

# 台湾中文古籍数字化成果特色谈

许红健

(南京图书馆, 江苏 南京 210018)

**摘要:** 通过网络调查、成果分析, 详论了台湾中文古籍数字化成果的特点, 即联合目录数据库收录范围跨越国界、资料库规模恢宏资源共享度高、合理分工合作成果产品多元化, 意在表明古籍数字化在保存与共享、利用与研究、文化传承等方面具有重大的意义。

**关键词:** 台湾; 中文古籍; 数字化成果; 特色

中文分类号: G250.7

文献标识码: A

文章编号: 1002-1248 (2009) 01-0130-04

## Characteristics of Achievements Gained in Taiwan Ancient Chinese Book Digitalization

Xu Hongjian

(Nanjing Public Library, Nanjing 210018, China)

**Abstract:** The paper has discussed three main areas of the characteristics of Taiwan ancient Chinese books' digital outcome through Internet survey and the analysis of results. Namely the scope of the joint directory database which across national borders, database magnificent scale and high resource sharing, and reasonable division of achievement products diversification. This is intended to show the great significances of preserving and sharing ancient Chinese books digitally, utilization and re-search of ancient books, cultural heritage and other aspects.

**Keywords:** Taiwan; ancient Chinese books; characteristic; digitalization achievement

台湾中文古籍数字化(数位化)的实践起步较早,早在20世纪70年代末台湾方面就开始了中文文字的信息处理技术研发,80年代开始较大规模、有规划的古籍数字化资源的开发与建设,如:1984年“中央研究院”最早开始了“史籍自动化计画”,该计划主要为开发《二十五史》全文资料库而成立,90年代中期以后古籍数字化工作纳入“国家典藏数位化计画”等大型规划中,更是取得了令人瞩目的成绩。经过20多年的发展,台湾古籍数字化工程在资料库建制管理、信息处理技术规范以及资料库的数量、质量、资源共享等方面均处于领先地位。通过网络调查、成果分析,总结出台湾古籍数字化成果的特点主要有以下3个方面。

### 1 联合目录数据库 收录范围跨越国界

古籍联合目录资料的建制不仅是图书馆开展古籍服务的工具,也是古籍整理的基础建设工作,认识到其重要性,台湾早在1970年及1980年分别编撰了纸

本联合目录《台湾公藏善本书目名索引》以及《台湾公藏普通本线装书目书名索引》。但是纸载资料不能随着各馆收藏变动而即时更新。

台湾古籍书目数据库的建制始于20世纪80年代,最初主要是各个单位的自主研发行为。为了建设较为完整的古籍善本资料库,1998年4月由台湾“国家图书馆”牵头,联合省内善本古籍收藏较有规模的故宫博物院、台湾大学、“国立中央图书馆台湾分馆”、“中央研究院”史语所图书馆等10余家单位图书馆,共同制定了“台湾地区善本古籍联合目录建档计划”,逐成“台湾地区善本古籍联合目录资料库”。随后,另邀请大陆港澳等地区图书馆提供资料测试,采购中国国家图书馆古籍书目资料,汇整大陆、日本、美国、欧洲等地中文古籍收藏量居前的多家图书馆古籍书目资料库,2004年8月建成“中文古籍书目资料库”。现该库收录海内外29家图书馆所藏古籍书目资料55万余笔,为国际包含书目量最多的古籍资料库,并为全球范围内提供免费查询检索服

收稿日期: 2008-08-26

作者简介: 许红健(1965-),女,本科,馆员,曾在核心刊物上发文数篇。

务。

该资料库收录的古籍书目以 Metadata 格式表达,并选定若干核心书目栏位加以呈现,共有 14 个检索点,提供繁、简体查询界面,2 种字体均可输入检索,查询结果亦可依读者所需以详目或简目式呈现。同时,该目录资料库还可供各单位汇入 CMARC、CNMARC 以及 USMARC 等 MARC 格式书目资料,将来拟将进一步扩充系统,提供符合国际标准的 Metadata 书目资料汇出功能,实现资源共建共享。至 2006 年底“中文古籍书目资料库”所收录图书馆及资料数如表 1 所示。

