbings

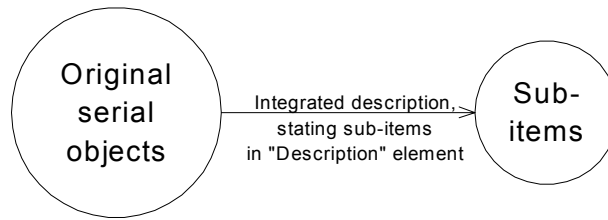


Fig.5-3: Attached relations between different rubbings: original serial objects and sub-items

Working out unique elements for rubbings: Based on the investigation among the end users and analysis of the characteristics of rubbings, five unique elements are worked out: Location, Collecting History, Materials and Techniques, Handwritings, and Original Object Identifier.

Content structure

In accordance with the Metadata Standard Framework of Peking University Digital Library, RBDL Rubbing Metadata Standard Version 1.0 is composed of descriptive metadata, administrative metadata and GIS metadata (see Fig. 6).

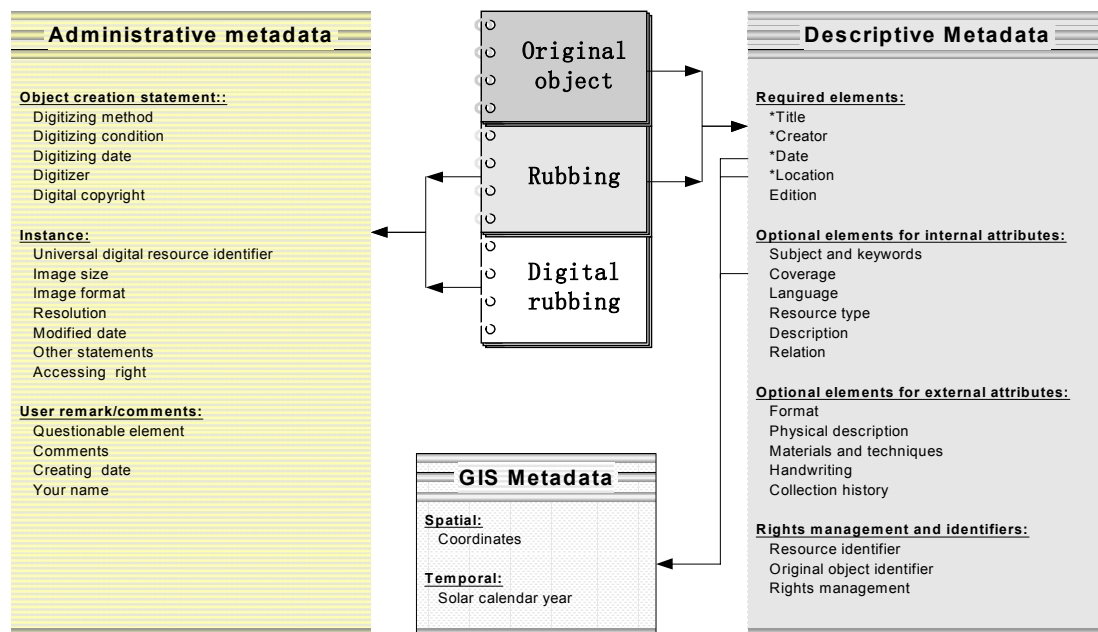


Fig.6: Content structure of rubbing metadata standard

Therein, the elements of administrative metadata and GIS metadata are the same as the Metadata Standard Framework. Particularly, the descriptive metadata has some differences and develops the Framework as following:

Core elements: It is generated based on Dublin Core with 12 elements: Title, Creator, Subject and Keywords, Description, Date, Resource Type (type of original object), Format (rubbing), Resource Identifier, Language, Relation, Coverage, Rights Management. There are 3 elements of DC to be deleted: Publisher, Source and Contributor.

Local core elements: Edition, Physical Description.

Unique elements: Location (place of origin), Collecting History, Materials and Techniques (materials of original object), Handwritings.

Related Technologies with Metadata

There are a few technologies related to metadata and to have to be considered while designing the rubbing metadata standard.

Applications of GIS: GIS metadata is designed for it.

A mapping list of Chronology: Applied for the conversion of Chinese Lunar Calendar and the Solar Calendar in Date and Coverage elements.

A mapping list of Chinese ancient and current place names: Applied for the place name authority file in Location and Coverage elements.

Networked knowledge organization systems/schemes/services (NKOS): An open thesaurus/subject/classification/taxonomy system will be worked out and applied in Subject element.

4, Conclusion

A metadata standard framework should be stipulated firstly at the initiative phase of a digital library, so that a specific metadata standard based on a specific object can be developed under the instruction of a certain standard. In this way, it is easy for a digital library to keep the integration and standardization of its varying metadata standards.

Varying metadata standards aiming at the different objects and their characteristics should be generated and suit a great diversity of needs. It's impossible to describe multiple kinds of objects by a uniform metadata standard.

Different digital libraries develop different application systems based on their own metadata standards. For the purpose of supporting the interoperability between heterogeneous systems, XML/RDF will also be applied as a technical standard at data level.

The development of digital library in China brings more and more researches on Chinese metadata standard. It is the hope of the authors that this paper can introduce some experiences and issues about the metadata of Peking University Digital Library, and will be helpful to foster the researches and developments of Chinese metadata standard.

Notes:

¹ Categories for the Description of Works of Art,

<http://www.getty.edu/research/institute/standards/cdwa/index.html>

² VRA Core Categories, <http://www.gsd.harvard.edu/~staffaw3/vra/coreinfopage.htm>

³ Government Information Locator Service,

⁴ Dublin Core Metadata Initiative, <http://dublincore.org/>

-
- ⁵ Content Standard for Digital Geospatial Metadata,
<http://fgdc.er.usgs.gov/metadata/metadata.html>
- ⁶ The Making of America II, <http://sunsite.berkeley.edu/MOA2/index.html>
- ⁷ Encoded Archival Description, <http://lcweb.loc.gov/ead/>
- ⁸ Electronic Text Encoding and Interchange,
<http://www.tei-c.org/Guidelines2/index.html>

References:

- 1, W3C homepage, URL: <http://www.w3.org>
- 2, Dublin Core Metadata Initiative, URL: <http://dublincore.org/>
- 3, Chen, Chao-chen, The Design of XML/Metadata Management System: an Brief Introduction to the Framework and Function of Metadata. In: Documentation and Information Center of Academy of Science (Mainland China), Institute of Library and Information Science Education (Taiwan). *Proceedings of the Conference of Fifth Library and Information Science: August 28-30, 2000, Cheng Du*. Beijing: Documentation and Information Center of Academy of Science, 2000. p. 209-231
- 4, Hsieh, Ching-chun, et al. Compilation of Technical Proposals of Digital Museum Project. Taipei: National Science Council, 1999.
- 5, Zhang, Xiaolin. Standard Framework for the Development and Application of Metadata. In: Documentation and Information Center of Academy of Science (Mainland China), Institute of Library and Information Science Education (Taiwan). *Proceedings of the Conference of Fifth Library and Information Science : August 28-30, 2000, Cheng Du*. Beijing: Documentation and Information Center of Academy of Science, 2000. p. 69-74.
- 6, Liu, Jia. Researches on Metadata: dissertation. 2000.
- 7, Wu, Jianzhong, et al. Dublin Core Metadata. Shanghai: Shanghai Science and Technology Press, 2000.
- 8, Zeng, Marcia Lei and Xiao, Long. Mapping Metadata Elements of Different Formats. For: *National Online Meeting, 2001*