

KAITUO YU CHUANGXIN



开拓与创新

宋史学术前沿论坛文集

包伟民 戴建国 主编

中西書局

图书在版编目(CIP)数据

开拓与创新：宋史学术前沿论坛文集 / 包伟民，戴建国主编. —上海：中西书局，2019.6
ISBN 978-7-5475-1600-3

I. ①开… II. ①包… ②戴… III. ①中国历史—宋代—文集 IV. ①K244.07-53

中国版本图书馆 CIP 数据核字 (2019) 第 127620 号

开拓与创新

——宋史学术前沿论坛文集
包伟民 戴建国 主编

责任编辑 吴志宏

装帧设计 梁业礼

出版发行 上海世纪出版集团

中西书局 (www.zxpress.com.cn)

地 址 上海市陕西北路 457 号 (邮编 200040)

印 刷 上海天地海设计印刷有限公司

开 本 700×1000 毫米 1/16

印 张 28

字 数 442 000

版 次 2019 年 6 月第 1 版 2019 年 6 月第 1 次印刷

书 号 ISBN 978-7-5475-1600-3 / K·303

定 价 98.00 元

本书如有质量问题,请与承印厂联系。电话: 021-64709974

目 录

序言	包伟民	1
近四十年辽宋夏金史研究学术回顾	包伟民	1
宋朝分权制衡的治军特点与边防困境	陈 峰	21
盔甲裹诗书：宋代将士文化水平考察	程民生	36
大力开拓文献史料源 不断提升学术创新力	戴建国	64
张载对宗法制度的构思及其“民胞物与”的大同理想	范立舟	73
欧美地区的古代中国数字资源概述	耿元骊	89
《宋代登科总录》——创新的宋代精英数据库	龚延明	108
御殿听政：南宋前期中枢日常政务的重建与运作	韩冠群	116
南宋高宗朝科举试策中的“光武故事”	何玉红	132
绝对理念与弹性标准：宋朝政治场域中对“华夷”和“中国”观念的 运用	黄纯艳	148
宋史研究中的两个较大问题	姜锡东	174
晚宋理学家的学术视界与君臣理想——以《大学衍义》为中心的探索	孔妮妮	177
高遵裕与宋夏灵州之役的再探讨	雷家圣	189
端正学风厚培功底 宏观、微观研究并重	李昌宪	207
唐宋史研究应当翻过这一页——从多视角看“宋代近世说”（唐宋变 革论）	李华瑞	211
加强整体贯通研究 恢复西夏社会原貌	彭向前	254
辽、宋、夏、金时期的民族史学	史金波 关志国	259

欧美地区的古代中国数字资源概述

(耿元骊 辽宁大学历史学院)

欧美地区特别是美国、英国、法国历来有收藏、研究东亚文献、文物的传统,对中国古代图书、地图、文物等方面的收藏和研究尤为重视。据调查,美国现存中文善本线装古籍不少于 70 万册,^①总数量则有接近 400 万册。^②欧洲所藏中文古籍数量不及美国,粗略估算有 200 万册左右。^③而文物方面就更难给出一个相对准确的数量估计,据说全部海外藏中国文物约 1 000 万件,^④美国近 30 年流入约 230 万件,加拿大藏约 20 万件。^⑤

数量庞大又极其珍贵的古代中国文化遗产,长期以来多数时间只能静悄悄地躺在图书馆、博物馆的库房里。虽然欧美收藏单位一般不会居奇自守,学者利用这些资源称得上方便。不过随着时间流逝,古籍、文物均已不耐久阅,必须加以保护,更不适合每日流通,难以为更多的学者所利用,更难为非本地居民所使用。另一方面,单纯仅供阅读的文献(不仅指前述海外汉籍,也包括所有现存中文典籍)和仅供观赏的文物,很难完全呈现它本身所具有的学术价值,必须纳入到数据库当中并且尽可能公开才会发挥作用。要解决这两方面问题,必须利用现代科技。随着半个世纪以来计算机、网络技术飞速进步,数字化技术不断发

① 沈津:《美国主要东亚图书馆所藏中国古籍文献及其展望》,台湾《“国家图书馆”馆刊》2001 年第 1 期。

② 刘家真:《中华古籍数字化的国际合作》,《图书情报知识》2013 年第 5 期。

③ 钱存训:《欧美各国所藏中国古籍简介》,《图书馆学通讯》1987 年第 4 期。

④ 闻哲:《1 000 万? 中国文物流失海外知多少》,《人民日报海外版》2007 年 1 月 29 日。

⑤ 刘瑞平:《海外中小博物馆及重要私人收藏的中国文物举要》,《中国文物报》2016 年 5 月 31 日。

展,越来越多的欧美机构对古代中国文化遗产进行了数字化加工,获得一大批数字化成果,极大方便了学术界和社会大众的使用。

1978年,美国学者首先尝试用计算机编制中文古籍索引,这是中文古籍数字化的先行者。^①但是在之后20多年当中,古籍数字化工作并未取得明显进展。直到20世纪90年代中期以后,随着存储成本降低,互联网快速发展,技术进步带来强大数据处理能力和无限量存储空间,推动了各个公共图书馆(大英图书馆、美国国会图书馆等)、博物馆(大都会博物馆、盖蒂博物馆等)、大学(哈佛大学等)、企业及私人基金会,等等纷纷投身于古代中国文史资源的数字化建设工作。经过近40年的辛苦工作,同样形成了数量亦称得上庞大的古代中国文化遗产数字化资源。据笔者的初步统计,欧美目前约有数百种各类型含有古代中国内容的数字资源(王国强所编《中国(汉学)研究开放获取学术资源集》多有收入,可在其微博@陆浑戎置顶免费下载),其中尤以古籍书目、古籍文本数字化、文物图像数据库、关系型数据库、数字人文等几方面最具优长。当然,这样分类较为简略,很多数据库包括多方面功能,既有目录的作用,也同时提供文本全文或者图像数据,有的还附带专用数字人文工具。以下简略概述之。

一、古籍书目数据库、文物目录数据库

这方面数据库为数甚多,几乎有中文古籍收藏的图书馆均开展了相关数据库建设工作。早在20世纪90年代初期,美国图书馆研究学会(Research Libraries Group, RLG)就开始进行“中文善本书目数据库”计划,^②当时预计将中国清代嘉庆以前古籍全部进行机读编目,到2005年已收集书目约25000条,^③后RLG与OCLC(Online Computer Library Center)合并,2009年,该项目转移到中国国家图书馆,建有中华古籍善本国际联合书目系统。^④由OCLC管理的World Cat是世界上最大的联合目录数据库,^⑤目前有170多个国家、地区的2万多家

