

ZHONGWEN GUJI
SHUZHUA YANJIU



中文古籍 数字化研究

王立清◎著



国家图书馆出版社

中文古籍数字化研究

王立清 著

国家图书馆出版社

PDG

图书在版编目 (CIP) 数据

中文古籍数字化研究 / 王立清著. —北京: 国家图书馆出版社, 2011. 5
ISBN 978 - 7 - 5013 - 4518 - 2

I. ①中… II. ①王… III. ①数字技术 - 应用 - 古籍整理 - 研究
IV. ① G256.1 - 39

中国版本图书馆 CIP 数据核字 (2011) 第 003913 号

责任编辑 邓咏秋

编辑助理 王丽丽

书 名 中文古籍数字化研究

著 者 王立清 著

出 版 国家图书馆出版社 (100034 北京市西城区文津街 7 号)

发 行 (010) 66139745, 66175620, 66126153

66174391 (传真), 66126156 (门市部)

E-mail btsfxb@nlc.gov.cn (邮购)

Website www.nlcpress.com (投稿)

经 销 新华书店

印 刷 北京华艺斋古籍印务有限责任公司

开 本 710 × 1000 (毫米) 1/16

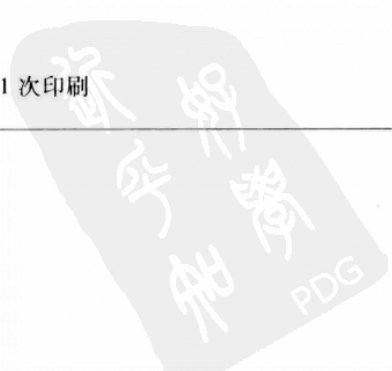
字 数 210 千字

印 张 15.75

版 次 2011 年 5 月第 1 版 2011 年 5 月第 1 次印刷

书 号 ISBN 978 - 7 - 5013 - 4518 - 2

定 价 38.00 元



自序

中文古籍数字化发端于20世纪70年代末80年代初，至今已走过了30年的历程。古籍数字化将中国最古老的文献与世界最先进的数字化技术完美地结合在一起，蕴载着中华数千年之文明，彰显着现代信息技术之优势，成为互联网时代中华文化展现和传播的一道亮丽风景。古籍数字化经历了从零星制作发展到规模开发，从初期的图形扫描发展为字符数字化，从做目录索引发展为做全文，从制作单机版光盘发展为制作网络版或因特网数据库。古籍数字化有效地解决了古籍保护与利用之间的矛盾，使传统古籍走出“深闺”，传本扬学，惠及士林。借助数字化强大而完备的检索功能，我们可以随意掇拾山中的一块卵石，可以尽情把玩海底的一粒细沙。

随着古籍数字化实践的发展，理论研究也陆续展开。初期的古籍数字化研究主要涉及古籍数字化现状、古籍数字化产品、古籍数字化技术、专门古籍的数字化等，研究成果体现出较为明显的描述性和总结性特征。20世纪末，文渊阁《四库全书》电子版的出现，极大地推进了古籍数字化实践发展，引发了古籍数字化理论研究的一个小的热潮。一些学者在总结古籍数字化实践历程之后，开始思考古籍数字化的基本理论问题，对古籍数字化的概念、性质、特征等基础理论和技术处理层面的相关理论进行研究。近年来，新的古籍数字化产品陆续面市，古籍数字化理论研究也不断走向深入。本论著力图基于已有的研究成果，全面系统地研究中文古籍数字化，以期对中文古籍数字化理论研究有所贡献，对古籍数字化实践工作有所启示和参考。

本论著特色之一：古籍数字化研究的宏观层面和全面视角。论著较为

全面地研究了中文古籍数字化的理论和实践，研究内容涉及背景分析、现状剖析、产品分析、主体分析、选题策略、评估方法、支撑体系、相关影响等，研究视野广阔，可以在一定程度上弥补国内研究之不足。

本论著特色之二：古籍数字化实践的系统梳理。论著深入地总结了中文古籍数字化三十年来的工作，展现和分析了中文古籍数字化的主要成果，对重要产品进行了个案分析，讨论了中文古籍数字化中存在的问题，推进了古籍数字化工作的梳理。

