



# 数字化转型背景下高校图书馆数据资源体系建设的思考与探索

——以复旦大学图书馆为例

侯力强 伏安娜 袁玉红\*

**摘要** 在人工智能驱动科研范式变革与国家数据战略深入实施的双重背景下,高校图书馆正经历从“文献资源收藏者”向“数据驱动创新支撑者”的历史性转型。文章梳理了高校图书馆数据资源建设的理论演进与实践路径,以复旦大学图书馆为典型案例,深入剖析其构建“文献资源保障、特色数据资源、科研数据支撑、智慧图书馆与数字基座”四位一体数据资源体系的创新实践。研究表明,通过顶层设计重构、组织架构调整、技术平台智能升级及治理制度创新,高校图书馆能够有效打破数据孤岛,实现资源从“数字化”向“数据化、知识化、智能化”的跃迁。面对人才、经费及安全伦理等挑战,未来应聚焦 AI-Ready 语料库建设、领域大模型训练及全生命周期数据治理,打造开放、融合、安全的学术资源新生态,为高等教育高质量发展提供坚实的数据基座与智慧引擎。

**关键词** 高校图书馆 数据资源体系 复旦大学图书馆

**分类号** G250.7

**DOI** 10.16603/j.issn1002-1027.2026.02.001

**引用本文格式** 侯力强,伏安娜,袁玉红.数字化转型背景下高校图书馆数据资源体系建设的思考与探索——以复旦大学图书馆为例[J].大学图书馆学报,2026,44(2):5-14.

## 1 引言

当前,人工智能正以前所未有的深度重塑当代科研活动的底层逻辑。在新的科研范式下,数据已不再仅仅是验证假设的辅助证据,而是跃升为驱动科学发现与技术创新的核心引擎。这一根本性转变深刻重构了图书馆作为学术信息资源中心的功能定位与战略使命——从传统文献资源的被动收藏者,转变为数据驱动科研创新的主动支撑者与深度参与者。

近年来,国家层面围绕科学数据管理与数据要素化建设出台了一系列政策文件,为高校图书馆探索构建数据资源体系提供了坚实的制度框架与清晰的行动指引。2026年3月发布的《中华人民共和国国民经济和社会发展第十五个五年规划纲要》明确指出,数据是推动生产方式深层次变革和生产革命性跃迁的关键要素,是“数智化”发展的核心底座之一<sup>[1]</sup>。早在2018年,国务院办公厅发布的《科学

数据管理办法》便确立了“开放为常态,不开放为例外”的共享原则,为科学数据的规范流通奠定了制度基石<sup>[2]</sup>。2019年,党的十九届四中全会首次将“数据”与劳动、资本、土地等传统生产要素并列,从顶层设计层面确立了数据在价值创造与分配中的基础地位<sup>[3]</sup>。2022年,“数据二十条”创新性地提出数据产权“三权分置”制度,初步构建起数据要素市场化配置的基础制度框架<sup>[4]</sup>。2023年,国家数据局等十七部门联合印发《“数据要素×”三年行动计划(2024—2026年)》,专设“数据要素×科技创新”板块,明确提出“建设高质量语料库和基础科学数据集”以支撑人工智能大模型开发,并“加速科学研究范式变革”<sup>[5]</sup>。2025年,国务院《关于深入实施“人工智能+”行动的意见》进一步强调“打造开放共享的高质量科学数据集”<sup>[6]</sup>;同年,国家数据局《关于加强数据科技创新的实施意见》则部署打造“Data for Science(数据驱动的科学)协同创新中心”,促进科学数

\* 通讯作者:袁玉红,邮箱:cfyyh@fudan.edu.cn。



据的高效流通与深度开发利用<sup>[7]</sup>。

上述规划与政策从基础规范、要素化制度到前沿应用层层递进,共同勾勒出我国数据资源体系建设的宏伟战略图景,同时也向作为高校信息资源建设核心阵地的图书馆提出了明确的时代命题:如何在数字化转型的浪潮中,主动融入国家数据战略,构建能够支撑新范式下科学研究的数据资源体系?

国家“十四五”规划明确提出“积极发展智慧图书馆”,进一步推动图书馆从传统文献服务向数据驱动的知识服务转型<sup>[8]</sup>。而在“十五五”规划纲要中,更是明确提出构建国家数据资源体系,健全数据资源统计调查制度,建立全国数据资源“一本账”,旨在解决“有什么”和“怎么供”的问题,要求从制度上加强数据资源建设<sup>[1]</sup>。本文在系统梳理国内外图书馆数据资源建设理论与实践的基础上,以复旦大学图书馆的探索实践为切入点,尝试构建高校图书馆数据资源建设的体系框架,以期为数字化转型背景下的高校图书馆高质量发展提供参考与借鉴。

## 2 概念界定和文献综述

### 2.1 概念界定

在数字化转型与数智融合的时代背景下,高校图书馆的资源形态正在经历从“文献本位”向“数据本位”的深刻演变。在图书馆学视域下,数字资源又称电子资源<sup>[9]</sup>,传统定义是指“以数字形式生产和发行的信息资源,其信息包括文字、图片、声音、动态图像等,并以硬盘、磁带、光盘等介质及网络形式展现”<sup>[10]</sup>。然而,在学术新常态下,数字资源已成为高校图书馆资源与服务的核心,其内涵正日益丰富——从早期的商业数据库,逐步拓展至数字特藏、数字学术与研究数据等自主建设或特色资源<sup>[11-12]</sup>。