① 陈东辉:《二十世纪古籍索引编制概述》,《文献》1998年第2期。

② <http://www.oclc.org/research/activities/chineserarebooks.html>。

③ 沈津:《北美地区中文古籍文献整辑工作近况举要》,《汉学研究通讯》24卷4期,2005年。

④ <http://mylib.nlc.gov.cn/web/guest/zhonghuagujishanbenlianheshumuxitong>。

⑤ <http://www.worldcat.org/>。

图书馆加入,包括3亿条书目和21亿条馆藏目录,其中含有大量中国古籍目录,目前在中国大陆无法访问。美国国会图书馆的“online catalog”^①书目检索系统已稳定运行近20年,目前可以输入简繁体汉字进行检索,并以汉语拼音方式呈现,点击后可看到相应中文。其他有中国古籍收藏的公私图书馆也均有书目数据库建设,足供查询。大英图书馆与剑桥大学、杜伦大学、爱丁堡大学、利兹大学、牛津大学、伦敦大学亚非学院共同建立了“UK Union Catalogue of Chinese Books”,^②可同时检索七家机构所收藏中文图书,使用简繁体中文或者拼音查询,可显示中文。牛津大学图书馆有“Serica”^③中国古籍馆藏目,有部分古籍和地图图像版免费共享。纽约公共图书馆^④可以用中文检索,但不是所有的返回结果都以中文显示。柏林国家图书馆建设了“Cross Asia”^⑤项目,除了自建数据库之外,还大量购置世界各地的数据资源提供给成员单位,是合作共享的典范。

大学图书馆如哈佛大学燕京图书馆亦建设了自己的古籍线装图书编目,初期即有约1800种,详情可在哈佛大学中国研究导航网页^⑥查阅。剑桥大学建设了一个中国丛书综合目录,^⑦包括中国近代现代丛书目录、中国丛书广录、中国丛书综录、中国丛书综录续编等四部目录书。哈佛大学CBDB项目建设了一个中文文本电子资源目录跨库查询系统,^⑧提供了10万余条古籍出处数据,以后会成为综合各大馆藏目录和古籍数据库目录的开放资源。普林斯顿大学建设了所藏中国善本书目目录,^⑨并可查阅详细版本信息。斯坦福大学“Socrates: Stanford Online Catalog”,^⑩可以检索该校中国图书目录。加州大学(洛杉矶)的在线目录,^⑪可以检索该校收藏中国古籍。旧金山大学亚太中心利玛窦中西

① <http://catalog.loc.gov/>.

② <http://www.bodley.ox.ac.uk/rslpchin/>.

③ <http://serica.bodleian.ox.ac.uk/>.

④ <http://catalog.nypl.org/>.

⑤ <http://crossasia.org/>.

⑥ http://guides.library.harvard.edu/Chinese?url_type=39&object_type=webpage&pos=1.

⑦ <http://www.lib.cam.ac.uk/mulu/congshu.html>.

⑧ <http://oopus.info/ccat/>.

⑨ <http://gest.princeton.edu/rarebook.htm>.

⑩ <http://bodoni.stanford.edu/uhtbin/cgisirsi/?ps=m5YJm2FGAq/GREEN/23000013/60/668/X>.

⑪ <http://catalog.library.ucla.edu/vwebv/searchBasic>.

文化历史研究所建设了利玛窦研究所藏书楼书目(The Ricci Institute Library Online Catalog),^①著录了约 85 000 种图书,50 000 种数字典藏数据。

另外,台湾“国家”图书馆,建设了一个世界范围内的“中文古籍联合目录”^②。在欧美方面,目前可以分别检索到多伦多大学、法兰西学院、法国国家图书馆、芝加哥大学、哈佛大学、加州大学(伯克利)、耶鲁大学、哥伦比亚大学、康奈尔大学、普林斯顿大学、国会图书馆、华盛顿大学、梵蒂冈图书馆、伦敦大学亚非学院、莱顿大学、巴伐利亚图书馆、德国国家图书馆、莱比锡大学、捷克科学院、里昂图书馆、牛津大学、里昂第三大学、鲁汶大学、华沙大学、加州大学(洛杉矶)等欧美图书馆和大学、研究机构所藏中文古籍书目。慕尼黑大学、比利时皇家图书馆、波兰亚捷隆大学、匈牙利科学院所藏中文古籍目录正在该网站建设之中。

加州在线档案馆(OAC)^③免费向公众提供著录信息入口,信息本身来自加州大学 10 个校区以及全加州图书馆、档案馆、各类学会等 200 多个机构,包含大批中国资料,多以近现代内容为主,古代内容较少。由美国 Henry Luce 基金会支持,耶鲁大学为主建设了“丝绸之路项目:重聚高昌宝藏”(The Silk Road Project: Reuniting Turfan's Scattered Treasures)数据库,^④搜集了 3 310 条数据,提供标题、编号等,对高昌资料编目较为详细,同时提供了研究书目和高昌国政治体系等专门内容。在博物馆方面,如大英博物馆、^⑤盖蒂博物馆、^⑥大都会博物馆、^⑦弗利尔美术馆、^⑧巴黎吉美博物馆、^⑨维多利亚和阿尔伯特博物馆^⑩等藏有中国文物较多的博物馆,均有自己的收藏文物检索网页。

基本上,古籍书目导航、数据库和博物馆在线目录库初步定型,数据库设计方案已经较为完善。经过多年建设,各大图书馆、博物馆主要藏书和文物均已纳

① <http://riccilibrary.usfca.edu/>。

② <http://rbook2.ncl.edu.tw/Search/Index/2>。

③ <http://www.oac.cdlib.org/>。

④ <http://turfan.research.yale.edu/turfan/>。

⑤ <http://www.britishmuseum.org/>。

⑥ <http://www.getty.edu/>。

⑦ <http://www.metmuseum.org/>。

⑧ <http://www.asia.si.edu/>。

⑨ <http://www.guimet.fr/fr/>。

⑩ <https://www.vam.ac.uk/>。

入登记,并开放网络检索。只要能联通网络,初步知道欧美哪些地方有什么书和文物存在,已经不再是一件让多数学者无法完成的不可能任务,虽然仍然要付出一定时间和精力。未来需要进行建设的,主要是这些目录库的连接和跨库检索功能。

