

中国古代科技文献专题数据库建设现状 及改进策略*

郑翔¹ 关思雨¹ 耿廉璠²

(1. 武汉大学信息管理学院, 武汉 430072; 2. 湖北大学计算机与信息工程学院, 武汉 430062)

摘要: 通过文献搜集、网络调查等方式详细调研当前中国古代科技文献专题数据库的建设现状, 统计中国古代科技文献专题数据库建设的主体构成、数据类型及系统功能, 发现当前中国古代科技文献专题数据库建设中存在缺乏规范、访问受限、介绍不足、学科失衡、功能单一、互动薄弱的问题, 提出若干改进策略, 即实现统一顶层设计、开放获取渠道、及时更新介绍、保障学科齐全、探索多样功能、提供互动环境。

关键词: 古代科技文献; 专题数据库; 古籍数字化

中图分类号: G257; G259.2 **DOI:** 10.3772/j.issn.1673-2286.2023.12.007

引文格式: 郑翔, 关思雨, 耿廉璠. 中国古代科技文献专题数据库建设现状及改进策略[J]. 数字图书馆论坛, 2023(12): 66-77.

中国古代科技文献承载着中华先民对科学技术的不懈探索, 蕴含着丰富的科技思想与智慧, 是中华民族智慧的结晶与中华优秀传统文化的重要组成部分, 兼具实用性与人文性^[1]。例如:《太初历》《大衍历》《授时历》等较系统地反映了古人对日、月、行星运动的不断探索和对年、月、日、时、节气的细致安排, 代表着我国古代日渐成熟的历史法体系;《禹贡地域图》《水经注》《徐霞客游记》等记录了我国古代的地域分布、大川河流、地貌风土, 是我国古代地理学著作的优秀代表;以《汜胜之书》《齐民要术》《陈敷农书》《王祯农书》《农政全书》为代表的农学著作是我国古人农业生产智慧的结晶, 对当代农业科学研究仍有借鉴价值;《九章算术》《周髀算经》《墨经》《梦溪笔谈》等著作在数学、物理、化学等多学科领域展现了我国古代灿烂辉煌的科技成就;《黄帝内经》《伤寒论》《本草纲目》等医药类文献对当代疾病防治颇有启迪之效。所谓“疑今者, 察之古; 不知来者, 视之往”, 我国当今科技的飞速发展离不开前代科技思想与智慧的支撑。但我国古

代科技文献体系庞大、卷帙浩繁, 使用者在查找、阅读和利用过程中面临诸多不便。历经多次书厄, 不少珍贵的古代科技文献已失传或残缺不全, 其历史传承价值和科学文献价值受到严重影响, 对它们进行数字化保存与开发利用迫在眉睫。

数据库建设是文献整理发展到一定阶段后的必然趋势。2007年,《国务院办公厅关于进一步加强古籍保护工作的意见》提出要大力实施“中华古籍保护计划”,“制定古籍数字化标准,规范古籍数字化工作,建立古籍数字资源库”^[2]。随后,原文化部召开全国古籍保护工作会议,古籍保护工作全面启动。《中华人民共和国国民经济和社会发展第十三个五年规划纲要》也提出要“建立中医古籍数据库和知识库”^[3]。随着“中华古籍保护计划”力度的不断加大,以医学、农学等领域文献为代表的古代科技文献成为古籍数字化工作的重点对象和学术研究的热点。

古代科技文献涵盖天文、地理、农业、水利、数学、物理、化学、生物、医药、工程工艺技术等内容,内

收稿日期: 2023-08-27

*本研究得到教育部哲学社会科学重大课题攻关项目“中国古代科技文献整理与研究”(编号: 19JZD042)资助。

容全面且分布范围广泛。涵盖古代科技文献的数据库主要分为通用型数据库和专题数据库两大类。通用型数据库通常由数字出版企业单独构建,或由多个高校图书馆联合构建,古籍收录规模较大、系统功能也较为完善,例如:北京时代瀚堂科技有限公司开发的“瀚堂典藏”包含农家农书、术数命理、科学技术、中医药文献等版块;“鼎秀古籍全文检索平台”共收集了近3万部、50万卷的古籍数据,主要集中于兵家类、术数类、农家农学类、工艺类、天文历算类等版块。专题数据库则从研究古代科技史的角度,依据现代学科体系分门别类地收录某一学科的古科技文献。在特定的学科范围内,相较于通用型数据库而言,专题数据库所收录文献往往更为全面,能够为用户提供精准学科专题服务。

本研究主要调研国内古代科技文献专题数据库建设现状,在此基础上分析相关建设成果及特色,并结合其存在的问题,对古代科技文献专题数据库的建设提出若干对策性建议。

1 我国古代科技文献专题数据库建设现状

以6批共计203个全国古籍重点保护单位为切入点,对它们自建古代科技文献专题数据库的情况进行调查统计,并结合其订购数据库情况展开网络搜索,以获得数字出版企业的相关数据库构建情况。同时,有针对性地对与古代科技有关的行业性大学自建专题数据库的情况进行调研,以补充和完善上述调查结果。调查发现,高等院校、科研院所、图书馆和数字出版企业纷纷从地理、农业、水利、医药等学科角度构建了古代科技文献专题数据库。

1.1 地理数据库

古代地理文献从自然地理、人文地理、经济地理、政治地理、城市地理等方面综合记载了我国古代的山川地貌、物产风俗、矿物岩石等,数量众多、内容丰富、类型多样。此领域的数据库以方志类古籍数据库居多。常见地理类古代科技文献数据库如表1所示。其中,华东师范大学调查与数据中心构建的“数字方志集成平台”在资源集成、知识组织与数据关联的基础上,使用知识图谱等技术,可视化展现方志资源内涵。除方志类数据

库外,较有特色的地理古籍数据库还有上海睿则恩信息技术有限公司构建的“古地图数据库”和北京书同文数字化技术有限公司构建的“明清两朝边塞海疆地理文献汇编全文检索系统”。前者收录了3 000余张古地图,绝大部分图片清晰度较高,具有较高学术价值;后者收录了明清边塞海疆地理相关史料文献,包括世界地理文献、边塞海防文献、西藏事务文献、新疆西域文献、辽东疆域文献等,不仅提供古籍的在线浏览服务,还提供了校勘、标注、下载等功能。

