

中国国家图书馆的家谱收藏及其数字化

Chinese version of the Original Paper: "The Collection and Digitization of the Genealogies in the National Library of China".

Prepared by: 谢冬荣 (Xie Dongrong), National Library of China, Beijing, China.

谢冬荣 (Xie Dongrong)

National Library of China, Beijing, China.

肖禹 (Xiao Yu)

National Library of China, Beijing, China.

石剑 (Shi Jian)

National Library of China, Beijing, China.



This is a Chinese translation of "The collection and Digitization of the Genealogies in the National Library of China" Copyright © 2013 by **Xie Dongrong, Xiao Yu, and Shi Jian**. 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:

[摘要] 本文简要回顾了中国国家图书馆的家谱收藏历史，阐述了国家图书馆现存家谱的收藏概况，分析了馆藏家谱的特点及价值，重点说明中国国家图书馆对馆藏家谱进行了多层次的数字化，并在此基础上提供多种形式的读者服务：（1）编制基于 MARC 的家谱书目数据，揭示家谱的基本信息，用于 OPAC 系统，提供书目数据检索。（2）对家谱原本进行扫描或拍照，获取家谱图像数据，配合家谱书目数据和卷目数据，提供在线图像阅览。（3）制作家谱索引数据，主要是人物索引和篇目索引，揭示家谱中的人物信息和篇目信息，既可以单独使用，也可以配合家谱图像数据提供服务。（4）制作家谱文本数据，提供全文检索。（5）整理家谱谱系关系，制作谱系树数据，提供人物关系检索和定制显示。中国国家图书馆将上述家谱数字资源进行整合，在“中华寻根网”上提供一站式服务。

家有谱，犹如国有史、地有志。家谱与正史、地方志共同构成了中国悠久历史的重要基石。中国的家谱可以追溯到商代的甲骨文。宋代以后，随着私家修谱的兴盛，家谱大量涌现。此种现象一致延续至今，家谱文献的数量蔚为大观。中国国家图书馆始终重视家谱文献的收藏，现已成为华人家谱原件的主要收藏地之一。在提供传统阅览服务的基础上，国家图书馆进行了馆藏家谱的数字化，并将数字化后的家谱影像发布在

互联网上，免费提供读者阅览。本文拟简要回顾中国国家图书馆的家谱收藏历史、收藏概况

¹，重点阐述对馆藏家谱所开展的多层次数字化，并在此基础上提供多种形式的读者服务。

一、收藏历史

著名学者梁启超（1873-1929）比较早地意识到家谱的文献价值。他在《中国近三百年学术史》中专门谈及家谱，认为“我国乡乡家家皆有谱，实可谓史界瑰宝”，并希望“将来有国立大图书馆，能尽集天下之家谱，俾学者分科研究，实不朽之盛业也。”在1925-1927年担任国立京师图书馆馆长（中国国家图书馆的前身）期间，梁启超积极推进家谱的购藏。1928年5月出版的《北京图书馆月刊》第1卷第1号，介绍了建馆至1928年购书情况，其中第七点专门提及谱牒：“搜集谱牒，本馆实为嚆矢。计受赠者十七种，购入者三十四种。此外通函征求，亦不下数百通。惜家藏者碍于旧习，往往不肯捐赠，应者寥寥。而政治纷扰，交通不便，未能作大规模之搜集也。”²随后又在《北京图书馆月刊》第1卷第2号上刊载了《征求家谱启事》³，广泛征集家谱。不过由于战乱缘故，再加上传统观念一般认为家谱应该个人收藏，而且是秘不示人，因而征集来的家谱数量并不多，但是这开创了图书馆公开征集家谱的先河，表明家谱正式成为图书馆的入藏文献。