随着数据驱动科研范式的深入推进,学者们近年来对“数据化”与“数据馆藏”等概念展开了深入探讨。曾蕾等指出,实现从资源数字化到数据化的转型,将推动数字馆藏得到更好的共享、关联、丰富和复用,是形成“智慧数据”的基础<sup>[13]</sup>。孟祥保等则从建设实践出发,指出数据资源建设内涵包括以用户需求为起点,遵循规范化的建设流程,以及建设来源主要包括采购的商业数据与机构自产的特色数据<sup>[14]</sup>。李书娟在探讨数智时代图书馆资源形态时,进一步区分了“数字馆藏”与“数据馆藏”的本质差异:数字馆藏是实体馆藏的“数字镜像”,其“粒度”仍

然是文献;而数据馆藏的“粒度”则细化为文献中的“知识元”及相关数据,是结构化、智能化、可计算的活态数据;馆藏数据化正是推动图书馆实现数据驱动智慧知识服务的基础<sup>[15]</sup>。

基于此,本文所指的“高校图书馆数据资源体系”,是指在数字化、网络化、智能化环境下,图书馆通过采集、加工、治理、存储与服务,形成的涵盖传统文献数据化成果、原生科研数据、特色文化数据及用户行为数据等多源异构数据的全生命周期管理生态系统。该体系不仅包含数据资源本身,还涵盖支撑数据流转的技术平台、标准规范、组织架构及服务机制。

### 2.2 文献综述

高校图书馆数据资源建设的实践重心,经历了从“文献资源数字化”到“科研数据资源管理”,再到“数据智能赋能”的深刻演进。这一过程既折射出图书馆对信息环境变革的适应性调整,更揭示了数据本身从研究的“附属产物”逐步确立为“核心学术资产”乃至“关键创新要素”的范式转换。

#### 2.2.1 数字图书馆阶段:文献资源的数字化与基础整合

早期建设中,数字图书馆的重点是将传统馆藏的文献资源进行数字化转换,并实现基础性整合。丁申桃等的调查显示,彼时我国高校数字资源建设已初具规模,但面临严重的“资源孤岛”问题<sup>[16]</sup>。这一问题的根源,正如管计锁所指出的,在于资源建设理念尚未完成从重“拥有”到重“存取”的根本性转变<sup>[17]</sup>。孙秀丽进一步强调,提升资源效益的关键在于打破重复建设壁垒,建立有效的共建共享机制<sup>[18]</sup>。总之,数字图书馆在本阶段的核心成就在于,“文献”作为图书馆的传统资源完成了载体形态的数字化转型,但资源的范畴尚未突破出版物的边界,数据主要服务于人类阅读,尚未具备机器可理解的结构化特征。

#### 2.2.2 科研数据管理阶段:科研数据被确立为核心学术资源

随着科研信息化(E-Science)和数据密集型科研范式的确立,科研数据本身的价值被空前重视。开放科学运动强调,研究数据是验证研究成果、保障科学发现可重复性的关键<sup>[19]</sup>。这一认知变革促使图书馆的资源建设范畴发生根本性拓展:从管理“关于研究的文献”,延伸到管理“研究过程中产生的数据”。随着科研数据被正式纳入图书馆的核心资源



体系,高校图书馆开始通过构建新型数据资源建设模式、搭建研究数据管理平台、制定数据政策与标准规范、探索数据出版与引证服务等方式,拓展资源体系,开展数据服务<sup>[20-21]</sup>。由此,“数据”作为一种全新的资源类型,正式进入图书馆的资源建设版图,图书馆开始扮演“数据保管员”的角色。

### 2.2.3 数据智能赋能阶段:从资源管理走向价值创造

当前,在人工智能与大数据技术的双重驱动下,图书馆的数据实践正迈向以“智慧”和“赋能”为特征的更高阶段。这一阶段的核心特征,是从对数据的“有序管理”跃升到对数据的“深度挖掘”与“价值创造”。数据治理成为资源建设的基础保障。在人工智能驱动的科学(AI4S)的背景下,数据资源建设的对象不再仅仅是“人类读者”,更扩展到“机器分析与模型训练”。张晓林指出,“高质量数据集”已成为人工智能创新的核心载体,其建设标准对数据的准确性、完整性、一致性、时效性提出了前所未有的要求<sup>[22]</sup>。叶梓萌等提出的“智慧数据”概念,进一步揭示了数据资源建设的前沿方向,要求数据具备“场景化、专业化、关联化和知识化”特征,能够直接支撑智能决策与知识发现<sup>[23]</sup>。在此阶段,图书馆不仅是数据的保管者,更是数据的加工者、知识的提炼者和智能服务的提供者。

## 3 复旦大学图书馆数据资源体系建设实践探索

数据资源是国家战略资源,已经成为图书馆馆藏资源体系的有机组成部分<sup>[24]</sup>。复旦大学图书馆自2014年提出转型发展方向的三个维度以来,其中“建设数据图书馆”便是核心战略之一<sup>[25]</sup>。经过十多年的持续建设与完善,复旦大学图书馆形成了“多校区协同、纸电互补、特藏突出”的资源格局,具有规模宏大、结构多元、特色鲜明、技术领先的特点,全面覆盖文、理、医、工、管等学科,尤其在人文社科、基础科学和临床医学领域资源更为深厚。作为中国高校人文社会科学文献中心(CASHL)的两个全国中心馆之一,其资源体系不仅服务于本校的“双一流”建设,还承担着辐射全国的人文社科文献保障的重任。

