

大数据技术赋能事业单位档案工作研究

覃理玲

渠县政策研究中心 四川 达州 635200

【摘要】：大数据技术蓬勃发展为事业单位档案工作现代化转型开辟了新的途径，给事业单位档案工作现代化转型提供技术支持和方法论指导。本文对大数据技术在档案管理中的多方面应用价值进行了全面的论述，对大数据技术在数据处理效率的提高、信息资源智能检索、安全防护体系的加强、服务模式的创新等各方面的优势做了详细的分析。研究形成了以智能化平台创建、数据挖掘分析、服务流程再造为内容的综合技术赋能体系，对实施中会碰到的技术适配、数据治理、人才建设等主要问题给出了系统的解决办法。大数据技术深入运用之后，可以冲破传统档案管理方式中信息壁垒和利用阻碍，促使档案工作朝着数据驱动、智能协同、精准服务的方向发生深刻改变，给事业单位档案管理的数字化、智慧化转型赋予了既有理论前瞻性又具备实践操作性的解决办法。

【关键词】：大数据技术；事业单位；档案管理；数字化转型

DOI:10.12417/3041-0630.26.08.067

在数字化转型的宏观背景下，事业单位档案管理工作正迎来从传统模式向现代治理范式转型的重要契机。由于档案数据量持续增长、形态越来越多样、利用需求也日趋复杂，传统管理方式的不足就日益明显。大数据技术依靠自身的数据处理、分析能力，给破解目前的困境、重新塑造档案工作的格局提供关键支撑。不仅可以优化档案管理的基础流程，还可以对档案信息进行深层次的挖掘，从而给管理决策和服务创新赋予新动力。本文主要研究大数据技术对于档案工作各个环节的赋能机制，设计科学可行的实施路径，分析档案工作转型过程中遇到的关键问题及解决办法，给事业单位创建现代化档案管理体系、提升信息服务水平提供理论指导和实践方案。

1 大数据技术对档案管理工作的核心赋能价值

1.1 革新档案数据处理与组织模式

传统的档案管理一直采用人工的方式对档案数据进行著录、分类、编目等，处理速度慢、标准化程度低，不能适应大量的、多样的、异构的档案数据整合要求。大数据技术体系里的分布式存储框架（HDFS）以及并行计算引擎（Spark），给海量档案数据的处理提供底层架构上的支持^[1]。使用这些技术可以实现快速入库、集中存储数字化扫描文件、电子文档和多媒体档案等各种格式的数据。使用机器学习技术可实现档案内容的自动识别、智能分类和标签化标记，将档案管理人员从繁琐的低层次重复性工作中解脱出来，使档案管理人员把注意力集中到档案鉴定、开发、服务的工作上，从而大幅度提升整个档案数据的管理及智能化水平。

1.2 实现档案信息的精准检索与知识发现

档案检索是否方便，准确率如何，都会影响档案的利用。传统依靠关键字匹配的方式，查全率和查准率很难同时达到要求，尤其是对那些内容繁杂、关系错综复杂的档案资料，用户很难快速找到自己想要的信息。大数据技术把自然语言处理、语义分析、知识图谱创建等融合起来，可以很好地认识用户的检索意图以及档案内容的深层语义。系统可以实现智能语义检索、关联内容推荐、档案类别之间知识脉络整理。以档案实体和事件为基础创建领域知识图谱，系统就可以自动找到档案之间的潜在联系，并且以可视化的方式展示出来，给用户提供立体化、关联性的知识服务，大大提高档案资源的可发现性以及利用深度。

1.3 筑牢档案数据的安全防护与合规性屏障

大数据环境下档案安全的管理要由被动防御转为主动预警、全方位防护。简单来说，大数据技术能通过多种方式为档案数据打造一个从生成到销毁的、无死角的安全保护网。具体包括：用高强度的数据加密给档案信息上锁，确保即使数据被窃取也无法被轻易读取；通过精细的权限管理，严格控制不同人员只能访问其职责范围内的档案，防止越权查看；同时，系统会完整记录下每一次对档案的访问和操作，留下不可更改的“操作日记”，方便随时审计和追溯。这些手段环环相扣，共同构建起贯穿始终的安全防线。尤其是区块链技术的应用，可以在重要档案的元数据、流转过程信息中产生不可篡改的存证，从而保证档案的真实性、完整性。除此之外，利用大数据行为分析技术可以对档案的访问、操作行为进行实时监控，建立正常行为基线之后，可以识别异常的访问模式或者数据泄露

