

大数据环境下高职档案管理的挑战与创新路径探讨

唐燕

湖南汽车工程职业大学

DOI:10.32629/jief.v8i3.21732

[摘要] 随着大数据、人工智能以及区块链等技术的迅猛发展,档案管理工作已从传统纸质载体逐渐向数字化、智能化方向转型。文章聚焦《全国档案事业发展“十五五”规划》要求以及数字档案馆建设工作,围绕高职院校档案管理实际,对大数据环境下档案管理面临的挑战进行分析,然后从分布式存储、多模态数据整合、安全防护以及数据治理等多个角度提出具备较强可操作性的创新路径,旨在为高职院校及各级档案部门深度落实数智化转型工作提供有益参考。

[关键词] 大数据; 智能化; 档案管理; 安全防护; 数据治理

中图分类号: G222 文献标识码: A

Challenges and Innovative Paths of Archive Management in Higher Vocational Colleges Under the Big-Data Environment

Yan Tang

Hunan Automotive Engineering Vocational University

[Abstract] With the rapid development of big data, artificial intelligence, blockchain and other technologies, archive management is gradually transforming from traditional paper-based carriers toward digitalization and intelligence. Focusing on the requirements set forth in the 15th Five-Year Plan for National Archives Development and the construction of digital archives, this paper addresses the realities of archive management in vocational colleges and analyzes the challenges confronting archive management under the big-data environment. It further proposes practically-oriented innovative paths from multiple dimensions including distributed storage, multimodal data integration, security protection and data governance. This study aims to provide useful references for vocational colleges and archive departments at all levels to fully advance digital-intelligent transformation.

[Key words] big data; intelligence; archive management; security protection; data governance

档案属于国家以及社会各项活动的宝贵记录,做好其管理工作对促进治理能力现代化发展大有帮助。在职业教育领域,近年来随着办学规模不断扩大、产教融合模式的持续深化,教学、学籍、实训以及校企合作等一系列档案数据呈现出迅猛增长的态势,导致档案管理开始面临巨大压力。因此,有必要对大数据环境下档案管理的挑战与创新路径做出深入研究,从而为档案管理工作实现高质量健康发展进行助力。

1 大数据环境下档案管理的挑战

1.1 数据规模爆炸式增长存储压力大

高职院校当中的档案资料涉及学生学籍、成绩、就业以及人事、校企合作等一系列门类。近年来,在国内扩招政策与数字化校园建设战略不断深入推进的背景下,仅学生学籍档案资料每年实际新增电子文件量即可达到数万条,这对高职院校的档案存储架构提出了挑战。传统集中式存储普遍采用单机或

者是小型服务器阵列,想要扩容必须停机采购硬件,不但成本高而且周期长,面对每年高速增长的数据量,档案存储空间将会频繁告急。此外,历史档案资料的数字化扫描形成的图像文件体积过大,通常单个学生完整档案扫描件便能够达到数百兆,整体来看存储压力突出。存储架构如果扩展性不足能够引起归档延迟、备份不及时等问题,甚至会面临数据丢失等相关风险^[1]。

1.2 非结构化档案难以统一管理

大数据环境下,高职院校档案资源不再单纯地局限于文书、表格等各种结构化数据,开始涌现出海量的教学视频、实训录像以及在线课程录播、虚拟仿真课件等一系列非结构化数据信息。对于这些数据来讲,其格式各异、标准不一,而且分散于不同系统,包括教务系统、各二级学院个人电脑以及第三方教学平台等。导致档案部门在实际收集这些档案资料时会面临格式不兼容、元数据缺失或者分类困难等诸多挑战。例如,实训课程视频

通常每节约为45分钟、1080P分辨率,实际单文件可达到1GB以上,同时命名规则主要由任课教师自行进行设定,档案管理人员难以高效地判断内容归属与具体的保管期限,容易造成大量有价值的非结构化档案长期游离于统一管理机制之外,无法实现有效管控。

