Building Metadata Application Framework for Chinese Digital Libraries

Yunyun Shen
Peking University Library, China
yyshen@lib.pku.edu.cn

Long Xiao
Peking University Library, China
lxiao@lib.pku.edu.cn

Ying Feng
Administrative Center of CALIS, China
fengy@calis.edu.cn

Abstract

This project report describes the *Metadata Application Framework* (MAF) of the National Library of China (NLC)—one of the six sub-projects comprising the *General Rules of National Digital Library Metadata*. The MAF aims at enabling the applications of Dublin Core in the National Digital Library of China by developing a series of related metadata standards, criteria and a platform to meet the requirements of describing, organizing, managing, serving and preserving Chinese digital objects. It functions to support producing, processing, organizing, releasing, preserving and managing information resources in the digital library system of NLC in a manner that achieves interoperability and data sharing with other digital library systems. The report introduces the whole project, but explains in detail the metadata application framework and principles of NLC based on the work of DCMI and the other international metadata initiatives. Through this project, we are trying to define metadata best practices for developing digital libraries in China.

Keywords: metadata application; general metadata rules; National Library of China (NLC)

1. Background

Whether it is in traditional libraries or in the movement and development of the digital libraries, metadata has been applied widely. Metadata standards—as basic standards in the development of digital libraries—impact each phase of resource description, organization, administration, preservation and service. Thus, research on metadata and its applications are in-depth with hundreds of metadata standards and application projects appearing all over the world. Following metadata developments, Chinese metadata experts have developed a variety of metadata standards, as well as related application projects, for describing and managing the Chinese special resources. In these numerous applications, the use of DC is remarkably prevalent.

NLC is developing the National Digital Library of China (NDCL) with the support of a series of standard systems, each of which can be divided into many projects that can be taken up by different libraries. One of the most important projects is named *General Rules of National Digital Library Metadata* (GR). The GR is building a metadata application framework for the NLC. The NLC expects the achievements of this project to be applied not only in the NDCL, but also in digital libraries throughout China.

The development of the GR was entrusted to the Peking University Library (PUL). The series of metadata standards and related application guides developed by PUL have been applied in Chinese academic libraries.

2. Introduction to the Project

The MAF project mainly explores the application of DC in the Chinese digital environment and includes the following six sub-projects:

1. ***Metadata Application Framework (MAF) for NLC.*** This sub-project is the subject matter of this project report. The MAF involves development of the metadata application principles and framework for the CDL. The MAF work is based on research and

investigation of the main metadata applications in the world and the analysis of the digital resources in NLC. This sub-project is the most creative and difficult for the PUL project team. See Part 3 of this report for the description of the primary contents of the MAF.

2. ***Core Metadata Element Set of NLC.*** The Dublin Core Metadata Element Set (DCMES) has been adapted to describe all document-like objects located around the world. It is advisable that core elements are designed in accordance with DCMES. However, DCMES could not be adopted directly or completely because of language differences, conventional usage, and particular applications in the Chinese digital environment. Therefore, DCMES is localized by Chinese language and some local application rules, forming the NLC core metadata element set.

   The details of localized rules of equivalent adoption of the DCMES element set are:

   a. The names of elements maintain coherence with DCMES.
   b. The labels can be Chinese and can be changed on the basis of the characteristics of resource objects.
   c. The NLC core elements are not mandatory.
   d. In the adoption and localization of DCMES through the NLC core elements, it is allowable to refine the semantic definition of DCMES elements to accommodate local needs and the characteristics of resources, but those refinements cannot contradict the DCMES definitions.

   The NLC core element set comprises the minimal Chinese element set necessary to support resource retrieval and integration with other applications. Meanwhile, the descriptive rules and the XML schema of the NLC core element are set up.

3. ***Guidelines for Designing Metadata Standards of Special Resources in NLC.*** Designing metadata schema for different classes of resource objects is always based on the characteristics of those classes in order to achieve the necessary specialization and accuracy required. Design guidelines named the *Guidelines for Designing Metadata Standards of Special Resources* have been created to achieve coherence and interoperability in function, data structure, format, semantics and syntax among the different metadata standards developed. Using the *Guidelines*, metadata standards will be designed for special resources, such as rare books, ancient atlases, Chinese pedigrees, rubbings, Chinese e-books, journal articles, dissertations and thesis, audio and video materials.
4. ***Developing CNMARC XML***. For conversion and sharing of CNMARC, an XML exchange format will be developed using MARC XML and MarcXchange (ISO/DIS 25577) as references. This sub-project requires provision of a format definition and notes of CNMARC XML, the file of CNMARCXML.xsd, the exchange program, and example data between XML and ISO2709 of CNMARC.
5. ***Mapping and conversion between CNMARC, DCMES and NLC Core Metadata Element Set.***
6. ***Mapping and conversion between MARC21, DCMES and NLC Core Metadata Element Set.*** Since the MARC format has been applied in traditional and some digital resource cataloging, it is essential to build the mapping between MARC, DCMES, and the NLC core metadata element set, as well as metadata element sets for different resource classes in a manner that supports data exchange and interoperability among systems in and outside NLC. The conversion program has been developed to check the accuracy of the data exchange results.

