

Application of Big Data Technology in the Development and Utilization of Archival Resources in Public Institutions

Lina Wang

Yijinhuluo Banner Human Resources Comprehensive Service Center, Ordos, Inner Mongolia, 017000, China

Abstract

Against the backdrop of the continuous advancement of informatization and intelligent technologies, archival resources in public institutions are evolving from traditional static storage methods toward data-driven, networked, and intelligent utilization approaches. Big data technology possesses robust capabilities in data collection, integration, analysis, and mining, offering innovative technical pathways for the development of archival resources. This paper analyzes the characteristics and current utilization status of archival resources within the big data environment, with a focus on exploring application methodologies such as data fusion, knowledge association, intelligent retrieval, and decision support. Furthermore, drawing on cloud storage, knowledge graphs, and machine learning, the paper proposes measures for data governance, privacy protection, and resource sharing optimization. These initiatives aim to facilitate the transformation of archival management from a “storage-oriented” model to a “service-oriented” model, providing a valuable reference for enhancing the intelligent management level of archival systems in public institutions and maximizing the value of their data resources.

Keywords

big data technology; public institutions; archival resources

大数据技术在事业单位档案资源开发利用中的应用

王丽娜

伊金霍洛旗人力资源综合服务中心, 中国·内蒙古 鄂尔多斯 017000

摘要

在信息化、智能化持续发展的这样一种背景之下, 事业单位档案资源正从以往的静态保管朝着数据化、网络化、智能化利用的方向进行转变。大数据技术具有较强的数据采集能力、整合能力、分析能力、挖掘能力, 能够为档案资源开发提供全新的技术途径。本文对大数据环境下档案资源的特征、利用现状展开分析, 着重对数据融合、知识关联、智能检索、决策支持等应用方式予以探讨, 并且结合云存储、知识图谱、机器学习提出数据治理、隐私保护、资源共享优化举措, 以此推动档案管理从“存储型”向“服务型”转变, 为提高事业单位档案智能化管理水平、数据资源价值提供一定参考。

关键词

大数据技术; 事业单位; 档案资源

1 引言

档案资源对于事业单位的行政决策、业务运行、公共服务而言, 是至关重要的信息根基。在传统的管理模式当中, 档案数据呈现出分散状态, 标准也并不统一, 资源利用的程度不够, 这样的状况已然难以契合数字化治理的要求。伴随大数据、人工智能、云计算技术的不断发展, 档案管理逐渐从静态的保管转变为动态的开发利用。大数据技术具备整合多源异构档案信息的能力, 借助关联分析、数据挖掘来提高资源的价值, 促使档案从事实记录朝着决策服务的方向进行延伸。本文依据事业单位档案数据数量庞大、类型繁多、价

值层次复杂等特性, 探究大数据技术在资源整合、开发利用、智能服务方面的应用途径, 从而为建立高效且智能的档案服务模式给予参考。

2 大数据背景下事业单位档案资源的特征与发展现状

2.1 档案资源数据化与多元化趋势

事业单位档案资源有着行政文书、业务档案、科研资料、财务凭证、声像记录等多种类型, 具有很高的多样性、结构复杂性。在大数据环境当中, 这些档案渐渐达成数字化采集和在线存储, 建立起从文本到图像、音频、视频的多模态数据模式。伴随信息系统的广泛运用, 档案信息的生成方式变得更加自动化, 档案数据的实时性、更新频率有了明显提高。数字化转型让档案从静态记录转变成动态流通的信息资源,

【作者简介】王丽娜(1982-), 女, 中国内蒙古鄂尔多斯人, 硕士, 副高级研究馆员, 从事人事档案管理研究。

不过也带来了数据碎片化、格式异构与管理冗余这些问题。事业单位要在大数据架构下重新建立档案信息模式,达成资源的标准化、高效利用。【1】。

2.2 传统档案管理模式的局限性

传统档案管理主要是对纸质或电子档案进行分类、归档、借阅。其核心目的是实现完整保存、方便查阅。但是在信息爆炸的时代,这种模式没办法满足数据规模呈指数增长的要求。档案管理人员依靠人工进行分类和检索,效率比较低,使得信息的利用率不高。档案数据缺少语义关联和知识映射,不能够为高层次的分析和决策提供支持。不同部门之间的档案系统相互孤立,接口也不统一,形成了信息“孤岛”。档案资源潜在的规律和价值没有得到充分挖掘,急需建立数据驱动的档案服务模式。