此外,台湾还建有联合目录“数位典藏联合目录”,该联合目录乃是“数位典藏国家型科技计划”所建制,提供台地区数字化典藏藏品的检索与搜寻,是作为数位典藏计划成效的目录性展示平台。通过该目录的单一窗口,即可检索台湾近百组跨十余个学术领域的数位典藏内容,不仅可作为教育研究素材与产业应用增值目录,同时也是提供学术研究、教育发展与企业应用的最佳浏览平台。该目录依据数位典藏计划将其内容主题划分为:动物、植物、地质、人类学、档案、地图与遥测影像、金石拓片、善本古籍、考古、器物、书画等 12 个,目前拥有资料数 3 312 494 笔。系统采用后设资料开放存取协定(OAI-PMH)与标准通用表示语言(XML)建构,提供分类目录浏览与全文检索服务,后设资料进阶搜寻采用 DC 栏位定义。

## 2 资料库规模恢宏 资源共享度高

进入 20 世纪中期以后,台湾古籍数字化工作纳入大型的“国家典藏数位化计画”等规划中。有当局的大力支持,台湾古籍数字化的研发规模、资源共享

等方面均取得突破性进展。在数字化成果方面主要表现为资料库规模继续增大,资源共享度持高。据统计目前在网络上能够查找到的中文古籍资源中,台湾的拥有量为世界最多。

其中规模最大、在全球享有盛誉的资料库当属“中央研究院”的“汉籍电子文献”。该资源库起建于 1984 年,其时鉴于意识到科技对人文领域的强烈冲击,“中央研究院”开始规划“史籍自动化计划”,目的是选择对中国传统人文研究具有重要价值的古代文献,建立电脑全文资料库,作为学术研究的辅助工具。当年 7 月试验性的完成了《史记》中的《食化志》的输入,获得了相当的经验和成功。1985 年全部完成相继开发前四史,即《史记》、《汉书》、《后汉书》、《三国志》。到 1990 年全部完成“二十五史”,1994 年开始提供网络版。参与制作资料库的主要是“中央研究院”的属下五所:史语所、台史所、资讯所、近史所、文哲所,以及该院的计算中心。1995 年之后,“中央研究院”结束“单打独斗”的自行研发局面,积极开展国内外合作关系共享共建古籍资源,合作单位包括台湾中山大学、台湾中正大学、台湾师范大学、香港中文大学、伦敦大学、史丹佛大学、密西根大学等。

经过 10 多年的持续合作发展,“中央研究院”逐步建成了规模宏大、整理严谨的古籍全文资料库——“汉籍电子文献”。目前该资料库包含整部二十五史、整部阮刻十三经、诸子古籍 34 种、大正新修大藏经、上古语语料库、台湾史料、文心雕龙、清代经世文编、词话集成、新清史——本纪、乐府诗集等,合计 300 多个资料库共 2 亿多字,每年并以 5 000 万字的速度增加。除此之外,还提供一个特别的资料库——“人文资料库”师生版,是从上述资料

表 1 “中文古籍书目资料库”收录图书馆及资料数表

馆名	资料数	馆名	资料数	馆名	资料数
中国国家图书馆	319 937	美国国会图书馆	3 043	清华大学图书馆	149
东京大学东洋文化研究所	106 571	天津图书馆	2 910	上海图书馆	117
国家图书馆(台湾)	48 895	台湾政治大学图书馆	2 139	内蒙古图书馆	99
中央研究院文哲所图书馆	22 678	台湾师范大学图书馆	2 014	南京图书馆	98
台湾大学图书馆	15 771	柏克莱加州大学东亚图书馆	1 204	华东师大图书馆	65
中央图书馆台湾分馆	9 616	香港大学冯平山图书馆	783	浙江图书馆	51
东海大学东亚图书馆	8 594	北京师范大学图书馆	504	复旦大学图书馆	50
东海大学图书馆	6 415	澳门中央图书馆	458	北京大学图书馆	45
中央研究院傅斯年图书馆	6 295	山东省图书馆	310	山西省图书馆	21
	4 320	澳门大学国际图书馆	182		