### 3.1 数据资源建设概况

#### 3.1.1 数据资源建设规划与定位

复旦大学图书馆加强顶层设计和规划,在“十四五”期间,提出了秉承以读者为中心的服务理念,以

高质量文献信息资源和特色馆藏资源为基础,发挥核心竞争力,提升不可替代性,重点建设智慧图书馆,努力改善知识信息服务条件,为本校乃至国内外教学、科研、学科发展和知识创新提供文献信息服务的总体目标。国家“十四五”规划中明确将数据管理与研究作为发展重点,在数据资源、数据平台、技术研发、数据服务等方面不断推进<sup>[8]</sup>。在这一背景下,制定《复旦大学图书馆数据管理办法》和《数据中心整体架构与建设规范》,综合分析图书馆数据资源发展现状、应用现状以及复旦大学智慧图书馆愿景,明确图书馆的数据管理基本原则和建设规范。

面向未来,复旦大学图书馆“十五五”规划将持续建设高质量、有特色、具有保障力的馆藏,深化数字资源平台建设与应用,提升数字资源利用效率。规划明确提出持续完善智慧图书馆数字基座的“数据中心”和“AI服务中心”,通过引入通用基座模型,开展面向AI的学术资源数据治理,提升图书馆的知识管理和服务能力,增强对学科融合和研究创新的精准知识服务支撑。同时,聚焦垂直领域大模型构建,发挥图书馆特藏资源数据优势,结合学科研究方向,重点探索构建东亚语言大模型语料库、考古发掘文献AI-Ready语料库等图书馆特藏资源大模型语料库,为人文社会科学智能(AI4SS&H)的新范式提供坚实的数据支撑。

#### 3.1.2 组织架构与经费投入机制保障

为促进数据图书馆建设,复旦大学图书馆于2017年成立大数据研究院人文社科数据研究所,2019年成立“上海市科研领域大数据联合创新实验室”,目前为全国高校唯一一家省市级科研领域大数据实验室。以数据管理与技术部以及人文社科数据研究所为基础,突出其数据管理的职能,负责数据图书馆的前沿研究、项目规划和建设<sup>[26]</sup>。此外,成立复旦大学阿法迪智慧图书馆学研究中心,进一步丰富和扩展数据资源,围绕数据资源、数据管理、数据安全和数据共享等内容进行技术研发和关键技术的突破,建设创新数据平台。

在经费投入方面,图书馆多方争取资源,努力保障经费的平稳增长,并积极引入社会资源。通过与上海阿法迪智能数字科技股份有限公司合作,增加人力、经费等资源的积累,共同推进智慧图书馆建设。这种“校企共建、资源共享”的模式,有效缓解了单一财政投入的压力,为数据资源体系的持续迭代



提供了资金保障。

### 3.2 数据资源建设体系架构

复旦大学图书馆数据资源建设体系呈现出从“传统文献采购”向“数据驱动型知识服务”转型的鲜明特点。在确立“打造数据驱动型研究型图书馆”这一核心目标的基础上,复旦大学图书馆构建了由文献资源保障体系、特色数据资源体系、科研数据支撑体系、智慧图书馆与数字基座体系组成的四大支柱,形成了全生命周期、全要素覆盖的数据资源建设生态。

#### 3.2.1 文献资源保障体系

文献资源保障体系是高校图书馆的传统核心职能,近年来经历了显著的结构调整,特别是在应对经费压力下的精细化配置方面取得了突破性进展。

优化资源配置,提高经费效率。不断优化电子资源和纸本书刊的协调采购,在“十四五”期间重点加强了外文图书的纸电协调。通过少量缩减外文纸本经费,电子书协调采购项目,增加了外文馆藏,提高了新书的及时性,提升了读者使用的便利性。2020—2025年,纸质资源与电子资源经费比例稳定在3:7区间,2025年资源建设统计数据显示,电子资源经费占比已超70%。

多载体资源覆盖,维持连续性与完整性。维持中外文核心期刊和高水平学术专著连续性与完整性。同时,加大永久使用权资源的采买力度,如Bloomsbury、De Gruyter电子书,超星视频库等,加强本地化资源存储,构建“自有资产池”,以应对商业数据库停订风险。

捐赠资源数字化整合。建立规范的捐赠接收与加工流程,将实体赠书快速转化为可检索的馆藏数据。“十四五”期间,复旦大学图书馆通过捐赠渠道增加中外文图书约4.3万册,占“十四五”期间馆藏新增总量的8.3%,成为馆藏重要的补充。捐赠量占年入藏量比例显著提升,2025年达到了21%。

构建全评价理论模型。针对集团采购的数字资源,不再单纯依赖价格或使用量,而是基于“全评价”理论,构建了涵盖内容质量、成本效益、技术性能、服务水平等多维度的综合评价模型,为采购决策提供量化依据。

在当前经济形势下,复旦大学图书馆与其他高校图书馆共同面临着资源建设经费紧张的情况;同时,随着《中华人民共和国增值税法》的正式实施,

2026年起,国家对特定图书进出口公司享受的向科研单位、高校销售进口书刊免征增值税(国内销售环节)的优惠已经取消,如果该增值税转由高校承担或部分承担,那么高校图书馆的资源建设经费可能会更加紧张。面对这种情况,图书馆采取“保核心、停低效、推共建”策略:一方面通过高校数字资源采购联盟(DRAA)进行集团采购以降低成本;另一方面通过评估停订高成本低效用的部分外文数据库,集中力量建设图书馆文献资源保障体系,支撑学校的学科评估与“双一流”建设。

#### 3.2.2 特色数据资源体系

复旦大学图书馆高度重视特色馆藏建设,聚焦于古籍、特藏、红色文献等珍贵资源的活化利用,致力于将静态的纸质文献转化为可检索、可关联、可计算的高价值数据资产。

推行层级化著录标准。引入国际通用的开源平台ArchivesSpace,采用编码档案描述(Encoded Archival Description, EAD)标准,对大型特藏如《希腊铭文集成》、当代中国社会生活资料等进行“全宗—系列—案卷—文件”的多层级著录。这种“馆藏指南”(Finding Aid)模式,实现了从单件文献到集合资源的立体化揭示,极大提升了特藏的检索效率与研究价值。