本论著特色之三：古籍数字化多元主体的探讨。论著突破了现有论文仅对古籍数字化开发主体进行讨论的研究局限，将主体研究扩展为开发主体、利用主体、控制主体和研究主体，分析了各类型主体在古籍数字化活动中呈现出的不同特点和发挥的不同作用，提出应站在国家的层面，建立古籍数字化各主体之间的协调机制，促进古籍数字化的有序发展。

本论著特色之四：古籍数字化建设的国家控制与管理模式的首次论述。论著提出政府应依据不同的古籍数字化主体，对公益性古籍数字化和商业性古籍数字化实施不同的控制与管理模式。选题控制与管理模式和成本效益评估模式适用于国家对公益性古籍数字化建设的控制与管理，市场调节与控制模式则适用于商业性古籍数字化。

本论著特色之五：古籍数字化影响研究的阐述。论著从思想文化层面深入论述了古籍数字化对中华文化遗产、网络阅读、传统学术研究和古籍整理带来的影响。

在网络环境下，数字化技术发展日新月异，古籍数字化所涉及内容诸多，作者虽殚精竭虑，但由于水平和经验所限，书中难免有疏漏、不妥甚至错误之处，敬请各位专家、学者和读者批评指正。

王立清

2010年12月

目 录

自 序	1
第 1 章 绪 论	1
1.1 研究缘起 / 1	
1.2 研究目标和研究方法 / 3	
1.3 研究思路和内容框架 / 4	
第 2 章 古籍数字化背景分析	7
2.1 社会发展背景分析 / 7	
2.2 文献资源数字化发展背景分析 / 11	
第 3 章 古籍数字化基本理论问题	16
3.1 古籍数字化概念 / 16	
3.2 古籍数字化性质和定位 / 22	
3.3 古籍数字化的开发层次和类型 / 24	
3.4 古籍数字化原则 / 29	
3.5 古籍数字化意义 / 30	
第 4 章 古籍数字化现状总述	33
4.1 古籍数字化发展历程 / 33	
4.2 古籍数字化发展现状 / 35	
4.3 国内各地区古籍数字化分析 / 57	
4.4 国外中文古籍数字化分析 / 80	
第 5 章 古籍数字化主体分析	87
5.1 古籍数字化主体构成 / 87	

5.2	古籍数字化开发主体	/ 90
5.3	古籍数字化利用主体	/ 100
5.4	古籍数字化研究主体	/ 103
5.5	古籍数字化控制主体	/ 106
第6章	古籍数字化项目研究	110
6.1	大学数字图书馆国际合作计划(CADAL)——古籍部分	/ 110
6.2	《中国基本古籍库》项目	/ 112
6.3	《四库全书》电子版	/ 114
6.4	《四部丛刊》电子版	/ 121
6.5	《国学宝典》	/ 126
6.6	《汉籍电子文献》	/ 129
6.7	《汉达文库》	/ 131
第7章	古籍数字化的国家控制与管理模式	136
7.1	公益性古籍数字化的选题控制模式与成本效益评估模式	/ 136
7.2	商业性古籍数字化的市场调节控制模式	/ 151
第8章	古籍数字化支撑体系研究	160
8.1	古籍数字化标准体系建设	/ 160
8.2	古籍数字化技术的研发	/ 167
8.3	加强古籍数字化人才培养	/ 179
第9章	古籍数字化影响研究	185
9.1	古籍数字化与中国传统文化	/ 185
9.2	古籍数字化与读者阅读	/ 196
9.3	古籍数字化与传统学术研究	/ 204
9.4	古籍数字化与古籍整理	/ 215
第10章	结语	224
10.1	全书结论	/ 224
10.2	研究展望	/ 228

参考文献	231
附录 1 部分古籍书目数据库比较表	236
附录 2 中文古籍数据库调查表	239
附录 3 北京地区中文古籍数据库调查结果汇总表	240
后 记	243