库中精选摘出经典部分而建成,目的是为普及文史教育而向台湾大中小学师生免费开放。

“汉籍电子文献”拥有自行研发的软件工具,即“瀚典全文检索系统”,历经 1997 年 1.3 版、2.0 版、2007 年 3.0 版。最新 3.0 版全套系统以 Pure Java 撰写,采用全文检索方式,并有主题列表、作者列表、书名列表等浏览途径。此外,为了避免重复建设,该系统还提供“建置进程”、“建置书目”等栏目,定时提供资料库已经选收的书目与拟收书目信息,显示系统修建情况。“汉籍电子文献”在合作单位间是免费共享的,除此之外其他单位需要使用则采取机构申请或个人申请两种方式,资料部分不收取任何费用,但考虑本院的电脑主机及网络设施部分的成本,酌情收取一定的作业成本费用。目前,“汉籍电子文献”在海内外拥有大量的用户评价极高。

此外,免费共享的资源库,还有台湾“国家图书馆”的“古籍影像检索系统”、“善本丛刊影像先导系统——明人诗文集初编资料库”、“金石拓片资料库等全部古籍料库”,以及“中华电子佛典线上藏经阁”、“寒泉古典文献全文检索资料库”、元智大学的“网络展书读”、台北大学的“经学之家”等。这些资料库的内容规模仍在不断的更新与添增当中。

### 3 合理分工合作 成果产品多元化

台湾古籍数字化的研发很注重协调与合作,如在“数位典藏国家型科技计画”下,组织参与该计划的有 8 家单位:“中央研究院”、“国立自然科学博物馆”、“国立故宫博物院”、台湾大学、“国立历史博物馆”、“国史馆”、“国史馆台湾文献馆”、“国家图书馆”,同时又成立了 5 个工作小组,分别负责内容规划、技术研发、应用服务、训练推广、运营管理工作。8 家单位分担相辅相成、主题内容各异的数字化项目任务,5 个工作小组则负责统筹安排 8 家单位的建库规划,提供相应的技术、规范、人才支持,形成了“八横五纵”的分工合作格局。

这样的分工与合作,避免了重复建设,促进了数字化成果呈现多元化的趋势。在实现原典数字化的基础上,台湾古籍数字化成果多元化主要表现为:

(1) 资料的形式从文字资料扩展到多媒体资料,其产品特性、辅助功能也越来越丰富,响应了不同的使用需求,如“不朽的殿堂——汉代的墓葬化与文化”,结合文字、影像、原典与研究资料的综合文献,引用 3D 动画技术虚拟汉代墓葬文化,展现了多样

的文献资料风貌,同时具备史料研究、教学参考、大众普及等多种功效。

(2) 资料库从文献原典扩充为研究专题资料库、古籍影像资料库、多媒体教学系统等形式,如“汉学研究中心”的“明人文集联合目录及篇目索引资料库”、“国家图书馆”的“古籍影像检索系统”、元智大学的“《红楼梦》网路教学研究资料中心”等。

(3) 资料库的框架也从单一式发展到立体式,从而具备了多种用途功能,如元智大学的“网络展书读资料库”从原来一般全文检索资料库规划成 5 个部分:唐宋全文资料库;研究论著资料;词文分析资料库;DIY 个人工作平台;网络教学讨论区,这些资料库相辅相成,相得益彰。

尤为称赞的是元智大学罗凤珠主持研发的“《红楼梦》网路教学研究资料中心”,该系统代表了古籍数字化多元化应用的典范。系统利用多媒体、超链接等网络技术,多样化的整合《红楼梦》及其相关资料,强调呈现《红楼梦》文化内涵以及其他相关的中国文化知识。即系统内容除了《红楼梦》全文及研究论著资料,还收集了文化艺术图像及声音资料,包括《庚辰本脂砚斋重评石头记》全文影像资料,乐曲资料、饮食资料——红楼宴、建筑资料——大观园、相关文物资料、绘画资料、戏曲资料、《红楼梦》历代著作书影资料、服饰、书画、工艺资料等,相当于在网络上建构了一个虚拟的《红楼梦》数位文化艺术展览馆。该“资料中心”的成功研制,标志着台湾古籍数字化从研究型走向民众,特别是以大中小学生、教师为对象的教学、普及路线。

