Chinese Metadata Review

Wei Liu
Director of Digital Library Research Institute
Shanghai Library
Shanghai, China 200031
wliu@libnet.sh.cn

Abstract
According to the literature and the statistics from digital library related projects, there has been a metadata movement in China since 2002. This paper reviews and highlights all the main efforts on the research and implementation of metadata standards, specifications and applications, by the institutions of national science and technology, education, and culture, and as well as the private sector.

Keywords:

The need for metadata arises from the ubiquity of digital resources and digital libraries. The term “metadata” has become a buzzword in China in recent years, much like “digital library”, which inherited too much meaning and became almost meaningless. China is a large country with a great imbalance in the distribution of its population and resources, and therefore also in the development of the economy. Most of the industries, as well as the universities and scientific research institutions are located around the coastal regions. This results in varying levels of librarianship and information literacy. The development of digital libraries and the research and implementation of metadata is usually based in Beijing - the capital city - and some other big cities that are located in the coastal areas.

The majority of funding for metadata projects comes from the central government, mostly from the national foundations for science and technology R&D, and the funds earmarked for construction of national information infrastructure. A small part is from the provincial institutions and affiliated libraries. One new development is that digital library research institutes have been established by some big universities, large public libraries and corporations to deal with the increasingly common requirements for the integration of digital resources and the digitization of traditional materials. The metadata research is no longer just a part of big digital library project or other R&D project to provide a method for resource organization. The stand alone project “the Building of Standards and Specifications for Digital Libraries in China” (see below) was approved by State Science and Technology Ministry in October 2002. This brings the construction of national standards and specifications to the state level.

---

1 This paper was published with revision as a chapter in International Yearbook of Library and Information Management edited by Dr. Gary Gorman. All rights reserved.

2 So far there have been several digital library research institutes established by Beijing University, Tsinghua University, Shanghai Jiaotong University, Shanghai Library and Shanghai Changjiang Computer Group.
Brief History
Research into metadata in China began in 1997. Apart from some domain-specific research and development on Geographic and Educational Information System, “the National Pilot Digital Library Project (NPDL)”\(^3\) was the milestone for the metadata research in China. It was the first “digital library” project in China, and was initiated by the Ministry of Culture to State Planning Commission. The research partners include National Library of China and five other large public libraries. Before the project, there were few reports on metadata research, giving only a brief introduction to projects or standards from abroad.

In April 1998, the project task force proposed a draft metadata profile, which defines a core (minimum) set of metadata elements and provides refinement rules and markup guidelines. It suggested that the minimum metadata element set should adopt Dublin Core Metadata Element Set, provided the HTML4.0 and XML/RDF markup samples, recommended a two-tier metadata application model in which the first tier is DCMES and the second is MARC, TEI Header or other rich metadata format. The first tier does not have to physically exist; It can be transformed dynamically by the mapping or bridge mechanism. The profile calls for the implementation of the two-tier model in the resources development phase of the pilot digital library project.

Although it was not taken up in its entirety by the project, the metadata profile had a big effect on the subsequent metadata research, application, implementation and promotion. Since 1998, there have been many research and development projects related to digital library initiated by many universities under the Ministry of Education (formerly the State Education Commission of China), and the science and technology research institutes, which are led by the Ministry of Science and Technology (formerly the State Science and Technology Commission of China), and the Chinese Academy of Science. Chinese Vice Premier Li Lanqing commented that “the Library operation model in the future is the digital library” when he was visited in National Library of China in October 1998. After that, many key programmes or projects were lauched, and the metadata research and application has become a hot topic in China.

Bibliographic Analysis
A literature search was conducted on published papers collected from CNKI and Chongqing VIP Database.

According to the bibliographic statistics of metadata related documents\(^4\), there are only three papers related to metadata research in computer science arena before 1998. The number becomes around 10 in 1999 and 2000 each (refers to table 1). The majority of the papers are about geographical information system, data warehouse and other computer software subjects. In 2002 there seems to be an explosion of 68 papers on metadata research, and most of the increases is from the research of digital library.