二 ■ 古籍图像 ■ 文本数据库

在图像、文本数据库方面,数字化主力仍然是美国国会图书馆这样的大型图书馆、哈佛大学这样以文史见长的著名高校。国会图书馆^①早在 20 世纪 80 年代就开始用光学磁盘来保存文献,进行过全文图像的数字化处理工作。到 90 年代末期计划建立虚拟图书馆,对一部分中文古籍图书进行了电子化扫描,但是相对数量并不大。由美国国会图书馆和联合国教科文组织主导的世界数字图书馆(World Digital Library),^②收入了一定数量的中国古籍,中国国家图书馆提供了 171 种典籍书影,台湾“中央图书馆”提供了 131 种典籍书影。法国国家图书馆建设了 Gallica 数字图书馆,^③其中收录了少量的中国古籍。大英图书馆从 20 世纪 90 年代开始致力于数字化图书馆建设,目前已经形成了一批数字化影像,可以在大英图书馆专题网页^④上访问,其中提供了像素很高的 868 年印刷《金刚经》,超过 450 片甲骨、欧洲收藏数量最多的 24 卷《永乐大典》等古籍影像,阅读较为方便。德国巴伐利亚国家图书馆建有“DIGITAL EAST ASIA COLLECTIONS”,^⑤目前收有从 7 世纪到 19 世纪中文古籍约 1 141 册,共计 90 万张图片,提供在线高清图像无级缩放浏览,网站同时提供德文、英文、中文三种界面。美国国家医学图书馆建有“CLASSICS OF TRADITIONAL CHINESE MEDICINE”^⑥专题网页,提供了 13 部古医书书影。

哈佛大学在古籍数字化方面投入资金巨大,其自建包括古代中国内容的数

① <https://www.loc.gov/>。

② <https://www.wdl.org/>。

③ <http://gallica.bnf.fr/>。

④ <http://www.bl.uk/collection-guides/chinese-collection>。

⑤ <http://ostasien.digitale-sammlungen.de/cn/fs1/home/static.html>。

⑥ <https://www.nlm.nih.gov/exhibition/chinesemedicine/index.html>。

字化扫描项目有宝卷、中国旧海关资料、哈佛燕京图书馆中文善本特藏(与傅斯年图书馆、中国国家图书馆等合作)、中国珍稀方志、拓片收藏、满文古籍、蒙文古籍、纳西东巴经、韩南教授藏书、齐氏兄弟藏书、齐如山藏书等数十种,并提供了哈佛大学中国研究导航网页^①方便使用。其善本将全部扫描,上传网络供免费查阅。加拿大麦吉尔大学(McGill University)与哈佛大学及中国大陆高校合作,共同建设了明清妇女著作数据库^②专题网站。据该网站,截止到2016年12月15日,共收入了221部古籍,4939位女作家、诗人以及1398位男性作家。此数据库可以以书名、人名等方式查找,其个人介绍、作品出处,原文影像等均可一目了然。由蒋经国国际学术交流基金会资助,“中研院”历史语言研究所傅斯年图书馆、美国国会图书馆、哈佛大学、普林斯顿大学合作建设了古籍善本数位化资料库“A Digital Library of Chinese Rare Books”,^③可在线浏览。由哈佛大学德龙博士(Dr. Donald Sturgeon)创建的“中国哲学书电子化计划(Chinese Text Project)”^④收入超过3万部著作图像,还收录了哈佛燕京图书馆的5百多万页古籍善本图像。^⑤通过专门设计的OCR技术制定了有待勘定的全文,其页面设计完备,提供了中英双语页面。

海德堡大学建立了一个汉学文典,^⑥汇编了大量古代汉语文献,主要是为历史音韵学和中国汉字史提供资料。该校还建有“Digital Archive for Chinese Studies”,^⑦收录了各种类型的网络资源,但大部分是当代社会政治内容,少部分是古代典籍。宾夕法尼亚大学提供了一个“The Online Books Page”^⑧收入了超过200万册的网络免费电子书,可以检索到部分中文古籍。耶鲁大学建有耶鲁大学所藏中国善本数据库,^⑨收入了1796年以前出版有代表性的439部善本,

① <http://guides.library.harvard.edu/c.php?g=310134&p=2071022>。

② <http://digital.library.mcgill.ca/mingqing/>。

③ <http://rarebookdl.ihp.sinica.edu.tw/ezp-prod1.hul.harvard.edu/rarebook/Search/index.jsp>。

④ <http://ctext.org/>。

⑤ <http://ctext.org/library.pl?if=en&collection=139>。

⑥ http://tls.uni-hd.de/home_en.lasso。

⑦ http://www.zo.uni-heidelberg.de/boa/digital_resources/dachs/。

⑧ <http://onlinebooks.library.upenn.edu/>。

⑨ <http://web.library.yale.edu/digital-collections/chinese-rare-books-yale>。

对外提供了第一页的全文图像,并链接到耶鲁大学 online catalog,可查阅全部详细信息及版本情况。普林斯顿大学收藏中文医书最为有名,大约有 500 种,1 700 册,目前已有医书 53 种供免费阅览,详情可见葛思德东亚图书馆网页。^① 普林斯顿大学还建有另外一个“East Asian Library Digital Bookshelf”^②项目,与前者有所重复,目前有 147 种善本书可供在线阅览。不列颠哥伦比亚大学藏有蒲坂藏书和庞镜塘藏书数据库,^③前者共计 3 105 种 45 000 册,后者 64 种 796 册,目前有 126 种可在线阅读。哥伦比亚大学建有古巴华工调查录数据库,^④提供了影像电子文本。康奈尔大学建有“Wason Pamphlet Collection”,^⑤即华生收藏,是研究中国与西方关系的重要文献资源。共有 220 卷,1 200 册,在该校图书馆网站东亚页面^⑥有详细的介绍。

Adam Matthew Publications 出版公司利用华生收藏建设了一个专题数据库 China: Culture and Society — Wason Pamphlet, 1750—1929,^⑦收入该公司旗下的“Adam Matthew Digital”系列数据库,同在此系列数据库中的还有“海外收藏的中国近代史珍稀史料文献库”(China: Trade, Politics and Culture, 1793—1980),^⑧收入海外收藏的 1793—1980 年间中国与西方往来的珍稀史料,含大批档案及 7.2 万幅左右的图片。以及“中国、美国和太平洋数据库”(China, America and the Pacific),^⑨收入了从 18 世纪到 20 世纪的中国、美国、太平洋地区围绕贸易和文化形成的手稿、书籍、图像、地图等数字资源。其他商业公司也有不少提供包括中国 1911 年以前报刊和中外相关史料的数据库,兹不列举。

互联网记忆(Internet Archive)^⑩是非政府组织,致力于保存网页、文本、声

① http://library.princeton.edu/eastasian/diglib/?url_type=39&object_type=webpage&pos=1。

② <http://pudl.princeton.edu/collections/eal>。

③ <https://open.library.ubc.ca/collections/asian>。

④ <http://www.columbia.edu/cu/lweb/digital/collections/eastasian/texts/gubahua/index.html>。

