

# 古籍图像数据库建设常见问题及对策研究

——以首都图书馆馆藏古籍珍善本图像数据库建设为例

史丽君

**摘要** 建设珍善本古籍图像数据库是古籍收藏单位开展古籍保护和读者服务的重要方法。本文结合首都图书馆具体工作经验及当前同类数据库存在的普遍问题,论证了建立健全珍善本古籍图像数据库应拟定切实可行的古籍数字化工作方案,结合实际需求选择数字化标准、理顺古籍图像加工流程、加强图像色彩与内容管理、合理配置发布平台功能、完善图书馆古籍数字化等工作。

**关键词** 珍贵古籍;数字化;图像数据库

**分类号** G250.74

**Abstract** The digitization of ancient books is one of the most important methods to protect the rare books. Based on the experiences of Capital Library, and in order to improve the quality of database, the producers need to establish reliable digitization program, select the appropriate digital standards, streamline image processing process, strengthen the management of image content and color, build the publishing platform features rationally and perfect the digitization of ancient books.

**Keywords** Rare book; Digitization; Images database

**Class Number** G250.74

DOI:10.16384/j.cnki.lwas.2016.09.013

建设珍善本古籍图像数据库是保护与传承珍贵古籍的重要手段,通过数字照相或扫描的方式将珍善本古籍存储为图像数据并借助互联网等媒介供大众使用,实现了珍贵古籍历史文物性和学术资料性的重新配置,有效地解决了古籍珍藏保护和开发利用之间的矛盾。2007年,文化部明确提出了“制订古籍数字化标准,规范古籍数字化工作,建立古籍数字资源库”<sup>[1]</sup>的要求,随后,国内图书馆建设了许多各具特色的古籍资源库。首都图书馆也在古籍图像加工和数据库建设方面取得了丰硕成果,在图像加工方面,逐步完成了馆藏孤本、稀见本及入选国家珍贵古籍名录古籍的扫描工作,并为北京市属古籍藏书单位进行了珍贵古籍的扫描加工工作;在专题数据库建设方面,建成了以馆藏古籍为基础的“古籍插图图像数据库”和“古籍珍善本图像数据库”;在本馆古籍数字化工作顺利进行的基础上,首都图书馆还参与了国家古籍保护中心“中华珍贵典籍资源库”项目。

1 当前国内珍贵古籍图像数据库建设存在的常见问题

1.1 古籍珍善本图像数据库建设软硬件门槛高,资金投入大

就目前国内古籍珍善本加工扫描及平台建设而言,对扫描仪等硬件设备、系统搭建等软件建设、技术人员等人力资源需求较高,需要长期和较大的资金投入。如媒体曾经报道苏州图书馆的古籍数字化情况:“从页面扫描、文字录入到校对、生成数据库等,花费工夫非常大。他们曾测算过,三名熟练工作人员忙一整天,也只能录入1万字。平均下来,一页古籍人工成本要6元多。如果算上房租费用、水电费和软件开发费用,那成本就更大了。徐伟国说,当年光是古籍数字化软件开发,就花了30多万元。”<sup>[2]</sup>可以说,硬件、技术、人才、资金的高门槛,直接制约了当前珍贵古籍图像数据库的建设。

1.2 古籍珍善本图像数据库数量相对偏小,读者阅览权限低

国内建设的古籍专题数据库较多,如方志库、家谱库、中医药古籍库、农业古籍库等,然而专门收录珍善本古籍的数据库却所占比例较低。如据文化部统计,全国公共图书馆系统共收藏善本古籍229.5万册件。<sup>[3]</sup>而就笔者粗略统计来看,公共图书馆开发建设的主要收录珍贵古籍的数据库仅有上海图书馆“古籍善本全文光盘”、首都图书馆“馆藏古籍珍善本图像数据库”、长春图书馆“馆藏国家珍贵古籍全文数据库”、广东省立中山图书馆

“特藏文献全文数据库”、苏州图书馆“馆藏古籍数据库”等,而且这些数据库也并未将全部馆藏善本收录其中。就公共图书馆善本古籍藏量而言,这样的古籍珍善本图像数据库数量,无论较之高校图书馆还是国外图书馆的同类数据库数量,相对而言是比较小的。另就读者阅览权限而言,这些数据库可供读者免费开放获取的比例也比较低。建设数量小、收录典籍少、读者阅览权限低的情况固然与珍善本稀缺度高、数据库建设成本较高相关,但是由此也可见公共图书馆对珍善本古籍图像数据库建设重要性认识不足,古籍数字化从业者对古籍珍善本数字化工作的战略储备意义及传承文化作用的认知有待进一步提高。