1.2 农业数据库

古代农业文献从农、林、牧、渔等方面记载了古代农业生产过程,反映了我国传统农业历史特征^[18]。目前,我国农业类古代科技文献数据库主要为中国农业科学院农业信息研究所构建的“农业古籍数据库”^[19]、中国农业大学图书馆构建的“农书古籍图片数据库”^[20]和沈阳农业大学图书馆构建的“农业古籍书目数据库”^[21]。其中:“农业古籍数据库”分为古农书、方志、史书和类书4个子库,除收录兼具专业性与学术性的农业古籍外,还收录地理、医药等领域有关农业生产和农业史料等的古籍;“农书古籍图片数据库”对其馆藏高质量农业古籍进行数字化处理,通过保存古籍全文图片,保护珍贵农业古籍;“农业古籍书目数据库”以书目形式收录沈阳农业大学图书馆的馆藏农业古籍。但遗憾的是,以上数据库目前均处于访问受限的状态。

1.3 水利数据库

在水利领域,中国水利水电出版社系统梳理了我中国经典水利史料,并构建了“中国经典水利史料数据库”^[22],共收录近200篇水利历史文献、6 000余幅水利历史图片(包括中国水利古地图2 000余幅、中国古代河渠图2 000余幅等)、1 000余条谕旨奏章、300余份影印资料。该数据库目前主要包括水利历史文献、水利历史图片、谕旨奏章3个版块。其中,水利历史文献版块将水利重要古籍按照涉及流域分为黄河、长江、淮河、海河、松辽、珠江、西部、运河、太湖及东南、沿海、综合和其他共计12个部分,每部分均包含相关经点校整理、繁体排录、全文数字化的水利古籍文献。水利历史图片版块支持按朝代、图片大类、专题类别、版本特

表1 常见地理类古代科技文献数据库

建设单位	数据库名称	收录范围	收录数据量	数据库功能与特色	数据类型				访问方式
					书目型	影像型	全文型	知识型	
北京爱如生数字化技术研究中心	爱如生中国方志库 ^[4]	历代地方志类著作(含全国地理总志、各地方志、各类专志、各种杂志以及外志)	影像1 000万页、录文20亿字,压缩后数据量为350 GB	可全文检索,可实现校勘、标点、编辑、复制等一站式整理研究作业	√	√	√		付费购买
北京籍古轩图书技术有限公司	中国数字方志库 ^[5]	1949年以前的旧志书(含地理学通论、总志、专类地志、游记等)	旧志书10 000余部,近10万册,近1 000万页,总容量达2 TB	含稿抄本、刻本、石印本等版本,可检索全文、特征、目录、类别等	√	√			付费购买
万方数据股份有限公司、中国科学技术信息研究所	万方方志中国数据库 ^[6]	含新方志与旧方志,以1949年为界	新方志共计40 000余册,旧方志近80 000卷	可检索书名、条目、全文、编纂单位等,可根据地图查找方志,可下载		√		√	付费购买
北京书同文数字化技术有限公司	明清两朝边塞海疆地理文献汇编全文检索系统 ^[7]	含明清时周边地理著作,涉及边塞、海疆文献;续编含明代海防、清代海防、清代东北边防、清代新疆边防、清代西藏边防、清代西南边防文献		可全文检索,可下载,复制文本时可带文本来源信息,可实现字频、关联汉字、字典、成语典故、元数据和中草药词典查询,可对文本勘误、标注	√	√	√		付费购买
上海睿则恩信息技术有限公司	古地图数据库 ^[8]	世界范围内收藏的中国古地图,分为天文、坤舆、疆域、山川名胜、城市、建筑、水利河工等类型	古地图5万余张	可高级检索,含版本信息,图片清晰度较高	√	√			付费购买
中国国家图书馆	数字方志 ^[9]	国家图书馆藏地方志文献	清代(含清代)以前的方志6 529部		√	√	√		用户注册
泰州图书馆	泰州历史文献特色库(历代志书) ^[10]	以馆藏的纸质地方文献为主		提供全文图片	√	√			用户注册
萍乡市图书馆	萍乡古籍方志资源库 ^[11]	馆藏方志资源	共计45部	提供版本信息和全文图片	√	√			免费访问
湖北省图书馆	湖北方志库 ^[12]	收录1949年以前修纂的湖北地区各版本地方志	共计166部	可评论收藏,提供版本信息,对部分文献提供全文图片	√	√			用户注册
南京图书馆	南京图书馆藏稀见方志全文影像数据库 ^[13]	南京图书馆馆藏方志	共20部,28册	提供全文图片	√	√			免费访问
华东师范大学图书馆	华东师范大学方志库 ^[14]								IP限制
华东师范大学调查与数据中心	数字方志集成平台 ^[15]	全国师范院校图书馆联盟数字方志资源		可对方志进行统一管理、大数据分析、可视化展示和智慧型服务	√	√	√	√	免费访问
苏州大学图书馆	苏州方志数据库 ^[16]	以苏州大学图书馆古籍馆藏为主	共计209部	可实现简繁体转换与全文检索	√	√	√		免费访问
大连图书馆	馆藏稀见方志全文数据库*								IP限制
天津图书馆	天津地方志丛书 ^[17]								IP限制
东北师范大学图书馆	东北地方志数据库*								IP限制
哈尔滨师范大学图书馆	馆藏古籍方志数据库*								IP限制
吉林大学图书馆	东北地方志目次库*								IP限制
	东北地方志全文库*								
山西省图书馆	馆藏地方志书目提要*								IP限制
	山西地方志*								

注:*表示笔者无法访问且未找到详细功能说明的数据库,该类数据库不纳入后续数据类型统计。

点、绘制技法、流域分类等组配检索,并提供经深度阐述、修复的历史图片。谕旨奏章版块提供从文献中提取的相关谕旨奏章,按照流域分类以时间轴形式体现历代兴水利、除水害的重大举措。中国水利水电出版社还构建了“中国黄河水利古地图数字资源库”^[23],该库包含黄河水利古地图1 298幅、《行水金鉴》(黄河部分)2册、黄河流域灌溉工程遗产视频资料4份,可实现资源在线检索、内容定位、文本复制等功能。上述两个数据库均需付费购买才可完整使用。

此外,河海大学图书馆在收集整理水利古籍资源的基础上,构建了“水利古籍数字文献数据库”^[24],包括水利史、地方志、工程项目报告等古籍文献,总计2 216册、18万页左右。但由于版权保护等,仅在该校古籍阅览室局域网内提供服务,并未开通其他访问渠道。青岛市图书馆构建了“涉海文献信息库——古代时期”数据库^[25],但该数据库同样仅提供馆内局域网访问渠道。