⑤ <https://ecommons.cornell.edu/handle/1813/29701>。

⑥ <http://asia.library.cornell.edu/ac/Wason/index>。

⑦ <http://www.chinacultureandsociety.amdigital.co.uk/>。

⑧ <http://www.china.amdigital.co.uk/>。

⑨ <http://www.cap.amdigital.co.uk/>。

⑩ <https://archive.org/>。

音、影像、软件、图片,该网站有“eBooks and Texts”专栏,有图书 10 762 297 种,其中中文图书有 84 409 种,含有大量古代文献,并可提供全文下载,目前在中國大陸无法直接检索利用。美国犹他州家谱图书馆制作了在线家谱检索系统(Family History Library Catalog),^①并从北美、中国、东南亚收集了中国 1239—2014 年的族谱,^②共计 13 247 185 幅影像,可在线全文阅读。谷歌公司建有著名的“谷歌图书”^③搜索,早在 2008 年即完成 700 万册图书的扫描,所收图书大多有简单介绍及出处,但是中国古籍部分非常混乱,提供了一部分无版权图书的阅览和下载,在中國大陸无法访问。

国际敦煌项目(INTERATIONAL DUNHUANG PROJECT, IDP)^④由大英图书馆发起,联合中国、俄罗斯、日本、德国、法国、韩国的相关机构共同参加,通过国际合作开发各国敦煌文献并全部数字化,以促进共享,但是该网站的中国北京^⑤和中国敦煌^⑥两个部分,访问速度缓慢,似乎较久未更新。德国部分自 2013 年,法国部分自 2009 年后未再更新。由美国学者 Michael S. Hart 倡导的古登堡计划(Project Gutenberg),^⑦是一个图书共享计划,目前提供 53 000 种具有公共版权的免费图书,其中包括了一定数量的中文文本图书,大多数是中国经典和古典文艺作品,而且同时提供在线 HTML 文本、EPUB、Kindle、UTF-8 文本等多种格式,对于阅读来说极为方便。中美百万册书数字图书馆计划(CADAL)^⑧美国卡耐基梅隆大学、西蒙学院(Simmons College)、中国科技大学、浙江大学、清华大学等国内外高校先后参与,扫描图书总量达 250 万册,古籍部分超过 50 万册,管理主体目前已经转移到浙江大学,国内高校从校园网提供服务,CADAL 系统根据参建单位提供 IP 地址范围控制用户阅读权限。

欧美地区的中国古籍全文图像扫描和数字化工作,与中國大陸、中国台湾地

① <https://familysearch.org/>。

② https://familysearch.org/search/collection/1787988?url_type=39&object_type=webpage&pos=1。

③ <https://books.google.com/>。

④ <http://idp.bl.uk/>。

⑤ <http://idp.nlc.cn/>。

⑥ <http://idp.dha.ac.cn/>。

⑦ <http://www.gutenberg.org/>。

⑧ <http://www.cadal.zju.edu.cn/index>。

区相比,已经相对落后。除了哈佛大学等具有雄厚财力的机构之外,大多数藏有中国古籍的机构似已放缓了图像扫描和文字识别的步伐。

三 ■ 文物图像 ■ 照片和地图数据库

欧美地区古代中国资料数字化建设的重要一环是文物图像和地图,各机构在这方面投入了极大人力财力,建设了相当多数量的数据库。在图像免费共享方面,美国远远走在了世界各国的前面。文物图片自然以博物馆为最大宗,大英博物馆“Collection online”,^①以“Chinese”为关键词,可搜出 24 217 件藏品(含非中国文物),提供约 600 * 600 PX 图像下载。盖蒂博物馆在线“COLLECTION”,^②以“Chinese”为关键词,可以搜到 357 项藏品,目前免费提供最高可达约 3 000 * 4 500 PX 的高清图像下载。波士顿美术馆(MFA)在线“CLLECTION”,^③以“Chinese”为关键词,可以搜索到 11 127 项藏品,目前免费提供约 1 600 * 1 000 PX 的图像下载。大都会博物馆的“COLLECTION”,^④以“Chinese”为关键词,可以搜到 7 012 项藏品,目前免费提供约 3 000 * 4 000 PX 的高清图像下载。亚洲艺术博物馆的“ONLINE COLLECTION”^⑤目前可提供在线浏览中国古代的艺术品万余件,免费提供约 800 * 500 PX 的图像下载。弗利尔美术馆建有宋元书画专题数字收藏,^⑥目前提供 85 件作品约 500 * 500 PX 图像下载。谷歌有“Google 文化学院”(Google Arts & Culture),^⑦提供 1 000 多家顶级博物馆和档案馆中的精选内容,进行虚拟展览,含有大量中国内容,方便用户在网络上全景观看,目前中国大陆无法使用。

德国柏林普鲁士文化基金会建有“乾隆朝战争铜版画”^⑧图像数据库,有 64 幅图像,可提供高清大图浏览,这批数据也同时收入了世界图书馆的在线馆

① http://www.britishmuseum.org/research/collection_online/search.aspx。

② <http://www.getty.edu/art/collection/>。

③ <http://www.mfa.org/collections>。

④ <http://www.metmuseum.org/art/collection>。

⑤ <http://searchcollection.asianart.org/>。

⑥ <http://www.asia.si.edu/SongYuan/defaultChinese.asp>。

⑦ <https://www.google.com/culturalinstitute/beta/>。

⑧ <http://crossasia.org/digital/schlachten-bilder/>。

藏。^① JSTOR 数据库原本只收录刊物,但自 2000 年以后逐渐建立 ARTstor 图像资料库,^②以“Chinese”为关键词,目前可以检索出大约 2 000 条数据,通过 Adobe Flash 插件方式在线浏览高清大图。纽约公共图书馆的“Digital Collection”,^③以“chinese”为关键词,可搜到 3 021 件藏品,提供约 800 * 800 PX 的图像下载。由英国联合信息系统委员会(JISC)支持,不列颠学院(The British Academy)、英国艺术与人文研究理事会(AHRC)和布里斯托大学等建立的“Visualising China”数据库,^④收藏了从 1850 年到 1950 年的中国照片约 8 000 张,其中 1850—1911 年约 2 600 张。芝加哥菲尔德博物馆建有中国拓片数据库(Chinese Rubbings),^⑤提供部分拓片最大约 800 * 800 PX 图像下载。