### 4 结束语

总之,古籍数字化在保存与共享、利用与研究、文化传承等方面具有举足轻重的作用。

目前,古籍数字化的研发已从个别行为上升到有计划、有组织、规模化、规范化的发展过程,其相关规划纳入国家政策性的发展规划中,如台湾的“数位典藏国家科技型计画”等,对古籍数字化作出了相应的计划支持与政策引导,这对于古籍数字化未来的发展具有重大的意义。

#### 参考文献:

- [1] 罗凤珠. 台湾地区中国古籍文献资料数位化的过程与未来的发展方向. [2007-3-6]. <http://cls.admin.yzu.edu.tw/present/tarcf.htm>

- [2] 数位典藏联合目录网. [2007-3-8]. <http://catalog.ndap.org.tw/dacs5/System/Main.jsp>
- [3] 姚松. 计算机用于古籍整理研究的现状与展望. 中国典籍与文化, 1995(2)
- [4] 徐清等. 近十年古籍书目数据库建设研究概述. 图书情报知识, 2006(9)
- [5] 姜哲等. 大型中文古籍《四库全书》自动版面分析系统. 中文信息学报, 2001(2)
- [6] 孙钦善. 中国古文献学. 北京: 北京大学出版社, 2006
- [7] 罗志欢. 国外所藏中国古籍概观. 中国典籍与文化, 1994(2)

(上接第 116 页)

的应用研究, 建立各种数字资源, 提供网络服务。服务是图书馆的永恒的主题, 但网络必然走入图书馆, 打破馆员只是读者与馆藏书刊之间的中介人的状态, 馆员要成为网络专家、信息中介人和系统设计员。在高质量服务的同时, 图书馆的优良传统不能丢, 为读者找书, 为书找读者, 利用一切先进设备, 给他们提供方便。

参考文献:

- [1] 书香社会, 李东来主编[m]北京图书馆出版社, 2008, 168 页
- [2] 莫伟鸣, 何琼. 阅读热点的读者群分析 [N]. 中国文化报, 2001-5-23
- [3] 赵燕. 网络时代大众阅读方式的选择[J]. 图书馆学刊, 2006(4)
- [4] 谭泽红. 网络环境下大学图书馆的阅读教育[J]. 图书馆论坛, 2006(5)
- [5] 王桂兰, 余春. 图书馆功能转型与阅读形态变化之互动 当代图书馆, 2005(6)
- [6] 周英. 网络时代的阅读与图书馆阅读指导工作[J]. 科技情报开发与经济, 2005(4)
- [7] 吴茜. 大学生网络阅读及其策略[J]. 图书馆工作与研究, 2006(6)
- [8] 段汶君. 试论盲校图书馆导读的必要性和具体实施 图书馆工作与研究, 2006(4)

(上接第 129 页)

- [3] 吴锡平. 图书漂流: 让书香飘散开去[J]. 世界知识, 2006(4): 63.
- [4] 李萍. “图书漂流”为何漂去漂不回[J]. 出版参考, 2006(4): 11.
- [5] 萧山图书馆学会. 萧山图书馆工作简报, 2007(1).
- [6] 袁征. 图书漂流考验读者诚信[J]. 中国消费者报, 2006-01-11.
- [7] 图书漂流网. <http://www.tspl.cn>. 2007-01-07.
- [8] 程焕丈, 周旭长. 图书馆精神—体系结构与基本内容[J]. 图书馆, 2005(2).
- [9] 黄俊贵. 提升阅读理论, 构建阅读社会 [J]. 图书馆论坛, 2005, (6): 40.