1.4 医药数据库

医药类文献是我国古代科技文献的重要组成部分,其利用与编纂也备受各界人士关注。国务院屡次发布文件,提出要“做好中医药继承工作”,“建立综合信息数据库和珍贵古籍名录,加强整理、出版、研究和利用”,“推动中医古籍数字化”^[26-27]。2019年10月《中共中央 国务院关于促进中医药传承创新发展的意见》指出要“加强典籍研究利用,编撰中华医藏,制定中医药典籍、技术和方药名录,建立国家中医药古籍和传统知识数字图书馆”^[28],从而促进中医药传承与开放创新发展。同时,在新型冠状病毒感染的防治中,中医药也发挥了重要作用。因此,医药类文献数据库建设更具现实意义。

相较于其他类型的古代科技文献数据库,从建设规模、收录文献数量与文献挖掘程度来看,医药类文献数据库目前取得了更大的成就。较为常见的医药类古代科技文献数据库如表2所示。

2 古代科技文献专题数据库建设情况分析

为厘清古代科技文献专题数据库的建设情况,从主体构成、数据类型和系统功能3个方面进行统计

分析。

2.1 主体构成

调查结果显示,共有55个国内单位或个人参与古代科技文献专题数据库的构建,构建主体可分为图书馆、学术机构、数字出版企业和个人4类,如表3所示。由于存在多主体联合构建数据库的情况,各类主体构建成果数之和并不等于总成果数。

由表3不难发现,在古代科技文献数据库构建过程中,图书馆和数字出版企业数量和拥有成果数均位居前列。其中:地市图书馆以本地方志类古籍为基础,构建了较多方志类数据库,而以农业类大学和医药类大学为代表的行业性大学图书馆则纷纷构建了相关行业的古代科技文献数据库;数字出版企业建设数据库的质量较高,但由于其建设重点为通用型数据库,其在古代科技文献专题数据库的建设数量和范围上仍有提升空间;学术机构数量与成果数均较少。此外值得注意的是,以个人为主体的相关专题数据库建设集中在医药数据库领域,相关成果虽质量有待提升,但表现出广大人民群众对于医药类古代科技文献的关注与迫切利用需求。

除无法获知具体地理位置的个人(共4位)外,参与古代科技文献专题数据库构建的51个机构分布情况如表4所示。我国古代科技文献专题数据库的构建主体分布极为不均衡,大多数机构和成果集中在我国经济较为发达的地区,如北京、江苏、上海等地。这可能是由于这些地区的高等院校、学术机构和数字出版企业众多,古代科技文献数字化的基础良好,具有数据库构建的基础环境。

2.2 数据类型

部分专题数据库需在内网环境下或需购买才可访问。因此,为排除主观因素的影响,仅统计可访问或有明确介绍的42个专题数据库的数据类型。根据数据类型(书目、影像、全文、知识),现有古代科技文献专题数据库可分为以下8种:书目型数据库;影像型数据库;书目与影像型数据库;书目与全文结合型数据库;书目、影像、全文综合型数据库;影像与知识结合型数据库;书目、全文、知识型数据库;书目、影像、全文、知识综合型数据库。具体情况如表5所示。

表2 常见的医药类古代科技文献数据库

建设单位	数据库名称	收录范围	收录数据量	数据库功能与特色	数据类型				访问方式
					书目型	影像型	全文型	知识型	
北京爱如生数字化技术研究中心	医书集成 ^[29]	包含历代医经、药典、脉法、医案	共计500余部	提供多种检索途径,可实现古籍原文对照、简繁体转换、批注收藏	√	√	√		付费购买
	中医典海 ^[30]	包含医经、本草、诊法、方书、针灸、临症各科、养生及医案、医话、医论等内容	共计1 000余部	可图文对照、全文检索、编辑、下载、打印	√	√	√		付费购买
上海微谱信息技术有限公司	医灯续焰中医古籍医案数据库 ^[31]	包含医案、医话医论、方剂、本草、古籍、医家、温病、肺炎、专业术语、中药成分等10个子数据库	中国古籍医案记录12万余个、中医方剂30万余个、中草药8 000余种	提供多关键字跨平台检索、药方分析、相似药方比较、方剂演变分析等功能	√		√	√	付费购买
北京书同文数字化技术有限公司	中医中药古籍大系数据库 ^[32]	历代经典中医中药著作及《日本医书》	历代经典中医中药著作104部及《日本医书》100部	可图文对照、高级检索、在线打印,字形保真					付费购买
	中医中药古籍大系续编	明清两朝官修和名医所撰的经典医书的注释、衍义、新编、辨证、直解、精要、合编、补遗等,大量收录国内稀有中医孤本	共计126部		√	√	√		
	中医中药古籍大系三编	清代已经或即将失传的方剂医著	共计289部						
	中医中药古籍大系四编	选自曹炳章主编的《中国医学大成》							
南京海维可信数据服务有限公司	千年医典——域外中医古籍丛书 ^[33]	来源于《域外中医古籍丛书》	共计2 376部,600 000余页	支持全文阅读、章节搜索、在线批注、社区交流、高级检索	√	√			付费购买
北京神黄科技股份有限公司	中医智库——中医药大数据分析平台 ^[34]	中医药知识库,包含古籍、医案、方剂、中医百科	共计1 800余部经典中医古籍的点校版、4万例病案、30万个方剂,涵盖9 000种证候、20 000种症状、40万条中医药相关百科信息	支持书名、作者、全文检索,可构建个人图书馆	√		√	√	付费购买
深圳市科信源网络空间科技有限公司、海南电子音像出版社	中医古籍类书全文库 ^[35]	由中国古代医学类书整理而成,含医学理论、临床病症、治疗经验、药物本草等内容		可全文检索、组配检索和模糊检索	√	√	√		付费购买
西安鼎浩科技有限公司	中医经典古籍辞书库 ^[36]	含方药本草、医理医案、诸科论治、针灸推拿、养生导引等古籍		可检索、下载	√		√		付费购买
中国医药科技出版社、北京神黄科技股份有限公司、药师在线	中医古籍全文数据库 ^[37]			提供提要与校注说明	√	√	√		免费访问
中医资源网	四库全书子部医家类 ^[38]	提供《黄帝内经·素问》《普济方》《证治准绳》《本草纲目》《景岳全书》《御纂医宗金鉴》等文献	共计97部	可实现简繁体转换,可下载	√		√		付费购买
	中医古籍 ^[39]	提供中药、方剂等古籍	共计1 102部	可实现简繁体转换,可下载	√		√		付费购买
中国中医药学会、湖南电子音像出版社、宏宇科技开发有限公司	中华医典光盘版	含医经、诊法、本草、方书、针灸推拿、伤寒金匮等历代中医古籍	共计1 156部,4亿多字	可实现简繁体转换,内设中医辞典,可下载	√	√	√		付费购买
山东搜搜中医信息科技有限公司、中国中医科学院、山东中医药大学	博览医书——中西医百科数据平台 ^[40]	含基础古籍库、专科古籍库、珍稀古籍库、域外古籍库、病种医学库、历代名医库等		提供古籍简体版全文和多个已出版版本	√	√	√		付费购买

