The Collection and Digitization of the Genealogies in the National Library of China

Xie Dongrong
National Library of China, Beijing, China.

Xiao Yu
National Library of China, Beijing, China.

Shi Jian
National Library of China, Beijing, China.

Abstract:

The history of the genealogies collection of NLC has been reviewed and the current situation of the collection is introduced as well. The character and value of the genealogies in NLC is analyzed and the digitalization of the collections in different levels is emphasized including the various services to the readers basing on these digital products. 1) The genealogies have been catalogued in MARC format and the fundamental information of the collections is described. 2) The images of the original genealogies are acquired by scanning or photographing, which are provided to the readers with the cataloging data and the volumes information of the genealogies. 3) The genealogies have been indexed, the indexes to the titles and family members are focused on which described the information of the peoples and articles in the documents that can be used separately or accompanying with the images. 4) The images are transformed into the text format and the full text search can be achieved. 5) The people relations in the pedigrees in the genealogies are sorted out and the data of the family trees is acquired. The readers can search the database by people relation and specify the searching result. The digital resources are integrated so that the one station services can be accessed through website.

Keywords: Genealogy, National Library of China

Genealogy is the record of a family, along with the national history and the local history composing the general history of China. The earliest genealogy could be traced back to the Shang Dynasty. The genealogies compiled privately begun at the Song Dynasty and becoming prosperous from then on, a large amount of genealogies were complied, that has been continuing to nowadays. The collection of the genealogies had been emphasized by the

National Library of China (NLC) that is one of the most important institutes collecting Chinese genealogies. The genealogies belonging to NLC can be read in the traditional way, besides that, the documents have been digitalized and the images are published on the internet which can be accessed to freely. The history and situation of the genealogy collection of NLC are reviewed in the article and it is emphasized that the digitalization of the documents in different levels and multiple services basing on that.[1]

1 THE COLLECTION HISTORY OF GENEALOGY IN NLC

The famous scholar Liang Qichao(1873-1929) realized the documental value of the genealogy in the earlier days. In his monograph *Zhong Guo Jin San Bai Nian Xue Shu Shi* (the Academic History of China in the Recent 300 years) , it was proposed that the Genealogy was the treasure of Chinese history that could be found in every family of every village in China. Liang hoped that in the future when the national library was built up, the genealogies scattered around China should be collected and conserved and preserved in it waiting for the scholars in different disciplines to research on, which would be the immortal achievements. From 1925 to 1927, as the director of the Capital Library (the former of NLC), Liang promoted the purchasing and collecting of the genealogies positively. In *Bei Jing Tu Shu Guan Yue Kan* (Journal of the Beijing Library) vol. 1 num. 1, May, 1928, the book acquisition of NLC was introduced. The acquisition of the genealogies was recorded that NLC initiates collecting genealogies. 17 titles are donated, 34 titles purchased, more than 100 letters asking for collection have been sent. However, because of the social tradition for the long time, folks are not willing to donate their genealogies, and the instable political situation as well as the traffic problems, the genealogy collecting could not be carried out in large scale.[2] The Announcement for Collecting Genealogies was published on the *Bei Jing Tu Shu Guan Yue Kan* (Journal of the Beijing Library) vol. 1 num. 2, 1928[3]; the collection of genealogy facing the society began. The genealogies collected at that time were not satisfied in quantity, because that it was war time and the people at that time thought genealogies as the private documents that should not be open to the society. It was the initiation of genealogy collecting in public, and genealogical documents became one of the important components of the collection of NLC.

During the Chinese Japanese War (1937-1945), a large amount of documents were damaged in the battles around China. The collecting of genealogy was emphasized by the Beiping Library (the former of NLC). In the Report of the Ancient Book Acquisition it was said that several titles of genealogies of the Ming Dynasty were purchased[4]. The Capital of China was moved from Nanjing to Chongqing in the Southeast China. In order to preserve the documents against the war, NLC decided to collect the documents resolutely[5]. Genealogy was the important component of the collection. In the documents acquired by Wan Sinian, the

---

[1] The articles on the collection of genealogy: Zhang Zhiqing: Summarizing of the Genealogy Collected in NLC, Journal of Research on Genealogy vol. 3 (Bibliography Publisher, Dec. 1992); Yang Yanping: Summarizing of the Genealogies of Common Ancient Books Collected in NLC, Journal of Literature, vol.2,1999

[2] Li Zhizhong, Zhong Guo Guo Jia Tu Shu Guan Guan Shi Zi Liao Chang Bian. Beijing: National Library of China Press, 2009, pp80-81.

