

## The Preservation Practice of EBooks in the National Library of China

### Wei Dawei

Deputy Director of the National Library of China, National Library of China, Beijing, China  
weidw@nlc.gov.cn

### Ji Shiyan

Information technology department, National Library of China, Beijing, China  
jisj@nlc.gov.cn

### Dong Xiaoli

Information technology department, National Library of China, Beijing, China  
dongxl@nlc.gov.cn



Copyright © 2014 by Wei Dawei, Ji Shiyan, Dong Xiaoli. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License:

<http://creativecommons.org/licenses/by/3.0/>

---

### Abstract:

*The National Library of China takes the responsibility of the collection and preservation of domestic and foreign publications with the purpose of cultural inheritance and serving the whole society. In recent years, National Library of China has collected many kinds of eBooks, and has gathered some experiences from the practice in eBooks preservation and protection and utilization. This paper introduced the source of eBooks in the National Library of China, the laws, regulations and standards of eBooks. The reliable preservation and sustainable utilization of massive eBooks are introduced in detail.*

**Keywords:** The National Library of China, preservation practice, eBooks.

---

According to the definition from the Chinese National Standard<sup>[1]</sup>, eBooks means authorized or non-authorized digital literatures, which are usually in searchable text format and looks like the printed books. In accordance with this definition, the National Library of China details the eBooks definition which is applicable for library industry: Such information as images and texts are stored in devices of light, electricity and magnetic materials and spread through computer network and mobile terminals. eBooks are a kind of digitalized books that can be read with the help of computers and similar devices.

## **1 The status of the National Library of China eBooks**

The National Library of China (NLC) serves as the repository of the nation's publications, a national bibliographic center, as well as a national center for the preservation and conservation for ancient books<sup>[2]</sup>. With the development of the National Digital Library Project<sup>[3]</sup>, it also becomes the world's largest Chinese digital information preservation base and service base. The NLC has been producing eBooks since 1999. In 15 years, the NLC gradually accumulates relevant experience for the construction, preservation, service and management of the eBooks.

### **1.1 The source of eBooks and the rules for the collection**

According to the principle of collecting all available books in Chinese and excellent books in foreign languages. The NLC expands its eBooks collection by self-built, imported and united-developed eBook resources, by the end of 2012, the NLC's eBooks collection totaled 3,994,000 volumes(items)(3,538,000 titles)<sup>[4]</sup>.

#### **1.1.1 The self-built eBooks**

Since 1999, the NLC started to carry on the digitalized construction of characteristic collection in a planned way. The construction scope includes the printed modern Chinese books, Minguo books, rare books, ancient books and special collections.

The NLC only chooses the official publication in the digital construction of modern Chinese eBooks, and considering the orientation of Public Library, during the digital construction, the NLC not only selects the ordinary Chinese books, but also chooses Chinese books for children, the disabled and other special groups. At the same time the NLC also buys digital eBooks from the publishers.

After accumulation of many years, the NLC gathers rare books, ancient books, manuscripts of famous, the materials in Minguo and other ancient literatures. Due to the long history, environmental pollution, paper aging and other reasons, some of the ancient literatures have been damaged, every time when users read the literature it will increase the damage of them. Therefore, the NLC started the digital construction of the ancient literatures to regenerate them since the middle of 90th. The NLC led the program of "Preservation of Materials in Minguo"<sup>[5]</sup> with other seven literature collection organizations to digitalize, collect and arrange materials in Minguo. In order to protect the literature, the NLC uses two digitalization strategies. One is to digitalize the original literature directly, such as chronicles, genealogy. The other strategy is to microfilm the ancient literatures first, and then digitalized the films, such as the materials in Minguo and the collection from the Song dynasty.

#### **1.1.2 Imported eBooks**

According to the principle of collect all available books in Chinese and excellent books in foreign languages, the NLC imports many commercial eBook resources. Base on the copyright, all the resources can be used in the library; some of them can also be used on the Internet. Not only the users in the library, but also the card users and the users who registered on the internet can also use the resources. All of these efforts make the NLC achieve the goal of the Public cultural services.

#### **1.1.3 United-developed eBooks**

The NLC collects local resources, chronicles, genealogy, the minority resources and the resources for the children from national libraries, the members of the national digital library construction and service conference since 2010<sup>[6]</sup>.

