

·理论园地·

中文古籍数字化的主体构成及协作机制初探

李明杰 俞优优 (武汉大学信息管理学院 湖北武汉 430072)

摘要: 通过网络和文献调研,将中文古籍数字化主体归为图书馆、学术机构和数字企业三大类型。以表格的形式展示了各大主体的古籍数字化成果,分析了它们各自的数据类型、系统功能及选题分布方面的特点,指出它们之间存在很强的互补性。以理论探讨的方式,提出了古籍数字化协作机制的构建途径。

关键词: 古籍 古籍数字化 主体构成 协作机制

中图分类号: G255.1

文献标识码: A

文章编号: 1003-6938(2010)01-0034-11

A Tentative Study on Subject Formation and Cooperative Mechanism of Chinese Ancient Book Digitization

Li Mingjie Yu Youyou (School of Information Management of Wuhan University, Wuhan, Hubei, 430072)

Abstract: The subjects of Chinese ancient book digitization are libraries, academic institutions, and digital enterprise through network and document investigation. This paper reveals, in the form of tables, the respective digital achievements of the three kind of subjects, analyzes their characteristics of data type, system function, and topic distribution, and points out the strong complementarities of them. The construction way of cooperative mechanism for ancient book digitization was put forward with theoretical probe.

Key Words: ancient book; ancient book digitization; subject formation; cooperative mechanism

CLC number: G255.1

Document code: A

Article ID: 1003-6938(2010)01-0034-11

中国古籍数字化始于20世纪80年代初,历经近30年的发展,已完成了以下三个转变:数字内容由单一的书目数据向全文数据、图像数据和图文数据相结合的转变;载体形式由单机版向光盘版、网络版的转变;系统功能由检索工具向知识工具、研究工具的转变。迄今为止,不少学者已撰文对中国古籍数字化的总体发展作了总结,如李国新《中国古籍资源数据库的进展与任务》^[1]、潘德利《中国古籍数字化进程与展望》^[2]、陈力《中国古籍数字化的现状与展望》^[3]、段泽勇《古籍数字化的回顾与展望》、吴家驹《中文古籍数字化的进展与主要成果述评》等^[4]对中国古籍数字化的发展阶段、成果类型、存在问题等作了宏观考察,并对古籍数字化发展方向进行了展望。鉴于此,本文不再赘述以上内容,而是从古籍数字化的主体构成入手,分析每种主体古籍数字化成果的数据类型、系统功能、选题分布等方面的特点,探讨各种主体协作开发数字化古籍的可能性及具体的协作机制。

笔者通过网络调研和文献调查发现,参与古籍数字化的主体主要由三类机构构成,即图书馆、学术机构和数

字企业。从构成数量来看,图书馆最多,有79家,占65.3%;学术机构其次,有33家,占27.3%;数字企业最少,只有9家,占7.4%。以下是对这三类主体的古籍数字化产品的具体调研情况。

1 图书馆及古籍数字化成就

根据调查结果,我们将国内外各类型图书馆所建古籍数据库的情况,按来源、数据库名称、数据类型及数量制成表1。

表1显示,各类型图书馆共建有古籍数据库121种。从数据类型来看,纯书目型数据库最多,有72种,占59.5%;图文型数据库有13种,占10.8%;纯文本的全文型数据库、纯影像数据库及书目与影像相结合的数据库均为12种,占9.9%。这说明图书馆主要是提供书目数据,部分能提供全文与影像信息。从功能来看,图书馆所建古籍数字化系统大多只能进行一般的书目检索和全文或影像的浏览。这表明,图书馆的古籍数字化水平还比较低,仍停留在揭示馆藏的层面,但不容否认,图书馆在古籍数字化领域的

