

# 民国时期文献目录数据平台的构建与实践<sup>\*</sup>

蔡迎春

**摘 要** 当前民国时期文献目录编制及数据库建设方兴未艾,但存在对建国后整理出版的“新版民国时期文献”目录编制关注不够、目录库建设缺失等问题,针对这些问题进行研发的“民国时期文献目录数据平台”即应运而生。文章介绍“数据平台”的结构框架、元数据构成以及功能实现,构建了一个既包括原版民国时期文献目录,又包括新版民国时期文献书目和新版民国时期文献子目的目录型“数据平台”。同时,探讨名称规范档、GIS技术和可视化技术在“数据平台”中的应用与实践,将传统数据库检索结果的平面式输出,转化为立体化的全方位时空呈现,将有助于文献研究、出版研究以及其他学科研究的开展。

**关键词** 数字人文 民国时期文献 民国目录 平台构建

**引用本文格式** 蔡迎春.民国时期文献目录数据平台的构建与实践[J].图书馆论坛,2019(3):121-127.

## Construction of the Bibliographic Platform of the Republic of China Documents

CAI Yingchun

**Abstract** With great importance attached to the bibliographies of the Republic of China documents and the construction of related databases, little attention has been paid to those newly compiled editions of the Republic of China documents after the foundation of the People's Republic of China. Therefore, the Bibliographic Platform of the Republic of China Documents is desired, in which bibliographies of both original and newly compiled editions of the Republic of China documents are included. This paper gives an introduction to the Platform's structural framework, metadata composition and function realization. Meanwhile, it explores the application of name specification, GIS and visualization in the construction of the Platform, and transforms the traditional planar output of retrieval results into a three-dimensional output, aiming to promote academic research in disciplines of literature, publishing and so on.

**Keywords** Digital Humanities; the Republic of China documents; the Republic of China bibliographies; platform construction

## 0 引言

民国时期文献,是指1911年辛亥革命至1949年中华人民共和国成立这一特定历史时期

的各种文献,为便于研究和论述,本文称之为“原版民国时期文献”。相对而言,本文将1949年之后汇编、再版或影印出版的民国时期文献,包括图书、期刊(含报纸)和档案等,称为“新版

<sup>\*</sup> 本文系国家社会科学基金项目“新版民国时期文献总目编撰(1949-)”(项目编号:17BTQ081)的阶段性研究成果。

民国时期文献”。

目前,针对“原版民国时期文献”开发的数据库产品比较多,如上海图书馆“民国时期期刊全文数据库”、尚品大成“大成老旧刊全文数据库”、爱如生“中国近代报刊库”、青苹果“华文报刊文献数据库”和“维库民国电子资源数据库”、国家图书馆出版社“民国时期文献总库”以及CADAL“民国时期文献大全”等<sup>[1]</sup>。虽然大部分图书馆的OPAC系统可以直接检索到馆藏民国时期书刊,但一些颇具影响力的大型综合性目录,如《民国时期总书目》,以及数量众多的专科目录,却很少被纳入相关数据库或检索系统,不能不说是一件遗憾的事情。目前已有的140余种专题目录中,仅有郑阿财等主编的《敦煌学研究论著目录(1908-1997)》正在制作书目数据库,其他目录仅以文本形式编印或出版,尚未进行数字化处理。特别地,针对“新版民国时期文献”而言,目前还没有一个全面覆盖1949年后“新版民国时期文献”所包含图书、报刊和档案等内容的目录数据库,以至于“新版民国时期文献”尚没有系统的检索途径可循,势必对民国时期文献的深度研究造成不便。