### **1.2 The characteristics of eBooks**

The NLC eBooks has the characteristics of long history, vast data, various formats and high speed development.

### 1.2.1 Vast data

By the end of 2012, the self-built eBooks include 177,662,000 pages (539,000 titles) of Chinese books; 4,961,000 pages (23,000 titles) of Minguo books; 3,340,000 pages (6,868 titles) of chronicles; 162,000 pages (513 titles) of the collection from the Song dynasty; 15,000 pages (38 titles) of the Yongle encyclopedia; 2,589,695 pages (2,392 titles) of genealogy.

The imported eBook resources include 927,000 volumes (items) (471,000 titles) of Chinese books; 2,496,000 volumes (items) (2,496,000 titles) of foreign language books, totaled 3,243,000 volumes (items) (2,967,000 titles).

United-developed eBooks include Chinese books, Minguo books, chronicles and genealogy from Xinjiang, Qinghai, Jilin, Heilongjiang, Shanghai, Jiangxi and other 20 public libraries, totaled 4,584,341 pages (17,262 titles).

### 1.2.2 Released most of the eBooks

After the collection and preservation of the eBooks, the NLC has made most of the eBooks available online.

More than 80% of the self-built and united-developed eBooks have been released, 82% of Chinese books have been released; 95.7% of Minguo books have been released; 99.3% of chronicles have been released. All of the imported resource can be used in the library, and 78% of them can be used online.

### 1.2.3 Various formats

Based on different sources and types, eBook resources have various formats such as PDF, JPEG2000, TIFF, TXT, PDF, SWF for preservation and release. Even the same type of resource, due to digital processing in different periods, they may have the different format for storage, or different versions of the same format.

## 2 Research of eBook standards and regulations

The NLC is responsible to collect and preserve the resources. So as to preservation and utilization of the eBooks, the NLC is involved in research, promotion and tracking of the relevant laws, regulations and standards.

During the National Digital Library Project, the NLC developed 34 standard projects, declared 17 items of the cultural industry standard of the construction and sharing for digital resources. By the end of 2013, 18 projects completed and published. These standards will not only used in the NLC, but also granted to the national libraries and research institutions. These standards will be useful for the digital library construction and the digital resource construction in related industries.

The NLC released *the Rules of the NLC Electric Books Metadata Specification and Description* for the metadata of eBooks<sup>[7]</sup>.

The NLC compiled *the NLC Digital Resources Long-term Preservation Management Regulations* for the long-term preservation of eBook, and *the NLC Metadata Long-term Preservation Specification* for the preservation of metadata.

With the rapid development of digital and network technology, many changes have taken place in work creation and dissemination. In order to adapt the progress of technology development and protect the copyright, the revision of "copyright law" started in 2011 March<sup>[8]</sup>. NLC actively fellow the amendments and suggestions of the "Copyright Law (Revised Draft)", and provided its own recommendations.

## 3 The preservation practice based on the eBooks life cycle

In accordance with related laws and principals of resource construction, preservation, utilization, the NLC developed preservation system based on life cycle. As the same time the

NLC established resource preservation strategy and technology protection mechanism to protect the long-term preservation, long-term utilization, permanent inheritance of eBook resources.

### 3.1 The principals of eBooks preservation

Valuable digital resource was preserved totally thanks to some principals were made by NLC as preservation is priority, utilization and service, highlight characteristic, which will contribute to utilization digital resources.

#### 3.1.1 The principals of standard data preservation

The identification of resource is very important for both preservation and utilization of eBooks, therefore the NLC constructed a unique identifier system and complied the naming rule. Each book will have a unique, unchanged identification in the process of eBooks production, and this *Unique Identifier* will be used through the whole process of its life cycle.

The NLC identifies the copyright information of eBooks in space, time and content. The resources will be saved by batches during the preservation. Each batch will have a copyright file, and every book will identify the copyright information by its *Unique Identifier* in the file.

#### 3.1.2 The principals of metadata preservation

In the preservation of resources and services of library, the collection, preservation, integration and release of the metadata are very important. Therefore, the NLC establishes preservation strategies for the metadata from different sources, and constructs the metadata warehouse.

