

# 数字化,让古籍从历史深处走出来、活起来

11月12日,国家图书馆(国家古籍保护中心)等20家单位联合举办古籍数字资源共享发布活动,在线发布古籍数字资源7200余部(件),免费服务大众阅览和学术研究。这是国家图书馆(国家古籍保护中心)组织的第4次联合发布活动,也是参加单位最多的一次,不仅有省级馆,而且有很多市、区、县级公共图书馆,以及博物馆、高等院校、科研机构等系统的古籍收藏单位和私人藏书家。全国古籍资源发布总量已超过7.2万部(件),提前超额完成《“十三五”时期全国古籍保护工作规划》古籍资源发布任务。会上,与会专家就古籍数字资源的发布共享工作进行了深入座谈,围绕古籍数字化提出了各自的看法。现将部分专家的发言摘登如下,以飨读者。

## 数字化关系到国家文化命脉

荣新江

现在这个时代确实是一个数字化的时代,但是就中国来讲,我觉得起步确实有一些慢。我从事敦煌学研究的时间比较长,敦煌这个领域,过去我们是到国外去抄文献,后来一些出版社有了动力,就做了大量黑白的图录,如英藏文献、法藏文献、俄藏文献等。但是在中国做图录的时候,英国就已经开始做IDP(国际敦煌项目)。其实,英国人在敦煌学的研究上并没有很大贡献,但是在敦煌文献的数字化上确实有非常大的贡献,有超前的意识。

在“十三五”古籍出版规划讨论的时候,各个出版社都策划要在5年之中出多少本整理本古籍,实际上没有那么多古籍可以整理。我提出,与其这样,为什么不搞古籍数字化,将国家用于古籍整理的经费支持中华书局等单位,牵头开展数字化整理。当时有3位老总站起来说我们的书卖不出去,不能搞。但现在开放了,非搞不可了,中华书局也已经开始牵头搞了。实际上,古籍数字化整理的东西跟纸本是两码事,并不影响纸本古籍的发行,而古籍数字化一旦弄好了,各大图书馆必须买,没有不买的,所以它实际上可以产生很好的经济效益,而且古籍数字化的大部分经费来源于国家和各类基金支持。古籍数字资源非常便于学术使用,但与国外相比,我们的古籍数字化工作已经相当慢了。

然而,我们中国人多力量大,所以一旦着手开展这项工作的话,很快就能上去。当时讨论“十三五”古籍出版规划的时候,我就强调一点,数字化是一个国家战略,是一个国家的命脉。比如IDP的负责人魏泓要建成一个敦煌的数字帝国,而这个帝国都由她来把控,就是要把所有的数字化资源汇聚到英国。中国国家图书馆自己上传数据,但其他一些国家或较小的单位觉得没必要,就直接将图片数据交给英方上传,一旦设在英国的服务器坏了,我们就都看不到了。这就涉及国家战略问题,因为一旦由英方掌握数据上传权限的话,原数据由对方掌控,那边出点什么事,我们就没有办法做工作了。

所以我把当时的发言稿整理后,由程康先生先给李克强总理写了一封信,表明数字化是一个国家战略,所有的东西得由我们自己掌握,自己上传,这个原数据必须掌握在我们自己手里。虽然古籍是全世界的共同资源,但不管谁做过,我们都还要自己做。北京大学历史系的刘俊文先生是很有眼光的,他非常早就意识到了这个东西的重要性。现在他建立的古籍数字资源巨大,虽然是个人工程,质量存在一定问题,但毕竟他有这么大一个数据库,国外相关机构全都购买。

这是我第一点想说的,数字化是一个国家战略。

第二点,古籍的数字化会带来一个古籍重新发现

的时代。过去我们只知道大的馆藏,但还有很多小的馆藏,过去是看不到的。而现在一旦数字目录发布了,甚至全文发布以后,我们不论做什么研究,都有可能在这里无意之中发现很重要的研究素材。

我指导学生研究中外关系史,这种古籍文献的新发现对摩尼教、景教、三夷教的研究都非常重要。特别是对摩尼教和景教的研究,多年来在文献上几乎没有新的发现。在沙畹写完《摩尼教流行中国考》,陈垣写完《摩尼教人中国考》后,几乎就没有新的材料。我在《历代法书记》里面找到一条摩尼教徒和景教徒、佛教徒打仗的资料。外国人说你怎么样找到的?我说就在1924年的《大正藏》里,但是你们在读基督教、摩尼教文献的,你们不读禅宗文献。实际上,在《四库存目丛书》印出来之后,用功的学生就会好好翻,因为这里有很多陈垣先生当年没有翻过的书,而陈垣把《四库全书》里的书基本上都翻过,是找不到摩尼教材料的。但现在古籍数字化以后,我们就有可能在小的馆藏里面发现一些珍贵的材料。