The problems involving the application of the metadata general framework and design principles, core element set, metadata structures (content, semantic and syntax), open mechanism, namespace, application workflows, and mapping are discussed in Part 3.

3. Metadata Application Framework and Principles for NLC

In support of the lifecycle of digital resources, different kinds of metadata such as descriptive metadata, administrative metadata, preservation metadata and service metadata are necessary to support sufficient description and management. Thus, the metadata application framework and related principles in NDCL includes:

- Metadata, metadata standards, and their definitions;
- The relations and functions of different types of metadata and metadata standards;
- The metadata application framework: principles, models, core element set, structures, open mechanism and interoperability, etc;
- The metadata application workflow;
- Other related issues.

The following sections describe key concerns of the Metadata Application Framework and Principles for NLC.

3.1. Types and Relations of Different Metadata

Figure 1 defines the metadata systems and types in relation to the different phrase of the lifecycle of resources in the digital library.

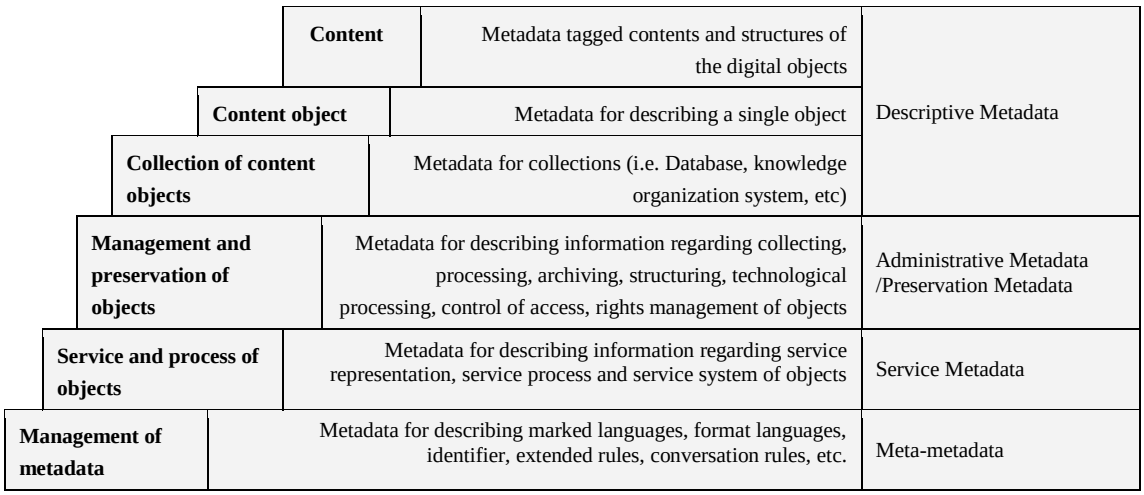

FIG. 1. Metadata system and types[1]

As theories of the lifecycle of information resources demonstrate, all digital objects go through the processes of creation, collection, description, use, preservation and removal. Figure 2 illustrates metadata and metadata standards relationships in the lifecycle along object and lifecycle dimensions.

Each metadata standard supports different phases of the lifecycle, interworking with the different functions but never isolated one from the other. For example, preservation metadata retains information from descriptive metadata, administrative metadata and service metadata. Meanwhile, because different types of metadata deal with objects on different levels (see axis Y in Figure 2), designing metadata standards needs to consider the requirements of the objects on different levels. Administrative metadata permeates the entire lifecycle and will be designed on two levels—objects and collections.