作者简介：覃理玲，出生年月：1987年2月，性别：女，民族：汉族，籍贯：四川，学历：本科，职称：馆员，研究方向：档案。

的风险,实现安全威胁的早期发现和快速响应,满足越来越严格的数据安全和隐私保护法规的要求。

2 大数据驱动档案管理转型升级的技术架构与实施路径

2.1 构建集成化、智能化的档案数据资源平台

推进档案管理数字化转型的第一步,就是创建一个可以汇集、管理、分析多种档案数据的中心平台。该平台应该采用云计算架构,具有弹性可扩展的计算和存储能力。核心是构建集成化的档案数据存储与服务平台,把来自各个业务系统、不同时期、不同格式(结构化、半结构化、非结构化)的档案数据统一摄入、集中存储,打破信息孤岛。平台要整合光学字符识别技术、自然语言处理技术和图像识别技术,做到纸质档案的数字化成果自动著录、自动提取内容。该平台用标准的数据服务接口,给上层各类档案应用(检索系统、分析系统)提供高质量的数据供给,形成数据底座、智能引擎、业务应用的分层解耦架构,提高系统的灵活性和可持续演进能力^[2]。

2.2 深化档案数据挖掘与分析应用

大量的档案资料当中存着可以体现出组织运作规律、业务发展趋向以及服务需求特点的诸多知识。用大数据分析技术可以从历史档案中找到有决策支持价值的信息。用时间序列分析可以预测出某种类型档案的查询与利用高峰,从而给存储资源的动态优化提供指导;用聚类分析可以发现用户群体以及其关注的档案主题,给个性化服务的推送提供依据;用关联规则挖掘可以发现不同的业务、不同的事件之间隐性的联系,帮助进行历史经验总结或者政策效果评估。将数据分析结果以可视化的仪表盘的方式呈现给管理者,可以由档案保管变成档案赋能,让档案部门由成本中心变成价值创造中心。

2.3 创新档案信息资源服务模式与体验

大数据技术应用最根本的目的就是改善用户体验,拓展服务模式。从服务模式来说,可以构建基于用户画像的个性化档案信息推送系统,达到信息找人的目的。在服务渠道上要发展移动化、场景化的档案服务,即开发档案移动应用,利用LBS技术,给外勤人员提供现场所需要的档案信息支持。更重要的是推动档案数据在合规的前提下实现开放共享,采用标准的API接口同单位内部OA系统、业务系统以及外部政务服务平台开展数据互通互联工作,使档案数据同各项业务流程有效融合起来,为跨部门协作提供支撑,从而改进行政效能和提升公共服务质量。

3 实施大数据档案管理系统的关键挑战与应对策略

3.1 技术选型适配性与实施成本效益平衡

事业单位在采用大数据技术的时候,要仔细考量自身的技术根基,数据体量以及业务诉求,防止出现技术“过度设计”或者“水土不服”的情况。采用分阶段、渐进式的实施策略,从业务痛点突出、数据基础好、见效快的场景入手进行试点,成功之后再逐步推广。技术选型要兼顾开源技术和成熟商业方案的优点,以功能、安全为前提控制好初始投资和长远运维费用。充分利用云服务的弹性以及按需付费的特点,可以有效地降低基础设施的投入门槛和运维难度,使得事业单位可以把更多的精力投入到档案业务应用中去。

3.2 档案数据治理与质量保障体系建设

高质量的数据是发挥大数据价值的前提。事业单位要创建起涵盖数据全生命周期的治理体系。首要的任务就是制定统一的档案数据标准,包括元数据标准、数据格式标准、数字化加工标准等,保证数据从源头上就是规范、一致的^[3]。数据集成的时候要安排数据质量核查与清洗工具,对数据的完整情况、正确程度、一致性展开校验并加以修正。对历史档案数字化工程制订严格规定、验收标准,保证数字化成果的真实性、可用性。创建长期固定的档案大数据分析结果数据质量检验评价制度,是保证档案大数据分析结果可信度的必要条件。

3.3 复合型档案人才队伍的培养与组织变革

技术转型要成功,人的转型不能缺少。大数据环境下对档案工作人员能力的要求已经发生改变,由原来的档案管理员变成了懂档案、懂技术、懂分析的复合型人才。事业单位应该制订出系统的培养人才计划,用内部培训、外部引进、和高校或企业联合等多种方式提高现有队伍的数据素养、数据分析工具的应用能力以及数据安全意识。此外还需要对档案部门进行组织结构的改良,增加数据分析员、系统管理员这样的新职务,并且要同技术部门密切合作。培育重视数据、利用数据的组织文化,鼓励基于数据的业务创新,是档案工作持续数字化转型的内在动力。