续表

建设单位	数据库名称	收录范围	收录数据量	数据库功能与特色	数据类型				访问方式
					书目型	影像型	全文型	知识型	
中国中医科学院 中医药信息研究所(图书馆)	国医典藏中医古籍数据库 ^[41]	先秦至清末民国的 历代中医典籍	共计500部(包括综合性丛书20部), 2 500册	提供古籍的版本和 提要信息, 可下载		√			付费 购买
陕西师范大学 出版总社	中医药文献数字图书馆 ^[42]	收录中医药及其相关文 献, 包括医经、基础理 论、伤寒金匮等79类	截至2017年9月, 中 医药文献目录库收 录文献约2.2万篇; 中医药文献图版库 收录文献原件4 914 篇, 约470万页	提供古籍文献的版 本信息, 可下载	√	√			付费 购买
中国国家图书馆	中华医药典 籍资源库(测 试版) ^[43]	馆藏中医文献	共计221部	可根据标题、责任者、 善本书号等检索	√	√	√		用户 注册
古联(北京)数字 传媒科技有限公司	海外中医 古籍库 ^[44]	《海外中医珍善 本古籍丛刊》	古籍善本427部, 影印217 044页	可提供全文检索、 版本对比、印鉴检 索、浏览检索记录 查看、批注等功能	√		√		付费 购买
中国中医药出 版社有限公司	中医数字图 书馆——中 医古籍 ^[45]			支持图书检索、全文 检索、高级检索、书 内检索等检索方式	√		√		付费 购买
浙江省科技信息研 究院、浙江省科技 发展战略研究院	中医古籍文 献数据库 ^[46]								IP限制
江西中医药大学 图书馆	中医药古籍 书目检索 ^[47]				√				IP限制
台北故宫博物院	中华古籍医 学菁华电子书 资源平台 ^[48]			提供打印、批注、 分享等功能	√	√			IP限制
南京中医药大学 图书馆	馆藏古籍全 文数据库古 籍数据库 ^[49]	主要收录馆藏中医 古籍和综合性汇编 丛书中的中医古籍			√		√		IP限制
	气功古籍 提要库 ^[50]	基本囊括了不同时期 的气功经典著作	共计758条数据		√				IP限制
梦境中华站	医道心源—— 中医古籍 ^[51]	部分经典中医古籍		可根据朝代、内容 查看中医古籍	√		√		免费 访问
医药网	中医古籍 ^[52]	部分经典中医古籍	共计33部	可全文检索	√		√		免费 访问
中华中医网	中医古籍在 线阅读 ^[53]	部分经典中医书籍	共计21部	可查看简体全文	√		√		免费 访问

注: *表示笔者无法访问且未找到详细功能说明的数据库, 该类数据库不纳入后续数据类型统计。

表3 古代科技文献专题数据库的构建主体

类 型	构建主 体数/个	构建主体 占比/%	成果 数/个	成果 占比/%	平均成 果数/个
图书馆	20	36.36	24	40.68	1.20
学术机构	9	16.36	7	11.86	0.78
数字出版企业	22	40.00	25	42.37	1.14
个人	4	7.27	5	8.47	1.25
总计	55	100.00	59	100.00	1.07

由表5可以发现, 当前最为普遍的古科技文献专题数据库是由数字出版企业构建的书目、影像、全文综合型数据库。仅包含书目或影像的数据库较少, 且集中于图书馆领域, 这可能是图书馆在古代科技文献专题数据库建设过程中资金、人力不足所致。此外, 古代科技文献专题数据库从少数机构的零星构建拓展为多机构的规模化开发, 其所含古代科技文献资源也从书目形式转化为书目、全文、影像等数据类型, 并逐渐演化出

表4 古代科技文献专题数据库构建主体地区分布

地区	北京	江苏	上海	山东	台湾	吉林	陕西	江西	辽宁
机构数/个	17	5	4	4	3	2	2	2	2
成果数/个	23	6	4	3	3	3	2	2	2
地区	湖南	浙江	山西	海南	深圳	黑龙江	湖北	天津	
机构数/个	2	2	1	1	1	1	1	1	
成果数/个	1	1	2	1	1	1	1	1	

表5 古代科技文献专题数据库类型统计

构建主体	数据库数量/个								总计
	A	B	C	D	E	F	G	H	
图书馆	1	1	4	1	3				10
学术机构	1		1			1		1	4
数字出版企业			4	3	14	1	2		24
个人				5					5
合计	2	1	9	9	17	2	2	1	43

注：A为书目型数据库；B为影像型数据库；C为书目与影像型数据库；D为书目与全文结合型数据库；E为书目、影像、全文综合型数据库；F为影像与知识结合型数据库；G为书目、全文、知识型数据库；H为书目、影像、全文、知识综合型数据库。

具备更高信息密度的知识型数据库。

2.3 系统功能

在系统功能方面，上述地理、农业、水利和医药类古代科技文献专题数据库已基本具备书目和全文等内容的查阅和检索功能，能够满足读者按朝代、类别、书名、作者浏览古籍的需求，并可实现书名、作者、朝代、全文等字段的单一检索或组配检索。部分提供全文的数据库还可实现古籍全文的简体与繁体转换，便于读者查阅和浏览。