构建领域本体与知识图谱。基于“实体—关系”模型,针对档案类资源构建专属本体,提取人、地、时、事等关键元素并建立语义关联。这不仅实现了数据的结构化存储,更为后续构建知识图谱、开展数字人文研究奠定了坚实基础。

强化红色资源保护与挖掘。系统整理陈毅藏书及资料、陈望道藏书及资料、延安时期红色出版物、美军飞行员藏延安鲁艺版画等红色特藏,结合先进的纸张检测与脱酸技术,确保物理载体的长久保存。同时,通过数字化手段深度挖掘其思政教育价值,打造沉浸式思政教育基地,为落实立德树人根本任务提供鲜活教材与实践平台,实现知识传授、能力培养与价值引领的有机统一。

古籍与特藏数字化建设。持续推进馆藏善本、民国文献、手稿的数字化。完成古籍善本特色文献1200多种64万余页的数字化工作。与大象出版社合作,对复旦大学图书馆藏《诗经》进行数字化和编目,借用企业的力量,推动古籍数据资源建设。

建设专题数据库集群。建设“印度古典学文献



集成”(梵文语料库)、“陈毅专题数据库”“冀汾文献”等垂直领域库。建设专题数据库和平台有助于图书馆间的合作共建与资源共享,通过平台的数字化、数据化呈现,发现研究线索,洞察研究先机。复旦大学图书馆于2021年启动建设“陈毅专题数据平台”,该平台在馆藏陈毅捐赠资料基础上,全面汇集与陈毅相关的文献、资料和数据。在内容上,平台在实现了多资源类型的文献汇集的同时,通过对《陈毅传》《陈毅年谱》、油印本《陈毅日记》等条目的拆解,形成多源异构数据的融合。

构建“上医医脉”知识库。整理百年医史档案、名医手稿,应用自然语言处理、知识图谱等数字人文技术手段,对特藏资源进行深度挖掘处理,实现规范化元数据标引。

技术赋能数据资源建设。引入AI技术进行古籍自动标点、繁简转换及实体识别,提升数据可用性。全面建成“复旦学术典藏”系列平台,构建起集学术成果数据采集、治理、存储、共享于一体的学术成果管理服务体系。实现复旦大学师生学术成果(论文、图书、专利等)的自动采集与人工审核双轨制。收录论文全文从数万篇增长至数十万篇,并向国家自然科学基金开放获取仓储平台(OAR)等平台自动汇交元数据,提升学校学术影响。

### 3.2.3 科研数据支撑体系

复旦大学社会科学数据平台于2014年正式上线,是国内高校首家人文社会科学数据平台。紧跟AI4S科研范式新发展,全面提高平台的自主研发性、安全可控性,引入变量数据库模式,探索FAIR评价等特色功能,采用前后端设计,实现消息实时提醒、主动推送数据信息,并支持多种访问方式,深度嵌入科研生命周期,支持数据驱动的科学发现。截至2025年底,复旦大学社会科学数据平台拥有数据空间203个,数据集1078个,文件4269个,访问量超2406万次,下载次数约72万次。响应国家数据战略,建设如“考古资料高质量数据集”,相关项目入选上海市数据局首批先行先试项目。积极推动教育科研数据共享,牵头建设“慧源上海教育科研数据共享平台”,打通校际数据壁垒,实现跨校数据资源共享与服务。探索数据治理规范,建立覆盖数据全生命周期的数据管理规范,包括数据采集标准、元数据标准规范、数据服务标准及隐私保护机制。

实施科学数据分级治理。借鉴哈佛大学 Datat-

ags系统的先进理念,结合本土实际,构建科学数据分级模型。通过差异化的访问控制、加密策略与脱敏处理,在保障数据安全与隐私的前提下,最大限度地促进数据共享。

建设高水平人文社科数据专题库。依托人文社会科学数据研究所、上海市科研领域大数据联合创新实验室,重点建设当代中国社会生活资料数据库、东亚语言数据库等特色库,填补国内相关领域数据空白,有力支撑数据密集型的人文社科研究。

### 3.2.4 智慧图书馆与数字基座体系

利用新技术重塑数据组织方式和服务触达路径,解决“数据孤岛”和“查找难”问题,拓展图书馆服务的广度与深度,不断提升读者的获得感和满意度。

系统架构迭代升级。构建以数字基座为基础支撑的开放、可自生长的“1+5+N”智慧图书馆服务平台架构——“复旦大学图书馆服务平台(Library Service Platform, LSP)”,实现图书馆管理系统的国产化和迭代升级,以支持更灵活的数据接口和业务扩展。

下一代资源发现系统。从传统的OPAC向“望道溯源”等资源发现平台演进,实现纸质、电子、数据等资源一站式检索。“望道溯源”是复旦大学图书馆的一站式学术搜索平台,通过平台检索,读者可以迅速发现、借阅或获取文献,包含Elsevier、Gale、Springer、Wiley、EBSCO、JSTOR、万方、维普等主要中外文数据库中的近7亿篇学术文章数据。

AI驱动的数据服务。基于大模型训练图书馆专属知识库“旦小图+AI”,提供准确率95%以上的咨询服务,支持答案溯源。

智能编目与审核。探索AI辅助编目及学位论文格式自动审核,提高工作效率。基于数据质量管理平台,数据审校工作思路从传统的“编目员思维”转向“大数据质量管理思维”:将自动化、流程化的图书馆全量数据质量监测和评价融入业务流程,并建立了“源头治理、闭环管理”的常态化数据质量管理机制,有效提升了工作效率和数据的价值。