基金项目 本文系国家社科基金青年项目“版本文化遗产保护与开发利用机制研究”(项目编号:08CTQ005)研究成果之一。

收稿日期: 2009-10-23;责任编辑:魏志鹏

表1 图书馆及其古籍数字化成果概况

来 源	数据库名称	数据类型及数量		
		书目	全文	影像
国家图书馆	中国古籍善本书目导航	29 万余条		
	碑贴菁华(中文拓片资源库)	23000 条		29000 副
	数字方志	6868 种		
	敦煌在线(IDP)资料库			38080 副
	西夏碎金	124 条		5000 副
中国科学院图书馆	古籍管理与检索系统	50 万条		
首都图书馆	明清北京城垣资源库		59 种 202 段落	358 副
	古籍插图库			10000 副
北京大学图书馆	数字图书馆古文献资源库	156080 条		
	馆藏古籍目录	20000 条		
	科技典籍目录库	13000 条		
	科技典籍全文库		100 种	
北京师范大学图书馆	线装方志书目数据库	2898 条		
	50 年来整理出版清代诗文集书目	2023 条		
	全元文篇名作者索引	33728 条		
北京语言大学图书馆	古典文学数据库		55 种	
北京大学图书馆、北京师范大学图书馆、南京大学图书馆、四川大学图书馆、香港中文大学图书馆、吉林大学图书馆及华东师范大学图书馆	学苑汲古——高校古文献资源库	20 余万条		部分或全文图像
首都医科大学图书馆	中医古籍电子全文数据库		近 100 种	全文图像
	农书古籍全文数据库		9 种	全文图像
天津图书馆	古籍善本图录	60 条		60 副
	天津地方史志丛书		25 种	
山西省图书馆	山西家谱书目数据库	94 条		
	谱牒资料库		建设中	
内蒙古自治区图书馆	古籍文献书目数据库	不详		
内蒙古师范大学图书馆	馆藏古籍书目数据库	7000 余条		
吉林大学图书馆	古籍音韵学书目库	478 条		
	古籍精品图片库			11 种
东北师范大学图书馆	特藏文献数据库	108 条		108 副
辽宁省图书馆	馆藏古籍书目数据库	24717 条		
大连市图书馆	明清小说全文库		40 种	
	古籍线装书目录	8910 条		
	善本书目库	1710 条		
上海图书馆	古籍书目数据库	129660 条		
	宋元善本全文数据库		20 种 19794 页	
	家谱数据库	16832 条		部分影像
复旦大学图书馆	古典文献数据库 :包括清人碑传索引、明人传记词典、古籍题记索引、近五十年古籍整理书目、四库系列图书综合引得、清人文集书目、明人文集书目、元人文集书目	包括复旦大学图书馆、国家图书馆、中科院图书馆等国内图书馆及台湾、日本、北美地区图书馆的收藏		
南京图书馆	中国传统体育图片数据库			1488 副
	馆藏精品			23 副

来 源	数据库名称	数据类型及数量		
		书目	全文	影像
金陵图书馆	中国古本小说目录	428 条		
	四库全书目录	3577 条		
	续修四库全书目录	5213 条		
	年谱丛刊目录	1200 余条		
	中国佛寺志丛刊目录	197 条		
	丛书集成目录	100 余条		
常州图书馆	古典文献库(普通古籍、地方文献、家谱)	15000 余册		
	地方志精选		4 种	全文影像
苏州图书馆	古籍善本目录	1710 条		
	古籍善本书录		12 种 502 卷	全文影像
	苏州地方文献联合目录	7648 条		
常熟图书馆	古籍地方志		2 种	全文影像
吴江图书馆	吴江五百年古代地方志		24 种	
扬州大学图书馆	古籍书目数据库	8747 条		部分书影
	精选古籍数据库	440 条		
苏州大学图书馆	清代图像人物研究资料数据库			5949 副
	中医药古籍文献数据库		156 种	全文影像
	中国汉代画像石砖数据库			6819 副
浙江省图书馆	浙江省图书馆家谱数据库	26574 条		
	家谱全文数据库		56 种	
	中国历代名人图像数据库(上、下)			38756 副
宁波图书馆	古籍善本书目	19 条		
嘉兴图书馆	馆藏家谱	38 条		
绍兴图书馆	绍兴方志		8 种	100 副
	地方碑帖	147 条		
	家谱数据库	553 条		
	馆藏善本数据库	698 条		
湖州图书馆	湖州家谱提要数据库	不详		
浙江大学图书馆	普通古籍数据库	9388 条		
	善本古籍数据库	1645 条		1645 副
宁波大学图书馆	古籍书目数据库	1 万余条		
山东省图书馆	古籍书目库	3 万余条		
济南市图书馆	古籍地方文献	51 种	2037 副	
	金石拓片		270 种	5000 余副
山东大学图书馆	古籍书目查询系统	4500 条		
安徽省图书馆	古籍善本目录	5418 条		
	安徽地方志资源库	610 条		
	安徽家谱数据库	763 条		
安徽大学图书馆	馆藏古籍题录数据库	不详		
福州图书馆	族谱目录	230 条		
泉州市图书馆	馆藏谱牒	590 条		
九江市图书馆	古籍善本书目	45 条		