[3] Li Zhizhong, Zhong Guo Guo Jia Tu Shu Guan Guan Shi Zi Liao Chang Bian. Beijing: National Library of China Press, 2009, p84.

[4] Li Zhizhong, Zhong Guo Guo Jia Tu Shu Guan Guan Shi Zi Liao Chang Bian. Beijing: National Library of China Press, 2009, p316

[5] Zhang Tingyin, Responsibility, Experience and Knowledge, a Case Study of the Document Collecting in the Southeast China by NLC during the Chinese Japanese War, Journal of NLC, vol.1 2005.

librarian of Beiping Library, *Mu Shi Huan Pu* （Genealogy of Mus）, *Li Jiang Duan Shi Zu Pu* (Genealogy of Duans in the Lijiang District), *Li Jiang Yu Shi Zu Pu* (Genealogy of Yus in the Lijiang District) were included[6].

After 20 years collecting, by 1950, 348 titles of genealogies had been collected according to the Biography of the Genealogies in the Beijing Library.

During the 1950s-1960s, the traditional idea of folks changed, they sold their genealogies conserved in families or ancestral halls. A large amount of genealogies were purchased during this period. About 1,500 titles of genealogies were purchased according to the incomplete statistics, which were 1/3 of the total amount of the genealogies collected now. Many bibliophiles donated their collections to the NLC including genealogies such as the 23 titles in the Xidi Collection from Zheng Zhenduo(1898-1958).

After 1980s, the re-compiling of the genealogies was popular in the mainland China. The new genealogies of NLC are collected mainly by donation. After 2012, considering the lacking of the collection is serious, besides accepting the donation, NLC began to purchasing the new genealogies positively.

After 1990s, ancient books began to be sold at auction which attracted more and more attention. Many bibliophiles sold their genealogies at auction, and bidding became one of the important ways for NLC to acquire genealogies.

2 COLLECTION OF GENEALOGIES

By the end of December, 2012, 7,678 titles over 50,000volumes of genealogies have been collected in NLC including 458 rare books, 4,680 thread bond ones and 2,540 titles of new genealogies. The total amount of the original documents is just next to that of the Shanghai Library, which takes the second place in China. Most of the genealogies are written in Chinese, some ones are in Mongolia, Yi character, Tibetan and so on. Most editions are movable type prints, parts of them are block prints or manuscripts. Most of the collections are Chinese genealogies, a few Japanese, Korean and Singaporean genealogies are also collected. In conclusion, the genealogies collected in NLC have the characters as below:

1) Early edition. Quite amount of the genealogies collected in NLC are early editions. 2 titles Song Dynasty editions, 1 Mongolia block print, 1 manuscript and 1 block print of the Yuan Dynasty, over 300 titles of the Ming Dynasty editions. It is seldom among the collecting institutions owning so many rare ancient genealogies. *Xian Yuan Lei Pu* and *Zong Fan Qing Xi Lu* are the transcripts of the Song Dynasty (960-1279) which are the earliest royal genealogies and unique existed now (some fragments of *Xian Yuan Lei Pu* are collected in the Shanghai Library). They are the earliest genealogies in paper existed in the world also.

2) Genealogies of the Celebrities. From the Qing Dynasty, many genealogies of the celebrities in political, economic, military science and culture have been collected in NLC. It is included that Kong Shangren, the famous dramatist whose representative work *Tao Hua Shan*, Ji Yun the chief editor of *Si Ku Quan Shu* (Complete Library in the Four Branches of Literature), the famous ideologist Wei Yuan, great politician Li Hongzhang and Liu Kunyi minister to Japan and printer of *Gu Yi Cong Shu* Li Shuchang etc.. The celebrities of the Republic Period include Zhao Erxun, Xu Shichang, Miao Quansun, Ye

---

[6] Li Zhizhong, Zhong Guo Guo Jia Tu Shu Guan Guan Shi Zi Liao Chang Bian. Beijing: National Library of China Press, 2009, p331.

