

# **Beginning and Development of China's Digital Library**

**Sun Chengjian & Liu Gang,  
National Library of China  
PRC**

With the rapid development of modern high and new technologies, Internet regarding computer technology and network communications technology rose almost overnight. Human society began to take pace into information society after experiencing agricultural society and industrial society. At the same time, digital library, representing a revolution and renovation in the history of library, appeared.

## **1. Basic Concept of Digital Library**

The digitization and great development of Internet led to globalization of economy, commerce and information dissemination, bringing deep effects on all levels of society. Mass information gushes into Internet. However, it becomes more difficult to search information on Internet because much information is in chaos. There is an urgent need to create a new model to manage these information resources. This is the background of putting forward the concept of digital library.

Currently, digital library is in development. But there is still no definite and complete definition of digital library in the world. In a technical sense, digital library is a digitized information resource system by using modern high and new technologies. It will become the managerial model of information resources for the next-generation Internet, making the scattered information on Internet in order.

Digital library is based on unified standards and norms with the support of distributed mass resource repositories. It is dealt with intelligent retrieval technology and managed by E-business. In addition, information will be transferred to a large number of households and families through broadband high-speed networks. It involves the whole process of producing, processing, storage, searching, transferring, protecting, using, documenting and deleting of digital information resources. It would become difficult for an individual library to do it without cooperation between libraries, museums, art galleries and archives all over the country.

Before the concept of digital library, people discussed the future development of library from some aspects of electronic library, library without wall and digitized library.

Electronic library refers to library that stores and transfers information by electronic media, such as electronic reading rooms and OPAC search. It puts emphasis on providing service in a fixed sphere by using computers. Library without wall refers to library of which readers can search digitized collection by using network communications systems. It puts emphasis on the importance of network transport. Digital library lay special emphasis on the digitization of literature. Readers can use digitized books and literature. It is more difficult to construct digitized library than electronic library and library without wall. Nevertheless, it is more convenient for us to use digital library.

### Contrasts between Digital Library, Traditional Library and Automatic Library

	Traditional Library	Automatic Library	Digital Library
Main Function	Collection	Collection	Users
Types of Collection	Publications	Publications and e-publications	Digital information resource
Forms of Work	Handwork	Automatic processing of bibliographic data and specific databases	Automatic processing of literature
Search Methods	Search cards by hand	Automatic search of bibliographic data and specific databases	Intelligent search
Service Object	Readers who come to the library	Mainly for readers who come to the library, providing literature delivery in a fixed sphere	Global readers on Internet
Collection Processing	No	No	Digitized collection can add value

## 2. Introduction to the Development of Foreign Digital Libraries

Western developed countries began the research in the late 1980s. Main projects in international digital library are:

**(1) The United States:** The United States plays a leading role in the construction of digital library, and puts increasing emphasis on it.

Since 1994, NCO for CIC of the American government issued an annual "Blue Paper". It is a formal official report about research and development. In recent years, the report has attached more importance to the strategic meaning of digital library.

In 1995, the Blue Paper "Technology for National Information Infrastructure" listed nine NII nation-level challenges. They are: digital library, crisis and emergent management, education and life-time learning, E-business, energy management, environment control and wastes minimization, health care, processing and products and government information for public access.

The report pointed out that digital library is the basis of knowledge center without wall.

In 1996, the Blue Paper "High Performance Computing and Communication: Basis for Future American Information" introduced 18 projects of digital library held in the United States and their IP address.

In 1997, the Blue Paper "Promoting the Forward Position of IT" listed digital library in valid technology.

In 1998, the Blue Paper "Technology Used in the 21<sup>st</sup> Century" put digital library in the first place of the six major projects of CIC (Calculation, Information and Communication).

Now, "American NSF/DARPA/NASA Digital Library Proposal of American Digital Library Project" has been completed. This project includes six subitems, covering large-scale literature repository, space images repository, geographic image repository and audio and video resource repository. The other project "American National Digital Library Project" is expected to be completed in 1999. It includes American history and cultural and scientific achievements. Fifteen research libraries and archives took part in this project.

## **(2) G8 Global Information Society Electronic Library Project**

It is made up of eight national libraries of France, Japan, the United States, Britain, Canada, Germany, Italy and Russia. This project covers essence of culture of these countries. It is going to be finished by the year 2000.

**(3). France:** The digitized construction of the National Library of France. Its digitized resources have reached over 3,000GB and have 8.3 million bibliographies.

**(4). Great Britain:** The National Library of Britain stores innovation proposals. It includes 20 projects, most of which have been finished and been served on Internet or LANs.