来 源	数据库名称	数据类型及数量		
		书目	全文	影像
江西师范大学图书馆	线装古籍书目数据库	4376 条		
武汉大学图书馆	善本书目录	794 条		
	史部书名目录	2637 条		
中南民族大学图书馆	线装古籍书目	999 条		
中南财经政法大学图书馆	古籍目录	2406 条		
湖南省图书馆	古旧文献数据库	67661 条		
广东省图书馆	古籍善本、线装古籍、影印古籍库	不详		
华南师范大学图书馆	新版古籍丛书书目数据库	29909 条		
中山大学图书馆	新影印古籍和新方志	不详		
重庆市图书馆	古籍地方志目录	82 条		
成都市图书馆	古籍善本目录	236 条		
广西师范大学图书馆	馆藏广西旧地方志		不详	
	馆藏古籍善本		不详	
贵州省图书馆	贵州省古籍联合目录	14080 条		
	贵州地方志全文数据库		127 种	
贵州师范大学图书馆	贵州省级地方志全文库(旧志)		15 种	
云南省图书馆	馆藏经典			47 副
云南大学图书馆	善本书目	1007 条		
陕西省图书馆	古籍珍品在线			64 副
西北大学图书馆	馆藏古籍书目数据库	不详		
西北师范大学图书馆	馆藏南阳拓片图录			100 副
兰州大学图书馆	敦煌遗书库	22000 余条		部分影像
宁夏大学图书馆	古籍善本目录	130 条		
台湾中央图书馆	古籍影像检索系统	37169 条		9097 副
台湾中央图书馆	明人文集资料库		17 种	部分影像
	中文古籍书目资料库	61 万条		
	台湾地区家谱联合目录	14945 条		
	金石拓片资料库	6462 条		12462 副
台湾成功大学图书馆	甲骨文全文影像数据库			41956 副
香港中文大学图书馆	汉达文库		近 1 亿字	18000 副
香港大学图书馆	日本见藏中国丛书目录	947 条		
香港大学冯平山图书馆	馆藏善本书录	727 条		
日本东京大学综合图书馆	汉籍目录数据库	数量不详		
日本京都大学图书馆	日本所藏中文古籍数据库	62 万条		部分影像
	东方学图书馆藏书目录		87 部	全文影像
韩国汉城大学图书馆	奎章阁档案馆汉文古籍库	数量不详		
大英图书馆、牛津大学、剑桥大学等六所大学及研究所	UK Union Catalogue of Chinese Books	英国 6 所大学及研究所收藏的中文古籍书目资料		
剑桥大学图书馆	Cong shu in Cambridge Libraries	1529 条		
美国国会图书馆	Library of Congress Online Catalog	包括美国收藏的宋元明清善本书、地方志和罕见的珍本图籍的书目资料		
哈佛大学燕京图书馆	HOLLIS Catalog	7900 条		
普林斯顿大学东亚图书馆	East Asian Library and the Gest Collection	3000 条		

潜力非常巨大。从选题来看,公共图书馆非常注重根据自己馆藏的地方特色来开发本土文化资源,特别是地方文献(方志及家谱等)的数字化得到了重视,如贵州省图书馆、重庆图书馆、绍兴图书馆、常熟图书馆、湖州图书馆、吴江图书馆等开发了地方志数据库,山西和安徽两省图书馆、绍兴图书馆、福州图书馆等开发了谱牒数据库,较好地满足了人们了解乡土文化的需求。一些具有专业特色的数据库较好地满足了人们专业研究的需求,如首都图书馆开发的明清北京城垣资源库、清华大学图书馆的科技典籍全文库、中国农业大学图书馆的农书古籍全文

数据库、南京图书馆的中国传统体育图片数据库、苏州大学图书馆的中国汉代画像石砖数据库等。从整个数字化成果的分布来看,以小型数据库居多,资源开发重复率小,公益性和地域性是其特色。

2 学术机构及古籍数字化成果

科学研究是实践的先导。早在上世纪70年代,一些学术机构(个人)就开始研制古籍机编索引。通过文献调研及对相关网站的访问,我们回溯性地列出了相关学术机构在古籍数字化领域取得的成果,如表2所示。

表2 学术机构及其古籍数字化成果

学术机构(个人)	数字化成果	时间	备注
德国汉堡大学	诗经索引	1975 年	由吴用彤编制的英文文本索引
[美]P.J.Ivanhoe 等	朱熹大学章句索引、朱熹中庸章句索引	1978 年	字、词索引
	王阳明大学索引、王阳明传习录索引		
	戴震原善索引、戴震孟子字义疏证索引		
挪威奥斯陆大学	先秦诸子百家全文检索系统	不详	全文检索
日本东京大学东洋文化研究所	汉籍善本全文影像资料库	不详	收入善本 3829 部,全文影像
	汉籍目录		7 万条书目数据
江苏省镇江市科委	红楼梦电脑处理系统	1983 年	彭昆仑先生负责
深圳大学	红楼梦多功能检索系统	1984 年	分语言、文学等 30 个专题,200 多个项目的专书检索系统
山东社科院语文所	论语逐词索引	20 世纪 80 年代	逐词索引
台湾中央研究院史语所	汉籍电子文献系统	1984 年至今	史书、经书与子书、宗教文献、医药文献、文学与文集、政书、类书与史料汇编等六大类
	傅斯年图书馆善本古籍检索系统		含善本古籍检索、人名权威检索和书目资料库(限馆内使用)
	傅斯年图书馆珍藏图籍全文影像库		24554 副原文影像(需授权使用)
	考古资料数位典藏资料库		收录与考古工作相关的各式照片、图表、记录等 80981 件
	内阁大库档案资料库		230359 副明清档案全文影像
	内阁大库档案疾病医疗相关资料库		120 条,人名、事由、关键词、官衔检索
	汉墓资料库		含汉墓、铜镜/玺印、汉画资料库
中国社会科学院	全汉字系统、论语数据库、诸子集成数据库、全唐诗检索系统	1984 年	由文学研究所栾贵明、田奕等完成。具有单字、人名、地名、援引、专有名词、补遗、衍文等 7 种索引形式
	寒山子诗索引	1998 年	由姚兆炜完成
	全唐诗、先秦魏晋南北朝诗、全上古三代秦汉三国六朝文、十三经、全唐文、诸子集成等数据库检索系统	20 世纪 90 年代	
陕西师范大学	辞书电脑处理系统	1987 年	在编撰《十三经辞典》和《十三经词语索引》的基础上展开
河南大学	宋人笔记检索系统南宋主要历史文献	1987 年	机构内部试用
南京大学、河南大学、苏州大学联合研制	计算机甲骨文信息处理系统	1987 年	采用义形四位等长码输入甲骨文,与激光照排系统连接