---

\(^3\) See: http://www.library.sh.cn/libnet/sztsg/fulltext/reports/1998/metadata.html

\(^4\) Based on the data collected from CNKI (http://www.cnki.net) and Chongqing VIP Database (http://202.119.47.6/), exhaustively but limited to the published papers, the result involved 43 professional periodicals in which at most issued 6 papers related to the topic in the year of 2002. But this does not imply by any means that the data is completed to this subject area, in which at least the proceedings and degree dissertations are not taken into account.
The increase in the number of papers related to metadata research in 2002 is extraordinary. Table 2 gives a content analysis (by sub-category) of the metadata-related papers of that year with up to three sub-categories given for each paper.

<table>
<thead>
<tr>
<th>Year</th>
<th>1997</th>
<th>1998</th>
<th>1999</th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
</tr>
</thead>
<tbody>
<tr>
<td>Num. of Entries</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>11</td>
<td>8</td>
<td>68</td>
</tr>
</tbody>
</table>

Table 1: Yearly number of metadata related research papers

<table>
<thead>
<tr>
<th>Subject</th>
<th>Entries</th>
</tr>
</thead>
<tbody>
<tr>
<td>The application of metadata</td>
<td>21</td>
</tr>
<tr>
<td>The general introduction of metadata</td>
<td>17</td>
</tr>
<tr>
<td>Computer science related (data warehouse etc.)</td>
<td>14</td>
</tr>
<tr>
<td>DC related</td>
<td>10</td>
</tr>
<tr>
<td>GIS related</td>
<td>7</td>
</tr>
<tr>
<td>Educational information system related</td>
<td>5</td>
</tr>
<tr>
<td>Metadata model</td>
<td>5</td>
</tr>
<tr>
<td>MARC and cataloguing</td>
<td>3</td>
</tr>
<tr>
<td>XML/RDF</td>
<td>3</td>
</tr>
<tr>
<td>Archiving</td>
<td>3</td>
</tr>
<tr>
<td>Metadata format</td>
<td>3</td>
</tr>
<tr>
<td>Preservation</td>
<td>2</td>
</tr>
<tr>
<td>Interoperability</td>
<td>2</td>
</tr>
<tr>
<td>Metadata capturing</td>
<td>2</td>
</tr>
<tr>
<td>Multimedia metadata</td>
<td>1</td>
</tr>
<tr>
<td>Administrative metadata</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2: Papers by subject categories

There also have been three monographs on metadata published in Chinese. They are: “DC METADATA” by Dr. Wu Jianzhong (the director of Shanghai Library), published in 2000 by Shanghai Scientific and Technological Literature Publishing House; “INTRODUCTION TO METADATA” by Liu Jia, published in 2001 by Beijing Huayi Publishing Company; and “THE RESEARCH AND APPLICATION OF METADATA” by Zhang Xiaolin (the director of the Library of Chinese Academy of Science), published in 2002 by National Library Publishing House.

There have been some practical achievements on metadata research in recent years, as well as the papers and books mentioned above. A number of metadata standards, specifications, guidelines, profiles and best practices which aim to diversify the use of resource description for a wide range of subjects have been issued by many kinds of institutions at different levels. Here below are some advanced schemes.

---

5 It allows at most 3 sub-subjects for one paper (but it’s rare) in accounting. And only the explicit topics would be implied to those subject categories. General topic is not considered to be a category. For instance, most of the papers refer to the Dublin Core Metadata, but they are not placed into the “DC related” category if they are not specifically about the DC metadata.
1. Metadata for Sustainable Development of China

Commissioned by the State Ministry of Science and Technology, the “Metadata for Sustainable Development of China”6 is worked out by National Fundamental Geographical Information Center (http://ngcc.sbsm.gov.cn/). Now the status of the standard is “draft for approval” and recommending for use. It is based on the ISO19115 “Geographical Information -- Metadata” (former the ISO15046-15) which is proposed by ISO/TC211 and refers to the Content Standard for Digital Geospatial Metadata v.2.0 (FGDC-STD-001-1998) approved by the Federal Geographic Data Committee (FGDC) of the United States in June 1998.