**(5) Japan:** Small-scale experimental digital library project. It includes the experiment of 8.8 million national union bibliographies, digital library practice experiment and digitization of 10 million pages collection in the National Diet Library of Japan.

Kansai - KAN project of the National Diet Library of Japan. This library will become the largest digital library in Japan and the delivery center of literature in Asia. The Japanese government invested 400 million US dollars. Moreover, the first section is expected to be finished in 2002.

Digital library construction must be supported by high and new technologies. In the construction of resource repositories, we must use highly efficient on-line analysis manipulation, data repository and data excavation. We also research on how to use data repository and data excavation in architecture, research on the definition and automatic extraction of metadata, compression and reduction of multimedia object data, realize rapid access of mass data in digital library. It is necessary for us to realize interaction in distributed resource repositories of digital library. It is important to make research of high-speed search engine and rapid access in super large-scale distributed resource repositories. We must come to the realization of real translation using multiple languages, the realization of multimedia-based search including retrieval, classification, search and automatic abstracts and transform matching, and the realization of flexible polyhedron search, requiring that the system can be applied in a national, global or individual sphere.

### **3. Research and Development of China's Digital Library**

#### **(1) Experimental Project of Digital Library**

In early 1996, the National Library of China submitted "Experimental Project of Digital Library" to the Ministry of Culture.

The project began with the image data repository of Chinese Doctoral Dissertations by using client/server model. It uses the index and search of the bibliographic data server management data, and the information of the image data server management digitization. The scanning images adopt 300-dpi resolution ratio, and compress and decompress according to CCITT Group 4 Standards. The images are stored in JukeBox. It will realize search by establishing multi-level indexes and multiple-repositories connection and will be able to provide service via networks.

## **(2) Research and Development Project for Future-based on Multimedia Information Search System**

This project is carried out between the National Library of China and the Computing Technology Institute of the Chinese Academy of Sciences. It has passed the technical authentication.

The project mainly makes research on feature-based image information search, brings about the search according to the textures of images, colors and figures, the full-text search of Chinese information, search and the research of information memory management methods, realizing search at the customer end at cross platforms.

## **(3) Application of SGML in Library**

The Ministry of Culture appointed this project in 1997. It is undertaken by the Modern Wenjin IT Research Center of the National Library of China and jointly developed by the Computer Institute of Peking University. SGML (Standard Generalized Markup Language) is a standard for information processing issued by the International Standard Organization in October 1986. At present, it has become one of the 12 international standards. In 1994, China determined it as national standard used in the fields of information processing of news publishing. HTML on Internet is its practice, while XML is its subset.

At present, the National Library of China provides readers with MARC data. Readers can find books by searching literature's title, author, call number and/or subject, but can not find the whole text. With the development of Internet, it is far from enough just to search bibliographies. The using of SGML can meet new demands under network environment. And SGML has been in use of the construction of many foreign digital libraries. It is one of the key technologies in the digital resource processing system of digital libraries and an essential condition in the data processing of digital libraries.

Therefore, the National Library of China developed an SGML-based applicable system with the light of practice of SGML usage in foreign countries. Users can directly access data in SGML applicable system through Web browsers on Internet.

The National Library of China began to provide readers with this system in February 1998, and has got good effects. The system was checked and accepted in May 1999.

## **(4) Experimental Project of China's Digital Library**

In May 1996, the National Library of China put forward the Experimental Project of China's Digital Library, and submitted it to the State Planning and Development Committee.

This project was appointed in 1997 as one of the state key technology projects. The National Library of China plays a leading role in it. Shanghai Municipal Library, Liaoning Provincial Library, Nanjing Municipal Library, Guangdong Provincial Zhongshan Library and Shenzhen Municipal Library joined in this project. The project includes several databases with overall and uniform technologies. These databases are: Image Database of Chinese Ancient and Rare Books, Image Database of Chinese Doctoral Dissertations, Database of Historical Atlas and Pictures, Full-text Database of Sun Yat-sen's Literature, Literature Database of Shenzhen Special Economic Zone, Multimedia Database of National and Overseas Tourism, Literature Database of Nanjing Government under the Republic of China (1911-1949), Graphic Catalogue Database of Northeastern China, and literature for International Digital Library.

The project plans to develop and complete a set of performance technologies of digital library keeping the track of international standards. The project plans to:

- Establish SGML-based object description and editing system;
- Organize digital object repository with open architecture and establish general repository-visiting protocols;
- Establish scheme and metadata set and uniform metadata;
- Make experiments to establish distributed scheduling systems of digital object with different sources;
- Design and realize convenient users' interface of network and practical interface of system management;
- Realize that the methods of descriptions for digital object must support distributed inquiry and search of different sources;
- Initially realize rapid search for super capacity database, and support screen interface with both Chinese and English;
- Make experiments to establish standard information repositories of manifold types in order to provide trial service on Internet.