少数数据库还提供了版本备注、印鉴检索、提要查看、校注说明、用户批注等功能。例如：“中国经典水利史料数据库”“医书集成”“国医典藏中医古籍数据库”“中医药文献数字图书馆”“海外中医古籍库”等数据库均提供古籍版本信息；“海外中医古籍库”还可检索藏书印鉴，并关联至相关古籍；“中医古籍全文数据库”提供古籍的校注说明，简要梳理了流传版本、校对过程与整理原则等内容；“海外中医古籍库”“医书集成”“国医典藏中医古籍数据库”等提供用户批注或浏览记录查看等功能。

值得一提的是，有的数据库不仅实现了目录浏览、

全文检索、图文对照等一般功能，还能对古籍进行知识加工。例如中国中医科学院主持开发的“博览医书——中西医百科数据平台”从350部常用中医古籍中抽取了病症、医案、方剂、本草4类知识体共计12.1万个，其中知识元40.7万个。该库支持中医知识的分类检索，当用户输入一个或一组中医概念时，系统会判断此概念的类型，并优先输出此类知识的检索结果，其他知识则按关联度排列在后。该库在基于知识元的中医古籍计算机知识表述方法的理论指导下，对古籍深度加工标引。此外，华东师范大学调查与数据中心构建的“数字方志集成平台”不仅提供全国方志的集成导览，还提供方志数据命名实体识别和知识图谱等可视化展示功能，可便捷实现对方志等古籍文本的数据挖掘与可视化。

3 我国古代科技文献专题数据库建设存在的问题

尽管古代科技文献专题数据库建设已初具成效，但从规范程度、访问权限、系统功能和涵盖学科等方面来看仍有较大的提升空间。

(1) 缺乏规划。由于建设过程缺乏统一规划，不同主体构建的古代科技文献专题数据库在底本选取、元数据制作、文献著录、文献标引等方面存在差异^[54]，导致相关数据库质量参差不齐，甚至出现文字转录错误、文献结构被破坏等现象。例如，数字出版企业由于具有营利目的，其构建的数据库涵盖范围较广，功能升级较快，且文献著录与标引较为规范。而个人缺少人力、物力或前期投入，所构建的数据库大多难以描述古籍版本或全文来源。

(2) 访问受限。部分古代科技文献专题数据库仅限特定局域网或特定用户访问，并不提供付费或其他访问渠道。例如，北京中医药大学图书馆、南京中医药大学图书馆、大连图书馆、山西省图书馆、天津图书馆等构建的古代科技文献专题数据库仅能通过局域网访

问,并不提供外部访问渠道,也未提供文献扫描或传递服务。这类访问限制给古代科技文献的传播和利用带来阻碍,提高了古代科技文献的获取成本。

(3) 介绍不足。古代科技文献专题数据库的系统介绍和资源目录对揭示数据库内容尤为重要。但遗憾的是,部分数据库并未提供可供公开访问的数据库介绍,不利于用户了解该数据库基本功能与涵盖范围,也为古代科技文献整合与利用带来困难。当用户想要查找特定古代科技文献时,到馆现场查找似乎成为唯一方式。但资源目录的外部访问限制又导致用户无法准确定位所需古代科技文献具体馆藏。此外,古代科技文献的介绍也存在更新不及时的情况,导致用户无法快速了解数据库最新的功能与资源情况。

(4) 学科失衡。古代科技文献涵盖范围甚广,除已经构建专题数据库的地理、农业、水利和医药领域外,还有天文、物理、数学、生物、冶金、纺织等领域的古代科技文献。然而当前既缺乏涵盖所有学科类别的古代科技文献数据库,又缺乏除上述专题数据库外的其他学科古代科技文献数据库。当读者存在文献查阅需求时,只能通过访问通用型古代文献数据库获取文献。但由于构建目的、用户群体等与专题数据库存在差异,通用型数据库无法保证涵盖所有的古代科技文献,这为用户查找与利用古代科技文献带来了不便。由于缺乏顶层调控,古代科技文献专题数据库建设常出现重复开发的现象。农学、医药等受到广泛关注的学科领域中古代科技文献数据库选题重复,而物理、化学、生物、工程技术等学科领域则普遍尚未构建相关数据库。

(5) 功能单一。调研发现,多数古代科技文献专题数据库的系统功能较为单一。由于古代科技文献搜求不易及数据库构建者版本意识有待加强,大多数专题数据库除提供古代科技文献目录外,仅单独提供影像或全文,未提供图文对照或简繁体转换等功能,也未对古代科技文献版本或提要信息作说明。然而古代科技文献流转过程中常出现误字、误注、误释等现象,其影像与版本信息的缺失会极大影响古代科技文献全文的可靠性和可信度,不利于古代科技文献的传播利用。数据库中古代科技文献影像与全文缺一不可,二者可互相对照,以便读者准确理解古籍文献的真正内涵。同时,现有数据库也较少利用序跋、印鉴、插图等副文本资源,尤其是缺乏对插图的检索或专题展示。此外,古代科技文献专题数据库对文献数字资源的利用率也明显不足,大多数数据库未对学科内容进行深入挖掘与

可视化,也无法实现不同文献间的知识关联。

(6) 互动薄弱。当前,仅有部分古代科技文献专题数据库提供评论、反馈等互动模块。用户在使用数据库过程中具有与其他用户交流且向数据库开发主体提供反馈的需求。一方面,评论功能的缺失使用户无法充分表达个人对数据库资源的评价,使其无法利用平台找寻志同道合之人;另一方面,用户反馈是开发主体感知用户体验最为直接的渠道,但用户难以通过在线反馈的方式联系开发主体反馈平台疏漏之处,这不利于数据库功能的改进提升。

4 我国古代科技文献专题数据库建设的改进策略

古代科技文献专题数据库应充分提取隐藏在不同类别下的古代科技文献,提升其可用性与易读性,从而有效解决缺乏规划、访问受限、介绍不足、学科失衡、功能单一和互动薄弱的问题。

(1) 统一顶层设计。古代科技文献专题数据库应统筹规划、统一设计,在详细调研我国现存古代科技文献的基础上,从学科角度整合已有古代科技文献专题数据库,避免资源浪费与重复工作。使用统一的分类法划分古代科技文献,构建统一的机读目录格式和著录规则,并根据古代科技文献特征,设计数据库实体及其属性,实现文献、作者、用户和管理员4类实体的关联。数据库的实体关系如图1所示。