第1章 绪 论

1.1 研究缘起

信息时代计算机掀起了惊涛骇浪，影响到经济、军事、文化、生活各个方面，人类文明面临新的技术环境和新的考验。2002年“第二届汉文化资料库国际学术研讨会”论文集，朱邦复先生序曰：“河有源、山有脉、城有基、树有根，中华文化，源远流长、基厚根深”。中华文化博大精深，积数千年之磨炼，载历代炎黄子孙之发奋努力，巍然屹立于世界文化之林。古籍数字化将华夏古国传统的文化载体与当代世界先进的计算机技术及其他信息技术珠联璧合。在因特网蔓延世界、西方文化席卷全球的今天，我心深为先辈们执着于弘扬中华传统文化的精神所感染，毅然斗胆选择了“古籍数字化”这一蕴涵古老、显现时尚的研究主题，真诚期望通过自己的努力，为推进我国古籍数字化发展尽绵薄之力。

当今，中文古籍数字化已不仅仅是一个技术层面的话题，更是涉及古籍整理出版、中华文化遗产、传统学术研究、阅读文化变迁、数字图书馆资源建设等方面的综合性研究课题。本书力图将中文古籍数字化研究置于广阔的社会背景之中，探讨中文古籍数字化自身的发展进程，阐释中文古籍数字化对中华文化遗产、读者阅读和传统学术研究的影响，思考古籍数字化过程中存在的问题，并试图提出相应的解决方案。

中文古籍数字化是古籍整理出版的新趋势，它体现了信息时代的特征，标志着古籍出版走向了一个新的时代。中华文化的传承、古代典籍的后传，如何与现代信息技术的发展紧密地结合、与时代的节奏合拍，这是信息时代的新课题。现代信息技术为传统古籍整理拓展了新的领域，提出了新的挑战，更带来了新的发展契机。高新科技的介入，将使得传统的古

籍整事业获得新的生长点,同时,对中文古籍数字化进程的理论总结,可以在一定程度上续写传统古籍整理出版的历史和现在,进一步补充和丰富古籍出版的理论。

古籍数字化是中文文献数字化的一个复杂特例,是数字图书馆资源建设的一个缩影。数字图书馆研究近年来一直是图书馆学界关注的课题,信息资源数字化、资源规模化、服务网络化是数字图书馆根本的特征。一些数字图书馆尝试着将部分古籍资源数字化,形成数字化资源的特色,如中国国家图书馆早在1999年就开发了中国数字图书馆实验演示系统,对古籍资源数字化进行了有益的实践,该系统建设了千家诗、中国古代建筑、北京故宫、海洋世界和宇宙结构等五个多媒体资源库。古籍的数字化对数字图书馆资源建设具有特殊的意义,通过对古籍数字化的剖析,可以映射资源数字化的理论。

古籍数字化顺应了网络阅读发展的趋势,改变了传统的纸质古籍阅读方式。网络阅读是网络环境下的一种新现象,古籍的数字化使古籍这些文化历史巨著走下神坛,亲近平民,让一般人能领略古籍的博大精深。透过对古籍数字化阅读的窗口,可以比较网络阅读与传统阅读的异同,研究古籍数字化对传统阅读的影响。

古籍数字化将在很大程度上推进中国传统学术的深入挖掘研究,从长远来看会影响一代人的思维方式和研究方法。它不仅延续了传统古籍的浏览功能,更具有强大的检索功能,对于相关问题的研究,可以做前人想做而未能做、前人想做好而未能做好的事情。古籍的数字化提供了便捷的资料获取方式,为中国传统学术的研究注入了新的活力,开创了中国传统古籍文献的定量研究。

古籍数字化是一项“功在当代、利在千秋”的大事,走过了30年的风雨历程,有值得骄傲的成就,也有不容忽视的问题。比如,数字化的标准和规范问题、重复建设问题等等。这些问题阻碍了我国古籍数字化的发展,如果不及时解决这些问题,势必会重蹈我国图书馆自动化发展之覆辙,造成人力、物力和财力的极大浪费。如何采用与国际接轨的标准,如何加强合作协调,集中有限的资金,采用先进的信息技术,从整体上规划

古籍数字化的发展,提升古籍数字化的质量和数量,已经成为目前古籍数字化实践中必须要探讨和解决的问题。

总之,中文古籍数字化是中华文化与以计算机为核心的高新技术相结合的产物,是传统与现代的融合,是信息时代图书馆学领域的研究焦点之一。通过对古籍数字化的探讨,可以预测古籍整理出版的新趋势,了解资源数字化的过程,并窥见由此带来的一系列社会影响。