视觉盘点。引入机器人进行馆藏资产可视化盘点,实现馆藏数据的实时动态更新。

推进数据治理,充分发挥数据价值。依托数字基座逐步完善数据治理体系,并着力统一数据标准、解决图书馆多源异构数据的关联和融合问题,建设一套完整、规范、高质量的数据资源体系。图书馆的



数据也汇聚到学校信息办,实现校内共享、一数多用,充分发挥数据价值。

嵌入智库建设与情报服务。推动医学图书馆深度嵌入科研团队,提供竞争情报分析、学科态势研判等高附加值服务,积极参与中国特色新型智库建设,发挥图书馆作为“思想库”与“智囊团”的作用。

以上资源建设体系相互支撑、有机融合。文献资源保障是基石,确保资源的优质供给;特色数字资产是亮点,彰显文化的深厚底蕴;科研数据支撑是核心,驱动科研的创新突破;智慧图书馆与数字基座是引擎,激发用户的内生动力。四者共同构成了复旦大学图书馆独具特色的数据资源建设体系。

#### 4 建设成效、挑战与展望

复旦大学图书馆通过构建数字基座、打造学术典藏平台、深化特色资源数字化及优化电子资源配置,实现了从“文献管理”向“数据治理与知识服务”的转型。在数据资源建设方面取得了一定成效。

##### 4.1 建设成效

建设成效主要体现在资源数量增长、资源结构优化、数据价值深度挖掘、服务模式创新以及治理能力现代化等方面。

##### 4.1.1 资源结构优化

在经费投入与增长方面,“十四五”期间,复旦大学图书馆电子资源订购费用1.29亿元,经费占比从2021年的52.5%增长至2025年的70.8%。在馆藏结构上,外文电子图书从219万册激增至659万册(增长率达200%),中文电子图书增至1496万册。重点加强了Elsevier、Cambridge等外文电子书项目的按需采购(EBA)模式采购。在资源揭示方面,完成了数据库导航系统建设,将76万余条电子资源信息导入Aleph系统,并实现了CASHL、全球文献资源传递系统(Rapido)等系统的深度集成。

依托CASHL、DRAA,通过集团采购降低数据库成本。同时,作为CASHL全国中心馆之一,复旦大学图书馆的资源共建共享能力辐射全国,提升了中国高校人文社科资源的整体保障水平。

##### 4.1.2 特藏资源数字化

复旦大学图书馆在古籍、特藏及红色资源的数字化与深度揭示方面取得了重要进展,形成了具有复旦特色的数字文化品牌。

特藏保护与研究并重。复旦大学图书馆特藏资

源丰富,近年来通过成立中华古籍保护研究院和文物保护创新研究院,加强对纸质文物资源的科学保护和高层次人才培养。对馆藏特色资源,如包括《诗经》等文献在内的善本及特色文献,进行数万页善本的高清数字化与自动标点工作。通过“故纸探新”开放课题,鼓励师生利用数字化特藏开展研究,实现了特藏资源保护从“被动封存”向“主动开发”的转变。累计新增上海医学院及上医学人史料特藏5000余件,并通过展览和新媒体矩阵(视频号播放量768万次)进行活化利用。

红色特藏体系化。系统整理了延安时期图书、鲁艺版画(George A. Hanlon 中校捐赠)、陈望道/陈毅藏书等红色文献,结合纸张脱酸等保护技术,实现了物理安全与数字存储。

##### 4.1.3 学术成果数据底座构建

经过近十年的努力,全面建成“复旦学术典藏”系列平台,构建起集学术成果数据采集、治理、存储、共享于一体的学术成果管理服务体系。目前平台存储全校2009年以来的学术成果数据总量突破50万条,其中期刊和会议论文元数据34万篇(全文12万篇)、图书1.4万种、学位论文11万篇、专利3万余项。与复旦大学信息办合作研发的“复旦学术成果统一认领平台”,年更新论文成果逾2万篇,年均参与认领的在教职工超过半数。2023年“认领平台手机端上线”被列入当年十大校园实事工程;2024年用户满意度调查显示平台功能认可度91%,使用便捷度97%,个人成果管理效果赞同度83%。《复旦学术典藏系列平台建设及多源异构学术成果数据融合清洗法的应用研究》获2024年华东地区科学技术情报成果奖二等奖。

“复旦学术典藏”的期刊会议论文和图书成果数据当前作为校内权威数据源,已通过复旦数据中心对接学校个人数据中心、教师个人主页、研究生导师综合管理系统、教工年度考核系统、文科科研创新服务平台、MOSS大模型等,实现了数据实时共享,为学校日常科研活动和管理决策提供了高质量的基础数据保障。2024年,“复旦学术典藏”数据顺利接入国家自然科学基金开放获取仓储平台(OAR)。这一举措不仅提升了复旦大学学术成果的开放共享水平,也为国家学术资源的整合与利用提供了重要支撑。



#### 4.1.4 智慧图书馆数字基座建设

复旦大学智慧图书馆数字基座是支撑高校图书馆数字化转型的新型数字基础设施,汇聚融合图书馆多源异构数据,沉淀公共业务,建立感知设施的智能互联,提供数据、业务技术和 AI 能力支持服务。数字基座综合运用大数据、人工智能、物联网等数字技术,在标准规范、运营机制和网络安全保障之下,通过数据中心、业务中心、基础服务中心、物联中心和 AI 服务中心的协同运行,使图书馆具备灵活应对未来多场景聚合、多技术融合和多应用汇集所需的能力。