Different formats of metadata include: Marc, Excel, MDB, MDF, XML and TRS. In the integration of metadata fields, considering the storage, release, retrieval of the metadata, the NLC defines the core field of metadata warehouse; corresponding relation between the fields and the core fields and extraction rules in different types of metadata database.

The metadata that will be stored in the warehouse need to be converted in character set. The system uses UTF-8 as the internal character set, and transforms the other characters into UTF-8.

In the management of metadata file, considering the data support, the NLC converts the metadata of different format into XML format for storage.

#### 3.1.3 The principals of object data preservation

The NLC defines three levels of preservation methods, and corresponding storage, media and backup amount.

Table 1. Different levels of the NLC eBooks preservation

Level	Media	Explanation	Amount
Permanent preservation level	magnetic type, disk	Permanent preservation	Three copies
File preservation level	magnetic type, disk	Non-Permanent preservation File preservation	One or two copies
Temporary preservation level	magnetic type, disk	Non-Permanent preservation	One or two copies

Management of eBooks preservation format:

- Metadata : all the metadata use CNMARC or XML file as unified standard.
- Object data : Self-built eBooks would be digitalized and stored by images. The imported and united-developed eBook resources, would be collected and organized in accordance with *the NLC Image Data Processing Standard and Operation Manual*<sup>[9]</sup>.

Long-term preservation object data would be stored by non-compression format TIFF or JPEG2000, object data to be published would be preserved and serviced as PDF file.

- Format of indexing and cited data: the NLC digitalizes its own collection of one class resources which takes one batch as the unit. For one batch of eBook resources, the related indexing and cited files are necessary including metadata and object data. Indexing and cited document can index the detailed relationship between metadata and object data, as well as indexing the information of contents, title etc, which have significant influence on indexing the current information of digitalized resources.

For the storage and organization of object data, eBooks would be packaged as one item and its unit of storage and utilization is .tar file. Self-built eBooks and united-developed eBooks would be digitalized and arranged as an organizational unit one batch which also would be preserved as a unit one batch. In other words, one batch preserved file contains a metadata file, an index file, all object data files.

Some eBook digital resources are not be involved in the NLC preservation due to authorization issue. The NLC only has access to use them instead of ownership.

### **3.2 The technical safeguard mechanism of how to preserve eBooks**

In order to guarantee that eBooks can be reliably preserved and long-term available, the NLC set down and adopted a series of technical safeguard mechanism to ensure the safety of digital resources and provide safeguard for making use of digital resources in future.

#### **3.2.1 Normalization management in the process of digitalized production and arrangement**

Digitalized production and arrangement both need to normalize eBook data and store eBook resources having complete structure and clear corresponding relationship in the preservation system.

For the self-built eBooks that can be preserved directly, the NLC normalizes digital production in accordance with *the Rules of the NLC Electric Books Metadata Specification and Description*, *the NLC Image Data Processing Standard and Operation Manual*, *Object Management Practices of the NLC Images Digital Resources*<sup>[10]</sup>.

For the imported and united-developed eBook resources, the NLC will digitally arrange these resources in accordance with its rules. First, systematically check these resources for their integrity, readability and correctness. The checked items include: whether the required fields are provided; whether the metadata content is normalized; whether the format of object data can be applied universally; whether the indexed data is complete; whether the corresponding relationship between metadata and object data is correct or wrong; whether the storage structure conforms to the NLC object data management specification. Every item of books will be given the Unique Identifier as its identification card. Only the correct data can be submitted and stored in the preservation system after data-checking.

#### **3.2.2 Data-checking before long-term preservation**

The NLC ensures the data's integrity and consistency in both ways of management rules and technical methods.

The National Library of China compiled *the NLC Digital Resources Preservation Management Regulations* which states the principle of digital resources preservation, preservation levels and format of diverse digital objects; *Hand-over Process of Digital Resources Long-term Preservation*, *Workflow of Digital Resources Long-term Preservation* and other practical operation specification; *the NLC Digital Resources Long-term Preservation Regulations* which defines data reception, data-checking, sampling and extraction in details.

Check the integrity of data by checking the MD5 code with the data inventory. The MD5 code is produced in the process before the data is submitted and the MD5 code is created again in the process of transmitting the data to the storage system. Check the integrity and consistency of preserved data by comparing the two MD5 codes. After the data is submitted to the preservation system, its catalogue, file names and the number of files are checked to ensure the integrity and consistency of preserved data.