4 大数据环境下档案管理流程的重塑与优化

4.1 档案前端控制与全生命周期管理的强化

大数据技术的应用不能只集中在后端数据的分析和使用上,还应该向前延伸到档案的生成和采集阶段,即加强前端的控制。将档案管理规则及元数据采集功能嵌入到办公自动化系统、业务系统中,实现电子文件的实时或者定期归档,保证重要档案数据的完整收集和规范著录。创建档案的形成、捕获、管理、保存、利用、处置全生命周期数字化管理流程,用大数据技术对每一个环节进行监测、记录、分析,可以改善归档策

略、科学确定保管年限、精准预测存储空间需求，从而实现档案资源的合理分配和高效使用。

4.2 基于大数据的档案鉴定与处置决策支持

档案鉴定属于档案工作的重要组成部分，其好坏直接关系到馆藏的质量以及保存的成本。传统的鉴定方法主要依靠人的经验来进行判断，主观性大、效率低。大数据分析可以给档案价值鉴定提供数据驱动的决策支持。根据档案的利用频率、用户关联度、内容主题热度、跨部门引用关系等各方面数据来创建档案价值量化评价模型。该模型可以辅助鉴定人员识别出具有长期保存价值、高利用率或者潜在研究价值的档案，也可以更准确地判定到期档案的处置方式（销毁或者继续保存），使档案鉴定工作更加科学、客观、高效。

5 大数据技术应用于档案管理的伦理与法律考量

5.1 隐私保护与数据安全边界的划定

利用大数据技术对档案信息，特别是包含个人信息、工作秘密的档案进行深度挖掘与分析的时候，要把隐私保护、安全合规放在第一位。要创建起严格的个人信息去标识化、匿名化处理流程，保证在数据分析过程中不会泄露个人隐私。在数据共享和开放的时候要按照最小必要原则来确定使用目的、使用范围和使用期限，并获取相应的授权。档案管理部门应该制定详细的档案数据安全管理办法、隐私保护政策等文件来规定内部人员的数据访问权限及行为规范，并使用技术手段对所有的数据操作实行全方位的审计和监督。

5.2 算法透明性与公平性的保障

当机器学习、人工智能算法被广泛地运用到档案自动分类、价值鉴定、智能推荐等场景的时候，算法的透明性、公平性就成为必须面对的伦理问题。尽量不要使用带有歧视或偏见的“黑箱”算法。对重要的自动化决策过程要保留人工复核、

干预的通道，保证最后决策的可解释性、可控性。档案工作者要认识所用算法的基本原理及局限，不能单纯依靠技术输出，而应该发挥专业判断力，保证技术在符合档案专业伦理和法律框架下使用。

6 结论与展望

6.1 结论

大数据技术给事业单位档案工作带来的是范式上的变革。经过本文研究证明，采用大数据技术可以对档案数据的管理效率、检索准确率、知识发现能力、安全防护能力起到系统性提高的作用。赋能路径在智能化数据平台的创建、数据挖掘分析的加深、信息服务模式的更新等各个方面都体现出来。但是成功的转型不能只依靠技术本身，还要应对技术适配、数据治理、人才建设、流程重塑和伦理法律等各方面的挑战。事业机构应该以业务为龙头，用大数据实现数据价值，坚持安全可控、持续演进原则，制订科学的合理的实施方案，分步推进，把大数据的潜力变为档案管理的实际效果，真正达到档案工作由保管到治理再到赋能的价值转变。

6.2 展望

大数据技术同档案工作相融合的程度会愈加加深，出现如下趋向，第一，技术融合越发紧密，人工智能、区块链、物联网等技术同大数据融合之后，会产生更加智能的档案感知、存证和利用场景。二是服务模式愈来愈主动化、个性化，深度学习用户需求预测、知识图谱驱动智能问答将会成为常态。三是档案数据开放共享的范围、深度更进一步，档案数据作为生产要素的价值会在更多的领域、更深的程度上被利用。四是档案管理的伦理规范和标准体系将会越来越完善，给技术的发展提供指引。档案工作者要不断学习，拥抱变化，在技术的浪潮中坚持档案工作的基本价值，促使档案事业向智慧管理迈进。

参考文献：

- [1] 张颖.大数据技术在事业单位档案管理中的应用分析[J].兰台内外,2023,(35):16-18.
- [2] 赵良辉,赵颖.大数据背景下事业单位信息化建设的路径选择[J].黑龙江档案,2023,(05):49-51.
- [3] 张春艳.大数据背景下如何提升机关事业单位档案管理效率[J].黑龙江档案,2023,(03):199-201.