学术机构(个人)	数字化成果	时间	备注
台湾师范大学中文系	红楼梦、水浒传全文检索系统	1987 年	陈郁夫教授
东北师范大学古籍整理研究所	贞观政要综合检索系统	1988 年	单字索引、专词索引
香港中文大学中国文化研究所中国古籍研究中心(汉达文库)	甲骨文全文数据库	从 1988 年至今	整理校勘约 100 万字甲骨卜辞
	金文全文数据库		来自社科院所编《殷周金文集成释文》
	竹简帛书出土文献数据库		重新标点校勘各重要竹简、帛书文献
	先秦两汉一切传世文献数据库		取旧刻善本未改者,重新标点校勘
	魏晋南北朝一切传世文献数据库		取旧刻善本未改者,重新标点校勘
	中国传统类书数据库		六千万字,将类书引文与传世文献对比
	古文献数据库现行计划之拓展与完善		传世类书数据第二期与竹简帛书新出土数据
	先秦两汉引录经籍研究计划		全面检索先秦两汉文献引录十三经文例
	先秦两汉互见文献研究计划		全面搜集先秦两汉所有典籍互见文献
陕西中医研究院	中医十三经通检	1994 年	药性相克相辅索引
哈尔滨师范大学	史记全文检索系统	20 世纪 90 年代	
湘潭大学	宋词别集索引三种	20 世纪 90 年代	采用 SYBJ 软件编制逐字索引、句子索引、人名索引、地名索引、专题索引
四川大学古籍所	全宋文、现存宋人著述总录、中国地方志宋代人物资料索引、朱熹集	1994 年	用 IDX 系统,按笔画、拼音、四角号码对著作、书名进行自动索引
广西大学	古今图书集成索引续编检索系统	1999 年	分类检索、标题检索、布尔检索、图谱检索及全文检索;具有古今地名之对照、特殊字及考证数据之查询等功能,全部共 1.44 亿字。
南开大学组合数学研究中心(与天津永川公司合作)	二十五史全文阅读检索系统	20 世纪 90 年代	按目录分类浏览,关键词全文检索
兰州大学敦煌学研究所	敦煌学数据库	不详	
南京师范大学	全唐五代宋词索引	2004 年	作者索引、词牌索引、词牌正名索引、首句索引、字频表,按作者、词牌、词牌正名和首句的汉语拼音编制索引目录
	全金元词索引	2004 年	
	全宋词索引	2000 年	
北京大学中文系	全宋诗分析系统	2005 年	包括《全宋诗》完整数据库,依《佩文韵府》、《广韵》等韵书建设并扩展的音韵库,诗歌模型库,古代诗人复合称名数据库,繁简字一对多关系库
首都师范大学电子文献研究所	中国古代文学史电子史料库、中华大典	2005 年	多途径检索和浏览
台湾汉学研究中心	明人文集联合目录及篇目索引资料库	不详	全文、书名、作者、馆藏地项目的检索
台湾大学佛学研究中心	佛学数位图书馆暨博物馆	不详	含中文、藏文、梵文、巴利文佛典,有书目数据 189226 条,全文 13193 种
台湾国立故宫博物院	故宫寒泉检索系统		十三经、二十五史、先秦诸子、全唐诗、宋元学案、明儒学案、朱子语类、资治通鉴、续通鉴等。不能浏览全文,但支持复合查询
台湾元智大学	"网络展书读"		以古典文学资料为特色,含多个独立的全文检索系统,诗经,可关键字、诗句、人物、地名、模糊诗句、全文检索等
台湾中央大学数学系	中国诗词文学选读		中国文学选集及佛教经典全文资料
台湾大学	玄奘西域行数位博物馆		玄奘丝路之旅、西域文物艺术、玄奘传等资料