抗战时期（1937-1945），中国各地都遭遇战火，大量的文献毁于兵燹。当时的北平图书馆（国家图书馆前身）依然注重家谱的收集：在1940年6-9月采访古籍报告情况中说，“此外并购入明刻本家谱多种。”⁴当时中国的首都由南京迁至位于西南部的重庆。为抢救濒临战火的文献，北平图书馆果断决定加强西南文献的收集。⁵家谱即是其中的一个部分。1941年在馆员万斯年采访的西南地方文献中，包括《木氏宦谱》一册、《丽江段氏族谱》一册、《丽江余氏族谱》一册。⁶经过二十多年的征集采访，截止到1950年，据馆编《北京图书馆藏族谱书目》记载，当时收藏的家谱达348种。

二十世纪五十、六十年代，随着观念的改变，众多的家谱一改以往秘藏家中、宗祠中的惯例，流出售卖。国家图书馆此时大量购买家谱，据不完全统计，这二十年购买的家谱数量在1500种左右，约占现藏线装旧谱总数的三分之一。而许多藏书家将自己毕生积累的藏书也纷纷捐赠给国家图书馆，其中就包括不少家谱，如郑振铎的西谛藏书中包括家谱23种。

¹ 此前专门谈论中国国家图书馆家谱收藏的文章有：张志清：《北京图书馆藏中国家谱综述》，载《谱牒学研究》第三辑（书目文献出版社1992年12月版）、杨晏平：《北京图书馆藏普通古籍家谱综述》，载《文献》1999年第2期。

² 李致忠主编：《中国国家图书馆馆史资料长编》，国家图书馆出版社2009年版，第80-81页。下引此书简称为《长编》，并只注明页数。

³ 《长编》，第84页

⁴ 《长编》，第316页

⁵ 张廷银：《收集地方文献须责任与识见并驾而行——抗战时期北平图书馆收集西南文献述论》《国家图书馆学刊》2005年第1期

⁶ 《长编》，第331页

二十世纪八十年代以后，续修家谱在中国大陆蔚然成风。国家图书馆主要采取接收捐赠的办法入藏新修家谱。2012 年以后，鉴于新家谱的缺藏量较大，在接收捐赠的同时，国家图书馆开始主动采购新修家谱。

二十世纪九十年代以后，随着古籍善本引入到拍卖公司，古籍拍卖越来越被人关注。很多藏书家也经常将自己所藏的家谱送到拍卖公司拍卖。通过拍卖竞买家谱成为国家图书馆家谱入藏的重要途径。

二、收藏概况

截止到 2012 年 12 月底，国家图书馆馆藏家谱达 7678 种，其中善本家谱 458 种、普通线装家谱 4680 种、新修家谱 2540 种，总共约 5 万多册。家谱原件收藏的总量仅次于上海图书馆，位居第二。国家图书馆收藏的家谱就所使用的文字而言，以汉文为主，兼及满文、蒙文、彝文、藏文等；就版本类型而言，以活字本为主，兼及刻本和抄稿本；就所属国别而言，以中国家谱为主，还有日本、韩国、新加坡等国家的家谱。此外，中国国家图书馆的家谱还有以下几个特点：

第一，早期版本的家谱较多。国家图书馆收藏有众多的早期家谱，宋代写本 2 种，蒙古刻本 1 种，元代抄本、刻本各 1 种，明代版本 300 多种。如此大量的珍贵版本家谱收藏，在各家谱收藏单位中实属罕见。其中，宋（960-1279）写本《仙源类谱》、《宗藩庆系录》为现存最早的宋皇室家谱，系海内外孤本（《仙源类谱》上海图书馆仅藏有部分残叶），又是现存较早的以纸质为载体的中国家谱。

第二，名人家谱较多。自清代以来，众多政治、经济、军事、学术文化等方面的名人的家谱，在国家图书馆都有收藏，清代的有传奇剧本《桃花扇》的作者孔尚任、主持编纂《四库全书》的纪昀、嘉道时期著名的思想家魏源、晚清重臣李鸿章、刘坤一等、刊刻《古逸丛书》的驻日公使黎庶昌等等，民国时期的名人有赵尔巽、徐世昌、缪荃孙、叶德辉、鲁迅、黄兴等，建国以后的名人有毛泽东、荣毅仁、董必武、徐向前等等。