Dehui, Lu Xun, Huang Xing etc.. The celebrities of the P.R.C. include Mao Zedong, Dong Biwu, Xu Xiangqian, Rong Yiren and so on.

3) The continuity of the genealogy. Along with the increasing of the family members, the genealogies need to be recompiled after some years. The periods would be varied from 20 years to hundred years. The evolution of a huge family lasting for decades even hundreds of years could be learned from the continuing genealogy that recompiled periodically which provide the valuable researching materials. Such genealogies include *Qian Jian Pu Shi Zong Pu* (Genealogy of Pus of Qian Jian) in the Wuxi county of Jiangsu province re compiled 4 times from the Qianlong Period of the Qing Dynasty to the Republic Period (the block print editions of 1748, 1820, 1871 as well as the movable type edition of 1931); *Rong Shi Zong Pu* (Genealogy of Rongs) in the Wuxi county of Jiangsu province recompiled 4 times also (the movable type editions of 1810, 1872, 1910, 1935).

3 THE DIGITIZATION OF THE GENEALOGY

In the recent years, the digitization of genealogy is developing rapidly, and some achievements have been made. Different from other documents, besides the libraries, achieves and other collecting institutes, digital companies as well as the clan relative organizations and individuals all attend the digitization of the genealogies. The digital products are various in forms, catalogue, images, text, encyclopedia items and so on included. The notable digital programs of genealogies include Genealogy Database of Shanghai Library, Comprehensive Bibliography with Abstracts of Genealogy Database and Full Text Database of Genealogy of Zhejiang Library, Union Catalogue of the Genealogies in Taiwan of the Taipei Center for Chinese Studies, Historical Genealogy Database of Beijing Erudition Digital Technology Research Center and so on.

The digitization of the genealogies developed from catalogue and index to image database. From 2000, the card catalogue are transformed into MARC, readers can search the catalogue on the website of NLC at home without visiting the library. After decades' research and practice, we found that the digitization focus on the description of the document could not satisfy the requirements of the readers.

1) MARC of the Genealogy
Catalogue is the basic of the collection, organization and description of the genealogy. CNMARC is adopted as the cataloguing format. Considering the specialties of the genealogy, the outside characters and the content are described including the title, author, edition, native place, origin place of the famous family, the name of a family, ancestor, and the ancestor of relocation. For example, *Xin An Su Shi Zu Pu*(the Su's Genealogy of Xinan).

```
001      312001068895
005      20121123094702.0
010      $b 綫裝
100      $a 20021005f17361795km y0chiy50      ea
1010     $a chi
102      $a CN
106      $a z
2001     $a 新安蘇氏族譜
         $9 xin an su shi zu pu
         $b 普通古籍
         $e 十五卷
         $f （明）蘇大纂
205      $a 活字本
```

```
          $b 木活字
210        $c 忠孝堂
      $d 清乾隆間[1736-1795]
215            $a 2 冊
       $c 圖及像
300        $a 版心題蘇氏宗譜
305        $a 9 行 23 字白口四周單邊單魚尾
6100       $a 家譜
6100       $a 蘇氏
6100       $a 安徽休寧
696        $a 傳 778.09
       $2 pgl
701 0      $a 苏大
       $9 su da
       $c 明
       $4 纂
801 0      $a CN
       $b NLC
       $c 19990914
905            $a NLC
       $b 96158
       $sJP3060
       $qFGPG
```

Now, the catalogue of the genealogies collected in NLC can be searched on the OPAC of NLC (http://opac.nlc.gov.cn), readers can use it visiting the website of NLC.

2) Images

From 2008, the scanning of the genealogies of NLC had begun. The technical standards were adopted as follow: 300 dpi resolution, RGB color type，24 bit color, storage in TIFF, JP2 and PDF format, 150 compressibility factor. The real condition of the origin book is required to be scanned. The top edge of the origin book is the benchmark for scanning, the adjacency as the middle mark. The top margin and the bottom margin of the origin page are maintained, the right and left margins of the original sizes are kept as much as possible. The page numbers of the images should be continued. Repetition, shortage, mistake, folding of the pages should be avoided (excepting the original shortage and mistake of the page). The complemented pages should keep the same size as the formal pages with similar color. The incomplete pages should use the back papers when scanned, avoiding the light and character reflection.