1.3 数据安全保护压力不断增大

高职院校档案数据涉及大量师生个人敏感信息,包括身份证号、家庭住址以及健康记录、成绩排名等,此外还包括学校财务预算、科研涉密内容或者是校企合作商业秘密等相关信息。大数据背景下,数据将会集中存储,并实施网络化传输,导致安全风险点大幅度增多。一方面,黑客攻击、病毒入侵等安全问题的发生概率持续提高,而高职院校档案管理系统普遍缺少专业性极强的安全运维团队,同时防火墙以及入侵检测系统等也存在更新滞后问题。另一方面,内部权限管理相对粗放,当前仍有部分院校存在账号共用、权限过大或者操作日志不完整等相关问题,留下了较大的安全隐患。此外,数据安全法、个人信息保护法等相关法律法规推行之后,对档案资料的收集、存储、使用以及共享开始有了更严格的合规要求,这导致高职院校在跨部门共享档案以及对外提供利用时同样会面临相对较大的法律压力^[2]。

1.4 档案利用效率低下

传统模式下,档案利用一般采用“人工检索+纸质调阅”模式,利用者必须前往档案库房填写申请单,相关管理人员凭经验查找一系列卷宗,档案调阅耗时较长,流程繁琐。当前部分高职院校尽管已经打造出档案管理系统,但相关检索功能依然停留于简单的关键字匹配层面,未能开发语义检索功能以及智能推荐功能,用户往往由于不知道准确档案名称而难以快速找到所需档案。这些问题的存在,会导致档案利用率无法提高,档案数据价值难以被充分挖掘出来。例如历年学生就业数据分散于多个不同年份档案内,无法开展纵向对比分析工作;教学评估材料以及专业建设档案内潜藏的规律与趋势无法被有效提炼,从而难以为学校制定相关决策提供充足可靠的档案数据支撑。

1.5 人才队伍以及制度标准建设相对滞后

进入大数据之后,档案管理工作强调从业人员不但要掌握档案业务知识,又要了解信息技术、数据治理以及网络安全等相关知识与技能。但部分高职院校当中的档案部门存在从业人员年龄偏大以及知识结构单一等相关问题,大部分工作人员仅了解传统档案整理流程,对于先进的数据库操作、数据分析工具以及元数据标准等缺乏充足的了解和认识。此外,学校层面未能打造出适配电子档案管理的相关专项制度,现有制度依然围绕传统纸质档案制定,导致电子文件的归档范围、保管期限以及格式标准等缺少明确清晰的规定。同时各业务系统之间也未能明确统一的数据交换标准,出现了教务、人事、财务以及学工等一系列系统各自为政的问题,形成“数据孤岛”,在这种情况下,档案部门无法实现高效自动化采集以及一体化管理,这对档案管理学数字化转型产生了不利影响^[3]。

2 大数据环境下档案管理的创新路径

2.1 打造分布式存储以及云原生档案管理平台

为有效应对大数据环境下存储架构难以承载海量档案数据的挑战,高职院校应分步落实好分布式存储以及云原生档案管理平台的建设工作,确保档案存储能力与大数据环境高度适配。

高职院校应落实好现有档案数据信息的全面普查工作,依照热数据(指的是近三年常用的档案资料)、温数据(三至十年内常用的档案资料)以及冷数据(超过十年归档未用的数据)进行分级;引入“私有云+对象存储”有机结合的存储技术架构,合理部署分布式存储集群,借助Ceph技术达到数据多副本冗余存储的目的,扩容期间仅需适当增加节点即可;应将冷数据有效迁移至蓝光光盘库当中完成离线归档,进一步降低在线存储成本;配套开发云原生档案管理系统,借助容器化部署模式,直接将档案全流程有效迁移至线上。实践操作中,可先将学生学籍档案充当试点,将一系列历史扫描件批量导入对象存储同时打造适配索引,验证稳定性后再逐步扩展至人事、财务以及科研等其他门类的档案,逐步达到数字化转型的目标^[4]。

2.2 创设多模态数据整合与智能分类体系

为有效应对数据类型多元化挑战,应打造出多模态数据整合与智能分类体系。详细编制《职业院校电子档案元数据规范》,借此确定好不同类型档案资料的必填元数据字段,例如对于教学视频应该标注课程名称、授课教师以及班级、录制时间、知识点标签等一系列信息,切实从源头上严格规范档案资料的采集工作;增设多格式解析的档案采集网关,借此与教务系统、在线教学平台以及实训管理系统等进行自动对接,负责定时开展电子文件抓取工作,能够对视频、音频以及图片等持续地自动开展关键帧提取、语音转文字或者是OCR识别等工作,并且生成能够进行检索的文本描述;对机器学习自动分类算法进行运用,通过已有档案资料持续开展模型训练工作,借此对各种新归档文件自动判断门类与具体保管期限,准确率超过90%后再实施人工复核。实践操作中,可优先处理高职院校的毕业设计以及实训视频等档案资料,为每门课程打造出标准化档案包,主要涉及教案、课件、视频、作品以及成绩等信息,利用统一标识符达到关联检索的目的^[5]。