陈刘钦智(Ching-chih Chen)主持建设了“Global Memory Net”,^⑥目前提供文化、历史、文物方面的图像,有一部分为中国古代相关内容,提供在线无级放大浏览。中国中古研究学会(美国)建有六朝陵墓数据库(Six Dynasties Tombs Database),^⑦收有六朝时期 1 739 座陵墓的相关数据。公共图片方面,Flickr 是雅虎旗下的图片分享网站,^⑧功能极为强大,可以使用标签和分组方式来管理图片,许多档案馆、博物馆、图书馆利用 Flickr 来分享图片。据王国强整理,有几十个专题,十万张以上中国老照片,如 Gary Lee Todd 博士拍摄收藏了中国各个历史时期图片,^⑨数量极为庞大,目前该网站在中国大陆无法访问。

在大学方面,哈佛大学建有中国拓片专题网页,^⑩有大约 2 600 方墓志拓片,目前可提供小、中、大、超大四种图像下载,最高可达约 1 100 * 2 400 PX。加州大学(伯克利)也建有中国拓片数据库(Chinese Stone Rubbings Collection),^⑪有超过 1 500 张拓片,提供最高约 1 100 * 1 500 PX 的图像下载。耶鲁大学的中

① <https://www.wdl.org/>。

② <http://artstor.org/>。

③ <https://digitalcollections.nypl.org/>。

④ <http://visualisingchina.net/>。

⑤ <http://archive.fieldmuseum.org/chineserubbings/index.html>。

⑥ <http://www.memorynet.org/>。

⑦ <http://www.earlymedievalchinagroup.org/sixdynastiestombs.html>。

⑧ <https://www.flickr.com/>。

⑨ <https://www.flickr.com/photos/101561334@N08/>。

⑩ http://vc.lib.harvard.edu/vc/deliver/home?_collection=rubbings。

⑪ <http://www.lib.berkeley.edu/EAL/stone/>。

国基督教大学图像数据库(China Christian Colleges and Universities Image Database),^①收集了 10 629 张老照片,其中大部分已经数字化。莱顿大学数字图书馆,^②数字化收藏了中国古代艺术品图像以及早期照片,数量不多,目前可提供约 1 700 * 2 600 PX 的图像下载。杜克大学数字馆藏(Digital Collections)^③以“china”为检索词,共有 5 353 条,其中包括了手稿和照片。在山东传教 50 余年的高第丕(Tarlton Perry Crawford)夫人 Martha Foster Crawford 的日记手稿亦在其中。

不列颠哥伦比亚大学的“Yip Sang Collection”,^④收集了温哥华叶氏家族各类文档 639 种,是研究早期华人移民北美的珍贵资料。哥伦比亚大学东亚图书馆建有“门神纸马数据库”(Chinese Paper Gods),^⑤收集了各类神仙图像 231 幅。普林斯顿大学的中国版画,^⑥有 30 幅,目前可提供大约 800 * 800 PX 图像下载。耶鲁大学丝绸之路数据库,^⑦为新拍摄的丝绸之路沿线重要古迹照片,从 2006—2010 年共拍摄了超过 11 000 张高像素照片。耶鲁大学有 Peter Parker 收藏, Lam Qua(林官、关乔昌)所绘的 80 幅中国病人患病图像,即 Peter Parker's Lam Qua Paintings Collection。^⑧ 华盛顿大学建设有“Silk Road Seattle”,^⑨主要是一个教育项目,但是搜集了较多的相关数据,其“COLLECTIONS”就搜集了大约 34 家图书馆、艺术馆的网络链接,^⑩同时可提供其他如丝绸之路城市、传统文化、地理等方面的图片,但像素不高。耶鲁大学美术馆的“COLLECTION”,^⑪以“Chinese”为关键词,有 8 477 件藏品,提供约 1 500 * 3 000 PX 的图像下载。

① <http://web.library.yale.edu/digital-collections/china-christian-colleges-and-universities-image-database>。

② https://socrates.leidenuniv.nl/R/?func=search-simple&local_base=gen01-disc。

③ <http://library.duke.edu/digitalcollections/>。

④ <https://open.library.ubc.ca/collections/yipsang>。

⑤ http://www.columbia.edu/cu/lweb/digital/collections/eastasian/paper_gods/index.html。

⑥ <http://pudl.princeton.edu/collections/pudl0030>。

⑦ <http://web.library.yale.edu/digital-collections/yale-silk-road>。

⑧ <http://cushing.med.yale.edu/gsd/collect/ppcdot/>。

⑨ <http://depts.washington.edu/silkroad/>。

⑩ <http://depts.washington.edu/silkroad/museums/srmuseums.html>。

⑪ <http://artgallery.yale.edu/collection/search>。

芝加哥大学“Digital Scrolling Painting Project”,^①收集了宋元明清时期的84幅手卷,提供在线高清大图浏览。里德学院(Reed College)建有一个19世纪台湾的图像数据库(Formosa: 19th Century Images),^②有数百张图像。明尼苏达大学图书馆的“Ming Gazetteer Images”,^③从明代各类地方志收集了郡县全图1010幅。加州大学建设了一个“World Images”^④数据库,以“Chinese”为关键词,有727件藏品,大部分内容是关于古代中国的。华盛顿大学建有“Robert Henry Chandless Photographs”^⑤数据库,收藏了1898—1908年间中国各地照片287张。欧柏林学院(Oberlin College)建有“Oberlin in Shansi digital collection”^⑥数据库,收藏有中国各个时期的图片810张,大部分是近代照片。

在老地图收藏和数字化上,哈佛大学也走在前面,建有哈佛大学地图收藏(Harvard Maps Collection)^⑦数据库。其中的“Army Map Service Series L500 of China”^⑧大多是20世纪50年代的老地图。哈佛大学“South Central China and Tibet: Hotspot of Diversity”专题,^⑨提供了10幅历史地图。德州大学(奥斯汀)建有一个“Historical Maps of China”数据库,^⑩收藏了1860到1967年间中国地图70余幅,提供最大约1500×1200PX图像下载。华盛顿大学也收藏了一批老地图“Skinner Map Collection”,^⑪其中中国地图有500多幅,多数为18、19世纪地图。美国国会图书馆地图数据库,^⑫收录了17733幅地图,以“China”为关键词,可检索出558幅,最早为1769年地图。

① <http://scrolls.uchicago.edu/>。

② <http://rdc.reed.edu/c/formosa/home/>。

③ <http://umedia.lib.umn.edu/taxonomy/term/725>。

④ <http://worldimages.sjsu.edu/>。

⑤ <http://digitalcollections.lib.washington.edu/cdm/search/collection/chandless>。

⑥ <http://dcollections.oberlin.edu/cdm/search/collection/shansi>。

⑦ <https://library.harvard.edu/libraries/harvard-map-collection>。

⑧ http://hcl.harvard.edu/libraries/maps/collections/series_indices/China_Index.html; 或 <https://www.loc.gov/maps/?fa=contributor%3Aunited+states.+army+map+service%7Clocation%3Achina&all=true&st=slideshow>。

