

数据驱动环境下档案信息资源的社会化协同利用机制探析

范立娟

(吉林省白山市健康教育所 吉林白山 134300)

【摘要】当今世界，我们正在经历一场以大数据为动力的大变革，大数据已经是人类社会创新和发展的重要动力。在这样的大环境中，我国的档案工作既迎来了新的发展契机，也迎来了新的挑战。在大数据时代，文档资源深度挖掘、动态集成和价值释放等需求已很难满足。我国档案信息资源是我国的一项重大战略任务，其发展与使用模式急需向开放化、协作化和社会化方向发展。研究构建“社会协作使用”的新模式，既是对我国档案工作本身发展的要求，又是服务于国家战略，服务于社会治理，推动文化繁荣的重要途径。

【关键词】大数据；档案资源；社会化共享；协同管理；数字化转型

【中图分类号】G279

Exploration of the Social Collaborative Utilization Mechanism of Archival Information Resources in a Data Driven Environment
Fan Lijuan

(Health Education Institute of Baishan City, Jilin Province, China 134300)

[Abstract] In today's world, we are experiencing a major transformation driven by big data, which has become an important driving force for innovation and development in human society. In such a big environment, China's archival work has not only ushered in new opportunities for development, but also new challenges. In the era of big data, it is difficult to meet the needs of deep mining, dynamic integration, and value release of document resources. The archival information resources in our country are a major strategic task, and their development and usage patterns urgently need to move towards openness, collaboration, and socialization. Studying and constructing a new model of "social collaborative use" is not only a requirement for the development of China's archival work itself, but also an important way to serve national strategies, social governance, and promote cultural prosperity.

[Key words] big data；Archive resources；Socialized sharing；Collaborative management；digital transformation

引言

根据 CNNIC2024 年公布的数字，到 2025 年，中国的数据产出有望超过 50 ZB，约占全世界的 30%。同时，根据《2024 年中国数字经济发展报告》，预计到 2025 年，我国的 GDP 增速将超过 20%。近几年在全国各地建立的“档案馆信息公开平台”的应用证明，通过“资源共享”可以有效提高“档案馆资源共享”的精度与覆盖范围。比如，在一些省份和地区，通过建立跨部门文件共享机制，将检索效率提高 40% 以上，这就是一个很好的例证。如何构建高效有序的社会化协作资源管理体系，充分发挥其应用价值，是当前迫切需要解决的问题。

1 数据驱动环境下档案信息资源的社会化协同利用机制探析的意义

1.1 激活数据要素价值，赋能经济社会发展
网络的新型公共服务平台，即基于“共享式”的新型服

务模式。据工信部《中国数据要素市场发展报告》预测，到 2025 年，中国的数据资源产业规模有望超过两千亿，平均每年以 24.6% 的速度增长。档案信息资源是我国最大的一笔公共数据资产，其深入发展与应用对于提高我国的生产力水平有着重大意义。从实现的角度来看，要构建一套完备的档案资料管理和利用系统。研究内容为：利用区块链的分布式记账方法，建立文件数据资产注册平台，对各类文件资源进行独立的电子识别，以达到对文件资源的确认和可追溯性。以上海市档案馆的“城市记忆”知识库为案例，已经创建了 500 多万个实体、2000 多万条文档的海量文档知识地图。通过建立规范的 API 界面及微服务框架，实现档案信息服务与政务服务、科研创新、文化交流等多个领域的无缝对接^[1]。

1.2 创新文化传播模式，增强文化自信力量

中国互联网络（CNNIC）于 2025 年 1 号公布的第 55 次《中国互联网络发展状况统计报告》显示，全国数字文化产业的总资产已经达到了 1.2 万亿人民币，而对这些数据的利用比例也在不断提高。在科技革新的层次上，要建立融贯互动的档案文化交流系统。主要研究内容为：利用高精度数字获取方法，对馆藏古籍进行毫米级精确的立体模型与光谱获

取,例如故宫博物院建成的“数字收藏室”,已经完成68000余个馆藏文件的超高清化处理;利用虚拟现实(VR/AR/MR)等虚拟现实技术研发档案文化体验产品,例如由中国一院主办的《明清档案浸入式展览》,利用MR/AR/MR等科技手段,使参观者能够与史料进行天然的互动;以历史文献为研究对象,以人工智能为主要研究手段,采用计算机辅助的方法,利用计算机辅助的方法,研究文档文本的智能化产生与推荐方法^[2]。

