



# 推进数据资源本地化服务 构建文献信息服务新格局

——2024 年数据资源本地化服务研讨会综述

□吴亚平\* 姚晓霞 任国华 李婉君

**摘要** 2024 年 5 月 14—17 日,“数据资源本地化服务研讨会暨 CALIS 第二十二届引进数据库培训周”在天津市成功举办,来自国内外的高校图书馆及相关行业的嘉宾、学者、馆长与馆员代表,数据库商、代理商等数据资源服务商代表,合作机构代表等 452 人参加会议。16—17 日会议围绕“数据资源本地化服务”的主题,展开研讨交流。文章综述了本次研讨会各位专家、学者以及数据资源服务商的研究成果和实践思考,在加快数据开发利用,推动数据要素发挥乘数效应的时代背景要求下,为高校图书馆数据资源本地化服务的高质量发展提供理论启发和路径思考。

**关键词** 高校图书馆 数据资源 本地化服务 长期保存

**分类号** G250

**DOI** 10.16603/j.issn1002-1027.2024.03.016

## 1 引言

“数据”已成为新的生产要素<sup>[1]</sup>,2024 年 1 月 4 日,国家数据局等部门印发《“数据要素×”三年行动计划(2024—2026 年)》<sup>[2]</sup>,强调要提供高质量科学数据资源与知识服务。2024 年全国数据工作会议指出要提升数据资源开发利用水平,强化数据安全保障能力,要提升数据领域国际合作水平等,更好发挥数字化在中国式现代化中的驱动引领作用<sup>[3]</sup>。数据资源是支撑高校教学科研的基石,推进数据资源本地化服务是教育数字化高质量发展的基础保障和重要途径,为深化数据领域国际合作提供了新契机。

为进一步推进数据资源本地化服务,探索数据赋能图书馆高质量发展的可靠路径,加快教育强国建设,由中国高等教育文献保障系统(CALIS)管理中心和高校图书馆数字资源采购联盟(DRAA)理事会联合主办、天津大学图书馆承办的 2024 年“数据资源本地化服务研讨会暨 CALIS 第二十二届引进数据库培训周”在天津市顺利举办,来自国内外的高校图书馆及相关行业的嘉宾、学者、馆长与馆员代表,数据库商、代理商等数据资源服务商代表,合作机构代表等 452 人参加本届培训周。培训周由两部

分组成,5 月 14—15 日为数据库培训,共有 23 家数据库商围绕数据库最新进展以及使用等内容开展专业培训。16—17 日会议围绕“数据资源本地化服务”的主题开展探讨,主办方特邀国内外图书馆业界和学界的专家学者、获奖研究课题组成员、数据资源服务商代表作了 20 余场专题学术报告,深入思考和探讨高校图书馆引进数据资源的本地化服务的理论指针与切实举措,就当前和未来一段时间内数据资源本地化服务的热点和难点问题进行了深入的交流研讨。

## 2 推进数据本地化服务,助力图书馆高质量发展

### 2.1 强化数据存取能力,夯实数据服务根基

DRAA 理事长、北京大学图书馆馆长陈建龙在主旨报告“推进数据资源本地化服务,支撑数字学术高质量发展”中指出,高校图书馆应把握数字中国与人工智能快速建设与发展的良机,服务于国家战略,根据国家政策确定发展方位和目标点。他将数据资源定义为“用户可资利用的数据(原生或再生的数据库等数据集、数据处理和使用的工具平台等基础设施),可经确权成为资产,可依法共享成为生产、生活或治理的要素;并分享了对于两个问题的思考,一是

\* 通讯作者:吴亚平,ORCID:0000-0002-4242-2434,邮箱:wuyyp@lib.pku.edu.cn。



数据资源及其服务的本地化存取问题,二是数字学术及其发展的高质量求解问题。第一个问题,陈建龙指出图书馆开展数据资源本地化服务的核心是建成数据“存”与“取”的服务能力,需要分别处理好与数据供给方以及与用户需求方的关系。图书馆作为用户需求管理者,在基础条件有限、难与用户需求匹配且存取服务能力不足的现实情况下,需思考如何让需求方可知、可信、可靠、可用、可测和可付;作为供给方的合作者,在数据资产属性胜过要素属性、营销模式可持续性挑战下,需思考供给方可盈利、可持续、可如愿等问题。第二个问题,陈建龙将数字学术描述为学人以多源数据的实时采集、动态管理和关联分析为基本活动的新型学术,并认为图书馆开展数字学术服务首先需关注在用户所需数据的求全和求准难度更大、数据基础设施运用条件和能力受限的现实情况下,如何帮助学人省时、省力获取与梳理相关信息,使其能将更多的时间用于辨析概念、凝练观点、建构理论和学术交流;还要解决在图书馆服务理念、能力和策略仍有待改进且“有组织”服务较弱的条件下,服务者如何建立与学人的信任关系,以共解难题、同创价值。最后,他以国家数字科技文献资源长期保存体系((National Digital Preservation Program, NDPP)北京大学分中心、CALIS本地化服务平台以及北京大学交融式数字图书馆建设的实践为例,介绍了在数据资源本地化服务上的有益探索。