民国时期文献数据库所包含的文献应该可以通过数据库本身提供的检索途径进行查询,但是已出版纸质文献的子目,因没有整理和编目,没有有效的路径可供检索,无法实现其编纂出版的价值。此外,已建成的“原版民国时期文献”数据库,主要以题名、责任者、关键词等简单检索和浏览为主,缺少统计、分析等功能,基本上仅对文献形式特征和少量内容特征进行标引,还处于比较原始的纸质替代状态。民国时期文献的开发研究仍采用传统的研究方法与模式,缺乏创新研究范式,导致当前规模庞大的民国时期文献数据与较低的文献深度利用率之间的矛盾比较突出<sup>[2]</sup>。更为重要的是,“新版民国时期文献”与“原版民国时期文献”之间的关联尚未被有效地揭示出来,众多出版编撰者无法分析掌握其出版动态。

为充分揭示民国时期文献及整理成果的子目内容和关联,“民国时期文献目录数据平台”(以下

简称“数据平台”)正在建设中,通过共享上海图书馆“人名规范数据集”,进一步扩展思路,建立“名称规范库”,从而突破原有数据库建设理念,提供海量和规范的数字化信息和数据,从不同角度实现对文本的分析统计功能,揭示信息和数据之间的关联性。同时,还将GIS技术以及可视化技术应用于“数据平台”建设中,不仅使民国时期文献的书目索引编制更加完整,而且将传统数据库检索结果的平面式输出,转化为立体化的全方位时空呈现,将大大方便相关文献研究、出版研究以及其他学科研究的开展。

## 1 民国时期文献目录数据平台构建

### 1.1 数据来源

“数据平台”除收入“原版民国时期文献”目录外,也收录自1949年以来至今在中国内地、港台地区和国外出版的民国时期文献,包括公开出版物、非公开出版物和一些综合性丛书。文献类型主要是图书、期刊、报纸和档案等。目前,借助各大图书馆的馆藏数据、各民国时期文献主要出版机构的出版目录、全国新书目、豆瓣网站和CALIS联合目录等,本项目已经采集了“新版民国时期文献”数据近1000种,数据也涉及了中国港台地区、美国、日本和欧洲等各大公共图书馆和高校图书馆联盟联合目录,尽量较全面和系统地收集建国后出版的民国时期文献成果。

### 1.2 平台构建

#### 1.2.1 设计思路

“数据平台”的主要目的是全面揭示“原版民国时期文献”和“新版民国时期文献”的全部子目信息。设计方法主要是依据《民国时期总书目》和《(1833-1949)全国中文期刊联合目录》及其补编本的分类排序规则,同时参照目前国家图书馆正在编纂的《民国时期文献总目(图书卷)》进行分类、标引与著录,使“新版民国时期文献”目录与《民国时期总书目》保持基本一致的体系和详尽的著录内容。由于民国时期档案存量非常大,收集、著录极为困难,故暂时未纳入平台设计中。

在平台的功能设计时课题组特别关注了如下几点:

(1)数据的后续更新和维护。不仅录入数据,便于查重、自动排序和索引编制,而且还可以随时补充新发现的相关有价值的信息;

(2)具有数据统计和分析功能。可按时间、人物、地点及出版机构等进行关联分析并以可视化的方式呈现,便于快速从海量数据中发现新的知识,发现事件脉络、人物关系及出版轨迹等。

### 1.2.2 结构框架

“数据平台”主要由“民国时期期刊目录库(1911-1949)”、“民国时期图书目录库(1911-1949)”和“新版民国时期文献子目库(1949-)”三个库组成,并且相互关联。平台先期录入《民国时期总书目》和《(1833-1949)全国中文期刊联合目录》作为“民国时期图书目录库”和“民国时期期刊目录库”的基础数据。然后,再通过检索基础数据,补充著录“新版民国时期文献”的书目信息,形成“新版民国时期文献子目库”。具体见图1。