2.3 大数据赋能档案开发的必要性

大数据技术借助分布式存储、数据挖掘、模式识别,给档案资源管理带来了全新的工具。其能够针对结构化和非结构化的档案数据进行智能分析,达成信息检索的精准性、知识发现的自动化。事业单位运用大数据分析技术,可以把零散的档案数据转变为决策支持资源,达成档案的再次利用、价值提高。大数据的引入不但提高档案信息流通的效率,还促使档案管理朝着智能化、服务化的方向转变,成为事业单位数字治理的重要支撑力量。

3 大数据技术在档案资源整合中的应用

3.1 多源数据采集与动态归档机制

在大数据环境下,事业单位档案的来源范围已从传统的纸质文件,延伸至政务办公系统、业务平台、电子公文、业务日志、智能终端等多个渠道,并且数据生成具备连续性与实时性的特点。为了提高归档效率,有必要建立一套面向多源数据的自动采集机制,借助 API 接口、数据中台等手段来达成不同系统之间的信息同步。在采集过程中,可以运用规则引擎、语义分析技术,以此来识别文件类型、提取核心内容并生成元数据,进而让档案在形成、流转、办结的整个过程中实现动态归集。对于关键业务数据,还应当设定归档触发条件、完整性校验规则,从而避免出现漏归、错归的情况。通过把数据采集与业务流程进行有效衔接,能够提高档案归档的及时性与准确性,促使档案管理从事后整理朝着全过程数据治理的方向转变【2】。

3.2 数据清洗与标准化集成

多源档案数据在格式、字段、编码、分类规则方面常常有着比较明显的差异,要是直接将其汇入统一平台,很容易出现重复记录、信息缺失、语义冲突的情况,进而对后续的检索与利用造成影响。在数据集成之前,需要建立规范的清洗与标准化流程,借助模式识别、规则校验、语义比对来识别异常值、重复项还有缺失字段,并且依据业务逻辑予以纠正或者补充。对于档案名称、形成时间、责任部门、分类

代码等关键字段,要建立统一的元数据标准、映射关系,以此实现不同业务系统之间的数据语义对齐。配合 ETL 技术完成抽取、转换、加载,能够形成结构相对一致、质量可控的数据资源库。标准化处理能够提高档案数据可信度,还对跨部门检索、关联分析、资源共享有益,为后续智能开发利用提供稳定的数据基础。

3.3 云架构下的档案数据存储与共享

云计算给事业单位档案资源的集中管理、跨部门利用给予了比较有力的技术支持。可以依靠云平台来建立统一的档案资源中心,依据数据规模、访问需求对存储与计算资源进行弹性配置,达成电子档案的集中保存、分布式访问、统一管理【3】。在共享过程当中,要依照档案密级、利用范围、岗位权限来设置分级访问控制,并且运用身份认证、传输加密、日志审计等举措来保障数据安全。针对重要档案操作,可以建立全过程访问记录、责任追溯机制,保证查询、下载、修改等行为能够被核查。云端架构能够解决传统档案受地域和物理空间限制的问题,让不同部门在授权范围内共享资源,提高档案利用效率。通过安全控制与共享机制的协同设计,能够为智慧档案建设、档案数据资产化管理提供基础条件。

4 大数据技术推动档案资源的智能化利用

4.1 基于知识图谱的语义关联分析

知识图谱具有这样的功能,它能够把分散在档案里的人物、事件、时间、地点、机构等要素都转变成实体节点,并且借助关系连接建立结构化的语义网络。将自然语言处理、命名实体识别技术结合起来,就能够针对非结构化文本进行实体抽取、关系识别、主题归类等工作,进而让不同载体、不同业务系统当中的档案信息建立起稳定的关联。和传统的关键词检索相比,知识图谱更为注重上下文关系和语义推理,能够依据用户的查询自动扩展相关的人物、业务事项、历史记录,以此提高档案检索的准确性与完整性。该系统还可以依据新增的档案持续更新实体关系,形成业务知识链、主题演化脉络,为档案编研、专题分析、知识发现提供支持,促使档案资源从单一的信息检索朝着关联化知识服务转变。

4.2 机器学习在档案数据分析中的应用

机器学习给事业单位的档案分类、内容识别、利用趋势分析给予了有力的技术支撑。监督学习能够依照档案文本的特征、文件的属性、业务标签来建立分类模型,达成电子文件的自动归类、辅助著录,进而减少重复性的人工操作。聚类算法能够识别不同业务档案间潜在的相似关系,为专题资源的整合、数字化加工顺序的确定提供参考依据。就档案利用数据而言,借助时间序列和回归模型,可以对访问频率、用户需求、资源使用变化进行分析,为服务资源的配置、开放策略的调整提供参考。在纸质档案数字化进程里,OCR 与图像识别技术还能够完成版面识别、文字提取、内容转换。