[1] See Xiao, Long and Zhao Liang, 2007, p11

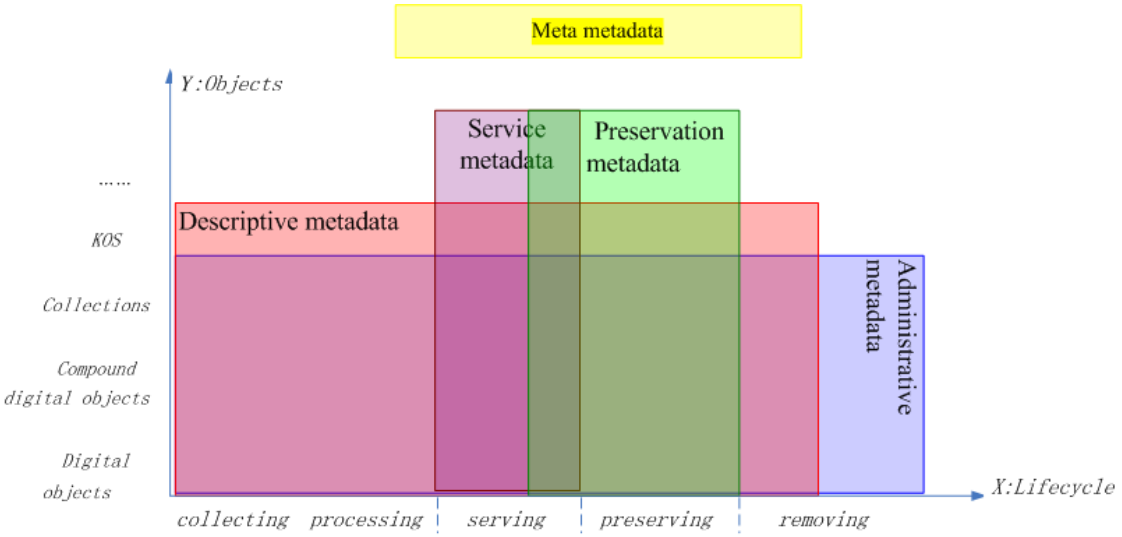

FIG. 2. Relations among metadata standards in the lifecycle of information resources

It is essential to make the relations among metadata standards clear for building the metadata application framework.

3.2. Metadata application framework

Firstly, the fundamentals of designing and applying metadata standards in the digital library are prescribed as open, modularized, and interoperable. Secondly, the model of the metadata application framework is set up referring to the Singapore Framework for Dublin Core Application Profiles, to standardize metadata structures and forms (see Figure. 3).

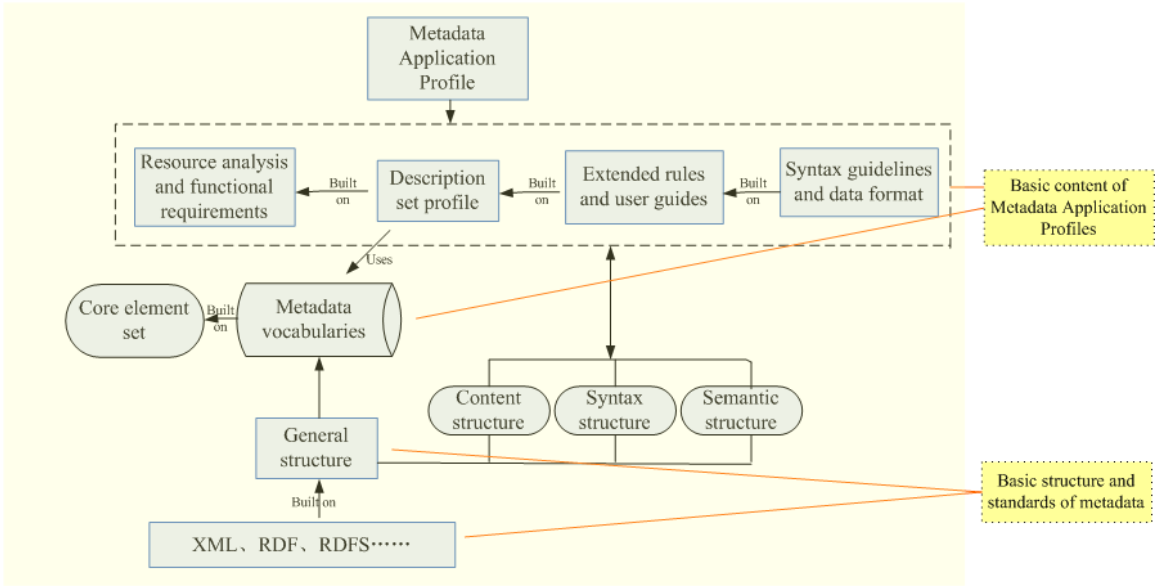

FIG. 3. Model of Metadata Application Framework (adapted from http://dublincore.org/documents/singapore-framework/)