(2) 开放获取渠道。限制资源获取渠道不利于用户利用古代科技文献。各建设单位应充分考虑用户需求:对于营利性数据库,应开放个人用户付费渠道,以单本价格或单页价格设计个人用户付费标准;对于非营利性数据库,应根据现实情况,考虑以合适的方式提供文献传递或文献扫描服务,使用户不必入馆查阅或是入馆使用内网便可获取所需古代科技文献。

(3) 及时更新介绍。古代科技文献专题数据库构建所需考虑的问题之一是用户能否真正利用。由于当前不少古代科技文献专题数据库仍在不断建设中,其收录范围与特色功能也在不断更新。因此,建设单位应及时更新数据库介绍,尤其是及时完善对其特色收录对象与特色功能的介绍,以图文结合的形式为用户提供数据库收录的资源目录,使用户充分了解数据库的收录范围与系统功能。

(4) 保障学科齐全。古代科技文献专题数据库应

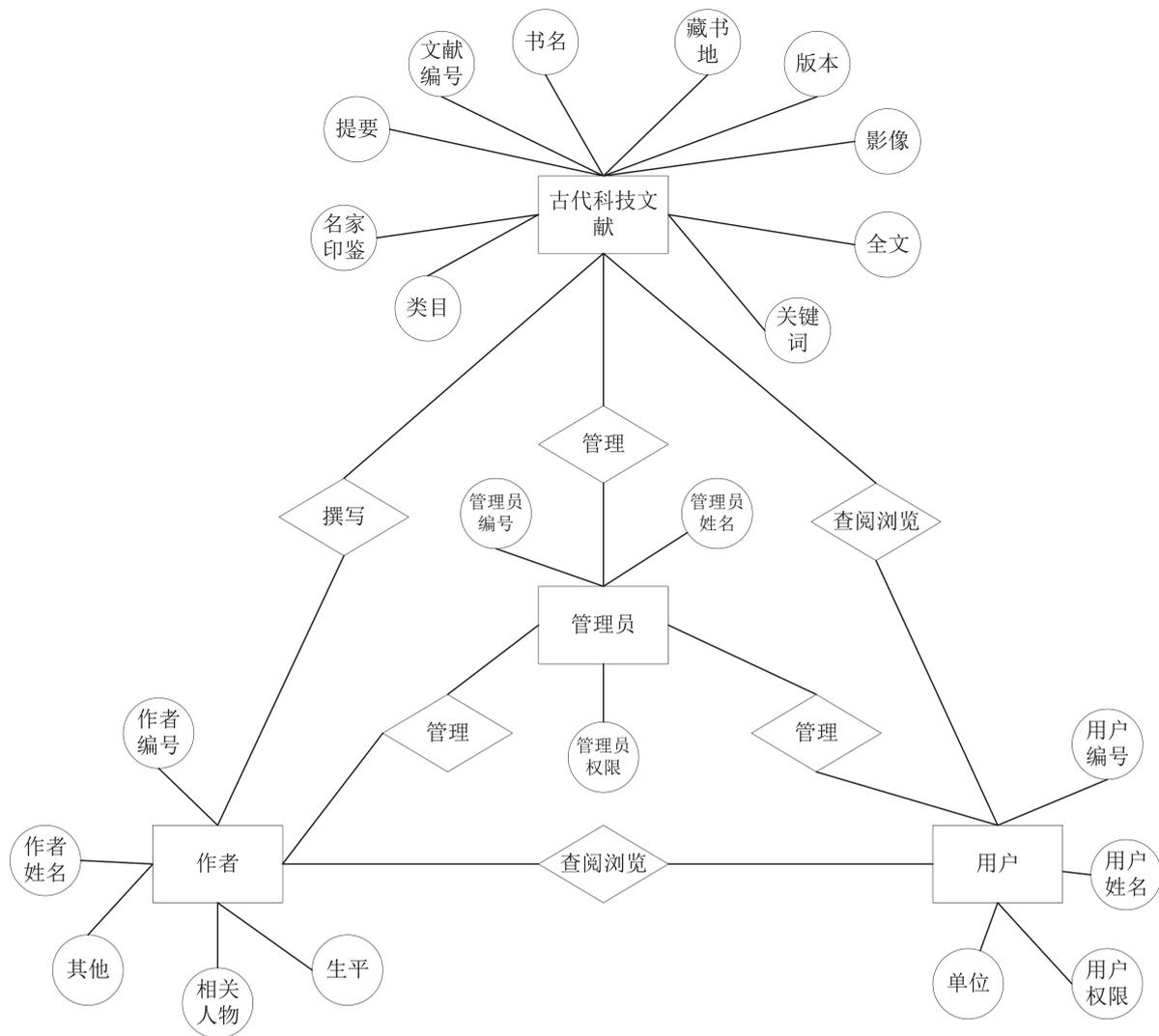


图1 古代科技文献专题数据库实体关系

涵盖所有学科类别（如天文、气象、地理、数学、物理、化学、生物、农业、水利、交通、纺织、医药、机械、建筑、勘探、冶金、造船、航海、造纸、印刷、陶瓷等），充分搜集并整理各学科类别中的现存古籍，由具备古汉语、文献学、图书馆学、计算机技术等知识的专业人才校对古籍原文，从而减少识别古籍文字过程中的数据失真问题，以充分反映我国古代科技发展状况，方便我国各领域古代科技史研究。此外，应根据不同学科开设不同版块，不再使用传统的四部分类法，而是充分考虑古代科技文献学科特征，根据现实情况设置类目，便于用户分门别类地检索、查阅古籍文献。

(5) 探索多样功能。古代科技文献专题数据库应具备完备的功能（如检索、收藏、批注、下载、打印等基本功能和字典查询、纪年换算等辅助性功能），并支持

电脑、手机等多设备使用。在系统功能方面，古代科技文献专题数据库应设置简单检索、高级检索、二次检索等检索方式，实现按书名、作者、提要、全文、藏书印鉴、版本、藏书地、插图等多字段检索，并增加可选择的语义联想功能，实现检索内容的智能联想与模糊匹配，以满足部分用户对极高检全率的需求；设立用户中心，保存用户浏览记录，提供用户收藏、批注等个性化功能，使用户在使用过程中能够根据个人阅读习惯选择阅读背景、字体、版本等；提供文献下载和打印功能，便于用户利用相关文献；提供字典查询和纪年换算等辅助性功能，使用户不必跳出数据库便可获知所需要的辅助信息，帮助用户更好地理解古籍文献内容。古代科技文献专题数据库也应重视文献副本作用，提取文献影像、文中插图（如疆域政区图、天文图、城市图、