通过把持续训练和人工校正结合起来,可以逐步提高模型的适用性,促使档案管理从经验判断朝着数据分析与智能辅助转变。

4.3 档案可视化与决策支持系统构建

档案资源有着多种类型,并且关联关系非常复杂,仅仅依靠目录、文字报表,很难直观地展现资源利用的状态。所以能够建立针对管理需求的可视化分析、决策支持系统。该平台能够借助图表、数据看板等形式,展示档案数量、利用频率、用户需求、资源分布、共享情况,如此一来管理人员就能及时了解档案的运行状态。结合GIS、时间序列分析,还能够呈现档案数据在区域、时间、业务领域里的变化规律,为专题研究、资源配置提供相关依据。决策支持系统能够进一步整合利用率、开放程度、数据质量、安全风险等指标,对档案管理绩效展开综合评价,进而识别出资源闲置或者利用异常的情况。通过将可视化分析跟指标监测结合在一起,能够推动档案管理从静态统计朝着动态研判、辅助决策方向延伸。

5 大数据驱动下档案资源开发利用的创新模式

5.1 档案资源价值挖掘与知识服务转型

在大数据环境当中,事业单位档案资源的功能正在从凭证保存、历史记录朝着知识服务、决策支持的方向延伸。借助自然语言处理、语义分析、知识关联等技术手段,能够针对分散的档案进行主题聚合、内容识别、关系重构工作,从而让原本处于静态的信息转变为可以检索、能够分析的知识资源。事业单位可以依靠数据平台建立专题档案库、知识图谱、业务关联网络,为行政管理、科研活动、公共服务提供相应依据。档案利用也不应该仅仅停留在查询和复制的方面,而需要加强数据加工、趋势分析、可视化表达,围绕具体的业务问题给予针对性的信息支持。通过提高档案数据的再利用能力,能够推动档案部门从资料保管角色转向知识服务角色,进而进一步释放档案资源的管理价值与社会价值。

5.2 档案安全管理与数据治理体系建设

档案数据进行集中存储、开放利用,使得资源使用效率得以提高,不过越权访问、信息泄露、数据篡改等风险也有所增加。事业单位需要建立起涵盖采集、存储、利用、共享、销毁环节的数据治理机制,依照档案敏感程度、利用范围、业务价值来实施分类分级管理。针对重要数据,可以运用身份认证、访问控制、加密传输、日志审计等措施,对非授权

操作加以限制,并且保留完整的访问记录。数据治理还应当强化元数据、分类编码、接口标准建设,以此保证不同系统之间的数据具备一致性、可追溯性。对于跨部门共享的档案,要明确使用权限、责任主体、数据边界,避免因管理职责不清晰而造成安全风险。通过将技术防护、制度约束、责任追溯结合起来,能够在扩大档案利用范围的情况下维持数据安全、管理秩序。

5.3 开放共享与跨部门协同机制

事业单位的档案资源常常分散于不同的业务部门、信息系统之中,要是缺乏统一的共享机制,就容易出现重复采集的情况,还会形成信息孤岛。依靠大数据平台,能够在统一的数据标准、权限模式之下,整合源自不同地方的档案资源,借助规范的接口达成跨系统查询、按需调用的目的,进而提高信息流转的效率。在共享的过程里,需要依据业务职责、档案的开放程度来设置不一样的访问范围,防止把资源共享单纯地理解成无条件开放。基于此,可以建立跨部门协同机制,围绕公共管理、政策研究、社会服务等具体业务展开数据联合分析,促使档案资源从单一部门利用转变为主主体协同开发。通过明确共享规则、数据责任、反馈机制,能够减少部门之间的信息壁垒,提高档案资源配置效率,并且为数字化治理、智慧公共服务提供更为完整的数据支撑。

6 结语

大数据技术用于事业单位档案工作,促使其从传统管理朝智能化服务实现深度转变。借助多源数据整合、智能分析、可视化决策,档案资源潜在价值被充分发掘,为公共管理和社会服务给予了坚实数字支撑。未来,事业单位要完善档案数据治理模式,强化隐私保护、标准建设,推动档案信息化朝着自主可控、高效协同方向发展。伴随人工智能、区块链与云计算进一步融合,档案资源会成为事业单位重要数据资产和决策基础,达成由“存档备查”至“数据帮助”的全面跨越,为建设数字政府和智慧社会提供持久动力。

参考文献

- [1] 申健.大数据背景下事业单位档案管理工作优化策略研究[J].兰台内外,2025,(34):4-7.
- [2] 周立强.大数据时代事业单位档案管理工作优化措施探析[J].兰台内外,2025,(32):60-62.
- [3] 程红.大数据技术赋能事业单位档案管理工作的创新研究[J].兰台内外,2025,(30):16-18.