The research into metadata standards is one of the outcomes delivered by the pilot project “the Information Sharing on China's Sustainable Development” (97-925), which started in 1997. It is one of the National Key Projects for Science and Technology of the Ninth Five-year Plan of China, to be followed by “the System R&D of Sustainable Development of China” 8 in the Tenth Five-year Plan. The primary goal is to build a information sharing system in a distributing heterogeneous environment so as to share the database of natural resources, environment protection and technology, biological diversity, comprehensive natural disaster, all of which are closely related to sustainable development of China. The information covers plant and crop species, natural disaster, environment and environment protection, geology and mineral resources, climate and weather, ocean resources and etc. Each of these domains has its corresponding descriptive metadata set to be built, for example, the Forestry Metadata Specification which is mapped out by the Resources and Information Institute of China Academy of Forest is one of the second level metadata standards for specific domain in its framework9. The main task of the project is to set up a central GIS and database server array, and decentralized with domain servers to provide information sharing services.

Metadata specifications in the “Metadata for Sustainable Development of China” are divided by two levels. The first level contains the core metadata entities and elements which can be considered as the minimum set that fulfill the basic requirements of description. There are about seventy entities and elements in the first level which can describe the digital object or collections in general or collectively. The second level contains all the entities and elements needed to describe an entire dataset file. There are seven sections and more than 100 entities and elements in this level. Except the property of entity or element, every entities and elements are conditional or optional. The detailed content of the seven sections in the second level should be declared in the metadata dictionary. The specifications regulate the range of value for the entity or element and the qualified code such as the code for responsibility, code of refinement classification for dataset, the code of the status of dataset, the code for data space type, the code for reference frame, etc.

2. 《Learning Object Metadata Specification Information Model》

7 Please see the website of Chinese TC211 working group: http://nfgis.nsdi.gov.cn/isotc211/default.asp
8 Please refer to the website of “Sustainable Development of China” http://www.sdinfo.net.cn/.
“Learning Object Metadata (LOM) Specification” of Chinese E-Learning Technology Standard\textsuperscript{10} is initiated and established by Chinese E-Learning Technology Standardization Committee (CELTSC) of the Ministry of Education, which at the same time acts as the Subcommittee for Educational technology of the National Standard Committee for Information Technology (TC28). It is expected to be a National Standard but the current status is “CD (Committee Drafts)”. Founded in 2000, the CELTSC is the former Modern Educational Technology Standardization Committee of the Ministry of Education, which is in charge of the research, preparing and implementation of the Distance Learning Technology Standards (DLTS). The Project changed its name to Chinese E-Learning Technology Standardization (CE LTS) after the Committee has changed its name. Lead by Zhu Zhiting, the professor of East China Normal University, the Committee members are coming from Tsinghua University, Beijing University and other famous universities in China.

The CELTS includes 27 specifications, which are grouped in five categories: the General, the Resources Related, the Audience Related, the Educational Environmental Related and the Service Quality Related. It aims to establish the entire system of modern distance learning standards with Chinese characteristics by introducing, adopting and learning from the international related standards and specifications, including the Learning Resource Metadata of IMS (Instruction Management System), Learning Object Metadata of IEEE Learning Technology Standards Committee, Dublin Core Metadata Set of OCLC, etc. It will promote the sharing of educational resources, the interoperability of e-learning systems and guaranteeing the quality of distance learning services. The CELTS mainly follows the IEEE 1848. The Specification for Learning Object Metadata of CELTS is correspondent with IEEE1484.12.