第三，接续久远的家谱较多。家谱修成之后，随着人口的繁衍、增加，每隔一段时间，短则二三十年，长则六七十年甚至上百年，都会重修，以接续前谱。通过这种多次重修、接续久远的家谱，我们可以了解数十年甚至上百年的家族历史演变过程，因而无疑具有重要的文献价值。此类家谱有：江苏无锡《前洞浦氏宗谱》，自清乾隆直至民国先后四次修纂，有乾隆十三年（1748）刻本、嘉庆二十五年（1820）年刻本、同治十年（1871）刻本和民国二十年（1931）木活字本；江苏无锡《荣氏宗谱》，自清嘉庆直至民国也先后四次修纂，有清嘉庆十五年（1810）木活字本、清同治十一年（1872）木活字本、清宣统二年（1910）木活字本以及民国二十四年（1935）木活字本等。

三、家谱数字化

近年来，中国的家谱数字化发展很快，也取得了一定成绩。不同于其他文献，家谱数字化的主体除了图书馆、档案馆等文献收藏机构和文献数字化公司以外，还有大量的宗亲会组织和个人。家谱数字化的形式也多种多样，如目录、图像、文本、百科词条

等等。目前中国影响较大的家谱数字化项目有上海图书馆的家谱数据库、浙江图书馆的家谱总目提要数据库和家谱全文数据库、台北汉学研究中心的台湾地区家谱联合目录数据库、北京爱如生数字化技术研究中心开发的历代谱牒库等。

国家图书馆的家谱数字化经历了从目录索引到图像数据库的过程。从 2000 年开始，中国国家图书馆的家谱目录从卡片格式向机读格式转变，读者即使身处千里之外，也可以进入国图网站，在线查询家谱目录，免去了奔波之苦。经过多年的探索与实践，我们认为以文献揭示为核心的家谱数字资源组织很难满足读者的需求，家谱的文献特性和使用方式决定了家谱数字化的特殊性。

1、编制机读目录

目录数据是家谱文献收集、整理、揭示的基础。国家图书馆编制的家谱目录数据采用 CNMARC 格式，依据家谱文献的特点，主要描述家谱的外在特征和内容特征，包括题名、著者、版本、谱籍地、郡望、堂号、始祖、始迁祖等信息。以《新安苏氏族谱》为例，该书的 MARC 格式数据是：

```
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

目前，中国国家图书馆已经完成全部馆藏家谱书目数据的制作，并在国图网站上发布（<http://opac.nlc.gov.cn>），为读者提供在线书目数据检索。

2、图像化

从 2008 年开始，国家图书馆的家谱开始全本图像化扫描。具体的技术标准是：采用彩色扫描方式，扫描分辨率为 300dpi，颜色模式为 RGB 模式，位深为 24 位真彩色，存储为 TIFF、JP2 和 PDF 格式，压缩因子为 150。要求扫描图像真实反映原书，以原书的上边沿为基准，以中缝折线为中心线，保持原书的天头、地脚的尺寸不变，左右两边的尺寸基本不变，扫描后的影像文件要叶码连续，没有重叶、缺叶，错叶、折叶等情况（原书缺叶、错叶除外）。补扫缺叶图像要与正本图像大小一致，颜色接近。原书有大面积的残缺时，要酌情添加衬纸（与原书的颜色接近），避免透光、透字。

扫描后进行必要的去图像黑边、图像拼接等处理，图像拼接后不得有明显的拼接痕迹。图像偏斜不超过 1°，扫描后每叶影像尺寸大小与原书一致，误差小于 1%。

现在国家图书馆已经完成 2392 种家谱的扫描工作，共扫描 257 万拍。这些图像全部在网上免费发布，供读者在线浏览。

3、编制索引

依据用户的需求和家谱文献的特性，国家图书馆编制了馆藏家谱的人物索引和篇目索引。索引数据要求完整、准确，文字错误率低于 0.3%，支持 unicode5.0 的字符集，支持检索和显示。例如：