⑨ <http://www.arboretum.harvard.edu/library/image-collection/south-central-china-and-tibet-hotspot-of-diversity/map-and-gazetteer/>。

⑩ http://www.lib.utexas.edu/maps/historical/history_china.html。

⑪ <http://content.lib.washington.edu/skinnerweb/index.html>。

⑫ <https://www.loc.gov/maps/>。

世界数字图书馆收藏了多幅珍贵古代地图,与中国相关的有清末北京全图、^①台湾及澎湖列岛全图,^②朝鲜的天下地图^③和朝鲜舆地图、^④七省沿海图、^⑤西藏全图、^⑥坤輿万国全图、^⑦江苏至北京运河全图、^⑧新疆地輿总图^⑨等数十幅,由美国国会图书馆等提供。莱斯大学(Rice University)所藏的“京板天文全图”,^⑩是1780—1790年间的古地图。台湾“中研院”建有美国国会图书馆典藏之中国相关地图文献查询系统,^⑪有美国国会图书馆所藏关于中国地图的78 151个数据记录,最大提供约800 * 800 PX阅读下载。

四 ■ 关系型数据库与数字人文工具

欧美关于中国古代内容的数字资源,最大贡献还不是各类文本、图像类数据库建设,而是在数据方法和数字人文思路的导向上。特别是哈佛大学发起的中国历代人物传记资料库(CHINA BIOGRAPHICAL DATABASE PROJECT, CBDB)^⑫项目,更是引领了关系型数据库的发展潮流,也是迄今为止自身最有学术含量的数据库,网站并提供中英双语同步更新。

CBDB数据库最早建设思路来自研究宋代社会经济史的郝若贝教授(Robert M. Hartwell, 1932—1995),他从80年代就开始把宋代人物的传记资料录入到计算机里面,同时自己编写软件来阅读和使用这些数据。与其他人不同的是,郝若贝极有远见地建立起人物之间的关系,所建立数据库里面包括各种人物的完整信息,如生卒年、亲属、门生、履历,等等,利用这些材料,学者可以进

① <https://www.wdl.org/zh/item/44/#languages=zho>。

② <https://www.wdl.org/zh/item/208/#languages=zho>。

③ <https://www.wdl.org/zh/item/2686/#languages=zho>。

④ <https://www.wdl.org/zh/item/4460/#languages=zho&page=2>。

⑤ <https://www.wdl.org/zh/item/314/#languages=zho>。

⑥ <https://www.wdl.org/zh/item/317/#languages=zho>。

⑦ <https://www.wdl.org/zh/item/4136/#languages=zho&page=2>。

⑧ <https://www.wdl.org/zh/item/7097/#languages=zho&page=5>。

⑨ <https://www.wdl.org/zh/item/11385/#languages=zho&page=7>。

⑩ <http://library.rice.edu/sites/default/files/pictures/ux/jingban2.jpg>。

⑪ http://webgis.sinica.edu.tw/map_loc/。

⑫ <http://projects.iq.harvard.edu/chinesecbdb>。

行深入的研究与分析。郝若贝教授去世后,他遗留的数据库捐赠给了哈佛—燕京学社。由包弼德(Peter K. Bol)教授领导组织,结合了历史学、文学、计算机、网络等多个学科的专家,不断更新与完善郝若贝教授的数据库,目前哈佛大学和北京大学、台北“中研院”合作,共同推动这个项目继续向深入发展。到2016年4月,已经收入了37万人的传记资料,同时仍然在不断扩充,目标是收录全部民国以前的人物数据。

CBDB项目收录的是人物信息,但是它并不是一个人物辞典,虽然它可以被当做一个辞典来查询,但是它的作用并不仅仅在于可供查询。从数据的角度来看,CBDB可以提供相对最大数量的数据,当数量足够多的时候,一些微小错误就不足以影响检索结果整体的正确性。CBDB以人物为中心,同时搜集姓名、时间、地址、官职、著作、亲属关系、社会关系等多个方面的数据。孤立看这些数据的时候,并不能发现规律性,但是大批数据集合在一起的时候,就可以对整体趋势进行判断和分析。同时也可以为我们经常心有戚戚,但是很难找到证据的一些看法提供证据或者反证。CBDB与前述查询检索类的数据库不同,它并不保留原文,甚至不收集原文,它只是从原文当中提取出来以“人物”为中心的数据。从这个意义上说,CBDB不是一个单纯的古籍资料数据库,它更是一个研究的基础工具。它给予收入数据库的每一个人一个单独代码,然后围绕“人”把不同的字号、官衔、地域、亲属、朋友,等等都编织到一个网络里面。这样的信息,就会提示给学者以新的思路。配合历史地理信息系统,可以在地图上把这些人都找到,看出他们的诸多关系。其他的古籍数据库,仍然是以原书为单位,所提供信息量并未超过纸质图书所提供的信息。换句话说,普通检索只是数量在快速增加,但里面信息含量并未增加,在检索方法的改良上也往往没有太大进展。

在CBDB数据库当中的检索,并不像其他文本数据库那样只有关键词的检索。它的检索是开放而且是多重的,可以进行复杂条件的检索。比如它可以查询出生在“苏州”的所有人,或者查询哪些人是通过何种方式获得官职,或者是某个职务的任职者,最低任职年龄与最高任职年龄,等等。通过这些查询,就可以为历史研究提供最基本的数据,特别是为社会网络分析提供最佳材料,同时可以基于所有现存材料去分析一组问题,乃至多重组的问题。比如可以讨论宋代的女性死亡年龄、生育年龄,以及她们与丈夫的年龄差,出嫁与家庭所在地的距离,

等等。因为得到的数据准确而且相对完全(局部错误不足以影响全局判断),学者就可以把更多的精力用于分析和思考。

与 CBDB 相配合的最佳工具是中国历史地理信息系统(China Historical GIS, CHGIS),^①由哈佛大学和复旦大学合作建设,中文版在复旦大学历史地理研究所的禹贡网。^② CHGIS 是一个为学术研究而建设的 GIS 数据平台,提供尽量简便的操作,可以更新相关数据,同时是一个可靠且开放的基础地理信息数据库。任何一个用户使用 CHGIS 基础数据,同时加入自己搜集的相关数据,就可以建设或者生成自己的专题地图,可以进行空间分析,按自己研究需要建设特定的统计模型,分析结果可以输出为表格、电子地图或者任何类型的数据。哈佛大学另建有免费在线 GIS 平台(World Map),^③是一个开放性的地图平台,任何人都可以登记建立账号创建地图并查阅地理信息,前述 CBDB 数据库在其中建有一个独立的频道“CBDB Spatialized Datasets”。^④同时其中已建有 6 877 种各类地图,中国相关的有 154 种,数据最丰富的当属“China Map”。^⑤如果学者有需要,也可以利用 World Map 所提供的基础功能,生成某个类型的专属地图。哈佛大学图书馆另外设有“Harvard Geospatial Library”,^⑥提供类似功能。密歇根大学建有 FREE MAPPING ONLINE^⑦是一个地图编辑工具,可以自行整理上传数据,绘出地图。