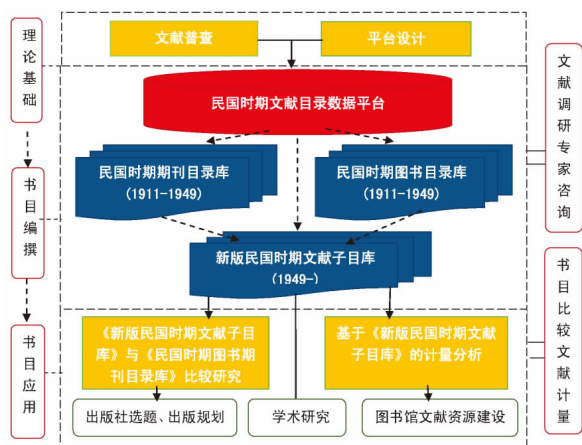


图1 民国时期文献目录数据平台的结构框架

此数据平台结构框架具有如下特点:

(1)实现“新版民国时期文献”书目数字化。“数据平台”的建设,有效扩展了民国时期文献书目整理的时间外延,全面普查1949年后民国时期文献整理出版成果,揭示已整理出版的民国时期文献中所有子目,弥补解放后民国时期文献的整理出版无书目可查的缺憾,并具有可持续性。

(2)发现文献整理的趋势与轨迹。将人文研究领域相关技术和成果应用到“数据平台”建设中,可以探究其出版特征和出版规律,揭示重复出版、资源收集不全等问题,考察现有民国时期文献整理出版的缺漏,特别是对较有价值的文献而又尚未整理出版的状况进行分析,为出版机构进行出版选题和规划提供参考。

(3)具有文献计量分析功能。对“新版民国时期文献”的图书、期刊(报纸)的出版机构、作者、出版时间等进行文献计量分析,便于分析研究其分布特征,为图书馆资源建设提供参考。

## 2 “数据平台”技术应用与实践

### 2.1 底层数据集设计及元数据的选取

“数据平台”的一个重要创新就是可以揭示“新版民国时期文献”中的子目。由于新版子目中的某些元数据与“原版民国时期文献”是相同的,例如题名、责任者、主题、摘要、总目分类、文献类型等标示文献的基本元数据,在原版文献揭示和新版子目揭示中都是一致的,为了避免重复录入,平台在底层数据集设计时,将这部分共同的元数据单独建库,称之为“基本数据集”。因此,在录入原版文献目录时,后台操作需要先后完成两个书目库的字段著录,先在“基本数据集”著录题名、作者、摘要等基础字段,然后再在“原版数据集”追加出版社、出版时间等与版本相关的其他元数据。如果版本较多,就追加多条原版信息。同样,在著录新版子目时,也是先检索“基本数据集”,并先与该库建立关联,再在“新版子目数据集”追加著录新版的书名、出版者、出版日期等相关元数据。

另外,由于新版子目著录可能会涉及到一些相同的著录项,如大套影印版丛书的新版书名、新版责任者等,为避免重复录入,数据平台建立了一个“新版数据集库”。录入时,先对包含子目的大套影印版丛书进行集中著录,当录入子目元数据的时候对这部分内容就可以直接检索并建立关联。