### 1.3 古籍图像数据库的图像色彩管理仍有提升空间

综合考察目前国内的珍善本古籍图像数据库,普遍存在古籍图像色彩管理有待提高的情况。这一问题主要表现在两个方面,一是出于古籍保护的原则,对于已有缩微胶片的古籍善本资源,业界通常优先使用缩微胶片进行数字化扫描,而使用缩微胶片制作的图像数据会存在图像明亮度、对比度等方面的问题。二是使用古籍原件扫描制作的图像数据存在色彩准确度低、色差较大等问题。图像色彩问题的存在,虽然不会直接影响古籍文献内容所传达的文字信息,但是有可能会大大影响读者的阅读体验,进而影响数据库的利用率。

### 1.4 古籍珍善本图像数据库网络能见度低,馆外地址访问率不高

国内公共图书馆建设的珍贵古籍图像数据库,普遍存在读者认知度低,馆外地址访问率不高的问题。如华东师范大学图书馆馆长胡晓明先生曾介绍过微博 ID 陆浑戎,“每天都有全中国乃至世界上的电子书爱好者自动到他的网页上报到,每天都会有源源不绝的新资源新链接新信息在他的微博上发布。”<sup>[4]</sup>陆浑戎曾通过微博推送过许多古籍文献类开放获取学术资源,但是此类资源中极少见到国内图书馆建设的古籍珍善本类图像数据库,此类数据库“养在深闺无人识”的情况可见一斑。

由于古籍珍善本的特殊性质,古籍珍善本图像数据库建设面临的困境和难点很多,需要从业者将现有理论成果和标准规范探索性地运用于数据库的建设,最大限度地突破以上瓶颈,为读者提

供更多的开放获取、内容丰富、图文并茂的古籍珍善本图像数据库。

## 2 加强珍善本古籍图像数据库建设的基本原则

### 2.1 成熟性原则

数据库建设的立项和开展必须建立在稳定的资金支持和切实可行的古籍数字化工作方案基础之上,开展古籍珍善本图像数据库建设必须坚持水到渠成的成熟性原则。

#### 2.1.1 全面考虑本馆基础设施建设,择机启动古籍珍善本图像数据库建设

在古籍图像数据库建设立项之初,图书馆需对本馆硬件和软件水平进行全面评估,对照各项标准来审核本馆仪器设备和人力资源配置,慎重评估本馆数字化建设的条件和时机是否成熟,避免对珍贵古籍造成不必要的过度利用甚至损坏。首图“古籍珍善本图像数据库”立项之初,曾全面梳理了首都图书馆的硬件建设、人员配备和前期成果,经过严格评估后审慎地通过了项目建设。

#### 2.1.2 深入了解本单位预决算制度及项目所需经费,为项目执行奠定坚实的经济基础

作为公益性事业单位,公共图书馆的预决算制度有其自身特点,数据库建设人员必须要深入了解各项财务制度,为项目执行提供经济保障。在“古籍珍善本图像数据库”立项之初,工作人员通过业界交流、专家咨询等多种方式,深入了解了当时古籍数字化加工和数据库建设的市场行情,制定了较为完善的经费预算,并视工作进度逐年合理申请经费,充分保证了该项目的长期平稳运行。

#### 2.1.3 全面评估古籍图像数据库制作商特点和能力,确保招标工作有的放矢

公共图书馆的大型数据库建设,必须要经过招投标的方式来确定乙方。立项之初,首都图书馆组织相关人员撰写了详细的招标公告,确保投标方能够清晰理解数据库的技术需求,并给出合理的制作方案。在招投标过程中,首都图书馆与招标公司的专家对各家方案进行了深入地地质询和分析,最大可能地确保了乙方的技术水平达到制作要求。

就首都图书馆的具体操作而言,首都图书馆自建古籍图像数据库建设开始于本世纪初。自2002年起,该馆利用本馆研制的智慧2000二次

文献标引及检索系统制作完成了古籍插图图像数据库,收古籍插图数据近2万条。在该数据库的建设过程中,该馆历史文献中心在插图扫描、图像标引等方面积累了丰富的工作经验,为开展更大规模的古籍数据库建设奠定了坚实的基础。2011年9月,该馆正式启动了“古籍珍善本图像数据库”的建设。在古籍珍善本图像数据库建设之初,并在明确数据库技术需求的基础上,经过周密调研,制定了切实可行的财政预算、招标方案以及操作规范,为具体工作的开展奠定了坚实的基础。