1.2 研究目标和研究方法

1. 本书的研究目标

(1) 较为全面宏观地研究中文古籍数字化,研究内容涉及背景分析、现状剖析、产品分析、发展模式、选题策略、评估方法、支撑体系、相关影响等。

(2) 系统深入地总结中文古籍数字化 30 年来的工作,展现和分析中文古籍数字化的主要成果,讨论中文古籍数字化中存在的问题,推进古籍数字化工作的梳理。

(3) 分析古籍数字化的各个主体及其特征,研究古籍数字化建设的国家控制与管理模式,以利于更好地从宏观层面推进我国古籍数字化的有序发展。

(4) 从思想文化层面深入论述古籍数字化所表现或隐含的各种影响,研究围绕中华优秀传统文化、网络阅读、传统学术研究的古籍数字化发展策略。

2. 本书的研究方法

本书主要采用理论阐述与实证分析相合的研究方法,力求准确地阐释有关古籍数字化的实践进展,深入分析古籍数字化的相关影响,提出我国古籍数字化的发展对策。具体的研究方法主要有以下几种:

(1) 文献调研法

文献调研是十分重要的研究方法,是我们开展研究的基础。我国古籍数字化工作开始于 20 世纪 70 年代末 80 年代初,30 年的发展历程积累了一定数量的文献资料。本书立足于近些年来主要研究成果,利用网络数

数据库、搜索引擎、公共联机查询目录等信息检索系统,查阅并参考了国内外相关文献千余篇(部),从文献的角度初步对古籍数字化问题的研究现状有所了解,针对已有研究的不足,展开了自己的论述。

(2) 实地调研法

古籍数字化具有非常强的实践性,本文采用了实地调研的方法,访问调查一些有代表性的古籍数字化部门或单位,如图书馆、公司、研究机构等,切实掌握古籍数字化的进展,了解古籍数字化存在的问题,力求使本论文的研究立足于实践的基础,为解决实际问题提供参考建议。

(3) 专家调查法

在文献调研和实地调研的基础上,采用专家调查法,对古籍数字化的相关专家进行了走访,认真听取他们对古籍数字化工作的看法,交流古籍数字化的成功与教训,集专家之智慧,为我国古籍数字化的发展献计献策。

(4) 问卷调查和网络调查

问卷调查的方法是学术研究过程中为寻求基本数据的支持而采用的一种比较基本的方法。笔者为此专门设计了中文古籍数据库调查表,对20家参与古籍数字化的单位进行了详细调查。同时,在因特网上对古籍数字化的相关网站进行了较为全面的调查,整理并分析了古籍数字化的网站。这些数据为本文的研究提供了重要的支持。

(5) 案例分析法

本文选择了有代表性的古籍数字化案例,对其进行深入细致地解剖和分析,研究古籍数字化过程中所涉及的相关问题。如《四库全书》、《四部丛刊》的数字化产品等。

1.3 研究思路和内容框架

中文古籍数字化是中华文化与以计算机为核心的高新技术相结合的产物,是传统与现代的融合。梳理和总结古籍数字化工作,探索未来我国古籍数字化发展道路,研究古籍数字化的影响,是本书的研究思路。本书的内容框架大致如下。

第1章为绪论,主要介绍了本书的研究缘起、研究目标、研究方法和内容框架。

第2章阐述古籍数字化的背景,分析了与之相关的社会发展背景、文献资源数字化背景和古籍整理的发展情况。中国经济文化的发展、信息技术的进步、文献资源数字化的趋势、数字图书馆的建设、古籍整理方式的创新,催生了中文古籍数字化。

第3章论述古籍数字化的基本理论问题。对古籍数字化的概念、性质与定位、开发层次和类型进行了探讨,并阐明了古籍数字化的原则和意义。本章是全书立论的理论基础。

第4章研究古籍数字化的现状。作者阐述了古籍数字化发展历程,从古籍书目数据库和古籍全文数据库两条线索梳理了古籍数字化现状,对国内外的古籍数字化工作进行了分析与总结。