作为国内首个落地“大平台、轻应用、以数据为驱动”的数字基座,解决了烟囱式系统导致的数据孤岛问题。建立了全量数据质量监测和评价机制,从“编目员思维”转向“大数据质量管理思维”,实现了多源异构数据的汇聚、标准化和资产化。支撑“座位预约系统”“智能问答机器人(旦小图)”“数字化教学参考书平台”等创新应用,并获 2024 年上海科学技术情报成果奖一等奖。

#### 4.1.5 情报服务与智库化转型

图书馆依托丰富数据资源和专业能力,学科情报服务提质增效,对接校内各需求方开展学科情报服务。“十四五”期间完成来自学校办公室、发展规划处、科学技术研究院、人才工作办公室、财务与国有资产管理处、国际合作与交流处等多个职能部门和科研机构的委托任务 80 余项,服务产出涵盖基础数据提供、数据统计分析、深度研究报告等形式,为学校学科发展、人才评估等工作提供多维度、全方位决策支持。陆续发布《复旦大学国际科研合作发展报告》《中国高校开放获取论文发展蓝皮书》《人文领域高层次人才画像初探》等科研产出成果。2021 年图书馆“复旦大学人文社科评价研究中心”作为牵头单位组织参加由中央网信办、国家发展改革委、教育部等 8 部门联合发起的国家智能社会治理实验基地申报工作,并成功入选教育特色基地。目前“复旦大学国家智能评价与治理实验基地”已实体化运行,推进学术与教育评价、科技与科创评价、社会与治理评价等全社会要素评价和智能社会治理问题的研究,进而为国家数字化智能社会转型做好综合性服务。

此外,通过可视化平台建设,推出“学术成果专题看板”(ESI 监测、科研合作专题),通过商业智能(Business Intelligence, BI)工具实现数据实时可视

化。与此同时,为复旦大学部分附属医院新增成果数据订阅服务,有效提升论文成果数据的准确性以及管理与应用效率。

#### 4.1.6 治理体系与人才队伍现代化

数据治理框架初步建成。从“传统管理”转向“现代治理”,提出了包含治理内容、治理过程、治理规范在内的数据治理框架<sup>[27]</sup>。总结了数据资源建设的总体流程;设立了专门的数据运营团队,将数据治理融入业务流程,形成了闭环管理机制。

复合型人才队伍崛起。培养了一批具备数据分析、可视化、编程能力的“数据馆员”和“学科情报专家”。馆员中硕士及以上占比超 84%,为图书馆的可持续发展提供了坚实的人才支撑。

复旦大学图书馆的数据资源建设成效,标志着其已从传统的“文献资源收藏机构”转型为“数据驱动的知识服务与创新平台”。在资源端,实现了从“大而全”到“精而准”的跨越;在支撑端,实现了从“分散存储”到“全生命周期治理”的升级;在服务端,实现了从“被动响应”到“主动赋能”的变革。这一系列成效不仅为复旦大学的“双一流”建设提供了坚实的数据基座和智慧引擎,也为我国高校图书馆的数字化转型提供了可借鉴的实践。

#### 4.2 面临的挑战与未来展望

复旦大学图书馆资源建设尽管取得了一定的成效,但在近年的实践中也面临着很大的挑战。

##### 4.2.1 面临挑战

在馆员队伍能力转型方面,面对 AI 技术的冲击,懂图情业务、通 AI 技术、知法律规范的复合型数据馆员不足,需不断提升馆员的数智素养。数智化转型与服务深度升级,对馆员的数智素养、学科专业能力、跨界服务水平提出了更高要求,现有队伍的知识结构、能力体系与转型需求存在差距,复合型、创新型人才供给不足,人才培养、评价与激励机制亟待优化。

在经费保障方面,资源建设经费与境内外头部高校相比优势不足。智慧图书馆基础设施、数据平台建设维护等都需要长期投入,而在经费紧张的大背景下,如何平衡传统资源采购与新兴数据基础设施建设之间的关系,是一个严峻的考验。

在数据安全与伦理风险方面,图书馆在数智化转型过程中,面临着数据合规、隐私保护、内容安全、技术安全等多重风险防控挑战。随着数据开放的深入,数据隐私保护、版权清除、算法偏见等问题日益凸显。



如何在促进数据共享的同时,构建符合社会伦理和法律规范的安全防线,是未来必须解决的难题。

#### 4.2.2 未来展望

复旦大学图书馆在“十四五”期间奠定了坚实的数据基座和学术典藏平台,实现了资源的数字化和管理的数据化。展望“十五五”,其核心战略将全面转向智能化与知识化,旨在通过 AI 技术重塑数据资源的组织方式(细粒度、知识图谱),将图书馆从传统的资源提供者转变为大学学术资源中心和 AI 驱动的科研创新基础设施,以支撑学校在“双一流”建设中的学科交叉与原始创新。面向“十五五”,复旦大学图书馆将围绕 AI 赋能、知识组织细粒度和学术资源基础设施升级三大核心战略展开建设。

(1)战略升级:从“数据仓库”向“科研创新基础设施”转型

推动“复旦学术典藏”从后台数据仓库升级为集数据治理、共享服务、决策支持于一体的科研创新基础设施。系统性建设 AI-Ready 语料库,构建“基础语料库—专题数据集—领域知识库”三级资源体系,重点推进人文社科学术资源的工程化知识组织。联合院系与 AI 企业,训练专业领域大模型,支撑人文社会科学智能(AI4SS&H)和人工智能驱动的科学研究(AI4S)。

(2)技术赋能:全面引入人工智能重构业务流程利用多模态融合算法,实现跨媒体学术资源的深层语义识别与关联,构建高质量学术知识图谱。提供基于自然语言理解的智能检索、多模态实时响应及结果关联,实现复杂研究任务的自动化处理与知识产品的结构化输出。以智能体技术赋能,按需构建分布式学术信息资源加工平台。更换新一代智慧图书馆管理服务平台,以适应 AI 时代管理需求。