因此,“数据平台”的底层数据集由基本数据集、原版数据集、新版数据集和新版子目数据集四个子库构成。其结构设计如图 2 所示。

在数据平台的元数据构成上,四个底层数据集均有各自特有的元数据元素,而对于四个子库所涉及到的共同元数据则通过相互的“链接”进行共享,具体见表 1。

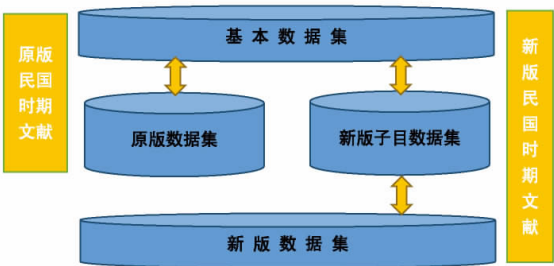


图 2 民国时期文献目录数据平台的底层数据集及关联

表 1 民国时期文献目录数据平台的底层数据集元数据及链接

名称	基本数据集 (基本元数据)	原版数据集(原版元数据)	新版知识库(新版元数据)	新版子目数据集 (新版子目元数据)
题名	原版书(刊)名	链接“基本元数据”	书名	链接“基本元数据”、“新版元数据”
责任者	原版责任者	链接“基本元数据”	新版责任者	链接“基本元数据”、“新版元数据”
分类/主题	分类、主题	链接“基本元数据”		链接“基本元数据”
内容摘要	内容简介(摘要)	链接“基本元数据”	内容简介	链接“基本元数据”
文献类型	图书、期刊、档案等	链接“基本元数据”		链接“基本元数据”
版本		出版地、出版者、出版日期	出版地、出版者、出版日期	链接“新版元数据”
丛书项		丛书名、丛书责任者	丛书名、丛书责任者	链接“新版元数据”
形态附注		卷册、页码、开本	卷册、页码、开本	
馆藏信息/文献来源		馆藏地		来源文献(第几册,第几页)
其他		总目号	ISBN	

“基本数据集”元数据的构成要素是同一种书在不同时期、不同出版机构的版本都相同的基础字段数据,包括原版书(刊)名、原版责任者、分类、主题、内容简介(摘要)、文献类型等。在“基本数据集”著录完成后,通过它提供的链接界面,添加原版或新的版本,从而形成完整的“原版数据集”和“新版数据集”。“基本数据集”的设计及其以此为基础的原版和新版链接著录方式,可以避免不同版本间相同字段的重复录入,使同一种书的不同版本通过链接实现相互关联。

“原版数据集”元数据除链接“基本数据集”的题名、责任者、摘要等外,还包括新版的版本项、丛书项、形态项及馆藏信息、总目号等与其他版本有所区别的字段信息。

“新版数据集”的元数据构建,由新版单行本或大型丛书、汇集本的书名、责任者、出版地、出版者、出版日期、摘要、丛书名、丛书责任者、卷册、页码、开本、ISBN 组成。

“新版子目数据集”除“文献来源”元数据

外,其余均是通过链接“基本数据集”以及“新版数据集”的相关元数据完成。以链接方式完成的“新版数据集”元数据构成,在避免相同字段数据重复录入方面功效显著。比如,国家图书馆出版社 2015 年出版的《民国文献类编》收录民国时期文献 4000 余种,如果没有实现元数据间的直接链接,那么相关信息就得重复录入 4000 次,而建立链接关系后,只需勾选,并进行点击确认,来源文献的所有信息就会自动添加到“新版数据集”的相关字段位置上。

2.2 规范档的建立及名称规范数据集的应用

为了实现“数据平台”的统计分析功能,除元数据选取尽可能完备之外,在平台数据录入时,各种规范档的建立就显得尤其重要。除了常规的主题和文献类型需规范外,民国时期文献的出版具有一定的特殊性,责任者、出版机构名称的变更比较普遍。此外,民国期刊多有停刊、复刊、出版周期不固定等现象。因此,在“数据平台”的建设中,名称规范档的建立就显得非常必



要, 主要包括人名规范档和出版机构规范档等。建立名称规范档的目的就是把同一名称的所有文献都集中在该名称词条下。例如, 同一作者可能存在多个笔名, 以茅盾为例, 其笔名多达一百多个, 只要建立一个规范档, 就能将茅盾以不同笔名的著作都集中在茅盾的词条下。

目前在互联网环境下的人文研究领域, 名称规范档的概念和图书馆传统意义上的规范控制工作迥异。2008 年国际图联发布主题规范的推荐意见时, 首次提出“人作为一个实体”的概念。把人当作实体之后, 人就不仅仅是一个名称, 而是囊括出生年月、与其他人物之间的关系、生平大事、任职经历等诸多内容。要将同一人的很多信息集中在一起, 就是要对这个人进行唯一的、可被机器读取的标志符的设置。有了唯一的标志符, 就可以用该标识符来代替这个人。唯一标识符用 URI(统一资源标识符)表示, 它在互联网上是唯一的, 也是唯一的定位符。“数据平台”通过互联网被标识、被定位、被访问, 对平台中的人名、出版机构等建立相关名称规范档, 如人名规范数据集、机构名称规范数据集等, 而每一个规范名称都有 URI, 并且这些规范档是开放的。因此, 在“数据平台”相关元数据著录时, 对于已标识的同一人或物的不同名称就可以直接选用规范名称的 URI。对于没有标识的名称, 可以通过开放的规范档数据入口进行规范标识后再选取。