从表2可以看出,学术机构开发的古籍数字化系统具有以下特点:一是时间跨度比较长,参与的机构和个人数量比较多;二是注重满足教学和科研的实际需要,选题多集中在古典文学和历史研究领域,非常具有针对性和实用性;三是系统功能参差不齐,早期开发的系统一般只有简单的字词索引功能,而后期的古籍数字化系统功能比较强大,有的甚至突破了全文检索的信息提供模式,如北京大学开发的《全宋诗》分析系统具有重出诗提取、格律诗标注、字及字组的频率分布统计、用户自作诗的格律分析等功能,带有智能化特点,在数据深层挖掘和知识发现方面具有开创性意义。该系统提供多维的检索分析方式,为中国古代文学、古代汉语、文献学等研究领域提供了可靠的分析数据。它标志着计算机科学在中文信息处理应用方面由全文检索的信息提供模式开始转向智能分析模式。

3 数字企业及古籍数字化产品

随着信息技术的发展和出版体制的放开,一些新兴的数字企业凭借其掌握的数字出版技术,借助现代企业运作模式和管理手段,很快就在古籍数字化领域立稳了脚跟。表3展示了国内数字企业及其古籍数字化产品的基本情况。

表3调查的结果显示,数字企业开发的古籍数字化产品以大型的综合数据库(如中国基本古籍库、国学宝典)和丛书数据库(如《四库全书》、《四部丛刊》、《二十五史》等)为特色,比较成系统,功能强大,基本没有书目型数据库。除了能进行多途径的全文检索外,一般都配备有各种知识辅助工具,如联机古汉语字典、年号与公元纪年对照表等,有的还提供字数、字频、词频的统计信息,行文风格特点的概率统计,参考数据的汇聚;有的还能提供不同版本比勘校对的接口、异体字的汇聚显示、读音的自动标注和朗读、在线标点断句、背景音乐等功能。这些知识工具极大地改善了古籍研究条件,也带来了研究者研究思路和方法的革新。作为古籍数字化的一支新兴力量,参与古籍数字化的数字企业数量虽然不多,但取得的成绩却令人刮目相看。这与数字企业的灵活的市场运作模式有很大关系,它们或依托图书馆的古籍善本资源,或吸纳文史专业研究人员参与研发,或与高校学术机构联合成立电子文献研究所,通过大规模、成系统地常用基本古籍数字化,很大程度上满足了专业研究人员需要。但需要指出的是,它们开发的数字化产品,选题重复情况比较严重,一味追求资源的规模和总量,有地方特色和专业特色的数据库不是很多,彼此之间缺少协作。

4 跨行业的古籍数字化协作机制初探

主体身份的多重性,带来的必然是古籍数字化价值取向的多元化。作为事业单位的图书馆,其从事古籍数字化活动更多的是出于保存民族文化的职责,追求的是社会效益;作为科研单位的学术机构,其从事古籍数字化研究和实践,更多的是出于科研和教学的实际需要,追求的是学术价值;而作为商业单位的数字技术企业,追逐的当然是商业利润的最大化和产业规模的扩张,即所谓的经济效益。通过对上述三类古籍数字化主体的分析,我们发现它们之间存在很强的互补性。对于公益性的图书馆而言,丰富的古籍馆藏是其优势。据不完全统计,全国图书馆系统保存的古籍有2750万册,其中可列入善本的超过250万册。足以弥补传统古籍出版存在的版本低劣重复的缺憾。图书馆拥有的古籍鉴定与整理的资深专家,亦堪称宝贵的人力资源;学术研究机构在古籍资源的甄别选择、选题策划、内容整理、系统功能的规划等方面是无可替代的;数字企业则在数字出版技术、资金的运作和投入、电子商务营销等方面具有不可替代的优势。而且,图书馆和科研机构可以成为古籍数字化产品最大的用户群。它们通过合作与融合,完全可以达到资源共享、优势互补、互惠共赢的目的。但关键是要有一个全国性的跨行业的协调机构,它既要跨行业,又要有一定权威性。为此,笔者建议仿照1958年成立的“国务院古籍整理出版规划小组”,在文化部社会文化司下成立“古籍文献数字化规划小组”,由它来统一组织和协调全国范围内的古籍文献数字化工作。