第5章论述古籍数字化主体。在本章作者界定了古籍数字化主体,探讨了各主体之间的关系,并分别对古籍数字化的开发主体、利用主体、研究主体和控制主体展开了分析。

第6章梳理和分析古籍数字化项目。对电子版《四库全书》、《四部丛刊》、《国学宝典》、《汉籍电子文献》等有代表性的古籍数字化产品等进行了分析。

第7章研究古籍数字化的国家控制与管理模式。提出政府应依据不同的古籍数字化主体,对公益性古籍数字化和商业性古籍数字化实施不同的控制与管理模式。选题控制与管理模式和成本效益评估模式适用于国家对公益性古籍数字化建设的控制与管理。市场调节与控制模式是指利用市场机制、运用各种经济杠杆的利益诱导作用,促使古籍数字化机构从经济利益上关心自己的活动,是一种间接组织协调、控制古籍数字化活动的手段。

第8章就古籍数字化支撑体系中的标准、技术、人才等,提出了解决方案或策略。提出应继续修订和完善相关古籍著录标准,建立古籍数字化规范数据库,最大限度地实现古籍数字化资源共享。推进古籍数字化技术研发,推广 ISO/IEC 10646,加强关联检索、知识发现技术等古籍数字

化产品中的应用。建立多层次、多方式的古籍数字化人才培养体系，出台古籍数字化杰出青年基金计划，为古籍数字化事业培养人才、留住人才。

第9章探讨了中文古籍数字化的影响，从文化思想层面研究了古籍数字化的深远意义。古籍数字化创新了古籍资料发现的方式，改变了学术研究的方式，开创了计量研究的新局面。古籍数字化有助于中华民族优秀传统文化的传播，其中的关联检索功能为人们更好地理解古文提供了捷径，古籍数字化将进一步预示着网络阅读新时代的到来。中文古籍数字化应对中华文化的传承、传统学术研究方式的创新、网络阅读的发展有所贡献。古籍数字化作为网络环境下传统古籍整理的一种新方式，将改变着古籍整理的方向，影响着古籍整理的手段和成果形式。

第10章为结语，总结全书的主要观点，阐述需要进一步研究的相关问题。

第2章 古籍数字化背景分析

2.1 社会发展背景分析

2.1.1 传统文化热持续升温

经过30多年的改革发展,中国已经成为经济大国。中国的开放型经济实现了持续快速发展,在国际分工中的地位日益提高。中国经济的迅速崛起对世界产生了巨大的影响,从美国纽约到瑞士达沃斯,中国发展成为各种论坛的热门话题,正如诺贝尔经济学奖获得者迈克尔·斯宾塞在《世界格局中的中国经济》演讲中所说的那样:“在20年前世界上不会有很多人关注中国的经济发展,因为坦率地说,从经济的角度看,对于欧洲和美国的大多数人来说,20年前无论中国一年的经济增长率是8%还是10%都无关紧要。但现在就不同了,因为中国的发展是全球经济强有力的推动力之一,如果中国的经济发展取得了成功,全世界每个人都可以从中受益。”^①

中国经济的连年高速增长使中国经济的分量让人无法忽视,越来越多的人开始关注和研究中国经济、中国文化,出现了前所未有的中国热。据报道,随着中国对外开放的进一步扩大,汉语越来越成为世界范围内一种重要语言。目前,全球学习汉语的人数超过3 000万,100多个国家和地区的2500余所大学开设了汉语课程,汉语教材已进入美、英、日、韩等国的中小学课堂。^② 汉语作为越来越重要的交流工具,改变着人们对中国的认识。“中国文化热”裹挟着“汉语热”弥散全球。相当多的西方人士对中国文化有着浓厚的兴趣,自发地研学中国古籍,并在他们的国家传播中国

① 迈克尔·斯宾塞:世界格局中的中国经济。[2010-01-20]. <http://www.cas.cn/html/Dir/2005/12/15/6696.htm>

② 全球范围上演汉语热 国外学习汉语人数超3000万。[2010-01-20]. <http://chinaqw.com.cn/news/2005/1205/68/7983.shtml>

文化。中国文化的基本经典之一《道德经》的外语译本之多，在世界上仅次于《圣经》，单是国外出版的英译本就有一百余种。据闻在德国，《道德经》已出版了十几种不同的译本，平均每9个德国人就持有一本《道德经》，这个普及率恐怕还远远超过中国本土。德国人热爱《道德经》完全是自发自愿的，并不是出于中国人的请求或者强迫。^①