行宫图、河渠图、军事图、航海图、自然灾害图、机械图、植物图、动物图、纺织图等)、名家藏书印鉴等图片,设置相应专题或检索渠道,并将其与简繁体两种形式的全文相结合,建立图片与文字的关联,丰富数据库的知识内涵,满足用户利用古代科技文献图像的需求。

(6) 提供互动环境。数据库互动环境直接影响用户的使用体验。由于人力、物力和开发者专业知识有限,数据库建设难免会存在疏漏之处。因此,需充分设计开发者与用户交流及用户间彼此交流的版块:一方面,加强开发者与用户的互动,以众包、评论等形式提升数据库建设效果;另一方面,通过增强用户间交流,促进学科知识融合,增强用户黏性,提升用户归属感。

参考文献

- [1] 张秀红. 中国古代科技文献概况[J]. 图书与情报, 2005 (5): 25-29.
- [2] 国务院办公厅. 国务院办公厅关于进一步加强古籍保护工作的意见[EB/OL]. [2023-08-26]. https://www.gov.cn/xxgk/pub/govpublic/mrlm/200803/t20080328_32601.html.
- [3] 共产党员网. 中华人民共和国国民经济和社会发展第十三个五年规划纲要[EB/OL]. [2023-08-26]. <http://www.12371.cn/special/sswgh/wen/>.
- [4] 北京爱如生数字化技术研究中心. 爱如生中国方志库[EB/OL]. [2023-08-26]. http://www.er07.com/home/pro_87.html.
- [5] 北京籍古轩图书技术有限公司. 中国数字方志库[EB/OL]. [2023-08-26]. <http://x.wenjinguan.com/>.
- [6] 万方数据股份有限公司, 中国科学技术信息研究所. 万方方志中国数据库[EB/OL]. [2023-08-26]. <http://fz.wanfangdata.com.cn/index.do>.
- [7] 北京书同文数字化技术有限公司. 明清两朝边塞海疆地理文献汇编全文检索系统[EB/OL]. [2023-08-26]. <http://stw.lib.whu.edu.cn/Web#/book/MQBHF>.
- [8] 上海睿则恩信息技术有限公司. 古地图数据库[EB/OL]. [2023-08-26]. <http://gdt.lib.whu.edu.cn/>.
- [9] 中国国家图书馆. 数字方志[EB/OL]. [2023-08-26]. <http://read.nlc.cn/allSearch/searchList?searchType=12&showType=1&pageNo=1>.
- [10] 泰州图书馆. 泰州历史文献特色库(历代志书)[EB/OL]. [2023-08-26]. <http://www.tzlib.com/Sub.html#!Function/Resource/Type/Detail/Column/f9855019-bb66-4133-9b4e-f5cf316ab986/Item/3b2212ff-8809-4aa9-a0a8-33b8a0377cf2>.
- [11] 萍乡市图书馆. 萍乡古籍方志资源库[EB/OL]. [2023-08-26]. <http://rrt.pxlib.cn:8122/LocalRecords/List>.
- [12] 湖北省图书馆. 湖北方志库[EB/OL]. [2023-08-26]. <http://gjpt.library.hb.cn:8991/index.html?ticket=aOiiGx8EO6hIXq6stmqjpA89x1z3XYoleEAnUfAxcfoff9XjW23dLWzCh7dFLiskkbQyoUu9NV0IQMbYv4WfYR80XQisAOuR2KTnFLmdpXNvVpW8ie>.
- [13] 南京图书馆. 南京图书馆藏稀见方志全文影像数据库[EB/OL]. [2023-08-26]. <http://www2.jslib.org.cn/was5/web/hdb.htm>.
- [14] 华东师范大学图书馆. 华东师范大学方志库[EB/OL]. [2023-08-26]. http://202.120.82.36/local_chronicle_books/.
- [15] 华东师范大学调查与数据中心. 数字方志集成平台[EB/OL]. [2023-08-26]. <http://fangzhi.ecnu.edu.cn/>.
- [16] 苏州大学图书馆. 苏州方志数据库[EB/OL]. [2023-08-26]. <http://library.suda.edu.cn/cb/7e/c4807a52094/page.htm>.
- [17] 天津图书馆. 天津地方史志丛书[EB/OL]. [2023-08-26]. <http://hhjyy.tjlib.cn/shwh/dfz/dfz1.htm>.
- [18] 范桂红, 赵纯洋. 基于知识图谱的古籍数字化研究前沿热点及演化趋势分析[J]. 出版广角, 2020 (11): 85-87.
- [19] 卢文林. 农业古籍数据库建立和著录实践[J]. 农业图书情报学刊, 2011, 23 (11): 23-26.
- [20] 塔娜. 古籍纸本资源数字化加工处理方法探索[C]//图书情报工作研究会第21次学术研讨会论文集. 北京:《图书情报工作》杂志社, 2009: 254-256.
- [21] 刘凤侠. 高校图书馆农业古籍书目数据库建设实证研究: 以沈阳农业大学图书馆为例[J]. 图书馆学刊, 2018, 40 (3): 42-45.
- [22] 中国水利水电出版社. 中国经典水利史料数据库[EB/OL]. [2023-08-26]. <https://sisl.digiwater.cn/>.
- [23] 中国水利水电出版社. 中国黄河水利古地图数字资源库[EB/OL]. [2023-10-26]. <http://hk.digiwater.cn>.
- [24] 河海大学图书馆. 水利古籍数字文献数据库[EB/OL]. [2023-08-26]. <http://lib.hhu.edu.cn/databaseshow/903.html>.
- [25] 青岛市图书馆. 涉海文献信息库: 古代时期[EB/OL]. [2023-08-26]. <http://192.168.16.139/>.
- [26] 中华人民共和国中央人民政府. 国务院关于扶持和促进中医