The Specification for Learning Object Metadata can be used to describe any “learning object” (i.e. learning resource) normatively, provide technical assistant to the educator or learner on searching, evaluating, obtaining, and making full use of learning resources, guarantee the sharability, reusability and exchangeability of the resources and the system. The specification contains nine categories of 47 metadata elements (sub elements not included) for learning objects. The categories are: general, life cycle, meta metadata, technical, educational, rights, relation, annotation and classification. There can be at most nine properties of each element: name, name in English, definition, obligation, range, order, example, range or value, data type.

3. 《The Establishment of Standards and Specifications for Digital Library Construction in China (ESSDLC)》 project

In September 2002, the project “Establishment of Standard and Specifications for Digital Library Construction in China” was proposed to the Ministry of Science and Technology by the National Science and Technology Library, National Library, China Academic Library and Information System (CALIS), Library of Beijing University, Library of Chinese Academy of Science, Institute

\textsuperscript{10} See: \url{http://www.celtsc.edu.cn/DOCS/CD/CELTS-3.zip}
of Scientific and Technological Information of China, and Shanghai Library etc. This is a real joint project in which the members are come from all the leading organizations related to digital recourses. Most of the each partner has led one sub-project and there were six sub-projects have been started at the end of 2002, three more projects are to start very soon and there are still several sub-projects on the way to propose. The six started sub-projects are:

- the Overall Framework and Strategy of the Standards and Specifications for Digital Libraries in China;
- the Opening Development Mechanism for the establishment of Standards and Specifications for Digital Libraries in China;
- the Standards and Guidelines for Digitization
- the Application of Unique Universal Identifier for Digital Objects;
- the Metadata Standard for Cross Domain DO Description;
- the Metadata Standard for Domain Specific DO Description;

The three other sub-projects which are about to start in this year are: the Metadata Standard for Digital Collection Description, the Standard for Searching, Retrieval and Accessing of Digital Resources, and the Open Registration System for Digital Library Standards and Specifications.

The goal of the ESSDLC is to generate a series of national standards and specifications regarding to the construction of digital libraries in China. It is also a feasibility study to realize in what degree the necessity is, for each of these standards and specifications in the sub-projects.

Although the project is proposed to the Ministry of Science and Technology and led by the National Science and Technology Library, the head of the project is Dr. Zhang Xiaolin, who is the director of the Library of Chinese Academy of Science. The members of the project are coming from the research front of digital library and related area in China. So far it is widely participated and highly opened. Before that, Dr. Zhang led another project titled “Open Description of Information System in the Distributed Environment” approved by Chinese Academy of Science, which is an initiative study for the build of National Digital Library of Science. It gives a framework on the standard system of digital library, which is needed for resources description application by the construction and services of digital libraries in the distributed open environment. The project is a practical preparation of the ESSDLC project.

In the proposal statement it says: “As a digital resource and service providing system on the internet, the digital library has to establish and follow a set of standards and specifications about digitization processing, resource description and organization, system interoperating and services. It should adopt all kinds of existing related standards and specifications of content encoding, data communication, intellectual property protection, system security, management, service and operation, so as to assure the usability, interoperability and sustainability of the resources and services of digital library.”

The ongoing project has already issued several reports in some of professional journals.

---

11 See Zhang Xiaolin etc. the Trends on Digital Library Standardization, the Journal of Chinese Librarianship 2002.6
4. **《Chinese Metadata Standard》**

The National Library of China (NLC) has made a great effort on standardization in developing Digital Library since 1996. In most of the digital library related projects which sponsored or initiated by NLC, such as the early project: National Pilot Digital Library Project and the most significant and costly project by now: the Digital Library Project\(^\text{12}\), the research and compiling of metadata standard and specification has taken a very important role. But NLC realized that it will not achieve any success on a National level standard without considerable openness and wide representation. NLC did a lot on popularization and training on digital library technology and metadata application. It helps to reach the consensus on the importance of the standardization and put together as many as the stakeholders to take part in the promotion activity.