(3)特色资源深化:数字化与知识化并重

在特藏深度开发上,不仅限于扫描数字化,利用 AI 技术构建特藏智能问答系统和推荐系统,建设可供教学研究的医学/古籍特色数据库。建立校内学术资源共建共享协议体系,制定符合 AI 伦理规范的数据分级开放标准。

(4)人才与素养:构建 AI 时代的新型服务体系

建立分层递进式人工智能素养培育体系,培养具备计算人文研究能力的复合型人才。优化馆员队伍结构,引进或培养具备数据建模、AI 项目管理能力的专业人才。

(5)强化风险防控:构建符合法律规范的版权清除与合理使用机制

从“被动防御”转向“主动治理”。构建全生命周期安全防线,实施动态分级分类保护策略。强化算法伦理与偏见审查;针对 AI 生成内容建立人工复核与伦理审查机制,防止算法歧视、幻觉信息及知识产权侵权风险。完善应急响应与容灾备份:开展数据安全演练,建立异地容灾中心,确保数据完整性与服务连续性。通过机制激活数据价值,技术提升服务能级,人才深化服务内涵,风险筑牢安全底线,四维协同推动复旦大学图书馆向“智能、开放、安全、高效”的现代化数据中枢跃迁。

## 5 结语

复旦大学图书馆数据资源体系构建的实践,是在中国式现代化和高校“双一流”建设宏大背景下的主动求变。通过顶层设计的战略重构、组织架构的敏捷调整、技术平台的智能升级以及治理制度的创新保障,图书馆成功实现了从传统“文献资源收藏机构”向“数据驱动的知识服务与创新生态”的转型。这一转型不仅提升了图书馆自身的生存与发展能力,更重要的是,重新定义了图书馆在大学创新生态系统中的基本功能与核心价值。图书馆不再是边缘化的文献仓库,而是成为了科研创新的孵化器、数据要素的集散地和智慧服务的策源地。

展望未来,高校图书馆应进一步深化人工智能与大数据技术的应用,不仅要成为学术资源的守护者,更要成为科研数据的治理者、AI 模型的训练者与知识创新的赋能者。面对数据要素市场化配置的新机遇,高校图书馆需主动打破围墙,加强馆际合作以及与科研院所、企业及政府部门的协同合作,构建开放、融合、智慧的数据资源新生态。唯有如此,图书馆才能在数字化浪潮中把握主动,为高等教育的高质量发展贡献独特的力量,为建设教育强国、科技强国、人才强国提供坚实的数据支撑与智慧保障。

## 参考文献

- 1 中共中央二十届四中全会. 中华人民共和国国民经济和社会发展第十五个五年规划纲要[EB/OL]. [2026-03-13]. <https://www.news.cn/politics/20260313/085af5de5a4b4268aa7d87d90817df2f/c.html>.



- 2 袁于飞. 让科学数据开放共享成为常态[EB/OL]. [2026-03-13]. [https://www.gov.cn/zhengce/2018-04/05/content\\_5279957.htm](https://www.gov.cn/zhengce/2018-04/05/content_5279957.htm).
- 3 中共中央组织部. 党的十九届四中全会《决定》学习问答 37. 为什么要健全劳动、资本、土地、知识、技术、管理、数据等生产要素由市场评价贡献、按贡献决定报酬的机制? [EB/OL]. [2026-03-13]. <https://www.dtdjzx.gov.cn/staticPage/zhuanti/ddsjxxwd/20200108/2640642.html>.
- 4 李婕. “数据二十条”对外发布, 构建数据基础制度体系——做强做优做大数字经济[EB/OL]. [2026-03-13]. [https://www.gov.cn/zhengce/2022-12/21/content\\_5732906.htm](https://www.gov.cn/zhengce/2022-12/21/content_5732906.htm).
- 5 国家数据局, 中央网办, 科技部, 等. 国家数据局等部门关于印发《“数据要素×”三年行动计划(2024—2026年)》的通知[EB/OL]. [2026-03-13]. [https://www.cac.gov.cn/2024-01/05/c\\_1706119078060945.htm](https://www.cac.gov.cn/2024-01/05/c_1706119078060945.htm).
- 6 国务院. 关于深入实施“人工智能+”行动的意见[EB/OL]. [2026-03-12]. [https://www.gov.cn/gongbao/2025/issue\\_12266/202509/content\\_7039598.html](https://www.gov.cn/gongbao/2025/issue_12266/202509/content_7039598.html).
- 7 国家数据局. 国家数据局关于加强数据科技创新的实施意见[EB/OL]. [2026-03-13]. [https://www.gov.cn/zhengce/zhengceku/202512/content\\_7053003.htm](https://www.gov.cn/zhengce/zhengceku/202512/content_7053003.htm).
- 8 中共中央十九届五中全会. 中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要[EB/OL]. [2026-03-13]. [https://www.gov.cn/xinwen/2021-03/13/content\\_5592681.htm](https://www.gov.cn/xinwen/2021-03/13/content_5592681.htm).
- 9 张若楠, 刘琪. 图书馆数字资源研究热点与发展态势分析[J]. 图书馆学研究, 2024(11): 2-13.
- 10 DRAA 秘书处. 标准规范[EB/OL]. [2026-03-13]. <http://www.libconsortia.edu.cn/Standard/list.action>.
- 11 马越. 高校图书馆数字资源精准化采购决策支持模型研究[J]. 图书馆学研究, 2023(7): 49-63, 82.
- 12 杨蕾, 崔琼. 国外高校图书馆数字资源本地化保存与服务的经验及启示[J]. 图书馆学研究, 2024(11): 60-70.
- 13 曾蕾, 谭旭. 数据的语义增强——解读图档博支持数字人文的新动向[J]. 数字人文研究, 2021, 1(1): 65-86.
- 14 孟祥保, 高冕. 图书馆数据资源建设: 内涵、价值与路径[J]. 新世纪图书馆, 2023(5): 48-54.
- 15 李书娟. 数智时代图书馆数据馆藏的内涵特征、逻辑构成与构建路径[J/OL]. 图书与情报, 1-8[2026-03-14]. <https://link.cnki.net/urlid/62.1026.G2.20260304.1905.001>.
- 16 丁申桃, 庞跃霞, 曹丽娟. 我国高校图书馆数字化资源建设现状分析[J]. 图书馆论坛, 2005(6): 200-203.
- 17 管计锁. 论传统图书馆向数字图书馆过渡时期的资源建设[J]. 中国图书馆学报, 2001(2): 40-43.
- 18 孙秀丽. 高校图书馆数字资源建设与利用的调查分析[J]. 大学图书馆学报, 2008, 26(6): 45-50.
- 19 OECD. Making open science a reality[EB/OL]. [2026-03-14]. <https://doi.org/10.1787/5jrs2f963zsl-en>.
- 20 Fearon D, Gunia B, Lake S, et al. Research data management services. (SPEC Kit 334)[EB/OL]. [2026-03-13]. <https://publications.arl.org/Research-Data-Management-Services-SPEC-Kit-334/>.
- 21 伏安娜, 汪东伟, 程蕴涵, 等. 中国高校图书馆研究数据管理服务十年实践与思考[J]. 图书情报工作, 2025, 69(4): 23-33.
- 22 张晓林. “人工智能+”背景下的高质量数据集建设: 图书馆的机遇与挑战[J]. 中国图书馆学报, 2025, 51(1): 4-17.
- 23 叶梓萌, 钱力, 刘细文, 等. 面向图书馆数据资源建设的智慧数据研究: 内涵界定、技术应用与研究启示[J]. 图书情报工作, 2024, 68(24): 3-15.
- 24 孟祥保, 高冕. 图书馆数据资源建设: 内涵、价值与路径[J]. 新世纪图书馆, 2023(5): 48-54.
- 25 陈思和. 高校图书馆传统功能外延的三个拓展方向[J]. 大学图书馆学报, 2018, 36(5): 12-16, 22.
- 26 袁玉红, 叶燕, 杨梦, 侯力强. 转型发展背景下复旦大学图书馆青年馆员队伍建设实践探索[J]. 大学图书馆学报, 2023, 41(4): 5-12.
- 27 薛崧, 张计龙, 苗青, 等. 高校图书馆数据治理框架构建——复旦图书馆的实践与思考[J]. 大学图书馆学报, 2025, 43(2): 84-94.