```
<?xml version="1.0" encoding="UTF-16" standalone="true"?>
<tag source_id="SZJP00005-10002410-00001.jpg">
  <persontag>
    <person sex="男" surname="吴" region="2694,633,2940,1788" id="1">
      <firstname>八府君</firstname>
      <hui>疎</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>
```

人物索引的范围包括家谱谱系图、谱系图表、谱系描述文字中的人物，人物索引包括人物编号、在图像中的位置、姓、名、讳、字、号、别称、谥号、性别、生、卒、世代、支派等字段。按照支派、世代的先后顺序，同一支派的同一世代按照从右向左的顺序依次索引。

篇目索引的范围包括家谱中的全部篇目、标题、插图等，篇目索引包括篇目编号、在图像中的位置、题名、著者、著者朝代、关键词（关键词表由甲方提供）等字段。按照篇目标目出现的顺序依次索引。

4、文本化

文本化是将家谱图像中的文字转换成文本数据，并保留必要的版式信息。谱系是家谱文献的主要内容，通常采用谱系图、谱系表、描述文字等形式。如果对谱系部分进行文本化，只能借助全文检索来查找人物信息，不能体现家族内部人物之间的关系。因此，我们对家谱文献的非谱系部分采用全文版式还原方式进行数字化，与其他文献相同；谱系部分（包括谱系图、谱系表等）都按照“谱系树”方式进行文本化。

5、谱系树

谱系树是将每个人物作为一个节点，每个节点包含该人物的主要信息和父子、兄弟等亲属关系，同时将像赞、荣恩、家传、艺文等与该人物进行关联。依据每个节点的直系亲属关系可以将各个节点串联起来，形成树状拓扑结构。“谱系树”中的每一个人具有唯一的 ID 号，以结构化数据的方式记录每一个人物的姓、名、字、号等信息和父子、兄弟等关系。如图 1 所示。