## 2.2 标准化原则

在古籍数字化工作过程中,工作人员发现关于古籍数字化的规范和标准较多,有些标准之间还存在一定的差异。由于没有强制标准,首都图书馆古籍珍善本图像数据库建设之初,曾遇到过选择与执行何种标准规范的问题。近年来,随着古籍数字化工作的日益标准化和规范化,关于古籍图像数据库的建设,可以参照的国家标准和行业标准有:

《图书馆馆藏资源数字化加工规范第3部分:图像资源》(标准号:GB/T 31219.3-2014;2015年1月1日起实施;发文号:中华人民共和国国家标准公告2014年第22号),给出了文献资源、原生图像数字资源和派生图像数字资源的加工标准和工作规范。

《图像数据加工规范》(标准号:WH/T 46-2012;2012年12月1日起实施;发文号:文科技发[2012]25号),规定了图像数据加工的适用范围、规范性引用文件、有关术语和定义、加工标准、格式体系、数字资源应用级别、图像数据加工流程、图像数据命名规则。

《古籍元数据规范》(标准号:WH/T 66-2014;2014年4月1日起实施;发文号:文科技发[2014]1号),主要技术内容包括对古籍描述性元数据的元素集的规定,以及对已规定的各元素及元素修饰词的著录规则。

《图像元数据规范》(标准号:WH/T 51-2012;2012年12月1日起实施;发文号:文科技发[2012]25号),本规范将图像元数据元素集分为核心元素、资源类型核心元素、个别元素三级结构,共设计了21个元素。目的是通过制定统一的图像元数据规范,实现行业内对图像资源的统一揭示、深层挖掘。

《数字资源长期保存元数据规范》(标准号:

WH/Z1-2012;2012年12月1日起实施;发文号:文科技发[2012]25号),制定图书馆数字资源长期保存元数据的数据字典,并制定相关应用纲要,实现与其他主要保存元数据框架的映射。同时根据中文数字信息资源的特点和数字资源的流程制定专有的文件格式定义技术元数据。

### 2.2.1 统筹兼顾,取法其上

对于加工级和发布级图像的制作标准,各项行业标准提出了不同的标准,对于加工级别图像制作标准的选择,需要根据本馆的资金能力、服务器建设、图像扫描设备、古籍原件特点等等具体情况,在统筹兼顾、取法其上的情况下,具体问题具体分析,选择最适合本馆实际情况的标准。

### 2.2.2 适时迭代,与时俱进

古籍数字化在资源建设方面涉及到大量的标准,而这些标准是随着科技发展而与时俱进的,这就要求古籍数据库建设在标准选择上要具有兼容性和开放性,并且依据数据库建设规划,制定建设标准适时迭代的原则,及时跟进世界先进国家的动态标准和最新技术,尽最大限度保障数据库与时俱进。

### 2.3 统筹性原则

图像数据库建设是一项多项步骤环环相扣的系统工程,在建设过程中必须要坚持统筹性原则。具体而言,统筹性原则在图像加工阶段的主要表现是理顺古籍图像加工流程。在古籍数字化工作中,根据国家古籍保护中心发布的《古籍数字化工作手册》(试用本)和本馆实际工作经验,可以将工作流程分为四个重要步骤实施:古籍数字化准备工作、古籍元数据著录、古籍图像采集制作、数据包审核及发布。而在每个步骤下,工作人员定期沟通,及时解决常见问题并完善工作机制,保证了工作稳步开展,尽可能地避免了反复取用古籍善本原件。

## 3 加强古籍图像数据库功能建设的对策和措施

### 3.1 加强图像色彩与内容管理

国家古籍保护中心下发的《古籍数字化工作手册》(试用本),对数字图像的采集、处理和检查都提出了具体要求,涉及纠偏处理、图像拼接、图像剪裁、尽量减少对图像文件的后期处理四个方面。这些标准的设计,保障了古籍图像可以如实反映珍贵古籍的文本内容,即学术资料性。根据

对珍贵古籍特性的分析,图像制作还应考虑到如何体现古籍的历史文物性和艺术代表性,这就对图像的色彩和内容提出了更高的要求。

### 3.1.1 加强色彩管理

为了更全面地反映珍贵古籍的质感,需要在图像扫描和图像加工阶段采取多种手段加强色彩管理。其一,在选择图像加工设备时,需要甄别图像采集设备的各项技术参数及实际表现,为色彩管理提供扎实的硬件基础。其二,在扫描制作图像的时候,应充分考虑客观环境中光线、背景等对图片效果影响,为图像采集营造干净的氛围。其三,在数据加工环节,建议增加标准色彩导表,尽量保证图像色彩的高精度还原。