由莱顿大学等合作开发的码库思(MARKUS),^⑧是一个自动标记工具,它可以快速在古籍文本上标记出人名、地名、年号、官名。这个工具本身没有任何文本收录,所有需要标识的文本都需要用户自行上传。当面对大量文本需要批量找出其中人名、地名、年号、官名的时候,码库思就可以发挥作用。当然机器自动标识的错误极多,需要人工阅读修正。网站本身所提供的修改界面操作较为繁琐,不适合大批量快速操作。CBDB 项目经理王宏魁设计了基于 Chrome

① <http://www.fas.harvard.edu/~chgis/>。

② <http://yugong.fudan.edu.cn/views/chgis/index.php>。

③ <https://worldmap.harvard.edu/>。

④ <https://worldmap.harvard.edu/maps/CBDB>。

⑤ <https://worldmap.harvard.edu/chinamap/>。

⑥ <http://calvert.hul.harvard.edu:8080/opengeoportal/>。

⑦ <http://spatialdataonline.org/>。

⑧ <http://dh.chinese-empires.eu/beta/index.html>。

的快捷操作插件,可以大幅度提高工作效率。The Sieve Online^① 是码库思的一个插件,可以用来在线对比同书的不同版本,对比发现文本中字词和通用蒙学书籍用字范围的不同,提供《百家姓》、《千字文》等以供选择。加州大学(尔湾)由 Andrew W. Mellon 基金会支持,建立了一个明代政府官名的众包翻译项目(Ming government official titles: A Crowd-Translation Project),^②通过网络整理、编辑网络免费版《明代官名辞典》。

美国有一位佛学爱好者,开发了佛光山南天大学佛教文本阅读器“NTI Buddhist Text Reader”,^③主要是帮助西方人阅读《大正藏》,打开相应文本后,会提供拼音、英译还有注释,也包括各种佛教名词的标准拉丁文转写。Christian Wittern 开发了慢读(Mandoku)阅读工具,^④可以同时阅读和对比不同版本的古籍。德国马普研究所陈诗沛主持建设了“Local Gazetteers Project”,^⑤使用自行研发的一系列工具来从事中国古代地方志研究,仅对其所内用户开放。由大众基金会支持,哥廷根大学等共同建设了 Wissenschaftssprache Chinesisch/Modern Chinese Scientific Terminologies (近现代汉语学术用语研究)数据库,^⑥试图把1600—1911年的哲学、逻辑、物理、化学、地理、政治等方面的名词翻译为德文。加州大学(美熹德)建有宋代数字地名数据库“The Digital Gazetteer of the Song Dynasty”,^⑦可以作为学术研究参考工具。苏黎世大学建有“China and the West: 1245—2000”数据库,^⑧其中有少部分资料数据,但大部分是研究著作。宋代研究工具书刊指南(Song Research Tools),^⑨是同名书籍的在线版,提供了相关古籍书目。电子佛教辞典(Digital Dictionary of Buddhism),^⑩可以提供大部

① <http://dh.chinese-empires.eu/beta/sieveOnline.html>。

② <http://mingofficialtitles.lib.uci.edu/#/>。

③ <http://ntireader.org/>。

④ <http://www.mandoku.org/>。

⑤ <https://www.mpiwg-berlin.mpg.de/en/research/projects/departmentschaeferspcms/LocalGazetteers>。

⑥ <http://kjc-sv016.kjc.uni-heidelberg.de:8080/exist/apps/wsc/modules/search/index.html>。

⑦ <http://songgis.ucmerced.edu/>。

⑧ <http://idoasdb.uzh.ch/oas/>。

⑨ <http://www.songyuan.org/SongTools/>。

⑩ <http://www.buddhism-dict.net.ezp-prod1.hul.harvard.edu/ddb/index.html>。

分佛教词条的英文解释。汉日韩越—英辞典(CJKV-English Dictionary),^①在线提供四种语言词汇的英文解释。以上所介绍都是与中国古代研究直接相关的一些工具,在数字人文方面还有很多通用工具,例如可用于中文语法分析的 Stanford Parser,^②用于数据可视化的 Palladio,^③用于语言分析的 Classical Language Toolkit^④,等等,兹不详述。

五 ■ 欧美古代中国资源数字化建设的几点思考

欧美地区特别是美国图书馆、博物馆等相关机构对中国图书、文物的收藏和利用,在世界上首屈一指。美国也是亚洲以外收藏中文图书最多的地区,欧洲则收藏了一批珍贵古籍版本,特别是敦煌文献,大部分都收藏在欧洲。在数字化方面,美国的中国古籍数字化也起步发展最早,欧洲在这方面虽然投入不如美国,但是也建设了一批中文数字典藏。总体上看,在中文古籍文献总目、图像扫描和文字识别方面,中国大陆和中国台湾地区逐渐后来居上,但是欧美的中国古代数字资源建设还是有自己的特点与优长,值得多加学习。

首先是欧美中国古代数字资源大多免费,除商业公司外,图书馆、大学等均免费向社会提供文献资源,特别是用公款投入的几乎全部免费。这既值得赞赏,又应该大力推广并学习。国内比较成熟的大型数字化资源多为商业公司建设,几乎没有大型古籍数字化建设项目免费向公众开放,免费古籍数字化网站大多是网友自发建设,也几乎没有图书馆、大学、博物馆将馆藏古籍、文物扫描拍照并无偿向公众开放。古籍、文物都是中华民族祖先创造的伟大遗产,而且是人类文明的瑰宝,无论保存在哪里,都应该加以保护和传承。特别是由政府投入的项目,理所应当应该全部免费向社会公众开放。建议今后凡是有政府资金投入的古籍整理、文物图像扫描数据库项目,必须免费向社会公众开放,这应该成为国内学术界获得政府资金的一项基本规则。