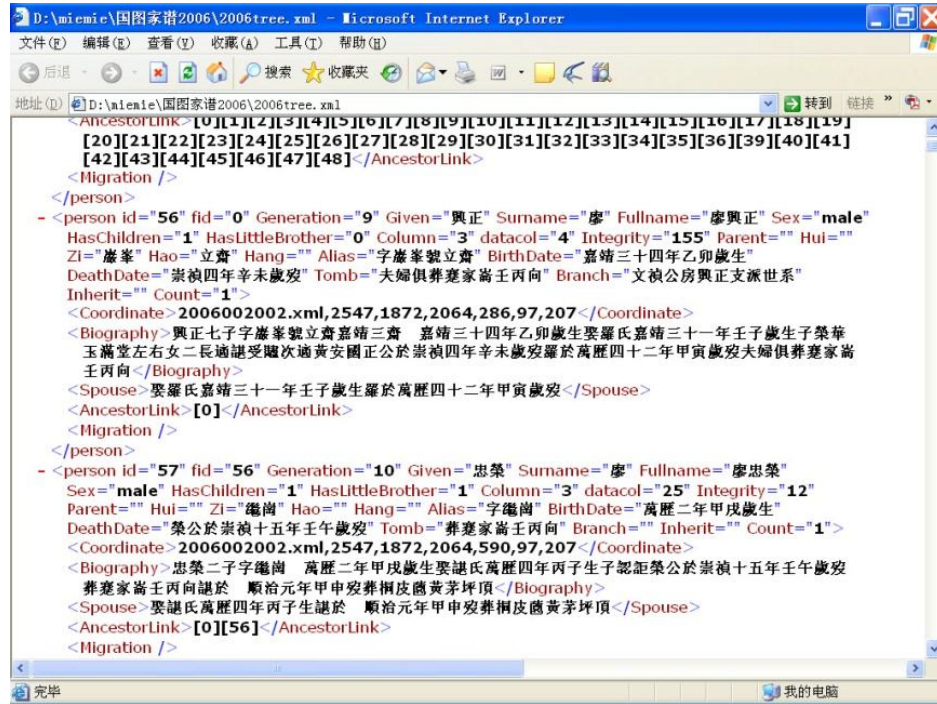


图 1

以“谱系树”为基础，系统可以生成比较规整的谱系图，用文字来描述人物的信息，用线段来表示人物之间的关系，如图 2 所示。同时依据用户的需求，可以采用多种方式显示“谱系树”。

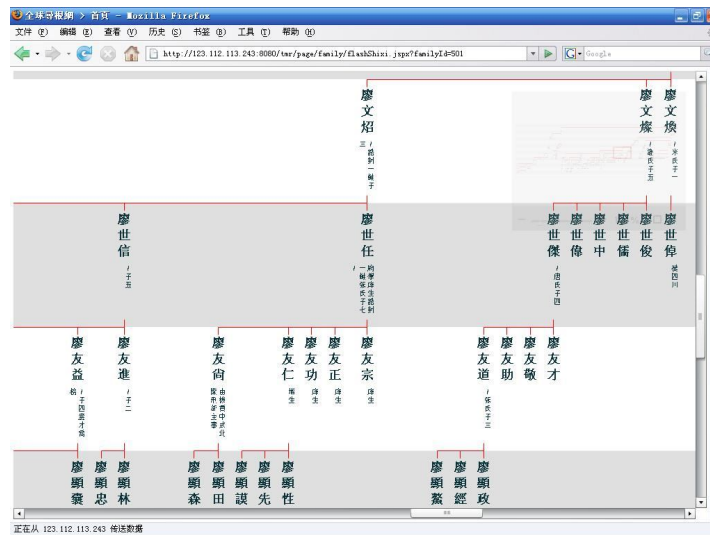


图 2

6、开放式服务平台

在家谱文献数字化的基础上，国家图书馆建立了开放式家谱服务平台“中华寻根网”（以下简称寻根网，<http://ourroots.nlc.gov.cn/>）。⁷寻根网项目始于 2008 年，系中国国家图书馆与澳门基金会合作的文化项目，2011 年 3 月上线提供服务。寻根网在互联网

⁷ 关于“中华寻根网”可以参见尚禹《Lib2.0 环境下谱牒文献的收集与整理——以全球中华寻根网项目为例》（《科技情报开发与经济》2010 年第 29 期）一文。

上提供中华家谱目录检索、家谱全文阅览、家谱谱系分析、家谱编纂互动、网上参考咨询、寻根百科等服务。该项目将分期建设，目前已完成项目的第一期，建立了寻根网系统模型、数据模型和软件系统，提供国图藏 2392 种 257 万多叶家谱文献的在线浏览，提供国内外 3 万余条家谱书目数据检索，还提供 500 余条姓氏数据查询，同时还为读者提供家谱网站导航、家谱知识百科和咨询服务。

寻根网的后续建设计划还包括：（1）开放式目录，允许用户上传自己所藏家谱的书目信息；（2）用户资源共享空间，用户可以将自己拥有版本的家谱数字影像或数字化家谱文件上传，与他人共享；（3）家谱编纂工具，用户利用此工具可以较快地编纂属于自己家族的数字化家谱。此平台还将向图书馆、宗亲会等机构开放，提供资源的跨库检索连接，并与“中国记忆”等项目展开合作。

以“中华寻根网”为平台，中国国家图书馆的家谱采访与编目、读者服务、整理与研究等方面的工作将开拓新的局面，并依此充分挖掘与发挥家谱的价值，为普通民众和研究人员提供便捷、专业的服务，从而进一步弘扬家谱这一中国悠久的传统文化。