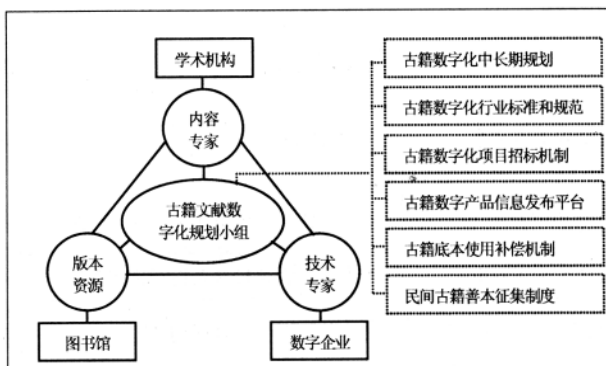


图1 古籍数字化协作机制示意图

如图1所示,古籍数字化主要由版本资源、内容专家、技术专家三大要素构成,分别由图书馆、学术机构、数字企业提供。由于古籍数字化仍属古籍整理和学术研究的范畴^[6],因此在整个数字化过程中,内容专家应处于主导地位,应由他们提出选题并进行科学论证,对古籍数字化对象进行版本鉴别和文本校勘,对系统目标和功能进行

表3 数字企业及其古籍数字化产品

数字企业	数据库名称	数据类型及系统功能		
		全文	影像	功能
北京书同文数字化技术有限公司	历代石刻史料汇编	15000 篇, 1150 万字	原文版本影像与数字化文本并存, 两者既可独立浏览, 也可对应切换	逻辑检索、联机字典、文字代码页与原书图像页关联、复制打印、添加注释、纪年换算等
	十通	3 千万字		全文检索
	四部丛刊	504 种, 9 千万字		书名、作者、分类及全文检索; 个人笔记、联机字典、干支公元换算、古今纪年换算
	康熙字典	收 47035 字		可按单字、部首、笔画、笔顺、拼音、注音查询; 提供中、日、简、繁、异等汉字关联代换检索技术
	大清五部会典	康熙、雍正、乾隆、嘉庆、光绪五朝会典全文		汉字关联检索, 中西日历转换, 加注笔记, 标注书签, 标点和校勘
	大清历朝实录	4441 卷		
北京国学时代文化传播股份有限公司	国学宝典	4000 余种, 10 亿字		标题、分类、专书、全文检索; 国学字典、人名词典、书名词典、国学字箴
	国学备览	81 种	2000 余副	内嵌即时古汉语字典功能
	书法备览	39 种	100 余副	按书家姓名、字号、爵里检索
	绘画备览	33 种	300 余副	全文多条件检索、复制、打印; 联机古汉语字典
	兵学备览	28 种		
	蒙学备览	75 种		
	唐诗备览	50000 余首		
	宋词备览	词 20430 首, 词论 420 篇		
	元曲备览	100 部元曲, 10 万字元曲史		
	篆刻备览	论著 59 种		
	小说备览	100 种	400 余幅印款	
	宋会要辑稿	366 卷		
	中国历代基本典籍库·隋唐五代卷	136 部近 1 亿字	2000 副图片	
	古代小说典	1000 种 1.5 亿字		
	国学备要	280 种经典		
	通鉴全编	4 种		
	清稗类钞	13500 余条		
	全唐文	20025 篇	有插图	
	文献目录典	40 多种 2000 多万字	1000 副书影	
	中国古代文学史	600 多种	数千副插图	
	中国古代戏剧专辑	310 种		
	崔东壁遗书	20 余种	有少量图片	
	飞鸿堂印谱		3500 方印章	
	三国演义电子史料库	全文及上千万字的相关文献资料	有插图	全文检索, 自由复制和打印, 文本对比、图文对照、同词脱文分析、文本差异分析、相似程度分析、小说历史对照
	中国历代笔记	1150 种 1.5 亿字		全文检索、逐字查询, 配备人名、书名、地名等数据库
	全上古三代秦汉三国六朝文	741 卷		可全文检索、复制、打印, 增加了联机字典、历代帝王纪年表等
	古代文论典	200 多种		
	香艳丛书	335 种	有图片资料	
	国学精品图库		1000 余副	