① <http://www.buddhism-dict.net.ezp-prod1.hul.harvard.edu/dealt/>。

② <http://blog.csdn.net/cuixianpeng/article/details/16864785>。

③ <http://hdlab.stanford.edu/palladio/>。

④ <http://cltk.org/>。

其次,欧美古代中国数字资源建设最早是图像扫描和文字识别,引领了八九十年代的古籍数字化潮流,但由于批量扫描工作的学术含量不足,后续利用情况又不尽理想,导致很多大学、图书馆纷纷停止或者放缓了批量扫描工作的步伐。但这是非常可惜的事情,只有扫描了足够多的书,并加以文字识别,才会形成更多更好的文本数据或专门的数据库,才有更多的可能去推动学术研究进步。如果没有数量上的逐步积累,就很难形成学术品质上的不断突破。希望欧美图书馆、大学等机构能继续加强古籍数字化建设,将全部馆藏扫描,并提供长期免费的下载。同时也希望国内的网友按需阅读,理性下载,不要导致国外机构采取技术手段反制,影响其他人的使用。

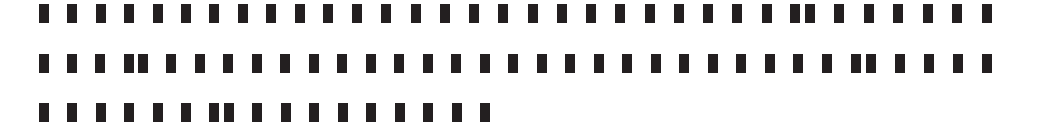
再次,欧美在图像数字化方面投入巨大,也取得了相对更多的成果。特别是照相术发明以来,大部分早期中国影像均依赖于欧美探险家、政府官员、传教士、学者拍摄并保存至今。所以中国图像数字化一直是欧美图书馆、大学的热门项目。但是这些图像极为分散,如果能有中国相关学术机构能和收藏机构合作,建立一个类似古籍联合目录的全球统一检索系统,则对学术研究工作大为有益。但是这样的工作既辛苦又需要丰富学术经验,而个人的学术成果收获甚微,需要有一个机制去解决这样的问题。另外,如果能有中国机构进行最基本的统计工作,把所有散落世界各地的中国文物统一编制目录,提供网络检索,则最有学术价值。更希望中国各博物馆均能将馆藏文物拍照,发布高清图像免费浏览。

第四,美国仍然引领着中国古代文化遗产数字化的前沿发展趋势。特别是关系型数据库建设和数字人文的发展方向,仍然由欧美学术机构引领,中国大陆的学术机构在古代资源数字化方面暂时还无法总体性超越美国。虽然中国大陆和中国台湾地区在扫描图像和文字识别方面发展势头迅猛,但是古代资源数字化的发展方向是关系型数据库建设和数字人文,而这些方面中国大陆相对还较为落后。希望中国大陆学术机构也能建设出类似于 ■■■■ 这样优秀而免费的关系型数据库,或参与合作,推动更多的高校和学者投身到数字人文,建设出更多更好的数据库和工具。当然,欧美中国文化遗产数字化和数字人文的发展,海峡两岸暨香港、澳门华人在其中发挥了重要作用,贡献极大,居功甚伟。几乎每个取得成功的项目背后,都有一群华人在默默奉献。

总之,欧美在中国古代文化遗产数字化建设方面仍然居于领先地位,希望中

[illegible]

欧美地区的古代中国数字资源概述



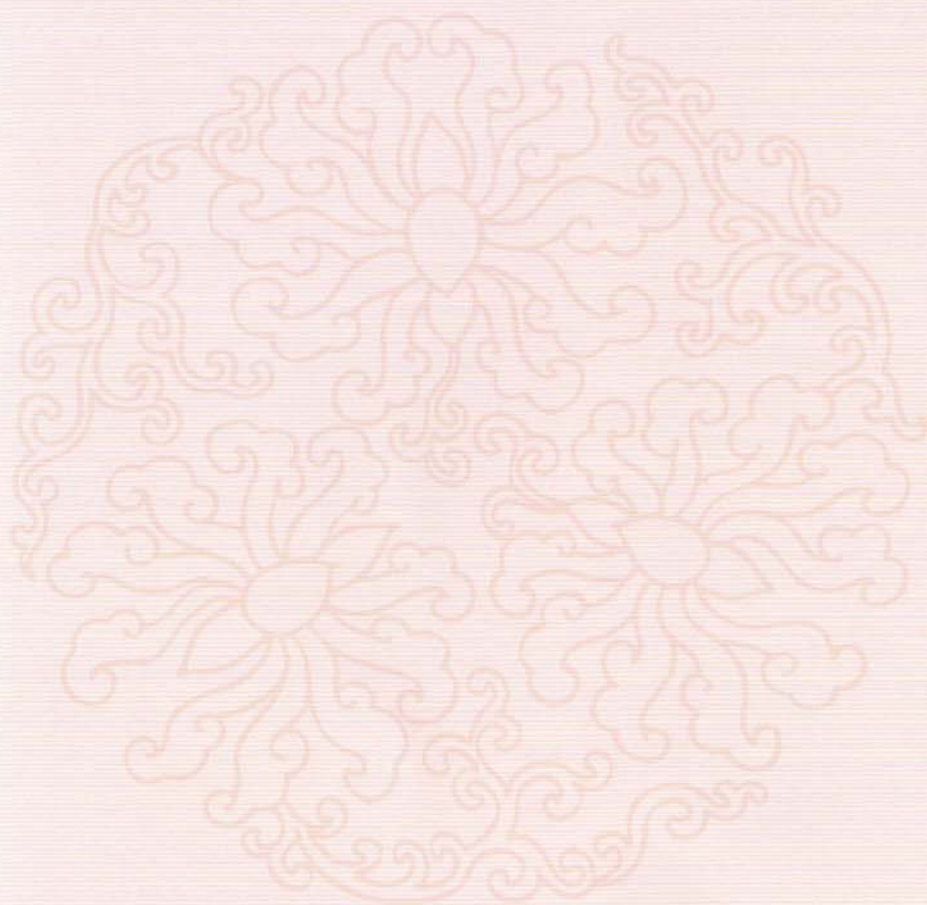
(本文撰成之后，曾请徐力恒、王宏魁、陈诗沛兄审阅一过，多有指正，深表谢意，当然文中错误仍由本人负责。)



峡两岸暨香港、澳门华人在其中发挥了重要作用,贡献极大,居功甚伟。几乎每个取得成功的项目背后,都有一群华人在默默奉献。

总之,欧美在中国古代文化遗产数字化建设方面仍然居于领先地位,希望中国的相关主管机构和学术研究机构能多多学习欧美经验,建设更多的免费数据库,发展出由中国主导的中国古代研究方面数字人文项目,建设更多的关系型数据库,推动中国的学术进步。

(本文撰成后,曾请徐力恒、王宏甦、陈诗沛兄审阅一过,多有指正,深表谢意,当然文中错误仍由本人负责。)



欢迎关注中西书局官方微信

上架建议 中国史

ISBN 978-7-5475-1600-3



9 787547 516003 >

定价：98.00元
易文网：www.ewen.co