略谈我国古籍数字化建设的现状、问题及建议

严如月

（北京美斯齐文化科技有限公司　湖北　武汉　430000）

[摘　要]我国古籍数字化事业经过三十余载的努力发展现已取得了不小的成果，但仍然存在着一些关键性问题。本文从管控机构、行业开发者、技术研发者、资源利用者、培养单位等各方面进行综合考量，提出加强宏观调控统筹、统一数字化标准、加强版权保护、提升数字化水平、培养专业人才等有效解决建议。

[关键词]古籍数字化；问题；建议

【DOI】10.12252/j.issn.2096-6288.2020.06.750

在中华民族数千年的历史长河中流传着浩如烟海的古籍文献，虽然在此过程中有不计其数的古籍历经散佚、毁损、淘漉，但如今存世的古籍数量依然不可胜计。为尽力保护这些珍贵的典籍，大多藏馆单位都不得不将其束之高阁鲜少示人，但如此一来，对于古籍“藏”与“阅”的矛盾便日渐凸显了出来。

得益于互联网和计算机技术的发展与完善，自20世纪八十年代起，我国的古籍数字化事业也随之起步。所谓古籍数字化，就是“从利用和保护古籍的目的出发，采用计算机技术，将常见的语言文字或图形符号转化为能被计算机识别的数字符号，从而制成古籍文献书目数据库和古籍全文数据库，用以揭示古籍文献信息资源的一项系统工作”[1]。简言之，就是利用信息处理技术将古籍文献转化为计算机可识别和处理的数字信息的过程。

一、古籍数字化发展现状

（一）政策支持

经过三十余年的发展努力，我国的古籍数字化事业取得了不少成果。这首先得益于国家对古籍保护事业的高度重视，尤其是近些年来对古籍保护的相关政策逐渐细化到了古籍数字化的层面。

2007年1月，国务院办公厅印发了《关于进一步加强古籍保护工作的意见》，标志着我国首次在古籍保护领域中由政府主持进行国家级重要文化工程建设。[2]2011年3月，国家文化部下发了《关于进一步加强古籍保护工作的通知》，加强推进“十二五”期间古籍保护计划的落实。[3]同年5月，文化部、财政部联合印发《关于实施“数字图书馆推广工程”的通知》，计划于“十二五”期间在全国实施数字图书馆推广工程。[4]针对古籍数字化的一些问题，文化部委托国家古籍保护中心进行多次调研，并在2012年初拟定了《中华古籍数字资源库建设方案》（征求意见稿），提出合作共建、资源共享的重要原则。[5]同年又相继下发《国家“十二五”时期文化改革发展规划纲要》[6]《文化部“十二五”时期文化改革发展规划》[7]，再次强调了古籍保护工作的重要性。2013年12月，国家主席习近平总书记在主持中共中央政治局第十二次集体学习时强调，“要系统梳理传统文化资源，让收藏在禁宫里的文物、陈列在广阔大地上的遗产、书写在古籍里的文字都活起来”。2017年5月，中共中央办公厅、国务院办公厅印发了《国家“十三五”时期文化发展改革规划纲要》，《纲要》提出了统筹推进古籍整理出版数字化，建设包括古籍资源在内的中华文化资源数据库等计划。[8]

（二）数字化成果

国家的大力支持为古籍数字化的发展提供了强有力的大环境支撑，各类古籍存保单位、高校科研院所、商业机构等作为主要开发主体在此领域内历经数十年的精耕细作，使我国的古籍数字化在理论研究、标准规范、技术研发、实践应用等方面取得了不少成果。据有关学者统计，近20年来，国内外通过计算机技术开发研制的各类古籍数字化资源就有近500种，其中包含有近270种古籍全文数据库，约140多种古籍书目数据库及80余种古籍电子索引系统。[9]