数字企业	数据库名称	数据类型及系统功能		
		全文	影像	功能
北京时代瀚堂科技有限公司	小学工具库	220 种	有原书影像	目录树浏览； 可全文、标题、出处、书目检索
	类书集成库	12 种约 3 亿字		
	出土文献库	6 个子库	图文对照	
	中医药文献库	700 多种	多有图文对照	
	经部集成库	6 个子库		
	史部集成库	近 1000 种	史料笔记有图文对照	
	子部集成库	5 个子库		
	古曲戏曲库	1000 余种	图文对照	
	佛教典籍库	8 个子库		
	道教典籍库	4 个子库		
	敦煌文献库	7 个子库	图文对照	
	专题文献库	8 个子库		
北京爱如生数字化技术研究中心	中国基本古籍库	1 万余种 18 亿字	1200 万副	分类、条目、全文及高级检索； 版式设定、字体转换、背景音乐、版本对照、放缩控制、 标点批注、阅读记忆、分类收集、下载编辑、原文打印； 版本速查、常用字典
	国学要籍系列	1 千种	原版影像	
	断代史料系列	1 千种		
	通俗大观系列	2 千种		
	诸书集成系列	2300 种		
	别集丛编系列	1 万种		分类、条目、全文及高级检索； 放缩、高清、翻转、连缀、设置、标注、书签、打印、下载
	地方文献系列	1 万种		
	中国方志库	1 万种 20 亿字	1 千万副	
	中国谱牒库	1 万种 20 亿字	1 千万副	
	中国金石库	10.1 万种 3 亿字	1 百万副	
	明清档案库	30 万件,10 亿字	3 百万副	
	中国类书库	800 部 5 亿字	3 百万副	
	中国辞书库	1 千种 3 亿字	1 百万副	
	中国经典库	1 万种 20 亿字	1 千万副	
	中国史学库	3 千种 5 亿字	3 百万副	
	中国俗文库	1 万种 10 亿字	5 百万副	
	历代别集库	1 万种 20 亿字	1 千万副	
	敦煌遗书库	3 万件 5 千万字	1 百万副	
	敦煌遗珍	3 千件	逐页对照原版影像	
	搢绅全书	近百种		
	永乐大典	813 卷		
	道教全书	2000 种		

数字企业	数据库名称	数据类型及系统功能		
		全文	影像	功能
北京爱如生数字化技术研究中心	历代碑志	1000种		
	全清经解	2000种		
	禅籍汇纂	500种		
	明清实录	27种 7356卷		
	千人年谱	1200种		
	清帝朱批奏折	10万件	逐页对照原版影像	分类、条目、全文及高级检索； 放缩、高清、翻转、连缀、设置、标注、书签、打印、下载
	明代日用类书	30余部		
	古今图书集成	6000余种		
	新编笔记小说大观	2000种		
	增订二十五史补编	500种	原版影像	分类、条目、全文及高级检索； 版式设定、字体转换、背景音乐、版本对照、放缩控制、标点批注、阅读记忆、分类收集、下载编辑、原文打印、版本速查、常用字典
	佚书合编	2000余种		
	增订四部备要	413种		
	增订丛书集成初编	100部丛书 4100种典籍	逐页对照原版影像	分类(编、部、类、目)、条目(书名、时代、作者、版本)检索、阅读记忆、分类书签、标点批注、原文打印
	全四库	9000种		
	丛书总汇	3000余部丛书 6万余种典籍		
	古版画		1万副	
	古籍提要便览	5万种古籍提要		快速检索和浏览、自由编辑
	古籍版本速查	5万种古籍版本		快速检索和浏览、自由编辑
	历代人物传略	5万个历史人物		
	爱如生拇指数据库	有500个产品,可量身定做		分类、条目、全文、高级检索、编辑下载、原文打印
数字方舟信息技术有限公司	古籍图书库	288种		全文浏览
北京天安亿友公司	二十五史、资治通鉴等			全文浏览
北极星书库	中华古籍库	1268部		按目录浏览全文
超星数字图书馆	电子图书	古今图书集成、清实录、二十五史、资治通鉴、中国古典名著大系等		按照中图法分类,通过超星浏览器可全文阅读
香港迪志文化出版有限公司 (与北京书同文和上海人民出版社合作)	文渊阁四库全书	3462种	原文图像对照	全文、分类、书名、著者检索、资料管理、单字字义查询、写字板、古今纪年换算、干支/公元年换算、八卦·六十四卦表、汉字关联、背景音乐、版面颜色

整体规划,而技术专家只是服务于内容专家,使这些目标和功能顺利实现。这好比建筑师与建筑工人的关系,由建筑师设计好图纸,建筑工人只负责施工。而“古籍文献数字化规划小组”处于整个古籍数字化协作机制的核心地位。在它的统一领导下,对全国古籍数字化工作进行整体规划,具体包括以下工作:

(1)制定古籍数字化中长期规划。古籍数字化是一项保护和弘扬中华文化的工作,具有强烈的公益性色彩,不能完全走市场化的道路,进行合理的规划是必要的。这样做既可以避免选题重复开发,也有利于古籍保护。在传统古籍整理出版领域,古籍整理出版规划小组先后制订了《整理和出版古籍十年规划(1962-1972)》、《古籍

整理出版九年规划(1982-1990)》、《古籍整理出版十年规划(1991-2000)》,对古籍整理出版工作进行全面安排和统一部署。古籍文献数字化规划小组可以借鉴此类做法,组织古籍专家对古籍进行定级和筛选,制定阶段性的古籍数字化规划,优先将文物价值、学术价值和艺术价值高的善本数字化,并有计划按步骤地推进该规划的实施。建议在古籍文献数字化规划小组下设立学术委员会,委员由一些学术机构知名的古籍整理专家担任。由他们负责制定古籍数字化中长期规划,提出各阶段古籍数字化的目标、措施,制订选题计划并遴选古籍版本,就系统的目的和功能提出建设性方案。