“The First Conference on Construction and Sharing of Chinese Documents” which held in Beijing on the June 2000 pleaded for the “Chinese Metadata Standard” to fulfill the needs on Chinese information resources construction. There are 62 delegates from 42 institutions attend the meeting. The delegates are from mainland China, Taiwan Province, Hong Kong, Macao, Singapore, United States and Holland etc. They argued that the Chinese Metadata Standard is the fundamental to build Chinese digital collections and provide the basis for the steady development of Chinese online information. The conference urged NLC to lead the work on drafting the Chinese Metadata Standard. The Ministry of Culture approved the project proposed by NLC. On June 2001, based on the tracking of the foreign research and its own experiences for a long period in this area, and integrated with the characteristics of Chinese document, NLC has finished the draft version of “Chinese Metadata Standard”\(^\text{13}\).

The Standard adopted an overall framework which followed the reference model presented by OAIS. As for metadata element set, it refers to the similar projects from the Library of Congress, the National Library of Australia, and the Cedars project, the NEDLIB project and Dublin Core Metadata Initiative. The core metadata set contains the majority of Dublin Core Metadata Set, but adds a lot of supplementary elements for the description of intellectual properties and abstract format of digital resources. The draft version of March 2002 consists of 25 core elements, their qualifiers and the encoding schema with DTD and RDFs. It has been considered to have the suitable coverage on characteristics of Chinese Documents with adequate descript ability, functionality, operability and simplicity.

<table>
<thead>
<tr>
<th>Element</th>
<th>Element in English</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 名称</td>
<td>Title</td>
</tr>
<tr>
<td>2 主题</td>
<td>Subject</td>
</tr>
<tr>
<td>3 版本</td>
<td>Edition</td>
</tr>
<tr>
<td>4 内容摘要</td>
<td>Abstract</td>
</tr>
</tbody>
</table>


\(^{13}\) Can be downloaded from: [http://www.cdi.cn/download/dmds.pdf](http://www.cdi.cn/download/dmds.pdf)
The Institute of Digital Library of Beijing University (IDLBU) was founded in 1999 by the Library of Beijing University, Center for Information Science of Beijing University and CALIS. It is a combination of computer science specialists, the librarians and the professional administrators. It has focused on metadata research from the start. The “Chinese Metadata Standard Framework and Its Application” is one of its achievements in 2001.

The need for metadata standard comes directly from the collection digitizing project of the Beijing University Library, which has the most valuable, numerous and unique ancient documents of all the universities in the country, including rare books, rubbing and manuscripts. It is a great challenge to work out a unified metadata profile for the description of such a rich and varied collection. The researchers of IDLBU started to solve the problem by introducing a methodology framework. This becomes the outstanding feature of the study and promoted the levels of Chinese metadata research and development.

The Metadata Standard Framework of IDLBU analysed the different description needs for

---

14 It can be downloaded from: http://www.idl.pku.edu.cn/pdf/metadata_framework.pdf
15 http://www.idl.pku.edu.cn/index.html
different information resources in the various phases of their life cycle, examined the whole procedure of generating and utilizing metadata, and made the restrictions for the related entities according to their different roles. It provides a few principles for the generating of practical metadata profile and outlines a three tier metadata elements structure and its extension rules. It also discusses theory and gives some application and implementation approaches for the framework.

6. Other metadata research

Like the Digital Library Institute of Beijing University, Shanghai Library has formulated a scenario of metadata implementation for the application of resources digitization and integration. Some other institutions, such as the Tsinghua University Library and the Institutes of Chinese Academy of Science, have done some work on metadata standardization and research but have not made their results public.

Many companies which involved in digital library construction are heavily involved in metadata research. Some of them, such as Tsinghua Tongfang Corporation16, Wanfang Data Corporation17, Chongqing VIP Information Corporation18, Beijing Superstar Corporation19 and Beijing Shu Sheng Corporation20 have developed simple metadata solutions.

Discussion and Trends

In May 2002, a Forum on Standardization for China Digital Library Project21 was hosted by the Office of Joint Committee of China Digital Library Project, and sponsored by National Standardization Committee of IT and Digital Library Subcommittee of Chinese Library Association. Its main purpose was to guide the China Digital Library along the right track of Standardization, as well as sum up past research, bring together the experts to co-ordinate national standardization, and plan for the future on cross-domain co-operation.