### 3.2.3 Data regular checking in long-term preservation

The NLC established *the Filing and Checking Regulations of Collected Digital Resources Long-term Preservation* and *the Regular Checking Workflow of Digital Resources Long-term Preservation* to ensure data and materials security, which especially defines the measures of how to preserve digital resources including the time interval, process and recording of regular checking.

The regular checking of data stored in long-term preservation system includes materials health checking and data checking. Regularly traverse the tape data and completely check the data stored in the materials in case of the damaged materials causing data loss. When the data was damaged, it would be recovered from the backup materials according to the disaster recovery strategy. To ensure the physical safety of storage devices and materials, our work staffs make the round of the storage machine room twice a day and check whether the machines work well and alerts occur, and the size of tape free spaces.

The regular checking of data stored in disks includes humidity, dust and magnetism elimination, regular work staffs operation and ensure that the storage environment meets certain requirements.

### 3.3 Three levels of storage system supporting eBooks preservation and utilization

The NLC has built its storage system whose capacity exceeds 2PB by using optical fibers channel storage system. The storage system adopts three-level storage mechanism, which meets the requirement of digital resources long-term preservation as well as the request and response of every application and services.

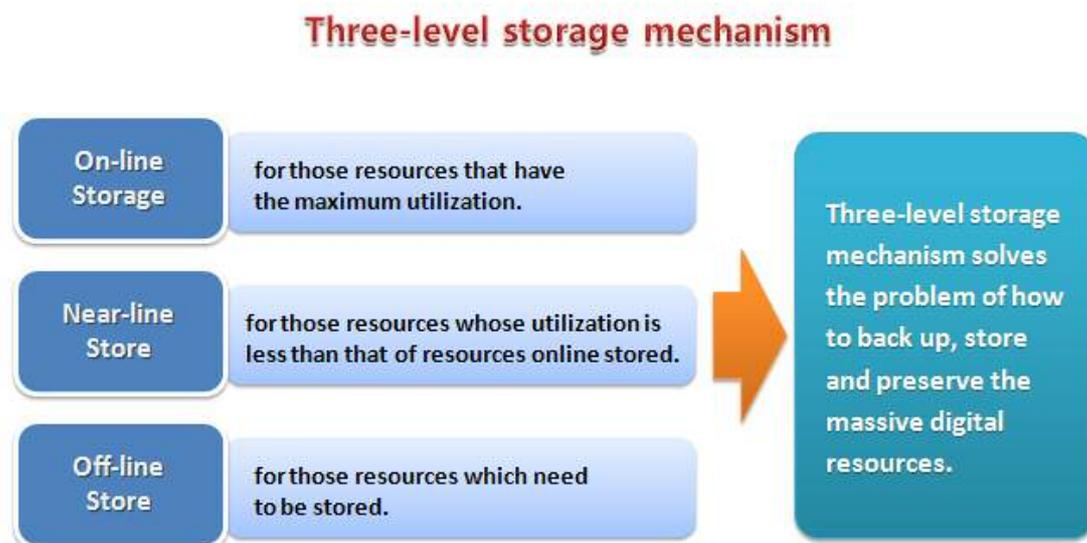


Figure 1. Three-level storage mechanism

The capacity of on-line storage and near-line storage system reaches 900 TB, using the disk arrays. Not only provide online service support for more than 20 important application systems including self-built eBooks Indexing system and imported eBooks database, but also provide near-line support for data format conversion. Disks and tapes constitute the NLC off-

line storage system equipped with 1,000,000 disks and 1PB enterprise-level tape library. The three-level storage mechanism effectively supports mass resources storage and long-term preservation.

#### **4 The NLC eBooks services**

The NLC provides many professional services based on the digital resources long-term preservation, users identity management and resources copyright control, which brings a kind of user experiences that is simple, quick, easy and effective.

##### **4.1 Resources revelation**

Users can access the NLC resources online wherever and whenever they are. Not all of the users have rights to access the whole eBooks. Resources access policy divide users into different groups and they have corresponding authorities to access resources.