### 3.1.2 增加图像内容

在古籍图像数据库中,建议增加古籍封面、古籍全貌(包含书匣、函套等)等图像信息,最大限度体现珍贵古籍的历史文物性和艺术代表性。目前正在开展的全国可移动文物信息普查,对文物拍照提出了一些规范,在古籍图像制作中可以充分借鉴文物拍摄的理念和方法,以全面展示珍贵古籍的价值。

## 3.2 合理配置发布平台功能

珍贵古籍的收藏遍布全球,世界各地不同的图书馆根据自己的馆藏建立了形式多样的珍贵古籍数字资源库,包括古籍影像数据库、古籍全文数据库、网络古籍资源的组织与检索平台等。这些网络资源的建设,既保存了珍贵文献,也传播了中华文化。综合分析目前网络上的古籍资源,结合目前本馆古籍数据库推广中遇到的问题,为提高发布平台的开放性和能见度,对古籍发布平台的优化提出以下建议:

### 3.2.1 规范和增加元数据款目

古籍元数据的著录规范多种多样,对于元数据规则的选择应兼容并包,在谋求古籍资源共享,为全世界提供汉学研究信息的基础上,博采众长,适时修订古籍著录规则、标题法、权威控制等,以利于图像数据库内容的揭示和查询。比如,在古籍元数据著录时,工作人员需全面审视和挖掘本馆藏本的特点,对序跋、钤印、题跋等内容进行详细标引和著录,突出本馆藏本特色,提高典籍的利用价值。

### 3.2.2 降低数据库的使用门槛,提高资源利用率

目前国内的珍贵古籍资源数据库对读者的开

放度较低,许多数据库只能在特定IP地址范围内全文阅读,对馆外读者仅开放部分内容甚至不开放。随着国家对图书馆经费投入地增加和图书馆公共服务意识的提升,今后应进一步提高公众免费获取图书馆珍贵古籍资源的权限,保证读者有效地利用珍贵古籍,充分发挥数据库的学术价值。

### 3.2.3 加强数据库的多向整合,提高数据库的网络能见度

古籍数字化越来越成为古籍保护的重要方式甚至主要方式,受此趋势影响,古籍数据库的数量会越来越多,为此国家应加强对全网数据库资源的整合,尽快建立统一的发布平台,避免各馆之间各自为政,人为地制造信息孤岛。

### 3.2.4 增强发布平台的开放性,及时发布新媒体服务端口

网络科技日新月异,各种平台各领风骚,图书馆古籍发布系统也应及时跟随时代脚步,积极主动接入各种发布平台,加强数字资源推广力度,提高优质资源的辐射范围,更便捷地为读者提供服务。

## 3.3 完善古籍数字化工作机制

公共图书馆的古籍数字化工作牵涉面广,如就数据加工而言,涉及古籍典藏、古籍修复等多个古籍工种,就平台建设而言,涉及馆内数字资源建设、网络管理等相关管理部门;由于委托加工的工作机制,图书馆还要面对良莠不齐的数据加工公司、流动性较强的工作人员,在这一形势下,图书馆采用项目制的管理办法,完善古籍数字化工作机制,可以大大提高工作效率。

### 3.3.1 完善图书馆古籍数字化工作制度

面对国家大力推动数字图书馆建设的新形势,古籍珍善本图像数据库的建设应积极融入其中,在为数字图书馆提供优质资源的基础上,充分利用本馆数字图书馆建设在管理制度、硬件设施和业务流程建设中取得的成果,快速理顺工程招标、项目管理、系统维护的工作流程。随着古籍数字化工作的常态化,图书馆还应设立古籍数字化工作组,固定人员配置,确定岗位职责,保证项目顺利实施。

### 3.3.2 建立工作人员定期培训机制

随着中华古籍保护计划的推进及“中华古籍数字资源库”的建设,国家图书馆和国家古籍保护中心对各馆古籍数字化的指导工作逐渐深入,各

图书馆应积极参与国家古籍保护中心举办的各类培训,提高业务能力。在许多普遍性问题得到及时解决的现实情况下,图书馆还应针对工作中出现的个性化问题对员工开展定制培训,如古籍编目、项目运营、色彩管理、数据库搭建、推广营销等相关知识的学习和储备能够极大地提高工作人员的工作效率。另外,面对项目外包已成为古籍数字化工作普遍方式的现实状况,对乙方工作人员的业务培训也应形成有效机制,以更好地保证古籍安全及数据质量并提高工作效率。