目前业界在人文研究领域已有开放的、较成熟的规范数据集, 如上海图书馆的“人名规范数据集”, 本项目与其合作, 共享人名规范数据, 弥补了“数据平台”建设中的人力、物力和技术实现上的不足。另外, 数据平台借助上海图书馆成功案例和技术力量, 也可以实现对出版机构等其他相关规范档的建立。本项目在“数据平台”数据录入时, 对于文献作者、题名、摘要中的人名, 直接调用了上海图书馆的开放数据接口, 或进入上海图书馆人名规范库获取其 URI, 然后进行著录。而对于有多种笔名或别称的作者, 其人名 URI 值都是相同的, 比如冰心, 原名谢婉莹, 笔名冰心女士、男士、素人, 所有这些名字的

URI 赋值都是一致的, 均为 [http://data.library.sh.cn/entity/person/05ebng66w4\\_qjnkkg](http://data.library.sh.cn/entity/person/05ebng66w4_qjnkkg)。因此, 无论文献的署名是笔名、原名还是别名, 因为具有相同的 URI 值, 只要输入其中一个名字, 就可获取该作者的不同署名的所有文献。同样, 对于同名作者, 因为 URI 值不同, 也具有了明确的区分度, 他们的作品会被归入各自名下, 不会被混淆同时检出, 让使用者难以判断<sup>[3]</sup>。

另外, 通过名称规范数据集的建立, 也可以揭示“数据平台”中人物之间的学术关系, 为民國年间彼此有过学术交往的人物建立关联。例如, 在获取这一机器可识别的 URI 值, 放入“数据平台”的相应字段后, 通过系统后台的算法和识别, 即可共享“上海图书馆人名规范库”中提供的包括笔名、别称、职衔、籍贯、作品、生平事迹等内容在内的人物简介<sup>[4]</sup>。同时, 通过对作者相关著作的合作者、编校者, 序言、弁言、书评等的撰写者乃至书名、题字等的题写者等进行关联描述, 对作者间的学术合作关系进行选择、分类及分析, 根据学术合作程度和合作方式, 列出不同关系的类型属性, 建立学者人物关系、学术联系图谱。用户可以选择与自己研究相关或感兴趣的人物进入关系分析页面, 查看人物简介、与该人物有关联的所有其他人物以及他们之间的关联关系, 自主选择建立进一步的多层级人物关系, 根据这些关联关系, 利用“数据平台”提供的可视化工具, 绘制人物学术关系图谱<sup>[5]</sup>。

### 2.3 GIS 技术和可视化技术的应用

GIS 技术在“数据平台”的应用主要通过“中国历史地理信息系统”(CHGIS)建立地理信息关联, 将地图的视觉化效果、地理分析功能与“数据平台”中的地名信息相结合, 实现时间和空间两方面的直观检索, 提供文献的出版地分析、作者的地域分布分析、出版的时空变迁分析等<sup>[6]</sup>。其通过地图直观显示文献的出版数据, 可按时间先后顺序自动生成地域出版文献数量、出版机构分布、出版机构迁徙流动路线图等, 实现检索结果、分析结果的电子地图呈现<sup>[7]</sup>。例如, 了解民国时期每个阶段哪些出版机构集中在哪一