90年代以后，大陆地区的古籍数字化事业也开始风生水起，“中国数字图书馆工程”“中华再造善本工程”“中华字库工程”等重点项目陆续推进。各类开发主体都在数字化建设方面取得了相当的成绩，如作为图书馆藏单位之首的国家图书馆，目前已建成包含36个子库的古籍资源库，其下的“中华古籍资源库”已发布超3.2万部在线古籍资源；作为高校科研代表的北京大学，研制出了“北大国学二十五史研习系统”，全国24家重点高校图书馆共同建成高校古文献资源库——学苑汲古；一些较有影响力的商业机构，如北京书同文数字化技术有限公司开发的《文渊阁四库全书》电子版、“《四库丛刊》全文检索系统”，北京爱如生数字技术有限公司开发的“中国基本古籍库”及“中国方志库”等等都是其中的典型代表。此外，古籍数字化的产品类型也日益丰富，各类专题数据库如中国社会科学院的《全唐诗》数据库，中国国家图书馆的“碑帖菁华”，湖北省图书馆的“湖北方志库”“湖北家谱库”，爱如生的“明清实录数据库”等等。[10]

数字化技术功能也在不断的发展完善中，从20世纪90年代开始，“新的计算机应用技术，如非键盘输入技术、中文数据库技术、多媒体压缩与传送技术、安全保密技术、自然语言理解技术，尤其是数据挖掘技术的出现，为古籍数字化事业提供了有力的支持”[11]。依托于技术的创新，古籍数字化在展现方式上现已包括全文展示、原书影像展示，以及二者相结合的图文对照展示方法。从开发者的角度来说，扫描技术与影像清晰度已经完全可以达到适应精度阅读的需求了，OCR文字识别技术在识别的精度与速度上也有了进一步的提升，汉字字库也有了相当程度的扩充，自动句读、标引及校对也依托人工智能技术在不断发展与完善之中。从读者阅读的角度来说，数字化阅读及交互界面的设计也更为合理与便捷，阅读中的各类小工具如纪年换算、联机字典、阅读批注等都相应上线。

二、面临问题及解决建议

我国古籍数字化事业在取得了上述成绩的同时也依然存在着一些亟待解决的问题。总的来说，主要有宏观调控统筹、数字化标准、版权保护、数字化技术、数字化水平、专业人才培养等方面的问题。

（一）宏观调控统筹不力

1、问题现状

我国现存古籍数量达十余万种且分散于全国各地，长久以来，由于缺乏业内权威机构的统筹管控，在对古籍资源的数字化开发与利用方面一直存在着不规范的状况。国内古籍数字化的开发主体主要分为三类：以公益性为主导的图书馆藏单位、以研究性为目标的高校科研院所和以盈利性为目的的商业机构。由于各主体的开发目的和对资源掌握的差异性，客观来说，在资源开发的过程中常常存在着选配不均与资金不足的问题；主观而言，在项目开发上还普遍存在着“趋热避冷”的现象，对于热点重复建设，冷门项目则乏人问津。开发者在对古籍数字化内容的选择上，也多侧重于文史一类的古籍，而较少关注如医学、数学、工技、艺术等其他领域的文献。这些都在很大程度上造成了古籍资源的开发不均和浪费。在古籍数字化过程当中，从版本的择取到资源发布管理，各主体单位也常常各自为政，即便是已经完成的数字化资源也无法真正于业内进行融合，形成更大的资源库。

2、解决建议

针对此类问题，需加强古籍数字化体制机制的创新，可经由国家权威机构牵头，建立专门的“古籍数字化管理委员会”，管委会成员可由各开发主体按照一定比例构成。管委会下可设立由各类开发主体群组成的若干行业协会对日常事务进行分管。统筹机构可依据我国行业领域的现状并兼顾各开发主体的优势特点，建立一套行业认可的统筹管理体系。管理体系中应包括业内统一的规章协议，选题、立项、开发、资源整合的管理方式，以及统一的数字化加工标准等等。[12]国家还应加大对数字化项目的资金投入，为不同的开发者提供不同的支持政策，如对中小微企业实行部分减免税费、增加项目补贴等。国家财政下拨的专项资金可由管委会根据行业及资源状况进行优化配置。具体执行计划可逐步纳入“十四五”规划的文化建设之中。