At the same time, this project also plans to finish the report "the Concept and Development Research of Digital Library", making comprehensive survey and study from theory to practice and from technology to application, and analyzing the concept of digital library, the relative major technologies, trends of digital libraries in foreign countries, perspectives of application and development, effects on society and economic benefits in an all-round way.

Now the participated libraries and companies have already made progress in relevant research and development.

### **(5) Knowledge Network - Systematic Project of Digital Library**

It belongs to the State 863 Project of 1998. It will be completed with the cooperation of the National Library of China and Beijing Shuguang Tianyan IT Corporation Limited.

This project aims at designing and developing the first-level architecture of digital library in order to set up an experimental digital library system of China. It will be built under Internet environment. Its architecture includes several distributed digital resource repositories. Artificial intelligence technology will be adopted focusing on the rapid inquiry of cross-repositories. It will greatly shorten the gap in the fields of research and practice of digital library between China and developed countries.

In technology the project must have the functions of network management, inquiry and search of multimedia information, storage and search of mass information, authorization of intellectual property. It will realize interaction on Internet. The methods of description for digital object can support distributed inquiry and search with different sources, and support the protection of intellectual property and tax-paying obligations stipulated by law. It can provide rapid search subsystem for digital object with super capacity.

This project will consult and follow the mainstreamed methods and fruits of research in the world. By combining with the features of Chinese resources, it will keep track with international standards in general structure and developmental technology. The main job of development will be centered on the establishment, maintenance and release of SGML/XML-based Chinese resources. We will accept and digest foreign mature technologies. The system puts emphasis on the practical functions of the next-generation Internet.

The fruits of digital library systematic projects will be used in the construction of China's digital library.

### **(6) Digital Library Project of Liaoning Provincial Library**

The project is carried out by Alpie Software Company of Northeastern University based on the IBM digital library system. The comprehensive management scheme for multimedia information in network environment of IBM digital library has five functions: establishment and attainment of contents, memory and management, authorization management, visit and search and information release.

Liaoning Provincial Library plans to use the system to realize digital processing for ancient books, information release via Internet, multimedia reading rooms and VOD for the

asking. It decides to use TDI (Time Delay and Integration) digital camera of IBM for the digitized processing of ancient books.

Liaoning Provincial Library chooses some multimedia information at earlier stage, including rare books carefully chosen (continuous images), historical pictures (single images), letters left by celebrities (images and full text), montage of films and television (audio and video).

The library has made some progress in the making of Zhang Xueliang and places of interest of Liaoning, getting precious experience.

#### **(7). Digital Library Project of the Ministry of Education**

The project shouldered by Tsinghua University, Peking University and Shanghai Jiao Tong University includes research on the structure, search mechanism and applicable standards of digital library; learning system for information of graphics and texts; and the shaping form of digital music library, and a small demonstration system of digital V.F. database.

In addition, Shanghai Municipal Library and the Literature Information Center of the Chinese Academy of Sciences are also making research and development of digital library.

### **4. China's Digital Library Project**

In 1995, the National Library of China began to keep track of the development trends of digital library in the world, learning the relative standards and technologies, and got the first-hand experience and had a general idea of the architecture of digital library. In July 1998, it put forward this project.

There are other reasons for the project. The first one is that the process of China's digital library is not satisfactory for the lack of funds. The second is that the gap will be widened if we do not take actions on the development of digital library. The third reason is that we do not have enough place for storage of books.

The aim of China's digital library is at the shaping of Chinese resource repository groups with large-scale and high quality on broadband IP network, supporting the development and forming of innovative architecture, and providing the whole country or even the whole world with services through national backbone communications network.

The construction principles are uniform planning, uniform technology standards, uniform operation principles, union construction and resource sharing, the usage of the existing public networks, preventing duplicate construction.

The cycle of the construction is ten years. The second section of the National Library of China covers 50 thousand square meters. We will construct super large-scale resource repository groups and finish practical technology of China's digital library.

### **(1) Characteristics of the Project**

- With visual networks, large-scale, open and distributed network architecture of information resources, with the capability of providing complicated information processing and access.
- Producing and storage of mass information;
- Information resources based on Chinese, including foreign languages and languages of the minorities of China;
- Compatibility and interaction, open and expandable network system;
- Network system, information resources and information system according with international standards.

### **(2) Technical Routes of the Project**

- Using SGML, URN (Uniform Resource Name) and CORBA (Common Object Request Broker Architecture);
- Adhering to relevant international standards and industrial standards for electronic information processing and exchange;
- Combining uniform architecture with flexible subproject;
- Adopting distributed object-oriented software technology under network environment;
- Combining the usage of domestic technology and introduction of foreign advanced technologies.