##### **4.2 Unified information retrieval**

The NLC built a professional resources discovery system using the distributed data-processing technology based on the metadata database. Users can quickly search more than 200 million items of metadata including 1,860,804 items of eBooks metadata. The discovery system has been popular with users since it started running in September 2012. About 7,000 unique users visit the website daily and the number of monthly average page view reaches 4,850,654.

##### **4.3 Integrated services**

The NLC has deeply integrated the library traditional services with digital library services in the service content and terminals.

###### **4.3.1 Service integration between the NLC physical collection and eBooks**

Visitors can not only read eBooks on line, but also search its collection information in libraries including the NLC, submit the application of Interlibrary Loan and Document Delivery or ask the consultancy librarian by accessing the Reference system.

###### **4.3.2 Services integration using multiple terminals**

The NLC provides the public with free services through its service system covering the Internet, Mobile Communication Network, Digital Television Network and Satellite network.

The NLC started its mobile services in 2006, and basically completes its mobile digital library service system named *the National Library on Your Palm*. In 2013, the NLC cooperated with national libraries and initiated *Digital Library Mobile Reading Platform* providing 47,000 items of Chinese eBooks for users. Digital television service named *the National Library Space* based on the mutual interactive network was launched in September 2009 and provides public cultural video programs and more than 38,000 graphic messages for 3,600,000 families in Beijing.

*Digital Library for Visual Impairment and Digital Library for People with Disabilities* started providing barrier-free services for 900,000 users in 116 countries and areas in 2008. The two platforms totally provide 4,975 items of eBooks by the end of 2013, having 84,000 daily average hits and 2,000 daily average visitors.

The Digital Library Virtual Network covering the whole nation as one of the Digital Library Promotion Program achievements has been under construction since 2011. 66 public libraries including the NLC have been connected and a collection of excellent resources including 1,000,000 Chinese and foreign books are shared with local libraries on the network. Users can access the local resources wherever they are.

#### **5 Future prospect**

The NLC will actively learn the advanced technology concerned with how to preserve and make use of eBooks, cooperate with the technology departments and information

institutions, and draw lessons from the successful experience. The library will also continue to explore the reliable eBooks preservation strategy, focus on the research of eBooks preservation laws and regulations, emphasize on copyright protection and the establishment and implementation of eBooks deposit system.

A disaster recovery center is under construction to ensure the data security and work consistency of the NLC utility systems. The NLC also applied for initiating the program of National Literature Strategy Storage Library, which plans to build the national literature storage library in the area around Beijing and ensure that all the precious literatures and digital resources can be safely preserved.

In the future, the disaster recovery center and the NLC local storage system constitute the security storage system. The national literature storage library and the two systems constitute the distributed storage system. A digital resources preservation system which has three centers in two places will be built. The NLC will also establish digital resources preservation centers in some developed areas around the nation to achieve the aim of preserving the digital resources nationwide through the platform of Digital Library Promotion Project.

## References

- [1] GB/T 13191—2008/ISO 2789:2006.
- [2] The National Library of China[OL] [2014-03-20] [http://www.nlc.gov.cn/dsb\\_footer/gygt/lsyg/](http://www.nlc.gov.cn/dsb_footer/gygt/lsyg/)
- [3] National Digital Library Project [OL] [2014-03-20] <http://www.nlc.gov.cn/newstgc/>
- [4] National Library of China. National Library of China Yearbook 2013[M]. Beijing: National Library of China Publishing House, 2013
- [5] Website of Minguo Documents Preservation and Conversation Program [OL][2014-4-2] <http://mgwxbh.nlc.gov.cn/>
- [6] National Library of China. National Library of China Yearbook 2012[M]. Beijing: National Library of China Publishing House, 2012:138-139
- [7] Zheng Qiaoying, Liang Huiwei, Chen Youhua. The Rules of the NLC Electric Books Metadata Specification and Description [M]. Beijing: National Library of China Publishing House, 2013.8
- [8] Copyright Law of the People's Republic of China [OL][2014-4-2] <http://baike.baidu.com/subview/275830/12268083.htm?fr=aladdin>
- [9] Zhu Qiang, Zhang Chunhong, Long Wei. The NLC Image Data Processing Standard and Operation Manual [M]. Beijing: National Library of China Publishing House, 2011.12
- [10] Xu Yalong, Long Wei. Object Management Practices of the NLC Images Digital Resources [M]. Beijing: National Library of China Publishing House, 2013.7