（二）数字化标准不统一

1、问题现状

由于开发主体的多元化，业内现虽已有类似《古籍描述元数据规范》《古籍数字化工作手册》《古籍著录规则》等规范性文件的出台，但在实际情况中，各单位机构往往依据自身需求来开发数据系统创建相应标准，这就使得古籍的信息著录、扫描、标引、检索等各方面的工作标准都存在着大小差异。仅从文件格式的存储与展示方面来看，常见的格式就有doc、txt、pdf、wdl、ceb、html、edk等十余种，各平台间的数字化资源格式常常无法兼容。缺乏标准的加工方式及存储格式给我国数字化资源的互通共享带来了很大的阻碍。而从国际方面来看，我国尚未形成一套能与国际通行标准体系接轨的数字化标准，这就在很大程度上阻碍了我国与其他国家在此领域内的交流合作。

2、解决建议

数字化加工标准也是当前亟待解决的问题，站在国际化的角度来看，在建立数字化加工标准时，首先应考虑到国际通行标准体系的兼容问题，在此基础上根据国内行业现状、兼顾各主要开发主体的特点建立一套通行规范的标准体系。一套完善的数字化加工标准需涵盖古籍资源的分类筛选、数据加工、文件储存格式、字符集代码、数据库检索以及数字化发布管理等各个方面。作为开发单位来说，在建设古籍数据库时，也应考虑到多种文件格式的兼容问题，而非取一排他，这样才能避免不必要的资源浪费与重复建设，为后期的资源共享提供便利。

（三）版权保护问题突出

1、问题现状

由于古籍作品的特殊性，对于古籍开发利用的版权问题一直存有争议：开发的古籍是否拥有版权？版权界限设在何处？保护的力度又该如何去把握？特别是互联网技术得到发展和普及后，网络盗版问题层出不穷，随着近年数字化的大发展，其产品的盗版现象也屡见不鲜。相关的版权争端案件频频出现，盗版者钻了法律的空子，对于盗版数字化产品更加没有后顾之忧。

古籍数字化的版权问题所产生的影响是不可忽视的，对于开发者来说，前期大量的投入，却因遭遇盗版而得不到相应的产出回报，这就大大打击了整个行业的积极性。另一方面，各单位出于对产权保护的考虑，在数字化资源的使用方面设立重重壁垒，读者往往只能通过指定的阅读器或局域网才能正常访问浏览古籍。这些都对数字化资源的集合、共享造成了很大的阻碍。

2、解决建议

针对数字化版权存在的问题，首先应该建立及完善相关法律制度，从法律层面提供坚实的保障，对数字化盗版行为要予以严厉的打击，从根本上遏制盗版风气。其次相关管理部门应加强全程监管、严防严控，对于相关的违规行为要采取相应措施，肃清业内风气。另一方面也可以借助相关技术来研发加密保护系统，增强保密级别，从内外两方面共同作用来保护数字化产权。

（四）数字化技术仍待突破

1、问题现状

近些年，数字化技术方面已有了不小的突破，但仍存在着很大的完善和发展空间。目前来看，古籍数字化的开发还处于相对表层阶段，只能实现基础的全文检索和基本阅读功能，语义内容层面的深度挖掘与高效利用仍有待实现。

就现有的数字化技术来说，在字符识别转换、文本加工、字库、检索等方面的技术也亟待进一步完善，录入中的错字漏字以及检索中的查准率不高等基础性技术问题一直存在。以OCR文字识别技术为例，当前的技术依然存在着扫描识别度低、单位成本高的问题，无法达到对竖排繁体文字的理想识别效果。古籍文献中存在着大量生僻字、异体字等难以识别的文字符号，而现有字符库还无法尽数将其囊括，计算机无法释读出正确的信息，从而对史料的真实性与价值性造成破坏，也对后续利用材料进行的学术性研究产生影响。