所以古籍数字化主要是两项,一是目录先发布,再是全文发布。这真是古籍发现的新时代,对于我们研究某一个方面的人帮助很大,这也是一个知识扩大的过程。我们可以在不同的地方来阅读。原来要去北海那边的国图看书,且不说从外地过去,即便从北大过去一趟也不是很方便。

现在古籍数字化建设也真是迎来了一个很好的时代,包括刚才上海图书馆介绍的“众包”的做法,把一些社会力量调动起来做古籍的数字化工作。

最后我想说的是,建议将来的古籍数字化工作更多地跟学术界合作,跟一些研究者合作。从研究者的角度来说,我们也有很多需求,哪些古籍先发布,哪些后发布,这其实也是一个国家战略。在某些研究上我们如何抢占一些制高点,比如说现在我们在做“一带一路”的文献,特别像海上丝绸之路,这跟广东、福建、浙江这一带的地方图书馆有相当大的关系,特别是家谱这类文献,如果国家优先用资金支持这些图书馆开展这类古籍的数字化,那将是我们占据这个学术制高点的一个关键。当然做任何古籍的数字化,都会对于某一个学科的学术研究有用。比如说《赵城金藏》,虽然有《中华大藏经》的影印件,但是我们还是想看《赵城金藏》原本,这对我们现在整理旅顺博物馆收藏的佛典就非常有用。因为我们可以看到原本,我们不知道《中华大藏经》是不是被描改过。所以,我希望图书馆界能够在未来的古籍数字化工作方面多跟学术界对话,大家在方向上勾画一个先后次序。

(作者系北京大学教授)

国家图书馆等20多家单位联合在线发布古籍数字资源,又为我们提供了很多新的古籍资源。这是非常振奋人心的成果。

第一,2007年“中华古籍保护计划”的实施对中华优秀传统文化的传承发展发挥了重要作用。其中一个重要成绩是对大量珍贵古籍进行了普查登记和抢救性修复;另一个成绩是建成“全国古籍普查登记基本数据库”和“中华古籍资源库”,并面向社会广泛使用。目前,“中华古籍资源库”点击率在国家图书馆网站排名第一,这反映了全民保护古籍的意识正被逐步唤醒。因此,我认为“中华古籍保护计划”不仅对古籍抢救保护工作意义重大,更重要还在于唤醒了全民保护中华古籍的意识,以及对传统文化和中华典籍的热爱。同时,“中华古籍资源库”使用率的大幅提升,与国家古籍保护中心十几年的努力分不开。为此,国家古籍保护中心深入高校开展了大量卓有成效的宣传工作,建设了从本科直至博士教育的古籍保护教育和课程体系,使得很多高校学生对古籍保护和古籍修复产生了兴趣。

同时,我认为国家古籍保护中心在贯彻习近平总书记关于传承发展传统文化中坚持创造性转化和创新性发展方针指示方面举措得当。古籍数字资源发布数量日新月异,增长显著。前两天国家图书馆古籍数字化的数量还是3.2万部,今天就到了3.3万部。原来全国在线数字化古籍有6.5万部,现在已经跃升到7万多部。

第二,古籍数字化对古籍阅读、古籍保护、古籍整理工作,包括对于古文献学科都产生了巨大影响,引起了革命性的变化。近三四十年来,我国古籍数字化工作突飞猛进。1983年,我国最早出现的是“红楼梦”全文检索系统”。上世纪90年代初,我们使用的古籍数字化成果还是在摊上买到的一些劣质光盘。当时我买了一套“二十四史”的光盘,非常高兴,但回来后才发现上当了,第一部《史记》不仅错字极多,而且有大量的删节,很多材料找不到,根本无法使用。在本世纪初,我国的古籍数字化就发生了根本性变化,出现了“国学宝典”“中国基本古籍库”“瀚堂典藏”等具有多种功能的大型古籍数据库。爱如生、书同文、国学时代等古籍数字化企业也具备了强大的开发和竞争能力。在公共图书馆的古籍数字化和线上阅读方面,诚如上述,成绩更为显著。就国家图书馆线上阅读而言,不但能阅览国内许多原来深藏秘藏的珍贵古籍,还能阅览美国哈佛、日本东京所收藏的中文古籍善本。在高校里,很多学生都能充分利用多种大型古籍数据库。其中,就包括国家图书馆在线的古籍数字化资源。