### 3.3.3 聘请业内专家为数据库建设提供专业咨询

科学技术日新月异,一线工作人员在工作中时常会遇到新问题,随时可以提供咨询的专家可以及时帮助解决工作中遇到的问题。为了解决数字化工作的实际需求和潜在困难,图书馆应聘请相关领域专家,成立咨询专家组,及时为古籍数字化工作提供智力支持和技术保障。

公共图书馆古籍珍善本图像数据库的建设从

(上接第 47 页)

与服务体系,基于这些交互过程让读者、用户体验图书馆的价值,从而实现图书馆发展权利的增权。

图书馆发展权利的增权路径是意义和功能的双重体现。它通过图书馆主体与读者、用户的交互作用,以创新的产品体系和服务体系诠释增权内容,是积极主动的权利实践过程。只有如此,图书馆才能与代理群体的需求发展相同步,才能创造出自己的核心竞争力,才能最大限度地实现图书馆权利的拓展与渗透。

参考文献:

- [1]刘意,文庭孝.我国 21 世纪新图书馆运动研究综述[J].图书馆,2012(4):46-51
- [2]李国新.21 世纪初年的“图书馆权利”研究与传播[J].中国图书馆学报,2014,40(6):4-11.
- [3]韩毅,刘丽.图书馆精神的历史积淀与层次性分析[J].图书馆,2006(1):14-17.
- [4]付国英,孔方礼.基于哲学图书馆学的增权及模式选择[J].情报科学,2007,25(7):995-997
- [5]范进学.权利政治论:一种宪政民主理论的阐释[M].济南:山东人民出版社,2003:序言.
- [6]燕辉,常安.论作为消极权利的图书馆自由[J].图书馆,2013(3):1-4.

无到有,从少到多,是一个实践和理论相互促进螺旋式上升的过程,及时反思和总结工作中遇到的问题势必能更好地服务于今后的古籍数字化工作。面对源远流长的珍贵典籍和快速发展的数字化技术,先行者有义务及时反思为今后工作提供参考,后来者亦应充分发挥后发优势提高数据库服务水平,共同建设具有国际影响力的中华珍贵典籍资源库。

参考文献:

- [1]国务院办公厅.国务院办公厅关于进一步加强古籍保护工作的意见[EB/OL].[2016-02-24].http://www.nlc.cn/pcab/ztl/gjbhzcfg/gfxwj/201511/t20151104\_107985\_1.htm.
- [2]施晓平.古籍数字化文字只能人工输 一页人工要花费 6 块多[N].苏州日报,2012-05-08(A10).
- [3]张志清.国家图书馆古籍保护的历史、现状和任务[M].杭州:浙江省古籍出版社,2008.7.
- [4]胡晓明.我的古籍电子书阅读小史[N].文汇报,2015-06-19(12).

作者简介:

史丽君(1980—),女,副研究馆员,首都图书馆,北京,100021。

- [7]常安,燕辉.论作为积极权利的图书馆权利[J].图书情报知识,2012(4):4-10.
- [8]陈树强.增权—社会工作理论与实践的新视角[J].社会学研究,2003(5):70-83.
- [9]丘柏林.我国图书馆权利研究述评[J].图书馆学刊,2014(4):129-134.
- [10]程焕文.图书馆权利的界定[J].中国图书馆学报,2010,36(2):38-45.
- [11]李国新.图书馆权利的定位、实现与维护[J].图书馆建设,2005(1):1-4.
- [12]范并思.论图书馆人的权利意识[J].图书馆建设,2005(2):1-5.
- [13]安栋.适应性视角下的社会工作增权理论分析[J].科协论坛,2010(11):184-185.
- [14]唐咏.中国增权理论研究述评[J].社会科学家,2009(1):18-20.
- [15]范并思.公共图书馆精神的时代辩护[J].中国图书馆学报,2004,30(2):5-11.
- [16]蒋永福.知识秩序·知识共享·知识自由—关于图书馆精神的制度维度思考[J].中国图书馆学报,2004,30(4):10-13.
- [17]程焕文.实在的图书馆精神与图书馆精神的实在[J].大学图书馆学报,2006(4):2-14.

作者简介:

杨晓琼(1974—),女,副研究馆员,重庆人文科技学院图书馆,重庆,401524。