2、解决建议

提升数字化技术主要体现在现有技术的不断完善及新技术的研发方面。对于现有技术来说，主要是要提高其效率及准确率，这不仅需要技术人员在基础技术方面的不断修正与完善，还需要技术研究者与人文研究学者以及读者三方加强沟通以不断调试。例如在提升全文检索的查准率及有效率方面，应在开发过程中加强古籍专业领域学者的深度参与，以达到检索的专业化、精准化。如对文献中的专有名词进行单独标注，对一些历史人文领域的关联词进行深度加工等。[4]而新技术的研发则更多地需要从使用者的体验和需求中寻求灵感，从输出和利用的角度进行创新和调整。从使用者的角度来说，一项好的技术应力求多方位为用户提供多样化、便捷化、个性化的使用方式，例如实现检索关键词的属性描述、范围控制以及多途径排检功能，进而达到满足用户自定义检索的需求。

（五）数字化水平不高

1、问题现状

我国数字化水平不高首先体现在开发广度上，即资源开发利用的范围较小，现有的数字化古籍资源不可谓不多，但仍有相当一部分古籍文献资源尚待进行数字化开发。而另一个更为突出的问题则体现在数字化水平的深度上。我国目前数字化水平尚停留于表层，数字化古籍的阅读门槛较高，能较好使用数字化古籍的人群必须具备一定的古文及历史知识，这就在相当程度上减少了受众群数量。从对数字化古籍的使用方式和程度上，我们大致将使用者分为两类：学术研究者与普通读者。其中学术研究者的使用比例占据了绝对地位。

要提升数字化水平就必须重视使用者的反馈，而通常研究计算机技术的专家以及从事数字化的工作者大多只专注于本职工作领域，他们对于数字化古籍在阅读和学术研究领域的实际利用效力知之甚少。而学术研究专家又只一心扑在做学问之中，对技术和数字化工作所知不多，更无法将自己的使用体验形成建议反馈出来。三方领域的沟通交流未成渠道也是导致数字化水平不高的一个重要原因。

2、解决建议

要全面提升古籍数字化水平，对于资源的开发利用只是时间和规范问题，更重要的是要根据古籍数字化工作的实践与需求来完善相应的体系以推进数字化的整体发展。数字化输出的结果是为了实现古籍文献的多方位高效利用，因此开发者应多从使用者的阅读体验角度出发，重视读者的意见与建议，不断改善现有的工作方法与呈现效果。要实现古籍数字化水平在新维度上的突破，就必须打破技术专家、数字化工作者与使用者之间的沟通壁垒，加强专业领域的交流互通。可以尝试定期组织开展多边领域的讨论交流会、建立问题的反馈平台等方法。

数字化的开发及工作者要以大众化的视角去推进古籍数字化进程，要多从普通读者的角度考虑，通过一些阅读工具的开发利用来降低数字化古籍的阅读门槛，帮助读者更好地去使用数字化古籍资源。例如引入自动句读、自动注音功能，增加历史文化常识注引超链接，甚至可以尝试加入语音阅读功能等等，来帮助读者扫除古文字、音、义方面的阅读障碍。另外，还应加大古籍数字化的宣传推广力度，如利用新媒体平台宣传或设计相关小游戏的方式来引起更多人的关注和兴趣，使读者群逐渐由小众变为大众，让更多的人去了解并学会使用这一巨大宝库，这样才能最大程度地实现数字化古籍的价值。