- 药事业发展的若干意见(国发(2009)22号)[EB/OL]. [2023-05-12]. http://www.gov.cn/zhengce/content/2009-05/07/content_6230.htm.
- [27] 国务院关于印发中医药发展战略规划纲要(2016—2030年)的通知(国发(2016)15号)[EB/OL]. [2023-05-12]. http://www.gov.cn/zhengce/content/2009-05/07/content_6230.htm.
- [28] 中华人民共和国中央人民政府. 中共中央 国务院关于促进中医药传承创新发展的意见[EB/OL]. [2023-05-12]. http://www.gov.cn/zhengce/2019-10/26/content_5445336.htm.
- [29] 北京爱如生数字化技术研究中心. 医书集成[EB/OL]. [2023-05-22]. <http://lib.jxutcm.edu.cn/yishu.rar>.
- [30] 北京爱如生数字化技术研究中心. 中医典海[EB/OL]. [2023-08-26]. <http://gj.ersjk.com/>.
- [31] 上海微谱信息技术有限公司. 医灯续焰中医古籍医案数据库[EB/OL]. [2023-08-26]. <http://db.yidxy.com/index>.
- [32] 北京书同文数字化技术有限公司. 中医中药古籍大系数据库[EB/OL]. [2023-08-26]. <http://guji.unihan.com.cn/web#/book/ZYYS>.
- [33] 南京海维可信数据服务有限公司. 千年医典: 域外中医古籍丛书库[EB/OL]. [2023-08-26]. <https://www.kanguji.com/>.
- [34] 北京神黄科技股份有限公司. 中医智库: 中医药大数据分析平台[EB/OL]. [2023-05-22]. https://www.zk120.com/inst/?inst_key=fimmu.
- [35] 深圳市科信源网络空间科技有限公司, 海南电子音像出版社. 中医古籍类书全文库[EB/OL]. [2023-08-26]. <http://newchinesebookyx1.szkxy.net/default.asp?inorout=out>.
- [36] 西安鼎浩科技有限公司. 中医经典古籍辞书库[EB/OL]. [2023-05-22]. <https://lib.sdutcm.edu.cn/bencandy.php?fid=44&id=288>.
- [37] 中国医药科技出版社, 北京神黄科技股份有限公司, 药师在线. 中医古籍全文数据库[EB/OL]. [2023-08-26]. <http://zygj.cmstp.com/static/reader/index.html>.
- [38] 中医资源网. 四库全书子部医家类[EB/OL]. [2023-08-26]. http://www.tcmdoc.cn/book_SKQSZBYJL/bookList.aspx.
- [39] 中医资源网. 中医古籍[EB/OL]. [2023-08-26]. <http://www.tcmdoc.cn/book/bookList.aspx>.
- [40] 博览医书: 中西医百科数据平台[EB/OL]. [2023-08-26]. <http://www.imedbooks.com>.
- [41] 中国中医科学院中医药信息研究所(图书馆). 国医典藏中医古籍数据库[EB/OL]. [2023-05-22]. <http://www.gydc.ac.cn:81/>.
- [42] 陕西师范大学出版总社. 中医药文献数字图书馆[EB/OL]. [2023-08-26]. <http://zhongyiyao.hanjilibrary.com/>.
- [43] 中国国家图书馆. 中华医药典籍资源库(测试版)[EB/OL]. [2023-08-26]. <http://read.nlc.cn/allSearch/searchList?searchType=81&showType=1&pageNo=1>.
- [44] 古联(北京)数字传媒科技有限公司. 海外中医古籍库[EB/OL]. [2023-08-26]. <http://books.ancientbooks.cn/banke/resource/1/index.htm>.
- [45] 中国中医药出版社有限公司. 中医数字图书馆: 中医古籍[EB/OL]. [2023-08-26]. <http://libtcm.cptcm.com/docZhongyi/pc/index.jspx>.
- [46] 浙江省科技信息研究院, 浙江省科技发展战略研究院. 中医古籍文献数据库[EB/OL]. [2023-05-22]. <http://www.istiz.org.cn/Portal/InfoResource/Detail.aspx?id=9765>.
- [47] 江西中医药大学图书馆. 中医药古籍书目检索[EB/OL]. [2023-05-22]. <http://lib.jxutcm.edu.cn/>.
- [48] 台北故宫博物院. 中华古籍医学菁华电子书资源平台[EB/OL]. [2023-05-22]. <http://library.njuem.edu.cn/do/bencandy.php?fid=47&id=2088>.
- [49] 南京中医药大学图书馆. 馆藏古籍全文数据库古籍数据库[EB/OL]. [2023-05-22]. <http://library.njuem.edu.cn/do/bencandy.php?fid=68&id=130>.
- [50] 南京中医药大学图书馆. 气功古籍提要库[EB/OL]. [2023-05-22]. <http://library.njuem.edu.cn/do/bencandy.php?fid=68&id=126>.
- [51] 梦境天华站. 医道心源: 中医古籍[EB/OL]. [2023-08-26]. <http://www.mst1739.com/yidao/yuedu.php>.
- [52] 医药网. 中医古籍[EB/OL]. [2023-05-22]. <http://zhongyi.shufaji.com/>.
- [53] 中华中医网. 中医古籍在线阅读[EB/OL]. [2023-05-22]. <https://www.zhzyw.com/zygjyd/>.
- [54] 颜艳萍. 数字人文背景下图书馆古籍数字化新实践: 以云南省图书馆古籍数据库建设为例[J]. 图书馆学刊, 2020, 42(7): 39-43.

作者简介

郑翔, 女, 博士研究生, 研究方向: 古籍整理与数字人文, E-mail: zhengxiang059@whu.edu.cn。

关思雨, 女, 博士研究生, 研究方向: 中国古代图书文化史。

耿廉鋈, 男, 硕士研究生, 研究方向: 计算机视觉。

Current Situation and Improvement Strategies of Construction of Ancient Chinese Scientific and Technological Document Thematic Databases

ZHENG Xiang¹ GUAN SiYu¹ GENG LianYun²

(1. School of Information Management, Wuhan University, Wuhan 430072, P. R. China; 2. School of Computer Science and Information Engineering, Hubei University, Wuhan 430062, P. R. China)

Abstract: This article conducts a detailed investigation on the current construction status of ancient Chinese scientific and technological document thematic databases through literature collection, online surveys, and other methods. We calculate the proportion of the construction entities, data types, and system functions of ancient Chinese scientific and technological document thematic databases. Based on the research results, we find that there are problems in the current construction of ancient Chinese scientific and technological document thematic databases, such as lack of standardization, restricted access, insufficient introduction, imbalance of disciplines, single function, and weak interaction. These databases should unify the top-level design, open up the access channels, update the introduction in time, guarantee complete disciplines, explore various functions of databases, and provide interactive environment.

Keywords: Ancient Scientific and Technological Document; Thematic Database; Digitization of Ancient Document

(责任编辑: 王玮)