作者贡献说明:

侯力强: 研究思路与框架设计, 论文统稿与修改

伏安娜: 部分内容撰写及修改

袁玉红: 参与论文框架设计, 部分论文撰写及修改

作者单位: 复旦大学图书馆, 上海, 200433

收稿日期: 2026年3月4日

修回日期: 2026年3月14日

(责任编辑: 王菲)



## Building Data Resource Systems in Academic Libraries in Digital Transformation : A Case Study of Fudan University Library

HOU Liqiang FU Anna YUAN Yuhong

**Abstract:** Driven by the dual impetus of AI-induced paradigm shifts in scientific research and the in-depth implementation of national data strategies, academic libraries are undergoing a historic transformation from “custodians of resources” to “facilitators of data-driven innovation.” This paper systematically reviews the theoretical evolution and practical pathways of data resource construction in academic libraries. Taking Fudan University Library as a representative case, it examines its integrated four-pillar framework: the resource assurance system, the distinctive data resource system, the research data support system, and the smart library and digital infrastructure system. (1) In the domain of resource assurance system, the Library has refined the coordinated acquisition of electronic and print materials, ensuring comprehensive multi-format coverage while maintaining the continuity and integrity of collections. Standardised workflows for donation acquisition and processing have been established, alongside a multidimensional evaluation model that provides a rigorous, evidence-based foundation for collection development. (2) With respect to distinctive data resource system, the Library prioritises the curation and activation of special collections, including rare books, archives, and red cultural heritage materials. This is achieved through tiered cataloguing standards, the construction of domain ontologies and knowledge graphs, enhanced preservation strategies, and large-scale digitisation initiatives supported by thematic database clusters. These efforts significantly improve the accessibility and scholarly value of unique resources. (3) In terms of research data support system, the establishment of a Social Sciences Data Platform has enabled lifecycle-oriented data management and tiered governance mechanisms, thereby providing strong infrastructure for data-intensive research in the humanities and social sciences. Meanwhile, in advancing a smart library and digital infrastructure system, the institution has developed an integrated Library Service Platform built upon a unified digital base. This platform marks a transition from traditional OPAC systems to next-generation discovery tools, enabling seamless, one-stop access to print, electronic, and data resources. AI-driven applications—including intelligent cataloging, automated review, visual inventory management, and enhanced data governance—further strengthen service capabilities and support embedded knowledge services. The findings suggest that through strategic design, organisational restructuring, technological upgrading, and governance innovation, academic libraries can effectively break down data silos and achieve a shift from digitisation to datafication, knowledge integration, and intelligent service delivery. To address challenges related to talent, funding, ethics, and security, future efforts should focus on developing AI-Ready corpora, domain-specific large models, and full lifecycle data governance. Such initiatives will help build an open, integrated, and secure scholarly resource ecosystem, thereby supporting the high-quality development of higher education.

**Keywords:** Academic Libraries; Data Resource Systems; Fudan University Library