（六）专业人才培养困难

1、问题现状

在古籍数字化领域内一直存在的另一个突出问题是专业人才的匮乏。一个合格的古籍数字化工作者不仅需要掌握传统历史文献学的专业知识，还需要对计算机信息技术有一定程度的了解，并且需要在实践中积累相当的工作经验。这种既要横跨文理学科又需要理论与实践相融合的要求，对于专业人才的培养来说是一个极大的挑战。尽管个别高校已经开始尝试开设古籍数字化专业，但招收人数少、培养体系不成熟，毕业后也面临择业性的人才流失，这些都造成目前从事古籍数字化工作的人才极度匮乏，且专业性严重不足。

2、解决建议

对于此类专业复合型人才的培养，更多的还是要依托高校科研机构的专业设置与系统教学，本科阶段应开设古籍数字化的相关专业，研究生阶段可在此基础上细化专业领域的研究方向，而在正式从事古籍数字化工作之前，也应有系统完整的岗前培训制度。一般来说，高校专业课程的人才培养更多的是向理论教学与研究方面的倾斜，故在此基础之上还需加强与一些开发机构的项目合作，让培养人才在接受理论学习的同时更多地参与实践工作，加强实践能力的培养，使理论与实践相互结合、相互促进。

四、总结与展望

古籍数字化是传统古籍整理方法的一种延续，更是一种创新，它不仅是对古籍再生性保护的一种有效手段，也是对深度发掘其历史文化研究价值和推动大众化阅读的一种重要方式。近三十余年来，在国家对古籍保护与数字化事业的大力支持下，我国的古籍数字化已取得相当成果，同时也存在着不少问题。这些问题主要涉及宏观调控、行业规范、技术完善、深度开发、人才培养等方面，要解决这些问题就势必要从管控机构、行业开发者、技术研发者、资源利用者、培养单位等各个方面进行综合考量。

传统历史典籍是一个巨量的宝库，其中包含了诸如政治、经济、文化、军事、科技、艺术等各种门类积聚了数千年的智识与经验。因此，古籍数字化工作不仅是对我国传统文化的传承和发扬，也是对我国各领域事业发展的强大助力。这就决定了未来古籍数字化的发展方向是以大众化阅读为趋向的发展方向。各类开发者及工作者也应从大众化普及与推广的角度去进行数字化资源的建设。我们相信，在厘清现存发展问题并制定出相应的解决方案之后，我国的古籍数字化事业必然会向着更加健康繁荣的方向发展。

参考文献

[1]毛建军.古籍数字化的概念与内涵[J].图书馆理论与实践，2007（04）：82-84.

[2]国务院办公厅《关于进一步加强古籍保护工作的意见》[J].时政文献辑览，2007（00）：840-843.

[3]文化部召开数字图书馆推广工程工作会议全面部署数字图书馆推广工程建设工作[J].中国图书馆学报，2011，37（04）：20.

[4]梁爱民，陈荔京.古籍数字化与共建共享[J].国家图书馆学刊，2012，21（05）：108-112.

[5]中办国办印发国家“十二五”时期文化改革发展规划纲要[J].城市规划通讯，2012（04）：9+12.

[6]《国家“十三五”时期文化发展改革规划纲要》[N].人民日报，2017-05-08（001）.

[7]刘志江.略谈古籍数字化的问题与对策[J].出版参考，2019（10）：50-52.

[8]余力，管家娃.我国古籍数字化建设现状分析及发展研究[J].数字图书馆论坛，2017（11）：41-47.

[9]史睿.论中国古籍的数字化与人文学术研究[J].北京图书馆馆刊，1999（02）：3-5.

[10]邵正坤.古籍数字化的困局及应对策略[J].图书馆学研究，2014（12）：32-34+79.

[11]朱锁玲，包平.我国古籍数字化进展与研究述评[J].图书馆理论与实践，2009（09）：18-21.

作者简介：

严如月（1991.12-），女，汉族，湖北武汉，北京美斯齐文化科技有限公司，古籍整理与数字化，硕士，秦汉史研究。