## 2.2 深化数据挖掘利用,构建科技文献知识底座

中国科学院文献情报中心馆务委员、数据资源部主任钱力在“构建科技文献知识底座,支持PubScholar公益学术平台”的报告中,提出新时期科技文献数据资源的五大新需求:多场景、响应及时的高质量基础科技文献数据资源,打造科技文献数据湖,解决数据孤岛的问题;多层次、细粒度的垂直化领域本体知识资源,解决知识推理、知识预测与知识生成的可信知识供应问题;AI就绪的高质量、标准化、权威可信的数据资源,解决当前大语言模型高质量语料不足问题;通用与专业领域大语言模型,按需生成知识资源,解决现有数据库知识量供给不足问题;多场景智能化知识引擎,解决复杂科学问题需求对敏捷数据资源的需求问题。并提出四维(数据、工具、模型、引擎)知识底座体系以及五层次(基础数据、本体知识、AI就绪数据、

大模型、场景化引擎)知识底座建设方法,介绍了PubScholar公益学术平台的实践。他表示,图书馆要面向新时代打造新质生产力场景,包括明确科技文献采购原则、转变科技文献采购思路、加强科技文献采购权益以及升级科技文献采购模式等。并在此基础上提出包括深入数据内容挖掘、学会认识数据价值、掌握专业领域数据、前瞻创新数据产品、研发掌握智能工具、积极融入数据生态的六大科技文献数据资源建设核心能力。

## 2.3 规范引进数据资源,助推本地化服务进程

DRAA是高校图书馆引进数据资源的核心渠道,也是推进数据资源本地化服务的重要力量,在集团采购谈判中将数据资源在国内的本地保存作为一项基本要求,积极支持和推动数据资源国家长期保存、联盟公共保存和成员馆保存及其本地化服务。CALIS管理中心副主任、DRAA副理事长、北京大学图书馆副馆长姚晓霞介绍了2023年度高校引进资源集团采购发展现状和规范化工作进展,包括集团采购数据库统计分析,按照《研究生教育学科专业目录(2022年)》更新数据库资源百科信息和使用统计指标,提升集团采购上线率,开展引进资源综合评价工作,尝试引进数字资源统计数据可视化等多个方面。解读了《高校数字资源采购联盟工作规范(2018修订)》对引进资源长期保存的明确要求:所购买数据库应有长期保存权,可采取国家长期保存、联盟公共保存或成员馆自行保存的方式,合同中应对产品的长期保存和永久使用权进行明确说明。《DRAA数据库集团采购方案大纲(2023)》再次更新并加强了长期保存与永久使用权方面的要求力度,明确了租用模式不能作为无法提供中国本土长期保存的理由,不断推进引进数据资源的中国本土长期保存解决方案。还介绍了DRAA和数据库商推进数据资源本地化服务的实践案例,如Nature、IOP、RSC、SAGE电子期刊、Brill电子书等。最后发布了2023年度引进数据库用户满意度调查的基本情况调查结果,为DRAA、数据库商、代理商、成员馆优化资源与服务提供重要参考。

## 2.4 推动数据多源融合,提升联盟服务效能

北京大学图书馆项目管理中心副主任冯英在“推动多源数据融合 释放联盟数据价值”的报告中阐述了数据资源、图书馆数据资源、图书馆联盟数据资源的概念,强调图书馆数据资源的重要作用。她表示,当前联盟数据资源以文献资源元数据、对象数据、服