The design's main tasks include:

1. Indentifying the NLC core metadata element set in accordance with Dublin Core (see Part 2(2));
2. Defining the metadata structures—content structure, syntax structure and semantic structure—including the extended rules. This is very important to support another sub-project, *Guidelines of Designing Metadata Standards for Special Resources in NLC*.
    a. Content structure consists of core elements, elements sharing applying to similar resources or special modules, and local elements;
    b. Syntax structure makes use of XML and RDF;
    c. Semantic structure uses the 15 attributes to define the elements and qualifiers, with references of ISO 11179 and the attribute definition of DC terms: *Identifier, Name, Defined By, Label, Definition, Comments, Type of Term, Refines, Refined By, Encoding Scheme For, Encoding Scheme, Data Type, Version, Language, Occurrence*;
    d. Extended rules are built for describing the detailed and more precise information about a certain object in some practical applications. The rules of extension include both horizontal (for element extension) and vertical (for qualifier extension) ones, complying with DC qualifiers and the metadata content structure.
3. Setting up the user application guides to specify the use of elements, qualifiers (element refinements), and the use of related authorities.
4. Building the open metadata and interoperability mechanisms, including identifying the metadata namespace, setting up the metadata registration system, the rules of selecting the elements from the other metadata namespaces, and mapping among the metadata standards in conformance with DCMI policies.

3.3. Metadata Application Workflows

Figure 4 illustrates the project metadata application workflows used in order to standardize the metadata application process:

1. Analyzing the metadata application environment and purpose, the resources objects, the users, the functional requirements, and the responsibility model;
2. Investigating, then selecting/designing metadata standards;
3. Designing the metadata application profile;
4. Creating the metadata records and control quality; and
5. Using the metadata records in digital libraries.

4. Further Information

Even though the project described is currently led by PUL, the achievements of the project are based on previous research results from the *Chinese Digital Library Standards* (CDLS) national research project in which PUL, NLC, ShangHai Library, Fudan University Library, and Tsinghua University Library joined in the effort. Many Chinese metadata experts joined such as Xiaolin Zhang, Long Xiao, Leon Zhao, and Qiaoying Zheng. In *CDLS*, the application of DC has been put forward and implemented in some Chinese libraries. In the work reported here, the work applying DC is improved further for NLC.

The project started in July 2008 and is basically complete. NLC will arrange for experts to examine and validate the project. It is sure that NLC will implement the rules to build the metadata repositories and structure the metadata retrieval mechanism in its digital library system

in the coming future. If the rules are applied successfully, NLC will push their application in more extensive manner—it is certain that the rules will spread gradually to all Chinese libraries.