The black edge, the split and joint of the image, and the traces of split and joint should be possessed, the deflection of the image must be under 1°. The size of the image must keep as same as the original page, the error rate should be below 1‰.

2,392 titles of genealogies have been scanned by now, 2,570,000 pages have been finished. All the images are published on the internet which can be browsed online.

3) Index

According to the requirements of the patrons and the characteristic of the genealogy, name index and title index to the genealogies collected in NLC are compiling, the complete,

correction of the index are ensured, supporting unicode5.0, searching and display. For example

```xml
<?xml version="1.0" encoding="UTF-16" standalone="true"?>
<tag source_id="SZJP00005-10002410-00001.jpg">
 <persontag>
   <person sex="男" surname="吳" region="2694,633,2940,1788" id="1">
     <firstname>八府君</firstname>
     <hui>竦</hui>
     <zi/>
     <hao>敬文</hao>
     <alias/>
     <shi/>
     <birthday/>
     <death_date/>
     <shidai>四代</shidai>
     <branch>堂陰世系圖</branch>
   </person>
 </persontag>
 <titletag>
   <title region="3459,793,3585,1125" id="1">
   <title>殷太公行略<reverse variant="1" type="reverse"/></title>
   <author>殷天正</author>
   <dynasty>清</dynasty>
   <keyword>傳記</keyword>
   <keyword2>殷太公 曹氏</keyword2>
   </title>
 </titletag>
</tag>
```

The describing scope of the people index includes the pedigree tree, the pedigree chart, the person recorded in the pedigree. The content of the index include the people number, the location in the image, the family name, given name, courtesy name, alternative name, posthumous title, gender, date of birth and death, pedigree branches and so on. According to the sequence of the pedigree, the persons of the same generation in the same branch are indexed from the right to the left in order.
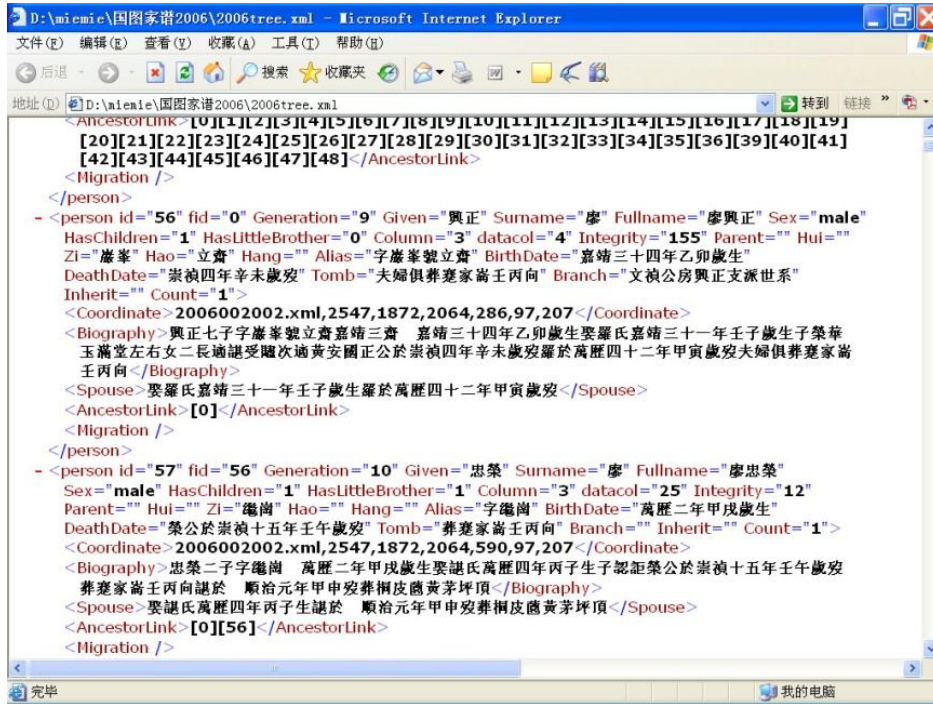
Index to the titles includes all the contents, titles and illustrates, serial number of the contents, the location in the image, title, author, dynasty, keyword(provided by client). Indexing is according to the sequence of the headings in the contents.
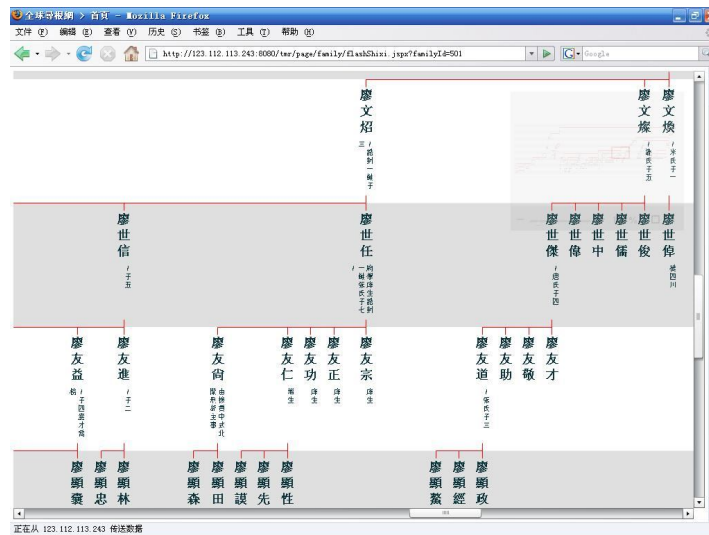
4 TEXTUALIZATION

