

浅析古籍文献数字化建设

——以苏州市吴中区图书馆为例

陈敏

(苏州市吴中区图书馆, 江苏 苏州 215000)

摘要: 对于学术领域而言, 古籍文献非常重要。古籍文献不仅蕴含了我国丰富的历史背景和相应的文化知识, 而且更具有自身的独特意义。因此, 在图书馆建设过程中, 对古籍文献进行数字化建设非常重要。通过古籍文献的整理, 以数字化形式, 不仅能够解决以往在文献保存使用中出现的各项问题, 如文献损坏以及、文献内容查找费时费力等, 还可以使古籍文献更好地发挥独特优势。通过数字化领域的建设, 能使古籍文献能够被及时取用, 发挥自身的独有价值。因此, 在本文的研究中, 将就古籍文献数字化建设——以苏州市吴中区图书馆为例展开讨论。

关键词: 古籍文献 数字化建设 图书馆 研究分析

DOI: 10.12319/j.issn.2096-1200.2022.22.10

一、古籍文献数字化建设的重要意义

古籍文献数字化, 利用文字、图片、音频等多种媒体, 将信息内容以数字化形式进行储存, 同时通过计算机系统, 对整个数据进行标记, 以便能够更好地进行取用。通过网络浏览、查询以及传输, 并结合权限管理技术, 保护知识产权, 通过Web发布数字化信息^[1]。因此, 数字化建设是图书馆在建设过程中的必然趋势, 以计算机技术、网络通信技术、数字技术、文献处理技术等为载体, 可以实现大型的数字化文献资源信息管理模式, 更好地发挥其独特特征。在苏州市吴中区图书馆, 古籍数字化出版成果《吴中珍藏》收录《王氏家谱》、《陆氏家谱》、《崇祯遗录》等, 这些古籍文献经过数字化建设, 可以更好地方便读者进行免费阅读、检索^[2]。

苏州市吴中区图书馆作为县级图书馆, 收藏古籍万余册, 其中有9部入选国家珍贵古籍名录, 15部选入省级珍贵古籍名录, 目前古籍数字库收录23部古籍。在专业技术支持下, 吴中区图书馆陆续对大批珍贵古籍文献进行数字化处理, 可以更好地使古籍文献发挥出自己的独特优势^[3]。古籍文献数字化是非常必要的, 在信息技术迅速发展的背景下, 依托网络, 可以解决古籍文献利用以及传承问题。古籍文献数字化建设是未来整理与利用的发展方向, 适应这种变化, 可以更好地结合我国目前图书馆建设需要, 对古籍进行全方位的要求, 同时还可以保护珍贵的古籍文本。古籍数字化有效解决古籍保存占用人力、物力、财力等问题, 对古籍研究工作将会产生巨大的推进作用。且古籍文献数字化的优势极多, 例如, 古籍电子版本可以无限地

复制, 取之不尽, 用之不竭。同时, 借助计算机网络, 可以使古籍文献实现远程共享, 避免了运输以及分配问题。且古籍资料汇集后, 产生的新信息对于研究工作非常重要。汇集大量的古籍加以整理, 仅靠人力无法完成, 但借助电子版本, 单人即可对古籍进行检索以及处理^[4]。

二、我国古籍文献数字化建设的现状

除苏州市吴中区图书馆建立古籍文献数字化外, 我国古籍文化数字化建设取得了相当成绩。例如, 目前出版的电子版有《四库全书》《四部丛刊》《康熙字典》等, 还有《中华文化通志》《汉语大词典》等, 这些古籍均是我国传统文化的精髓。中国基本古籍库是综合性的全文检索版大型古籍数据库, 是北京大学的重点科研项目。全套光盘库根据中国古籍自身特点, 按照国际通用图书分类法进行分类, 包含了哲科、史地、艺文、综合等4个子库, 共计20大类, 近百个细目。古籍科目范围涉及极广, 几乎可以满足我国科研需求, 且适用于中、日、英等多种应用操作平台, 并提供了多重检索功能。在实际使用中, 用户仅需要懂得基本电脑操作方法, 便可以查找到自身所需要的资料, 且在检索过程中每次检索时间可保证在5秒内完成。我国于2000年正式成立了中国数字图书馆, 以推进整个图书数字馆建设的工作。其中, 古籍图书馆数字化建设是我国的一项重点工程, 该项目在数据库中发挥了重要作用, 对珍稀古籍文献进行录入并实现检索, 可以更好地完成应用^[5]。