1.3 提升社会治理效能,服务国家战略需求

我国的档案信息资源社会化协作使用机理是提高我国社会管理效率的有效途径,为我国的重大战略需要提供有力支持。从技术实施方面来说,构建一个智能的文档数据分析和服务平台是非常必要的。主要研究内容为:以广东省“粤政图”为例,以27个省厅为例,开展了多源文档数据的联合分析研究;利用时间序列分析、社会网络分析等方法,从历史文献中发掘出具有一定规律的认知与发展趋向。在此基础上,以长江流域为例,以“历史记录”和“即时”信息为基础,构建基于“历史”和“动态”信息的“多维”信息集成的“多维”模型。江苏省政务服务中心的数据表明,到2025年,通过对全市98个智慧城市的应用情景的支持,降低了75%的重复提交资料,将办事的速度提升了50%以上,为实现我国的“数字化”和“治理的现代化”奠定了良好的基础。



图1 档案资源价值最大化

2 大数据环境的档案资源开放共享模式研究中存在的问题

2.1 数据标准化与互操作性不足

在大数据背景下,我国档案信息资源的规范化和互用性一直是制约其发展的瓶颈。2021年,中国档案局将公布《全国档案信息化建设调查报告》,其中17个以上的文件使用了

17种以上的元数据规范,28种以上的数据库体系,这给跨部门之间的数据交互带来了很大的困难。从技术角度看,上述问题主要体现为:各省市、省、市、县局分别采用DC、EAD、MARC等多种标准,即使是同一片区域内,不同单位所采用的标准也不尽相同^[3]。

2.2 数据安全与隐私保护挑战

随着我国信息化建设进程的加快,我国在信息化建设进程中面临着越来越多的信息安全和信息保密问题。根据《全国安全隐患分享平台》(CNVD)发布的2021年度报告,与文档信息系统有关的安全隐患较上年同期增加了32%,风险隐患占17%。从技术角度看,主要是由于文档信息的传输安全性不高,导致文档信息的跨网传递有被盗的危险。目前多数文件都是通过SSL/TLS进行数据加密,但是其加密程度高低不一,一些县档案馆仍然沿用TLS1.1等失效的通信协议。目前对海量文档数据库的存取控制与审核还不够完善。2024年,美国全国计算机安全检测报告显示,文档中有40%以上的数据库有漏洞,28%的文档没有配置防火墙。

2.3 技术架构与基础设施滞后

现有的文件资源开放共享体系结构和体系结构远远落后于现实需要,已经成为一个“瓶颈”。根据工信部2022年发布的统计,目前我国档案管理体系中,计算机网络老化率已达38%,5年以上没有更新的网络设备高达45%。从技术体系结构上看,大部分文档仍然是以传统的分布式体系结构为基础,很难满足大数据时代下的高并发性、大规模化的要求。全国各地区的馆藏数据总量只有2.3 petab,而全国的数据总量却以每年0.8 petab的速度增长,其数据量的增长速度已经达到了0.8 petab,这给我国的数据环境带来了很大的挑战。当前,由于GPU等多核处理器的配置比例小于15%,使得很多AI应用难以部署。例如,受计算机资源的制约,全国各地方档案馆平均每天只能处理30,000张文件,而全国共有200多万张纸质文件,要实现全数字化标识至少要18年。我国档案信息系统建设的薄弱环节较多,档案信息系统的上网速率一般都在1 Gbps以下,而部门间的专用线路则不到30%^[4]。

3 大数据环境的档案资源开放共享模式的有效研究策略

3.1 建立统一标准体系与增强系统互操作性

构建大数据环境下档案资源开放共享模式,即:1)建立大数据背景下的档案信息资源的开放共享。按照《数字中国标准体系建设指南》的要求,互联互通的标准系统是否完备,将对数字资源的使用产生重要的作用,而规范的建立可