务数据和公共数据为主要构成,联盟提供的数据服务主要包括资源与数据共享、资源发现与获取以及数字资源集团采购。并提出未来联盟将从三个方面创新联盟数据服务:一是助力图书馆馆藏建设,包括馆藏对比分析服务、馆藏质量分析与评估;二是进一步丰富联盟数据服务,包括丰富数据建设类型、扩大数据采集范围,对数据进行深度处理、智能组织,实现数据服务平台化;三是助力提升图书馆服务,包括制作资源画像、支持个性化、精准化服务。此外,联盟也将继续推进与数据资源服务商的合作,综合运用长期保存、本地保存、镜像保存等形式,根据用户需求进行本地化的组织管理,提供本地化服务,提高数据容灾能力。同时联盟还将拓展数据合作,一方面与行业上游合作,基于资源订购与使用数据提供分析数据,帮助出版社深入了解用户需求;另一方面,与博物馆、档案馆等合作,实现元数据的拓展。此外,她还介绍了CALIS外文期刊网与本地库的数据融合案例以及CALIS与CASHL的数据整合实践。

### 3 加强数据开发利用,推进图书馆数字化转型

#### 3.1 深化内涵建设,重塑服务体系

随着当今社会信息技术快速发展,高校图书馆作为知识传播和服务的重要机构,在适应和应对信息时代的变化上面临着许多新的机遇与挑战。制定适应教育数字化发展需要和服务本校事业发展目标的图书馆规程,是推动高校图书馆工作与时俱进的制度保障。天津大学图书馆馆长傅平青在报告中介绍了天津大学图书馆以用户需求为导向打造的智慧图书馆服务体系,该体系包括信息资源保障中心、学术支持服务中心、科技创新体验中心、自主研修学习空间和文化传播交流平台;并从加强先进技术研发应用、构建多元立体馆藏体系、促进资源空间多重融合、拓展高质高效信息服务、建设校园文化社会文化五个方面,深入阐述了天津大学图书馆的探索与实践:聚焦学校学科建设发展思路以及学校一流本科教育、新工科建设,开辟发展新赛道,全力打造“五新”特色服务,建立数字资源长期保存机制,构建发展新格局。最后,他分享了天津大学图书馆在新的时期,聚焦大学图书馆战略规划研究制定的九大关键任务,擘画了新的发展蓝图。

#### 3.2 以平台建设推动专利数据本地化服务转型升级

在我国创新驱动的发展战略下,知识产权作为

战略性资源,日益成为国际竞争力的核心体现。《知识产权强国建设纲要(2021—2035年)》中指出要“打通知识产权创造、运用、保护、管理和服务全链条”<sup>[4]</sup>。高校图书馆开展专利信息本地化服务,对于满足高校师生创业与地方企业创新孵化需要,助力高校学科发展与成果转化等方面具有重要意义。深圳大学图书馆金银雪介绍了深圳市高校专利大数据本地化服务平台的建设、服务产品及内容的设计与“四个面向”(面向科技创新、面向技术转化、面向产业发展、面向公众服务)的专利信息本地化服务体系。该平台设有高校专利库、检索分析、可视化展示和资讯简报等模块,遵循“以科研需求为导向”“参与科研活动全流程”的原则,灵活组织和开展专利查询、查新、分析、评价等多样化、个性化、跟踪式的专利情报服务,从“创造—保护”到“运用—管理”,打通知识产权信息服务全链条,为提高专利的经济效益、社会效益和生态效益提供了重要保障。未来将从内容和功能方面进一步优化深圳市高校专利大数据平台,尝试开发专利查收查引、价值度报告等自动化生成和导出模块,建立能实现技术需求和专利技术智能匹配的供需对接平台,实现简报自动生成等功能,以提供更加高效的专利信息服务。

#### 3.3 基于数据中台的图书馆数据资源本地化服务

数据资源的汇聚、管理与服务是图书馆应承担的职责<sup>[5]</sup>。2021年7月,教育部等六部门发布《关于推进教育新型基础设施建设构建高质量教育支撑体系的指导意见》,提出构建新型数据中心、促进教育数据应用。一个开放共享、技术集成的智能综合治理平台是图书馆服务数字化转型的重要途径。中南大学图书馆罗孟儒等认为,数据资源本地化服务是对图书馆所拥有或控制的数据资源进行集成管理、分析挖掘后,提供具有高附加值的数据资源产品与服务的过程。通过调研“C9联盟”高校图书馆和湖南省“双一流”高校图书馆的业务系统及其产生的数据特征,指出数据资源本地化服务目前面临数据资源非本地化、数据获取成本高、数据资源异构分散、组织利用困难、缺乏相关管理架构、数据价值转换功能弱等问题。进而提出基于数据中台的图书馆数据资源本地化服务架构体系,即包含业务数据化、数据资产化、资产服务化、服务业务化四大过程和数据规则、数据采集、数据集成与存储、数据开发、数据运营、数据发布与共享、数据再利用、数据反馈与评估八个阶段