三、古籍文献数字化建设的相关问题

(一) 古籍数字化标准不统一

各机构开发的数据库都有自己的数字化标准, 没有形

成统一的跨平台接口。古籍资源数字化的技术标准必须符合中文古籍的特点。古籍数字化流程中,数据采集标准规范、图像采集处理规范、文字采集规范、资源内容表示规范、古籍用字标准、专有名词标准、整理本古籍元数据标准、整理本古籍文献数据标注标准等等,都必须符合中文古籍的特点。目前国内对古籍数字化缺少统一的行业标准,各数据库自行制定的标准存在较大差异,导致各数据库的质量良莠不齐,也不利于不同数据库之间的合作开发与利用^[6]。目前,图书馆界缺乏统一的分类标准,其主要分歧意见包含两种,一种主张利用《中图法》进行分类;而另一种则主张利用《四库法》进行分类。古籍文献分类的不统一,对古籍文献数据库的建设共享带来了很大困难。《四库法》作为目前常用的一种分类方法,有完善的分类体系。在学术研究方面,可以更好地达到分类取用的目的,但分类较粗,类目较少^[7]。在分类过程中,针对近代科学方面的图书,就存在《四库法》无法录入的问题。而《中图法》科学性较强,针对于古籍文献编制技术先进,且不断修订,可以更好地适应未来发展需求,也符合广大读者的检索习惯。但由于古籍文献文哲史不分家,因此有的古籍不能有效地按照《中图法》进行分类。对于古籍编著而言,两种分类方法各有千秋,在古籍编制工作中,需要结合古籍的实际情况,两种方法进行合理配合取用,形成完整且合理的分类检索体系,保障分类更加趋于完善,满足读者的阅读检索需要^[8]。

（二）古籍文献数字化保存风险

古籍文献数字化在保存过程中存在着明显的风险问题,古籍文献数字化建设是一项非常复杂且系统的工程,古籍文献数字化产品无疑是学者的心血结晶。因此,古籍经过扫描、录入后,其数字化的保存较为重要。我国着重于古籍数字化,但对于古籍数字化后的保存相关研究较少,无法制定有效的保存规范。在以往的古籍使用过程中,由于其具有明显实体,因此可以把握损毁情况,并制定一系列的管制措施。但数字化古籍在录入完毕后,存于服务器、硬盘等载体中。在使用时,由于长期读取或服务器存在损坏风险,会导致载体物理裂变,出现数据资料丢失等问题。此外,也包含了不可预料的自然灾害或损坏、数据保存人员变动等,这些因素均会导致古籍文献出现难以逆转的数据丢失。一般出现数据丢失后,古籍需要重复录入,耗费了人力、物力。若对于濒危古籍或在录入后出现破损无法使用的古籍而言,便会导致文化以及历史背景永久丢失,造成严重影响^[9]。

（三）古籍文献在数字化建设中的人员问题

古籍文献在数字化建设中,呈现一定的人员问题,包含两大层面:

1.建设同步问题。古籍文献的建设必须走数字化之路,结合实际情况,实现多架构、勤同步的目标。但目前,我国古籍数字化工作体系缺乏统一的协调领导机制,且部分商业机构加入古籍数字化队伍中造成项目重复建设,而冷门项目无人问津的问题。古籍文献数字化重复投入所带来的不仅是资源浪费,也有可能导致古籍文化出现一定程度的问题以及风险。因此,必须走协同合作道路,建立全新的古籍数字化建设领导小组。将古籍有关的教学研究机构、收藏机构等进行融合,实现共建,以更好地统筹规划古籍数字化建设工作。

2.人才培养问题。古籍数字化工作过程中,在录入文本前需要对古籍进行整理工作。古籍人员不仅要掌握古籍文献的分类及其版本,同时还需要编制数据库,将计算机、多媒体等现代信息技术融合至古籍数字化建设领域。但目前,古籍整理专业人员缺乏对于信息技术的支持力度,而信息技术人员对于古籍知识又无法进行了解。例如,在进行古籍数字化录入时,由于录入人员自身的水平限制,将大量的异形字、异体字等视为不同的汉字进行录入,不仅增加了录入工作量,更对于检索系统产生了影响,使录入后的古籍无法使用。

四、古籍文献数字化建设的解决思路

（一）成立全新的工作领导小组

成立全新的工作领导小组,以便能够更好地发展古籍数字化文化。建设古籍文献数字化,需要多方协作,在古籍数字转换过程中,以吴中区图书馆为例,可以根据建设过程中出现的各种不足,设定一个全新的工作方针,解决目前在工作中出现的各项问题。例如,目前古籍数字化工作缺乏统一性的领导,针对这项问题,开发单位以及各图书馆需要整合自身的工作模式。且能有效解决冷门项目无人问津的问题,同时避免热门项目重复录入。如《四库全书》在国内便至少出现了三种不同版本的数字化资料,导致人力、物力、财力的浪费。因此,在编制过程中,需要整理文献数字化专门领导小组,负责审批古籍文献数字化,建设立项以及审批工作,可以更好地完成宏观指导。