以减少交易费用超过 60%。从现实的角度看,要以系统的方式来推动科技创新的发展。首先,根据 ISO 23081 《信息与文献—文件管理流程—元数据原则》的国际标准,制订《中国档案数字资源描述规范》,并根据国内的具体国情,建立国家统一的文档元数据标准。此标准要求定义至少 38 个元资料要素,包括识别符、创建者、时间、主题等必备要素,并对扩充后的元资料进行了说明。在上海市档案馆开展工作,力争在 2024 年前将实现多个部门间的信息互通,提高 75%。在技术实现层次上,应构建多层次的标准实施系统。其中,在数据层次上,使用 RDF 三元数据模式,使其具有可识别性;在交换层上,使用 JSON-LD 的方法实现了对数据的串行化处理,实现了嵌入式的数据结构;在界面方面,采用了 OAI-PMH3.0 技术规格,为元数据获取提供了一个统一的界面。在此基础上,提出了一种新的元数据映射方法,即通过 XSLT3.0 来完成从 EAD 向统一规范的转化;构建基于语法检查、语义检查和商业规则检查的标准化一致性检查平台,实现对数据的自动检验。浙江省档案馆在到 2025 年完成的标准化项目结果显示,在使用了自动设备后,可以将标准的执行时间减少到 6 个月。

3.2 构建多层次安全防护与隐私保护体系

大数据环境下档案信息资源社会化协同利用是促进公共服务和信息共享的一条重要途径。然而,随着数据开放程度的不断提升,信息安全和隐私问题也日益凸显,亟须建立系统的、多层次的安全保护体系。该系统需要涵盖数据采集、传输、储存、处理和销毁等整个生命周期,并结合技术手段和管理机制,实现全方位的防控。从技术上来说,可以通过加密算法和匿名技术来预处理敏感信息。如借鉴医疗健康领域的电子健康档案保护实践,采用公钥基础设施 PKI (PKI) 技术,实现用户的身份认证和访问控制。在数据存储阶段,可参照 GB/T 35273-2020 《信息安全技术个人信息安全规范》(GB/T 35273-2020) 对档案信息进行分类,并部署对敏感字段(如身份证号、病历等)的数据减敏机制。

参考文献:

- [1] 孙燕. 基于大数据的企业档案信息资源整合与共享机制研究[J]. 兰台内外, 2025, (06): 27-29.
- [2] 梁苑曼, 黄丽君. 大数据环境下档案资源数据治理体系的构建[J]. 太原城市职业技术学院学报, 2024, (10): 205-207.
- [3] 土红莉, 韩乔乔. 浅析大数据环境下档案资源数据治理体系构建[J]. 陕西档案, 2024, (01): 28-29, 24.
- [4] 蒋洪杰, 刘云朋. 大数据环境下政务数据共享机制的构建研究[J]. 科技创新与生产力, 2022, (01): 88-90, 93.
- [5] 杨博文. 数字档案资源共享机制构建研究[J]. 办公室业务, 2024, (16): 97-99.

作者简介: 范立娟(1970-),女,汉族,吉林白山人,研究生,研究馆员,研究方向:档案管理与应用。

3.3 升级技术架构与建设新型基础设施

随着大数据技术的深入应用,档案信息资源社会化协同利用成为推动公共服务创新的重要途径。为构建高效安全的档案资源共享系统,迫切需要对现有技术体系进行系统性升级,构建与之相适应的新体系架构。以医疗健康为例,电子健康档案跨机构查询和区域医疗信息平台建设对数据融合能力、交互规范和计算效率提出了更高的要求。在技术结构层次上,引入云原生与微服务的设计思想,对传统的集中式档案管理系统进行重构,使之成为一种松散耦合的分布式服务系统。首先,利用 Hadoop、Spark 等技术,构建基于 Hadoop 和 Spark 的数据湖,实现文档结构化和非结构化的统一存储。例如,某省级档案馆利用 HDFS 集群,实现 500 TB 以上病历数据的数字化入库;依托 ApacheNiFi,构建每天处理近 10 万份病历的数据管线。第二,在数据整合方面,需要按照国家标准《医疗卫生信息共享文档规范》(GB/T 35648-2017),采用医疗行业通用数据交换 (HL7 FHIR) 标准,实现多源异构文档资源的标准化映射和语义标记,保证跨系统协作过程中数据的一致性和互操作。

4 结语

综上所述,构建数据驱动环境下的档案信息资源社会化协同利用机制,为我国有效地盘活国家数据资产,提升公共服务水平,提升社会协同效率提供重要支撑。这一新的工作模式,既可以促进我国档案工作由单纯的“保存”,又可以促进我国的“知识”与“智能”的转变,更好地为我国的“数字化治理”和“数字文化”提供重要的支持。今后,要加强顶层设计,打破体制、技术的障碍,并在此基础上形成一个良性的、多方参与、价值创造的生态。通过不断深入的体制创新和实践探索,使我国的档案事业在“数字化中国”的过程中获得更大的发展,为我国的社会发展、人类的文化遗产做出了巨大的、无法取代的贡献。