Textualization is to transfer the words in the images into text with the necessary edition style maintained. Pedigree is the main content of the genealogy including pedigree tree, pedigree chart, and introduction in text and so on. If all the pedigree is converted in text, the personal information can be access to through full text search; however, the relationship between the family members cannot be reflected. Considering the specialty of the genealogy, the pedigree tree and the pedigree chart are converted into family tree structure, and the instructions in words are transferred into full text.

4) Family trees

Every person in the family tree is a node including the general information of the person and the relatives such as father and so on, brothers and so on. At the same time, the portrait, honor, biography and literatures are related on the node. The different nodes can be connected through the lineal relative relationship which leads to a tree structure topology. Every person in the family tree has a unique ID, the personal information such as name, alternative names as well as the relatives' information such as father and son brotherhood will be recorded in the structural data. See illustration 1

Basing on the family tree, the systematic pedigree tree is formed; names with brief introduction are related with lines like the illustration 2. Furthermore the result can be displayed according to the patrons' requirements.

1) Open service platform

On the fundamental of the digitalization achievements of the genealogy, the Global Chinese Roots Net(GCRN http://ouroots.nlc.gov.cn)[7] is set up, that is an open service platform of genealogy provided by NLC. GCRN has been begun in 2008 supported by NLC and the Macau Foundation, providing service online in March, 2011. GCRN is a project in order to conserving the common world civilization memory providing service and education on the digitalization result of the genealogy. A series of service can be accessed to on the GCRN including the catalogue, full text review, analysis of the genealogy, interact of the genealogy compiling, virtual reference, and encyclopedia of the rooting and so on. The first phase of the GCNR has been finished now; the system mode, data mode and the software of the GCNR are all set up. 2,392 titles of genealogies more than 2,570,000 pages can be browsed online; more than 30,000 catalogues of genealogy and over 500 family names can be searched as well. Navigation of the website about genealogy, encyclopedia, and reference are also provided.

The continuing of the GCRN will include: 1) Open cataloguing, the patrons is permitted to upload the catalogue of their private genealogies. 2) Resources sharing space for the patrons. Patrons can upload the images or digitalized documents of the genealogies if they own the copyrights. 3) Tools of compiling genealogy. Using the tools, patrons can compile their own digital genealogies conveniently. This platform will also be facing the libraries, clansmen association and other institutions providing searching links over databases and cooperate with other projects such as Chinese Memory.

GCRN is a project providing a platform for the innovative programs revolving the genealogies collected in NLC including acquisition, cataloguing, reader service, literature reorganizing and relevant research. The value of genealogies should be excavated and promoted basing on the project, convenient and professional service should be provided to both the researchers and the public as well, which will develop the Chinese traditional genealogical culture greatly.

---

[7] Xiao Yu, Discussion on the Collection and Arrangement of Family Tree Literatures under Lib2.0 Environment——Taking the Project of Global Chinese Roots Net as an Example, Sci-Tech Information Development & Economy, vol. 29, 2010.