

Construction of archive resource integration and sharing mechanism in the era of big data

Bin Liu

Organizational Structure Research Center of the Office of the Organizational Structure Committee of the Hanshan District Committee of the Communist Party of China, Handan, Hebei, 056001, China

Abstract

The advent of the big data era has created unprecedented opportunities and challenges for archive work. Archive resources are massive, diverse, and growing rapidly. How to efficiently integrate and deeply share them, explore potential value, has become an urgent problem to be solved. This article focuses on the background of big data, deeply explores the new connotation of archive resource integration and sharing, and analyzes the urgency of building a new mechanism. Based on this condition, explore the organization and collection of archival resources from the resource dimension; Analyze the application of relevant technologies from a technical perspective; From the perspective of management, explore management strategies and paradigms, systematically explain the practical ways of building a mechanism for integrating and sharing archival resources, and provide theoretical reference and practical guidance to help the archival industry develop towards dataization, knowledge-based, and intelligent direction.

Keywords

big data; Archive resources; Resource integration; Sharing mechanism; data governance

大数据时代档案资源整合与共享机制构建

刘彬

中共邯郸市邯山区委机构编制委员会办公室机构编制研究中心, 中国·河北 邯郸 056001

摘要

大数据时代的降临,为档案工作创造了前所未有的契机与难题,档案资源体量庞大、样式繁多且增长迅猛,怎样对其开展高效整合与深度共享,发掘潜在价值,成了亟需解决的难题,本文围绕大数据背景,深度探究档案资源整合与共享的全新内涵,分析构建新型机制的紧迫性。基于此条件,从资源维度,探究档案资源的整理与归集;从技术维度,剖析相关技术的运用;从管理维度,探究管理策略与范式,体系化地说明构建档案资源整合与共享机制的实践方式,为助力档案事业朝着数据化、知识化、智慧化方向发展给予理论参考与实践指引。

关键词

大数据; 档案资源; 资源整合; 共享机制; 数据治理

1 引言

档案作为社会活动的原始凭据,是国家和社会演进的宝贵信息资产。伴随信息技术革命的持续推进,人类社会已步入大数据时代,数据成为关键生产资源。在该背景情形下,档案资源的形式由传统的纸质载体快速拓展至电子文件、数据库、多媒体、网页档案等多样形式,其规模呈指数型增长。以往的、以实体管控为核心的档案工作模式,已无法满足海量、异构、动态档案数据的处理、分析和价值挖掘要求,破除“信息孤岛”,促使档案资源由分散保存迈向系统集成。从封闭私用走向开放互享,搭建契合大数据时代需求的集成

与共用机制,不只是档案工作自身发展的内在必要条件,更是助力国家治理体系和治理能力现代化、助力经济社会发展的战略抉择。

2 资源层整合: 夯实档案共享的数据基石

资源层整合是搭建共享机制的前提与基石,其重点在于把分散、异构的档案数据资源进行系统性聚合、标准化操作与规范化整理,形成逻辑一致、质量可管的数据集合。

2.1 档案资源的多源汇聚与全面采集

搭建整合与共享机制的首要环节是达成档案资源的广泛聚集。这需要冲破机构、地域和系统的壁垒,不仅要传统馆藏纸质档案实施数字化转型,还要强化对各类电子文件、业务系统数据、社交媒体信息、音视频档案等原生数字资源的整体采集与存档。通过设定统一的归档范畴与准则,保障有意义的档案数据能够被全面、准时地采集并归入统一

【作者简介】刘彬(1989-),男,中国河北邯郸人,助理档案管理员,从事档案管理研究。

治理架构，为后续的融合与分享筑牢丰富的资源根基。

2.2 档案数据的标准化与规范化处理

若海量异构数据未进行标准化处理，便无法达成有效的互联互通。必须构建一套包含数据格式、元数据、分类编码、目录体系等方面的统一标准准则，对源自不同源头、具备不同结构的档案数据开展清洗、转换、标引与格式化工作，让其符合共同的数据规范^[1]。这一过程是解决数据语义含混、提高数据水准、维持数据可理解度与可交互性的关键步骤，是达成跨库检索与深度共享的技术支撑。

2.3 档案知识体系的构建与关联组织

以数据标准化实现为前提，应当进一步向知识化范畴进行深度整合，借助知识图谱、本体建模等技术手段，对档案资料进行深度标鉴和知识萃取，识别并确立档案实体（如人物、事件、地点、机构）之间的内在关联。借助构建结构化的档案知识系统，将原本离散的档案文件演变成相互联系的知识结点，进而实现对档案资源的语义化整理与智能化关联，为用户的精准查找和知识发掘提供坚实支撑。