Although Chinese metadata practitioners learn a lot from foreign pioneers, they are not very active in the international forums, perhaps because of the language barrier. Inevitably they are repeating and debating many topics already discussed over and over again by their foreign colleagues, such as the contradiction between simplicity and functionality, and how to reach acceptable flexibility, extensibility and scalability, etc. After years arguing, they come to a consensus on the functionality of the metadata. But how to encoding metadata in an information system and make it functional as promised is another story. That’s why there are many metadata specifications and profiles but few metadata applications. The metadata myth still exists. Especially for librarians, who compare any other metadata format unfavourably with MARC. But they realize it is time to change.

There are two directions of current metadata research and application in China. First, there is the

16 See: http://www.cnki.net/daobao/cnkidaobao5/cnkibzh02.htm
17 See: http://www.wanfangdata.com.cn/
18 See: http://www1.tydata.com/productor/vipservice.htm
19 See: http://www.ssreader.com.cn
20 See: http://www.21dmedia.com
introduction of reference models for setting up co-ordinates to investigate if the metadata profile has suitable extensibility and enough inoperability, for examples, the Functional Requirements for Bibliographical Record (FRBR) model from IFLA and the Open Archival Information System (OAIS) model from CCSDS\textsuperscript{22}. The FRBR model is used for comprehensive systems that contain alternative objects with different formats in different period of its lifecycle. It provides a framework which is perfect for recording the intellectual properties of the different manifestations of the same object, while the later model is good at describing preservation system which is demanded imminently by a lot of traditional libraries and achieves. The two models provide a balanced methodology for choosing the appropriate approaches in all kinds of metadata applications.

Second, metadata standardization becomes more systematic and hierarchical. There are general, cross-domain as well as domain specific, special metadata standards or specifications. There are national and ministry level standards as well as just for institution itself usage. There are draft national standards, recommended specifications, guidelines, best practices and examples. So it can be expected that the metadata movement in China will be going on. And there will be many mature metadata applications in the following years.

**Conclusion**

As a developing country, China’s experiences with metadata standardization and implementation are comparable to those of many other countries:

1. The importance of information resources has become more and more evident after many countries built up their basic information infrastructure. For the sake of long term preservation of the digital resources, and for the purpose of widely sharing and reusable in a distributed environment, the standardization is very important.

2. Keeping track of research on digital library standardization by developed countries is more important than does research and development by developing countries themselves. Many standards are not as simple as they look like. They are supported by a large amount of investment on basic research. It’s no necessary to reinvent the wheel. Standards and specifications can be used directly. However, there is much work to do in localization and customization.

3. The popularization and training of the standards and specifications is of the same importance as the establishment of them. The current metadata standards in China are mostly well designed with rich description but less flexibility and operability, and limited in some disciplines. There should be efforts made to distribute them in the professional circles and improve the situation. Only by this means will it be possible to make amendments to the standards from practical applications.

**Acknowledgements**

The writing of this paper was encouraged greatly by Dr. Jianzhong Wu, the director of Shanghai Library. Also I was fortunate to have the support of my colleague, Mr. Leon Zhao, for occasional discussion, and I was helped with bibliographic data on the research in metadata and digital library in China by Cuijuan Xia and Xiangying Lou.

\textsuperscript{22} See: http://www.ccsds.org/documents/pdf/CCSDS-650.0-R-2.pdf
Reference
Zhang Xiaolin, Xiao Long etc. “The Framework of Standards and Specifications for the Construction of Digital Libraries in China”, from one of the reports of the project “the Construction of Standards and Specifications for Digital Libraries” by the National Science and Technology Library;
Wang Qiyun “The Investigation and Analysis on Thesis of Digital Library in Recent 6 years” The Modern Technology of Library and Information Science 2002.2