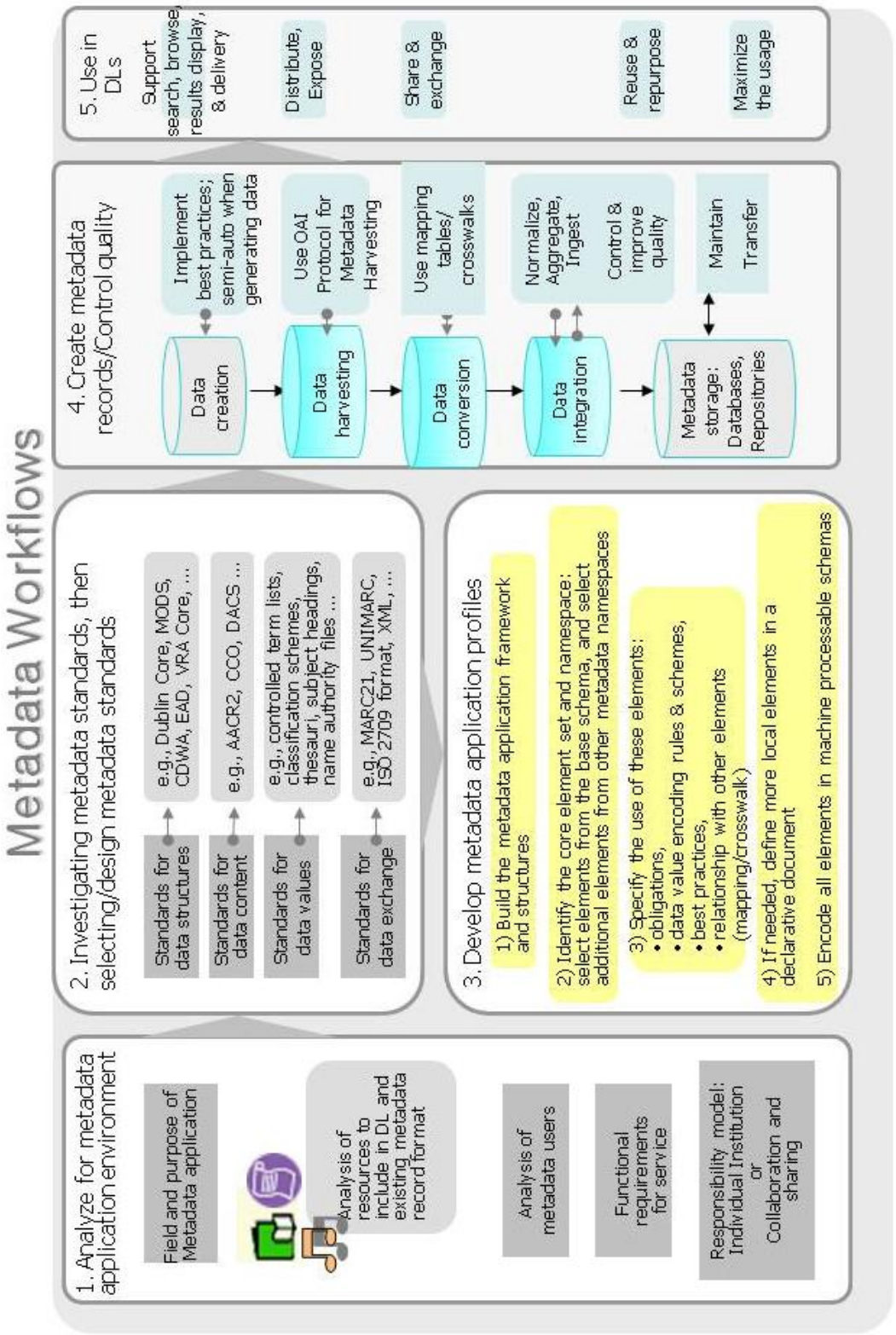

FIG. 4. Metadata Application Workflows (adapted from Zeng, Lee & Hayes (2009)).

References

Xiao, Long, et al. A Guide to Design Metadata Standards for Special Objects. Prepared for the Project "Chinese Digital Library Standards", 2003. (Text in Chinese). 肖珑等. 专门元数据规范设计指南. "中国数字图书馆标准规范建设"项目, 2003.

Zhang, Xiaolin. Digital Library Standards Development, PPT. 张晓林. 数字图书馆标准与规范建设.

Zhang, Xiaolin. Metadata Research and Application. Beijing: Beijing Library Press, 2002. (Text in Chinese). 张晓林. 元数据研究与应用. 北京：北京图书馆出版社.

Xiao, Long and Zhao Liang ed. An Introduction to Chinese Metadata: with Actual Examples. Beijing: Beijing Library Press, 2007. (Text in Chinese). 肖珑，赵亮主编. 中文元数据概论与实例. 北京：北京图书馆出版社, 2007.

Zhang, Xiaolin, et al. Standards Related to the Development of Digital Libraries. China Academic Journal, 2002(6): 7-16. (Text in Chinese). 张晓林等. 数字图书馆建设的标准与规范. 中国图书馆学报, 2002(6): 7-16.

Zhang, Xiaolin. Driving digital libraries forward by metadata development. DCMI 2004, Shanghai, 2004.

Chan, Lois Mai and Zeng, Lei, Metadata Interoperability and Standardization - A Study of Methodology. D-Lib Magazine, v.12, no. 6 (2006 June).

Namespaces in XML 1.0 (Third Edition). Retrieved January 7, 2010, from http://www.w3.org/TR/xml-names/

DCMI. Namespace Policy for the Dublin Core Metadata Initiative. Retrieved January 7, 2010, from http://dublincore.org/documents/dcmi-namespace/index.shtml

DCMI. The Singapore Framework for Dublin Core Application Profiles. Retrieved January 7, 2010, from http://dublincore.org/documents/singapore-framework/

Zeng, Marcia Lei, Jaesun Lee, and Allene F. Hayes. 2009. Metadata Decisions in Digital Library projects – Summary of an international survey. Journal of Library Metadata, Vol. 9 No. 3 & 4: 173 - 193.