### **(3) Technical Aim of the Project**

The aim of the project includes the realization of the framework of digital library in China, adoption of new and advanced technologies, establishment of resource repositories of digital library in China. Inquiry and search of all resource repositories are based on national broadband network systems and on Internet.

#### **Firstly, in the aspects of digital processing of information resources;**

- The digitized processing of books, magazines, recordings, videos, drawings and

pictures in all-level libraries;

- Collecting and arranging electronic publications and digitized audio and video materials for resource sharing;
- Promoting the usage of information resources and making levels;
- Expanding the fields and ranges of information resources, and promoting domestic coordination and international exchanges;
- Developing the reprocessing of information resources with manifold forms, multiple language and multiple texts;
- Strengthening the memory, search and standardization of digital information resources, realizing the process of intelligence step by step and promoting management.

**Secondly, in the aspect of network transport of information resources:**

- Making full use of the existing network system for information transfer;
- Using general networks with Internet TCP/IP and text forms;
- Fully considering the future of information structure to meet the development of interactive multimedia information;
- Realizing operative and interactive network transport on users' interface.

**(4) Social Economic Aim of the Project**

The project has broad domestic service architecture and exchange function for international information with distinguishing features.

**Firstly, the realization of service marketing of information**

It embodies in social effects and economic results. China has rich information resources and large potential markets. The competitors of markets come from foreign countries. The following information services are important:

Social information services of digital library;

Information release and advertising media services;

Buying and selling services of electronic publications;

Services for the asking of recordings, videos and television;

Exchange services of international digitized products of digital library;

Public notice services of interactive information between users.

**Secondly, internationalization of information institutes**

The project must make breakthroughs in the following areas:

Business coordination between international digital institutes;

Market complementarily between international marketing and service institutes;

Information sharing between international information marketing institutes;  
Technical exchanges between international software/hardware development institute.

Leaders at all levels and people in all walks of life attach great importance to China's digital library project. It is of vital importance for developing knowledge economy.

## **5. Experimental Digital Library of China**

The National Library of China has completed the experimental environment and pilot system in a period of only three months.

The pilot system makes progress and initial experience in the designing and processing of multimedia digital resources, metadata extracting and generating of metadata base, construction of object database, selection of scheduling system and discussion of cross-repositories search and users' interface. We have developed an software for processing, markup and management, an software for the uploading of metadata and object data and management, an software for dynamic pages including multimedia objects, and an software for supporting SGML and distributed database search, and an software for inquiry and the connection of metadata and object data.

The system is being run on the library-wide network of 1,000 Megabits, realizing real transfer.

It includes five multimedia resource repositories. They are Qian Jia Shi, Chinese Ancient Buildings, the Forbidden City, Aquarium World and Universe. It has information of 5GB. Two of the resource repositories are positioned in the Branch Library 10 kilometers away.

The Qian Jia Shi Repository digitizes the colorful drawings of the Ming dynasty, adding watermarks on them by using IBM's technology. There are pictures, notes and reading with music for 36 poems.

The Chinese Ancient Buildings Repository reflects palaces, temples, pavilions, grottoes and the Great Wall with pictures and explications.

The Forbidden City Repository gives us the glory and magnificence of the buildings and its art curiosas with pictures and explications.

The Aquarium World Repository selects many audio and video materials.

The Universe Repository aims at popularizing basic knowledge of universe, with pictures and explications.

According to the demands of the construction, these resource repositories are connected to the national bibliographic database and some encyclopedias, realizing multi-repositories connection, cross-repositories search and cross-regions and cross-platforms search.

Through the development of the experimental system we have gained some experience in the designing and processing of multimedia digital resources, the extraction of metadata and generation of metadata database, the construction of object database, the selection of scheduling system and cross-repositories search. We have developed software on the processing, markup and management of data, software on the loading and management of metadata and object database, interface software with the capability of supporting SGML, distributed database search, search software and the connection of metadata and object database, and software capable of generating dynamic pages that users can use Web browsers to browse.

After the completion of the experimental system, leaders from the State Council and the Ministry of Culture and some experts from various libraries came to the National Library of China for a visit. They thought highly of it.

By several years of hard work, the National Library of China has made great progress in the research and development of digital library. Nevertheless, there is still a long way to go to catch up with the developed countries. We will widen our cooperation with companies and institutes both home and abroad to quicken the construction of China's digital libraries.

**(Sun Chengjian: Deputy Director of the National Library of China Liu Gang: Deputy director of the Information and Network Department, the National Library of China)**