在内的基于数据生命周期的本地化服务模型,以推动图书馆掌握数据权属、提升数据价值、增强价值转换。最后,他介绍了中南大学图书馆数据中台本地化服务建设案例,并提出将 AI 技术与本地化数据资源结合,加速图书馆智能服务的优化和拓展。

### 3.4 数据资源本地化服务的国际经验与启示

数据资源的本地化服务是全球关注的重点,美国、英国、澳大利亚等国纷纷建立国家级数据资源长期保存机制,学术图书馆界和出版界也合作建立了 LOCKSS、CLOCKSS、Portico 等典型项目,高校图书馆也开展一系列实践项目,这些经验有一定的借鉴意义,武汉大学图书馆杨蕾等选择 2022—2023 年度 U.S. News 排名前十的顶尖大学为样本,重点考察其大学图书馆的数字保存与服务政策规划、平台基础设施建设以及系列增值服务开展情况。调研发现,除斯坦福大学图书馆、约翰霍普金斯大学图书馆外的 8 所高校图书馆在最新的图书馆战略规划中均提及了完善数字基础设施、优化资源发现获取、嵌入科研生命周期、推进开放学术研究等数字资源长期保存与服务战略方向。继而分析了 10 所图书馆商业数字资源和包括数字特藏资源、研究数据资源在内的自建数字资源的本地化建设实践,总结出确保商业数字资源的本地持续访问、提升数字特藏资源的可用性与影响力、开展科研数据全生命周期管理服务等若干经验。建议国内高校图书馆抓住教育数字化转型的发展趋势,从营造良好的数据资源建设环境、聚焦主流本地化保存策略、提升以用户为中心的服务体验、打造开放数据资源保障体系等方面学习国外经验,推进数据资源的本地化建设与服务创新。

## 4 多方联动,促进引进数据资源本地化保存

数据资源已逐渐成为支撑高校教学科研的基石,2022 年高校图书馆电子资源购置费占馆均文献资源购置费的 66.5%(2021 年是 64.7%)<sup>[6]</sup>。研讨会同时邀请了全球的数据资源服务商代表,包括 EBSCO、Springer Nature、科睿唯安、Wiley、ACS、Taylor & Francis、Elsevier、DeGruyter、Emerald、剑桥大学出版社、牛津大学出版社、CAS、IOP,带来数据资源建设的新方向和新思考。部分数据库商提出了数据资源本地化服务的新视角。Taylor & Francis 中国区期刊及电子产品客户经理石晓璟介绍了公司本地化服务实践情况,包括与北京大学图

书馆、国家科技图书文献中心(NSTL)签署 NDPP 长期保存协议,与本地代理商中国教育图书进出口有限公司、中国图书进出口(集团)有限公司合作实现电子书本地服务,并从数据传输的技术层面分析实现本地化服务的方法及解决方案。英国物理学会 IOP 出版社中国区经理徐涛分享了其本地服务的案例,IOP 定期向 CALIS 提供期刊数据,以确保国内高校用户对于 IOP 电子期刊永久访问权,并与中国科学院文献情报中心签署 NDPP 长期保存协议,不断创新平台功能提升用户体验。ACS 美国化学会出版部中国区商务拓展经理汪洋介绍 ACS 是 NDPP 合作保存机构之一,以 Dark Archive 的方式提供数据保存服务,为本地用户提供学术资讯、学术交流与培训等服务。

## 5 结语

“CALIS 引进数据库培训周”是中国高校图书馆资源建设领域的大型专业性研讨培训活动,从 2002 年迄今已经举办了二十二届。CALIS 管理中心和 DRAA 理事会组织图书情报领域的资深专家、数据库商和高校图书馆界代表共同交流,探讨高校图书馆数据资源建设的热点问题和行业发展策略,致力于完善引进资源服务体系,为用户提供丰富、优质的数字学术资源和服务。本届研讨会紧扣国家数字强国和教育数字化战略,聚焦数据资源本地化服务核心议题,数据资源的需求方、供给方、服务方代表充分交流,在推动数据资源本地化服务的共识下各抒己见,带来了数据资源本地化服务的顶层设计、优秀实践和最新国际趋势,共同探讨了未来的发展路径,推动建立和深化了多元国际合作,为高校图书馆在推进数据要素发挥乘数效应和教育数字化建设等领域的高质量发展提供理论启发和路径参考。