地域,或某一地域的某一时间阶段的文献出版数量、整个民国时期出版机构的迁徙集散情况,及民国时期的出版机构、出版事业的变迁和发展;新版文献地域分布分析,以及反映1949年后民国时期文献再版整理地域重心的变化情况;教材的出版发行地域分析,以及反映抗战前后以及抗战期间国统区、日据区和解放区教材在出版数量和内容上所呈现出的特色;地方志、游记、寺庙志、校史、图书馆史,以及各种社会调查材料等与地名相关的文献的地图呈现和地理数据分析。可以说,GIS技术在上述研究和分析中的应用,对民国时期的出版史研究、文献版本研究乃至各个学科的专题研究,都能提供较为直观的分析结果。一方面可以将传统数据库的检索结果,变成用户可以开展自助分析的基础数据,另一方面把传统数据库检索结果的平面式输出,转化为立体化的全方位时空呈现。

### 3 “数据平台”功能的实现

可以说,“数据平台”的建设以及人文领域相关技术的应用,使“数据平台”除常规的存储与检索外,还具备分析数据、串联知识、发现问题的功能<sup>[9]</sup>。在提供分析统计功能和可视化结果呈现的同时,可以有效地辅助文献研究、出版研究以及通过目录进行的地方志、教材、文学、宗教、历史等学科的研究。

#### 3.1 完整地呈现版本源流

通过“基本数据集”添加“原版信息”和“新版子目信息”的方式,可以使同一种书的不同版本的情况都能完整地体现,清晰呈现同一种书的版本源流。例如,通过书名或作者在“基本数据集”检索陈大齐编著的《哲学概论》一书,即可检测到民国时期该书共有三个版本;再如,解放社编《社会发展简史》一书,可检索到七个版本。类似案例,不胜枚举。

#### 3.2 强大的文本分析功能

虽然“数据平台”只是目录数据库而非全文数据库,但是通过名称规范数据集的建立以及GIS技术等的应用,通过基本著录信息、出版信

息以及内容提要等元数据的提取以及各个子库之间多重关联关系的建立,可以开展基于数据挖掘的文本分析。

“数据平台”规范的元数据包括文献目录(含内容提要)中的人名、地名、学科主题和出版机构等信息,用户可以利用平台提供的检索和分析功能,根据检索到的文献进行文献主题分布、出版地分布、作者分布、年代分布等分析,并且可以利用这些信息进行组配式的关联分析。比如通过对著作的主题、出版时间分析,揭示某一学术领域的研究或某一学术流派在民国时期的学术史;通过对作者及其著作出版时间的分析,可以勾勒出作者的学术轨迹和学术生平;通过主题与出版时间的分析,可以发现哪些学科的研究在民国时期受到重视,哪些学科关注较少,与此同时在建国后,哪些学科的文獻开发整理程度较高,哪些在民国时期相关研究文献较多的学科未受到足够的重视,这些为民国时期文献整理机构提供有价值和有说服力的数据分析支持;通过主题与作者分析,可以揭示某一学科领域的作者群,尤其对于一些新兴学科,可以分析出在学科发展之初,哪些学科学者最先介入或推动了学科的发展。同样,平台数据还可以用于民国时期电影、戏剧、海派绘画等发端于晚清或民国时期的行业发展史和学术史的研究与分析。这些统计和分析,有助于帮助学者发现新资料,开拓前人未曾关注的新领域,拓展新的研究视野。

应用可视化工具,还可以实现文本的可视化分析。以民国时期的敦煌学研究为例,通过学科主题结合时间范围的分析,可绘制出敦煌学在1908-1949年间学术论著发表数量柱状图,以此体现敦煌学研究的发展情况。通过发文量的可视化呈现,可以非常清晰地看出,敦煌学研究从民国初年(1911-1920)的18种,到民国末期(1940-1949)增长到317种,呈现从起步到越来越受到关注,研究性论著逐年增长的趋势。也可以明显地看出,即使民国时期发表论著是最多的一个历史阶段,但10年间的论著总发文量也仅有300余种,每年平均仅30种,这些证据表明

了整个民国时期的敦煌学研究的确只处于发端期。

3.3 揭示各版本数据的关联性

根据“新版子目数据集”与“新版数据集”形成的链接关系，可以很容易地判断民国时期文献的整理出版情况。通过“新版子目”链接，可以获知该书已在1949年后有过新版本出版，再通过数据平台提供的“文献来源”信息，可以准确了解该书的出版信息，详细知道文献被收录在哪家出版社的第几卷第几页。