3 技术层支撑：打造档案共享的智能引擎

技术是推动档案资源整合与共享的关键力量，大数据、人工智能、云计算等新一代信息技术为打造高效、智能、安全的共享平台提供了有力的支撑。

3.1 大数据存储与计算平台的建设

面对持续增长的巨量非结构化档案数据，传统的集中式存储和单体计算架构在存储容量、I/O吞吐以及并行处理能力上已呈现出显著的力有不逮。为应对数据规模从TB级朝PB级的不断扩大，有必要搭建具备横向拓展能力的存储基础设施，依托分布式文件系统或对象存储云平台的解决方案，能够以灵活、可延展的方式切实管理海量档案文件。此类架构不仅依靠数据多副本机制维护了存储安全性，还借助并行数据访问途径大幅提高了数据读写效能，为上层应用打造了坚实可靠的数据根基。

基于雄厚的存储根基，引入以Hadoop、Spark为代表的分布式计算框架，是达成高效数据处理的关键。这些框架可把海量的档案数据集拆分成多个数据单元，同时安排多个计算节点一同执行批量处理任务，由此将复杂的分析任务转化成高效的并行计算过程^[2]。无论是历史档案的批量格式转变、全文检索索引的创建，还是运用机器学习算法实施非结构化内容的智能辨别与分类，均可凭借这些框架得到强大的算力支撑。这种存储与计算协同拓展的架构体系，最终保障了在超大规模档案数据情形下，数据处理任务的可操控性与即时性。

3.2 智能处理与深度分析技术的应用

把人工智能技术深度融入档案管理环节，是达成档案信息资源从消极保管向积极赋能升级的关键办法。凭借自然语言处理技术，系统可自动剖析文本档案的语义信息，达成

精准的实体甄别、智能分档与关键词标注，且进一步产出内容概要，进而把人力从繁杂的基础编目事务中解脱出来。利用先进的图像识别算法，能够针对历史老照片、工程图纸等非文本类档案开展视觉元素剖析与内容解读，自动达成关键信息的提取与结构化编目，显著提升了档案内容的可检索性与可利用性。

基于此条件，借助数据挖掘与机器学习手段对海量档案数据开展深度剖析，能够从看似毫无关联的历史记载里发掘潜藏的关联联系、发展趋向与潜在范式。这种深度层面的智能剖析，不但极大提高了档案整理与描述的自动化程度，更关键的是把静态存放的“死档案”转变为体系化、结构化的“活知识”。这种知识化的变革，为搭建高品质的档案知识库、开展深度专题编研与个性化知识推送等高端知识服务筑牢了稳固根基，进而大幅拓宽了档案资源的服务范畴与价值水准。

3.3 统一共享服务平台与接口的设计

档案资源价值共享的最终达成，在很大程度上依靠一个用户亲和且功能卓越的集成化服务平台。该平台应当搭建成为一个整合了档案查询、线上阅览、资料下载、可视化呈现以及标准化API接口服务的综合性门户，其要点在于助力对分布式、异构数据库的跨库“一站式”查找，让用户可以借助单一接口和查询命令，穿透式地访问处于不同物理地点或结构不同的档案数据资源，从根源上消除了信息壁垒，为用户获取信息赋予了前所未有的便捷感。

基于基础检索，平台还应提供多种类型的数据可视化手段，诸如时间轴、关系图谱抑或是地理信息映射等，把抽象的档案信息转变为直观的图形界面，助力用户开展深度剖析与知识探寻^[3]。平台需要以明晰、公开的授权协议为依托，凭借一套设计精良、文档完备的标准化API接口，向政府部门、研究机构、社会公众等不同用户群体提供粒度和权限不同的差异化数据服务。这种集易用性、开放性与安全性于一体的服务架构，最终保证了档案资源在安全可掌控的条件下，能够达成高效、便利且广泛的价值流通与知识交互。

4 管理层保障：构建档案共享的制度生态

完备的管理架构是保证档案资源整合与共享机制可实现持续、安全、标准运行的基本保障，涉及政策法规、安全伦理和组织协调等多个层面。

4.1 政策法规与标准体系的协同完善

在国家与行业维度，急需加大顶层谋划与战略协同力度，颁布专门针对大数据环境下档案资源整合与共享的综合性政策法令。此类法规需明确划分数据所有者、管理者和使用者在资源共同建设与共享进程中的权利、义务及利益界限，为跨系统、跨层级的档案数据流转与业务协作提供稳固的制度支撑。只有建立起权责明确的法律体系，才可以有效攻克因部门壁垒或利益差异引发的数据孤岛困境，引领各类