## 参考文献

- 1 新华社. 中共中央、国务院发布《关于构建更加完善的要素市场化配置体制机制的意见》[EB/OL]. [2024-05-17]. [https://www.gov.cn/zhengce/2020-04/09/content\\_5500622.htm](https://www.gov.cn/zhengce/2020-04/09/content_5500622.htm).
- 2 国家数据局. 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]. [2024-05-17]. <https://mp.weixin.qq.com/s/YyhLQo4IZIFNMiyupdvO1A>
- 3 国家数据局. 深入学习贯彻习近平总书记关于数据发展和安全的重要论述 奋力开创数据工作新局面——2024年全国数据工作会议在京召开[EB/OL]. [2024-05-17]. <https://mp.weixin.qq.com/s/diEmFKPROI0nKNLPDZT3dg>.



- 4 国务院. 中共中央、国务院印发《知识产权强国建设纲要(2021—2035年)》[EB/OL]. [2024-05-17]. [https://www.gov.cn/gongbao/content/2021/content\\_5643253.htm](https://www.gov.cn/gongbao/content/2021/content_5643253.htm).
- 5 陈建龙,邵燕,张慧丽,等. 大学图书馆现代化指南针报告[J]. 大学图书馆学报, 2022, 40(1): 22-33.
- 6 吴汉华,王波. 2022年中国高校图书馆基本统计数据[J]. 大学图书馆学报, 2023, 41(6): 63-72.

作者单位: 吴亚平、姚晓霞、任国华, 北京大学图书馆, 北京, 100871

李婉君, 天津大学图书馆, 天津, 30007

收稿日期: 2024年5月18日

修回日期: 2024年5月19日

(责任编辑: 支娟)

## Promote Data Localization Service and Build a New Pattern of Information Services

——Summary of the 2024 CALIS Symposium on Data Localization Service

WU Yaping YAO Xiaoxia REN Guohua LI Wanjun

**Abstract:** The Symposium on Data Localization Service & 22nd CALIS Database Workshop was held in Tianjin of China during May 14 – 17 in 2024 attended by 452 delegates from libraries, academies and digital resource vendors from around the world. During the seminars held on May 16 and 17, it focused on the theme of “Data Localization Service” to share practices and experiences concerning the current services and prospects of data localization. The article summarizes the major discussions of the symposium with spotlights, opinions and practical experiences shared by the delegates and advocates the data localization service in the library community for further development. Spotlights during the symposium are summarized as followed. (1) In the keynote speech delivered by Prof. Chen Jianlong, Director of DRAA and Chief Librarian of Peking University Library (PUL), he emphasized the importance of strengthening data access capabilities and consolidating the foundation of data services according to the Digital China Initiative and the advent of artificial intelligence. He took the National Digital Preservation Program (NDPP), the CALIS localization service platform and the practice of PUL as examples to share data localization services. (2) Prof. Qian Li, from National Science Library, Chinese Academy of Sciences, put forward five demands of scientific and technological data in the new era, including data silos, the lack of reliable data, insufficient high-quality corpus of Large Language Models and the shortage of knowledge supply, etc. (3) Prof. Yao Xiaoxia, Deputy Director of DRAA introduced the current development and standardization of the consortium acquisition of DRAA in 2023, including data analysis, acquisition evaluation, visualization of the digital resource statistics and so on. Librarians from Tianjin University, Shenzhen University, Central South University and Wuhan University shared their experiences and prospects of data localization services. (1) To reshape the service pattern oriented towards user demands and transform the academic libraries as information resource hub, academic support and innovation center, cultural communication platform. (2) To improve patent big data localization service with platforms providing services from four aspects including scientific and technological innovation, technology transformation, industrial development and public demands. (3) To construct data middle platform based on the data life cycle so as to promote the value-added services. Delegates from EBSCO, Springer Nature, Clarivate, Wiley and other vendors and presses brought up new perspectives on data resource construction. As a nationwide professional seminar and training workshop, the CALIS Symposium has been held for 22 consecutive sessions since 2002. The symposium this year provided theoretical inspirations and guidance for the high-quality development of academic libraries in promoting the multiplier effect of data elements and the construction of digital education.

**Keywords:** Academic Libraries; Data; Data Localization Service; Long-term Preservation