（二）加强古籍文献数字化的建设开发力度

加大古籍文献数字化建设的开发力度,可以更好地体现数字文化遗产。古籍文献与其他资料相比,古籍文献自身要求读者有一定的阅读基础。因此,在开发、研究以及

利用过程中,需要解决目前存在的局限性,以便能够更好地贴合我国相关部门的要求,实现古籍文献数字化的建设工作。例如,在数字化建设开发力度中,我国给予了高度重视。山东图书馆早于1996年便完成了数字化古籍的建设工作,而对于苏州市吴中区图书馆而言,就需要贴合我国的发展目标以及古籍文献数字化的建设方向。加强古籍文献的宣传以及建设力度,扩充古籍文献数字化的建设范围以及古籍文献的开发利用渠道。挖掘古籍文献的价值以及财富,以便能够更好地传播中国文化。

(三) 注重古籍文献的数字化建设以及人才培养管理

注重古籍文化的数字化建设以及人才培养管理,古籍文献数字化建设工作是一项非常艰巨且复杂的系统工程。不仅要求研究人员充分掌握古籍文献的分类以及版本知识,同时还需要保障工作人员具有数字化专业素养,以便能够更好地对古籍数字化建设提供力所能及的帮助。因此,需要根据全新的数字化建设技术进行更新,如利用ISO标准码以及ICS3、ICS4等数据库进行编制。在编制过程中,根据古籍中含有的生僻字、异体字等字符,进行着重处理。这些异体字符在电脑系统的字库中不存在,要想更好地浏览古籍内容,就需要下载相应字库,并根据类型,保存到本机电脑中。例如,甲骨文库就必须下载ICS3以及ICS4字库,且在7万汉字环境以及国际标准内码、四字检索为核心技术中,可以打造全新的网络应用平台,解决字符翻译需要以及录入需要。与国际互联网接轨,完成浏览以及检索。且在古籍数据库中,可以更好地完成光学符号辨识技术。在进行数字化录入时,利用OCR技术,可以更好地将古籍进行扫描,并改变原有的排版格式,增强古籍阅读的真实感。例如,在扫描过程中,可以采取零边距或非接触式扫描仪。A3幅面的光学辨识度必须达到600DPI,感光元件不得低于5000像素点,且色彩位数需要保持在24BIT。在扫描过程中,使用无紫外线的同步冷光源。扫描时,必须注意扫描玻璃板上是否清洁到位,以避免扫描后图像出现异物。在扫描过程中,将古籍放平、摆正。若扫描效果依然不理想的话,可采取半叶扫描方式,我国珍贵古典需要将其文件储存为TIFF格式。在古籍扫描时,严禁拆除,避免透光、透字。在扫描时,如遇到虫蛀或漏洞时,需要以衬纸扫描。而针对于画幅扫描时,各扫描边框必须有三厘米以上的重复扫描区。且在扫描前,就需要进行描色,保证

扫描效果。

五、结语

综上所述,古籍文献数字化建设无疑是一项复杂且艰巨的系统工程。古籍数字化基本依然以专用型为主,在实用性、通用性、共享性方面亟待更新。且建设过程中,依然有较多问题值得深入探讨。通过建设大型古籍数据库,将古籍、历史朝代等文化成果进行数字化转化,形成一个庞大的数据库,完成数据库信息查询研究等工作,能保留我国传统文化精髓。肩负中华文明传承使命,给后代留下一笔伟大的精神财产,不断传承中华文明。

参考文献

- [1]李建勋,陈雪颖.地方古籍和民国文献数字化实践与思考——以江门市图书馆为例[J].传播与版权,2021(9):79-81.
- [2]靳伯云.古籍文献何以为本地文化续脉——以石家庄市图书馆为例[J].科技风,2020(14):240.
- [3]丁琴娥.古籍文献的保护,开发与利用——以甘肃省平凉市庄浪县图书馆古籍为例[J].农村经济与科技,2020(13):336-338.
- [4]潘煦.古籍数字化发展策略探析——以西南民族大学图书馆为例[J].中文科技期刊数据库(全文版)图书情报,2021(10):135-137.
- [5]颜艳萍.数字人文背景下图书馆古籍数字化新实践——以云南省图书馆古籍数据库建设为例[J].图书馆学刊,2020,42(7):39-43.
- [6]张秋,杨玲,毛李洁等.真人图书馆在一流大学建设中的实践探索与发展策略——以“学在清华·真人图书馆”为例[J].图书馆学刊,2020,42(3):85-92.
- [7]魏清华,孙林,胡文静.高校图书馆特藏文献数字化建设研究——以CASHL为例[J].大学图书馆学报,2020,38(1):81-86.
- [8]李明杰,方正,宋时雨.文献整理学术传统对古籍数字化的参照价值(之二)——以“述而不作”为例[J].图书馆论坛,2020,40(7):142-149.
- [9]张会芳.数字人文时代图书馆古籍文献资源的数字化长期保存机制分析[J].办公室业务,2020(21):191-192.