(2)制定和推广古籍数字化行业标准和规范。当前古籍数字化领域存在的一个突出问题就是数据格式繁多的问题。这种“各自为政”局面,不利于古籍数字化的发展及古籍数字资源的共享,一些新开发的系统不得不考虑与其他文档格式兼容的问题。但从长远来看,建立并规范古籍数字化的元数据标准是必由之路。从2002年开始,由科技部委托国家科技图书文献中心协调,中国科学院文献情报中心、中国科学技术信息研究所、国家图书馆、CALIS管理中心等21家单位联合进行了数字图书馆的相关标准规范研究,其中直接与古籍有关的包括《舆图描述元数据著录规则》、《舆图描述元数据规范》、《拓片描述元数据著录规则》、《拓片描述元数据规范》、《家谱描述元数据著录规则》、《家谱描述元数据规范》、《古籍描述元数据著录规则》、《古籍描述元数据规范》、《地方志描述元数据著录规则》、《地方志描述元数据规范》等。此外,还制订了不同类型数字资源的加工标准和规范,如对古籍、拓片、舆图、字画、手稿、文牋、契约以及甲骨、金石、竹简、陶器等实物的三维造型等。如果成立了古籍文献数字化规划小组,组织协调工作就可固定由它来做,包括这些标准和规范的推广教育工作等。

(3)建立古籍数字化项目招标机制。在制定了古籍数字化整体规划后,应向社会公开选题方案,通过招标的方式向全社会征集有资质、技术力量强、开发经验丰富的数字技术企业来承担该系统的设计开发。通过公平、公开、公正的有序竞争,使得各种资源得到最优配置,保证古籍数字化的质量。

(4)建立古籍数字化产品信息发布平台。由于缺乏相应的沟通、交流渠道,古籍数字化领域存在比较严重的重复开发问题。为提高古籍数字化效率,减少不必要的资源浪费,有必要建立国内外中文古籍数字化产品的信息通报机制。通过权威部分发布的信息,不仅可以共享产品信息,同时也可以对产品进行宣传和推广。

(5)建立古籍底本使用补偿机制。古籍作为超过版权法保护期限的特殊文献,其数字化复制不存在版权问题。但即使这样,相关障碍依然存在。2005年发生的所谓“苏图事件”给了我们一个警示:古籍的保藏与利用仍是困扰当前图书馆界的一个普遍问题。图书馆对古籍虽然没有版权,但拥有保管权,并为保管古籍付出了一定的经费和代价。图书馆常常借口保护古籍善本,或出台制度限制底本的使用,或收取高额的底本费。因此,笔者建议,可以在古籍数字出版的收益中提取一定比例,作为提供古籍底本单位的经济补偿,使之成为一项制度确立下来。这样,既可以保障古籍善本征集渠道的畅通,又照顾到图书馆保护古籍文献的付出,符合文献资源公平共享的原则。

(6)建立民间古籍善本的征集制度。除了各类图书馆、档案馆等官方文献收藏部门,有相当一部分数量的古籍流落到民间私人收藏者手中。如何对广泛存在于海内外私人藏书者手里的孤本、善本进行征集和数字化?这就需要对私人收藏者的利益有所考虑,就需要协调国家、机构与个人的利益关系,建立相应的补偿机制。国家应该从制度上着手,建立因底本使用而取得合理补偿的制度,并确定合理的缴纳报酬的比例,避免由于底本垄断而阻碍古籍数字化进程。由于在古籍数字化的利益分配过程中,部分古籍文献所有者无法得到合理的补偿,因而不愿意把自己掌握的资源与他人分享,这种情况在珍本、善本古籍文献上表现的尤为突出,严重阻碍了古籍文献数字化的进程。补偿制度的建立在很大程度上可以消除部分人的顾虑,有利于择优选择版本,提高数字化古籍的质量。

参考文献:

- [1]李国新.中国古籍数字化资源的进展与任务[J].大学图书馆学报,2002(1):21-26.
- [2]潘德利.中国古籍数字化进程与展望[J].古籍整理与研究,2002(7):117-120.
- [3]陈力.中国古籍数字化的现状与展望(上、下)[J].古籍整理出版情况简报,2004(4、5).
- [4]段泽勇,李弘毅.古籍数字化的回顾与展望[J].图书馆理论与实践,2004(2):37-39.
- [5]吴家驹.中文古籍数字化的进展与主要成果述评[J].南京师范大学文学院学报,2004(9):178-183.
- [6]李明杰.中文古籍数字化基本理论问题刍议[J].图书馆论坛,2005(5):97-100.

作者简介:李明杰,男,博士,武汉大学信息管理学院副教授,硕士生导师,研究方向:文献学、中国出版史与出版文化;俞优优,女,武汉大学信息管理学院2008级硕士研究生,研究方向:文献与出版。