近十年来,古籍数字化的发展给古籍整理和古文献学研究也带来了方法上的“革命”。比如,目前已经出现了计算机的联机自动校勘软件,在古籍数据库建设中实现了对古籍的自动校勘,以及校勘记的自动生成。又如,计算机对古籍自动标点的技术也已取得了明显成绩。首都师范大学电子文献研究所研制的古籍自动标点技术,采用前沿的人工智能算法,在若干次演示中其自动标点的准确率达到99%以上。还有,目前清华大学研制的中文古籍版本识别系统,也取得突破性进展;一些机构也正在研究计算机对古籍的自动辑佚、注释、翻译技术。这些方法的革新将为古籍整理工作带来全新的发展前景。

古籍的数字化技术对于古文献研究和学科建设

也带来许多新的理念。古籍数字化的新成果、新技术,要求古文献学给予及时的归纳和总结。首先,从文献学科来讲,首都师范大学电子文献所提出建立数字文献学,把它作为一个特色学科进行建设。其次,在文献载体材质上,应该更多地重视电子文献这一类型。其三,由于大量古籍数据库出现,其技术不断成熟,功能越来越多,产品质量也越来越高,如产出“国学宝典”的电子公司提出建设古籍数据库的“定本工程”,要求一些数据库的质量要达到零差错率。因此,在目录学书目著录上,应考虑增加新的分类,著录一批流通广泛、使用频繁的优秀传统文化数据库。如“中国基本古籍库”“瀚堂典藏”“国学宝典”等应属于多功能的综合古籍电子丛书;而像敦煌、家谱、方志等特色数据库,则应属于传统古籍电子丛书。版本学方面,不仅在传统的古籍版本中要增加电子版类型,而且在电子版中又可据产出机构的不同而分类,还可据古籍数字化的不同途径分为图像版、文字版或图文版等类型。至于校勘学、文献的e考据等,都可根据已有的新技术,总结文献考校的新方法和新途径,从而为古文献学增加新内容。我认为,今后古文献学教材要认真加以修订,为古籍数字化的发展和学理总结,增加大篇幅的内容。

第三,对古籍数字化发展充满期待。我个人对古籍数字化这一领域可以说很外行,没有专门的研究。虽然有时也会使用,也听到一些青年学生反映的问题,所以也想提一些粗浅的看法。首先,古籍数字化发展很快,成果很多,特别是公共图书馆上线的古籍数字化资源很丰富,但是我们希望更快更多,能看到其他未见的古籍善本。另外,希望能不断改进浏览和检索功能。

除此之外,在数字化古籍发展的新潮流中,国家古籍保护中心还要对新发展有所“预流”。陈寅恪先生讲,凡治学“得预此潮流者,谓之预流”。对于数字古籍这一新品种,也要纳入古籍普查、保护的视野,考虑数字化古籍的存储、保护,并逐渐形成《全国古籍数据库总目》,对国家数字化古籍的状况有所把握。这也是国家文化安全的需要。这个工作即使现在做不了,也要有预案。

古籍数字化联合发布会非常好,如果在联合发布后,再形成全国上线数字化古籍的联网,那就更好了。比如说,在国家图书馆网页登录之后可以进入全国公共图书馆的在线古籍数字化联网,则可使古籍数字资源得到更好的开发和利用。由此,我也考虑到还可以建设更多链接,更充分地利用海外中文古籍数字化的资源。目前国家图书馆网页已有美国、日本的一些数字古籍资源,但是还有很多没有利用,特别是像欧洲的一些中文古籍数字化资源。现在一些县级图书馆也有了上线的数字化古籍,非常了不起。由此我也想到,随着国学热和国家对传承发展中华优秀传统文化的推动,一些比较偏僻的地区也有大量热爱传统文化的群众,其中甚至有一些是国学高手,应该让他们也能充分利用古籍数字资源。所以,要尽量地考虑为他们提供更多的线上古籍数字资源服务。

最后,希望国家图书馆在线古籍数字资源能提供全文检索等更多功能。也希望今后国家图书馆能尝试和一些数字化企业合作,开发更多更好的古籍数字化数据库,共同推进我国数字化古籍和古籍保护事业的发展。

(作者系北京师范大学教授)

## 古籍数字化发展日新月异

周少川



中华古籍资源库



全国古籍普查登记基本数据库



上海图书馆家谱数据库

## 古籍数字化解古人未解之题

杜泽逊

我国当前的战略是由大国发展成强国。一个强国当然有若干配套要素,政治制度完善、文化教育繁荣、经济发展、科技领先等都是首要因素。这其中的文化教育繁荣和我们直接关系,我们是文化教育事业的参与者、建设者。