影印版信息，共 1 条。 显示 隐藏		原版信息，共 1 条。 显示 隐藏	
修改		修改	
流水号	MG098316	流水号	MG098316
原版书名	中国红十字会战地写真	原版书名	中国红十字会战地写真
原版责任者	沈敦和编	原版责任者	沈敦和编
影印书名	民国文献类编	出版地	上海
责任者	民国时期文献保护中心，中国社	出版者	中国红十字会
出版地	北京	出版日期	1911年出版
出版者	国家图书馆出版社	页码	70页
出版日期	2015	开本	24开
ISBN号	978-7-5013-5561-7		
丛书名	/		
丛书责任者	/		
卷册	第六册		
页码	309页 -		
开本	22cm		
版本附注	影印：韩永进，王建朗主编；		

图3 《中国红十字会战地写真》原版、新版信息著录细目

例如，通过“基本数据集”查到沈敦和编著的《中国红十字会战地写真》，可以看到该条目下有“原版信息1条”“影印版信息1条”的提示(图3)。通过点击“显示”，可查到民国时期的原版为1911年由红十字会出版，而建国后的影印版收录在国家图书馆出版社2015年出版的

《民国文献类编》第六册里。当然，某种书反复多次重新出版，或从未被整理出版过，也可通过平台数据检索而得到清晰的展示，为出版机构和文献机构后续的文献整理提供了较可靠的依据。

4 结语

民国时期文献目录数据平台的建设，以及数字人文领域相关技术在民国时期文献整理与出版领域的应用实践，可以说是为传统意义上的民国

时期文献研究注入了新的活力、提供了新的视角。虽然，数字人文相关技术的应用，使得“数据平台”更有助于民国时期文献研究领域的学者、出版者，以及图书馆的研究和决策提供多方面的依据和路径，但是，“数据平台”在建设过程中还有诸多问题需要进一步论证和深入探讨，例如，除人名规范之外的其他名称规范档应该如何选取、如何合作，是否考虑众

筹方式，依靠专业团队技术优势使得平台的功能更加完善，以及在目录数据库的基础上扩展数据内容，最终增加全文等。希望通过本项目的实战，能够抛砖引玉，进一步推进民国时期文献数字化开发的深度揭示，以充分挖掘民国时期文献的利用价值。

参考文献

[1] 段晓林. 民国时期文献数据库开发现状研究[J]. 图书馆学研究, 2016 (20): 42-45.

[2] 欧阳剑. 面向数字人文研究的大规模古籍文本可视化分析与挖掘[J]. 中国图书馆学报, 2016 (2): 66-80.

[3] 夏翠娟, 张磊. 关联数据在家谱数字人文服务中的应用[J]. 图书馆杂志, 2016 (10): 26-34.

[4] 夏翠娟, 刘炜, 陈涛, 等. 家谱关联数据服务平台的开发实践[J]. 中国图书馆学报, 2016 (3): 27-38.

[5] 崔春, 毕强. 关联数据在数字人文领域中的应用剖析——以关联爵士项目为例[J]. 图书情报工作, 2014 (24): 99-105.

[6] 夏翠娟. 中国历史地理数据在图书馆数字人文项目中的开放应用研究[J]. 中国图书馆学报, 2017 (2): 40-53.

[7] 范佳. “数字人文”内涵与古籍数字化的深度开发[J]. 图书馆学研究, 2013 (3): 29-32.

[8] 赵思渊. 地方历史文献的数字化、数据化与文本挖掘: 以《中国地方历史文献数据库》为例[J]. 清史研究, 2016 (4): 26-35.

作者简介 蔡迎春, 博士, 副研究馆员, 上海师范大学图书馆副馆长, shtq@shnu.edu.cn。

收稿日期 2018-05-03

(责任编辑: 沈丽霞; 英文编辑: 郑锦怀)