主体有序介入到国家档案资源体系的建设进程中。

应当迅速搭建与核心法规相适配的、成体系的标准系统，这包括但不限于数据采集与存档的格式规范、数据净化与整合的质量规范、数据存储与传送的安全保密规范，以及平台服务效能和用户满意度的绩效衡量规范^[4]。这一标准体系必须覆盖从数据生成、加工、存储到服务使用的全生命周期，打造一个逻辑紧密、相互配合的标准规范框架。该框架的设定，会给跨部门、跨区域的档案资源共享实践给予统一、清晰的操作指引与技术支撑，从根基上保障大规模数据聚合与交换过程中的质量连贯性、安全可监性与服务合规性。

4.2 数据安全与隐私保护的强化落实

在全力促进档案资源共享的过程中，应把安全与隐私保护摆在首位。这需要搭建一个包含物理环境、网络传输、数据本体以及应用接口的多层面、纵深型防护体系。对于计划开放的档案资源，务必严格落实专业且周密的鉴定审查机制，依照相关法律法规对涉及国家秘密、商业秘密与个人隐私的敏感信息进行精确辨认，进而在此基础上开展精细化的分级分类管理。针对确有必要共享却包含敏感内容的数据，应运用技术措施进行可靠的脱敏或匿名化处置，从起始点上把控隐私泄露与数据滥用的风险。

为进一步提升共享进程的安全可控度，可引入区块链等创新技术用以构建可信的审计追溯机制。借助将数据资产的授权访问、使用流转等关键环节信息以不可篡改的方式记录到链，可达成数据共享全生命周期的透明监管与事后追溯。这种依托技术赋能的管理模式，有力保障了在复杂多变的数据流通环境里，数据主权的归属始终清晰明确，各类用户的使用行径全程均可追溯，潜藏的安全隐患能够被迅速察觉并应对，从而在充分挖掘数据价值的同时，构建起牢固可靠的安全壁垒。

4.3 协同治理与人才培养机制的建立

档案资源的整合与共享作为一项牵涉多方主体、多种技术以及多重标准的复杂系统工程，其顺利推进绝非档案部门可独自完成。这迫切需要搭建一个由政府宏观引领、档案部门专业引领、各立档单位积极协作、并激励社会力量广泛融入的多元主体协同治理机制^[5]。在这一机制体系中，政府担当政策供给和跨部门协同的工作，档案部门聚焦于标准设定与平台治理，各立档单位肩负数据规范呈送与内部资源整

理的职责，而企业、研究机构等社会力量能够在技术研发和服务革新方面起到补充功效，由此形成职责明晰、优势互补的共建局面。

面对大数据、人工智能等新兴技术给档案工作造成的重大改变。现有人员队伍的知识架构与技能储备承受着严峻压力，需把复合型专业人才的培育与引入提高到战略层面，需借助体系化的在职培训和高校及科研机构实施联合培养，以及有目的地引入具备信息技术背景的专业人才等多样途径。着力打造一支既熟知档案管理核心业务与法规标准，又可灵活运用数据分析、智能处理与平台运维技术的专业化团队。这支队伍会成为整个协同体系达成长期稳定运转和持续革新的核心依托，为应对未来难题提供必不可少的人力保障与智慧根基。

5 总结

大数据时代为档案资源的开发运用拓展了广阔空间，然而也对传统的管理方式构成了严峻考验，搭建档案资源整合与共享的机制，属于一项牵涉资源、技术、管理等多层面的系统工程。应当把资源层整合当作基础，达成数据的规范与关联；以技术层面支撑作引擎，带动处理的智能与高效；以管理层保障作根基，维系共享的安全与可延续。仅当坚守系统思维，才可以突破阻碍，激活档案数据资源，最终搭建起一个开放、协作、智能的国家档案资源共享体系，让其在助力社会进步、弘扬人类文明中展现出前所未有的重大价值。

参考文献

- [1] 高丽丽.大数据时代档案信息资源整合与服务创新[J].中文科技期刊数据库(文摘版)社会科学, 2024(9):0089-0092.
- [2] 丁美香.大数据时代企业档案信息资源整合与利用[J].中文科技期刊数据库(全文版)经济管理, 2022(7):261-264.
- [3] 李进军.大数据时代档案数字资源整合的难点及对策探讨[J].中外交流, 2017, 000(018):8.
- [4] 丁训华.大数据时代高校数字档案资源的整合问题与策略研究[J].中文科技期刊数据库(文摘版)图书情报, 2024(8):0142-0145.
- [5] 何惠云.大数据时代企业档案信息资源整合与利用研究[J].兰台内外, 2021(24):2.