中国文化走过了数千年的历史，包括儒、道、诸子百家文化，以及文学、艺术、医学、武术等等，内涵深邃。中国文化崇尚和平，儒家主张“和为贵”，这与世界发展的潮流相一致。古籍是中华文化的重要结晶之一，我们有责任有义务通过对古籍的现代阐释，不仅让世界了解中华文化的历史涵义，更要进而认识中华文化的当代价值。

2.1.2 信息技术日新月异

古籍数字化是信息技术发展到一定阶段的产物。信息技术是指能够扩展人类信息器官功能的技术。它包括信息表述技术、信息处理技术、信息传输技术、信息存储技术和信息利用技术等。信息技术的发展经历了不同的阶段，语言的创造、文字的发明、造纸和印刷术的发明、电报电话电视等现代通讯技术的创造、电子计算机的发明和创造被称为是人类社会发展史上的五次信息技术革命。我们历经了以人工为主要特征的古代信息技术、以电信为主要特征的近代信息技术，步入了现代信息技术的发展新时期。现代信息技术更加强调计算机网络技术、多媒体技术、数据存储技术、通信技术和声像技术等综合技术，以微电子、计算机、软件、通讯、网络等技术为代表，具有高增值性、高渗透性等特征。伴随着社会的发展，现代信息技术已经成为信息社会科技发展的核心，应用广泛、发展迅速。信息技术的发展大大提高了人类的信息能力，尤其是信息与知识的交流范围，它的发展为古籍数字化提供了科技保障。

计算机技术的发展成为古籍数字化最基本的技术支撑。1946年世界上第一台计算机诞生，此后，计算机技术有了飞速的发展，获得了广泛的应用。数据库技术是计算机技术中发展最快，应用最广泛的重要分支之一，

^① 胡海龙. 弘扬国学不是民族主义. [2010-01-20]. <http://www.dlenglish.com/showthread.php?t=1359&pp=40>

它已成为计算机信息系统和计算机应用系统的重要技术基础和支柱。数据库技术萌芽于20世纪50年代后期,形成于60年代中期,成熟于70年代,广泛应用于80年代。随着计算机技术的不断发展,计算机在人文社会科学领域的应用日渐广泛。早在20世纪80年代初期,在西方发达国家,计算机技术开始进入个别人文科学领域(如经济学),当时钱钟书先生就提出,中国古代文献的整理和研究,应该尽快采用这种新兴的技术手段^①。这是富有远见的创意。计算机的海量存储能力,强大的计算功能,基于主题、关联的穷尽检索,数据库技术的发展,都为计算机在古籍管理与古籍整理方面提供了技术前提。

信息存储技术的不断更新给古籍数字化存储提供了可能。从我国几千年灿烂文化的历史进程,我们可以领略信息存储技术的变迁。殷商时期的甲骨文书、青铜器铭文、春秋战国时期的简策和帛文、汉代发明的纸张、现代的磁盘和光盘……信息存储技术的发展是随着社会生产力水平的提高和科学技术的进步而发展的。结绳记事是人类采用的古老的信息存储方式,造纸术和印刷术的发明为人们提供了一种新的存储信息的方法,纸的发明使古老原始的信息存储技术发生了一次革命性飞跃。正因为如此,承载中华文化数千年来的古籍才得以保存下来。电子计算机的发展催生了新的信息存储方式,近现代以来随着缩微品、磁盘、光盘等的出现,信息存储技术又发生了翻天覆地的变化,为古籍内容提供了新的载体形式。

因特网技术的发展为古籍数字化的网络传播提供了良好的平台。20世纪人类有两大发明对社会进程产生了极为深刻的影响,一个是原子弹,另一个就是因特网。因特网是全球性的、最具影响力的计算机互联网络,也是世界范围的信息资源宝库,它是众多计算机网络的集合体,用户可以超越不同形态的计算机网络,在世界范围内共享信息资源。今天,计算机网络技术已经并将对人类社会的发展产生重大影响。因特网技术构成了本世纪跨越地域最广的全球信息传媒,它突破了地域和时间的限制,以先进的电子技术手段向人们适时传播人类优秀文化遗产。网络提供了信息传播的

^① 田突. 古籍整理与研究的电脑化. 中国文化, 1994 (9): 85-89