巧妇难为无米之炊,文化教育工作者需要书籍,就像一个人要吃饭一样。有些书可以从书店购买,或者从网上下载;可是有些书不能轻而易举地获得阅读的机会,这类书籍主要是古籍善本。

古籍善本长期以来秘不示人,几乎成了传统,历史上有名的藏书家在书上盖有这样的印章:“霏及借人为不孝。”古籍既不能卖,也不能借。钱谦益是明末排名第一的藏书家,绛云楼不慎失火,大部分藏书化为灰烬。他的好友曹溶去安慰他,同时也批评他吝于把藏书示人,致使绛云楼所藏孤本绝迹人间。为什么不给人看呢?主要还是为了保护。

清代黄丕烈收藏的校影宋本《博雅》,经过宋定之借给了王念孙。王念孙研究《广雅》,写了一部第一流的著作《广雅疏证》。王念孙借黄丕烈的《博雅》当然是为

了校勘,可是6年没有归还。后来黄丕烈在京城,见到王念孙的儿子王引之,问到此书。王引之回答说已经还给了宋定之。于是黄丕烈又向宋定之追索,果然在他那里,第二年才追回来。一部书借出去7年才追回来。

杨氏海源阁收藏的明嘉靖芝城活字蓝印本《墨子》是稀世珍本。这本书光绪年间被潘祖荫借去,长期未还。杨保彝在刊刻他父亲杨绍和的《楹书隅录续编》时特别加了注释,说这本书“为潘文勤师借书未还”,直到新中国成立后这本书才在苏州出现,现藏于国家图书馆。

历代的藏书家以个人之力收藏保护典籍,是传承中华文化的功臣,他们也是最重要的古籍专家。他们最关心的是珍贵古籍的安全,不外借是合情合理的。借书的不愉快,无论如何会得到后人的理解。

在今天,古籍的安全仍然是头等大事,如何实现既保护又流通呢?古人采取了影抄、照抄的办法来解决。乾隆时期的《四库全书》也是把献来的书录副本,原本发还,可惜没有贯彻到底。

传世的嘉靖影宋刻本、明清的影宋抄

本、近代的影印本,都是往这个方向努力的成果。还有一种校宋本、校本,把宋、元本文字的不同校在通行本上,也是这个目的。这种行为有很大的局限,因为能够被影印、影刻的本子数量很少,既受物质条件的限制,也存在思想保守的因素。

今天我们有了数字化和网络传递的技术,可以跨国、跨地区,在自己家里、在研究室里,查看异地的藏品。今天在座的荣新江先生脚踏铁鞋,到世界各地访求敦煌吐鲁番文献,他的前辈王重民先生、姜亮夫先生、刘半农先生,以及更早的黎庶昌、杨守敬、董康等先辈,也都是采用走访这种方式。显然这种方法有很大的局限性,因为一般人无法像他们那样随时出去看书,更重要的是一般人看到了也看不懂。要克服这个难题,就要请荣先生这样的专家鉴别原物,进行学术性著录,然后再把图像挂出来。图像加上专家的著录就达到了古籍数字化的基本要求,或者说底线要求。读者上网看书,不仅看内容,还享受了专家鉴别研究的成果。

当然荣新江先生说了,数字化会带来新发现。北京大学教授张丽娟发现了著名

的十行本《尚书注疏》的元刊原印本,收藏在国家图书馆,这是过去不知道的。昨天一位青年同志说他发现杨复《仪礼图》元十行原印本,也在国家图书馆,这也是过去不知道的。这类发现会带来很多新认识,因为元十行本,从清代以来用的是明代历次修版,基本上未见到未修版的印本,原印本的出现当然会解决很多悬而未决的问题,甚至会改变原有的结论。这都是古籍数字资源发布给一般读者带来的发现。

古籍数字化及免费阅读并不是一帆风顺的,需要克服很多主观和客观的困难。国家图书馆作为全国图书馆的龙头老大,在数字化方面也起到了表率、带头作用。这次“中华古籍资源库”网上公布成果达到了3.3万部,带动了其他图书馆,甚至得到了私人收藏家韦力先生的赞成和参与。数字资源公布达到了7万余条,这一大好局面是国家古籍保护中心积极规划、努力推动达成的,这是为中国建设文化强国作出的重大贡献,是我国古籍史上的历史性转折,必然对学术文化的繁荣发挥长效的、不可替代的巨大作用。

(作者系山东大学教授)