2.3 落实全生命周期安全防护以及权限管控工作

为缓解安全与隐私保护压力,应打造档案数据全生命周期安全防护机制。档案资料收集环节,对涉及个人敏感信息的档案资料实施自动识别以及脱敏处理工作,例如对于身份证号的中间八位直接用星号替代,脱敏规则可灵活配置;档案存储环节,可以借助AES-256加密算法对一系列敏感档案实施加密存储,对于密钥应该由专人管理并推行定期轮换制度;档案传输环节,加大对HTTPS加密技术的运用力度,严禁通过任何未加密渠道直接传递各类档案数据;档案利用环节,重点推行基于角色的细粒度权限控制(RBAC)模式,切实依照部门、岗位以及密级等合理地分配权限,对于敏感档案还应该推行“双人审批”,全部操作做到全过程留痕,实际日志保存时间控制在六个月以上;运维环节,

主要引入数据库审计以及行为分析等相关技术工具,借此对异常下载、批量导出以及非工作时间访问等进行实时识别与自动告警。实践操作中,学校可对学生档案当中的身份证号、手机号以及家庭住址等各类字段默认实施脱敏显示,仅通过书面申请审批后方可准许查看完整信息,后续逐步覆盖至其他档案信息。

2.4 加大数据治理以及智能检索分析力度

为应对利用效率低下、价值挖掘不足等有关挑战,应站在数据治理以及智能应用两个角度进行发力。

数据治理层面,组建一个由档案管理部门负责牵头、其它一系列业务部门共同参与的档案数据治理小组,联合编制数据质量标准,并定期落实好数据清洗工作,重点解决重复档案、缺失字段或者是格式不统一等相关问题,同时落实好全校统一的档案数据资产目录建设工作;智能检索层面,重点升级系统检索功能,应该引入全文检索引擎(如Elasticsearch等),一般支持关键字、全文以及语义三种检索功能,用户通过直接输入自然语言便能够迅速匹配有关档案资料;胚胎开发可视化分析模块,对学生就业趋势、专业建设成效、科研产出分布等开展多维度统计分析工作,生成报表从而为校领导制定相关决策提供充足可靠的参考依据。在实践操作中,应打造“学生成长档案”应用模块,重点开展学籍、成绩、获奖、实习以及就业等各类档案数据的整合工作,生成学生成长画像,然后依照专业、年级等汇总分析学生的就业去向或者是课程通过率等,从而为学校人才培养方案的修订提供可靠支持^[6]。

2.5 落实好复合型人才培育以及标准体系建设工作

为应对人才与制度滞后压力,应落实好人才培养以及制度建设等相关工作。

人才培养方面,编制档案人员三年能力提升计划,坚持每年安排40学时以上的专项培训活动,培训内容主要包括数据库基础、数据治理以及网络安全、档案信息化标准等内容;鼓励和引导在职人员主动考取有关职业资格证书;通过优化薪酬待遇机制,引进1-2名具备计算机或者信息管理专业背景的人才来充实队伍,提高队伍整体数字化素质;与信息中心打造协作机制,定期组织实施技术交流工作。制度标准方面,紧随大数据时代发展变化对现有的《电子档案管理办法》《档案数据共享管理规定》《档案系统运维管理规程》等开展修订完善工作,确定好电子档案的归档流程、保管期限以及销毁程序等内容;此外,应明确业

务系统和档案系统之间的数据接口标准,统一设定数据格式以及交换协议,推动教务、人事以及财务等各类系统能够和档案平台实现高效自动对接。实践操作中,可由档案部门负责牵头编制《职业院校电子档案元数据与分类标准》,通过学校信息化建设领导小组审议后再进行发布,作为各业务系统建设的核心规范^[7]。

3 结语

大数据时代的到来为档案管理工作带来一定挑战的同时,也提供了一个极佳的转型升级机遇。在面对数据规模高速增长、类型日益多元、安全压力激增、档案资源价值挖掘利用不足以及人才制度滞后等一系列挑战时,高职院校应坚持问题导向,切实利用分布式存储有效打破存储瓶颈,通过多模态整合的方式应对数据复杂多元的压力,借助全周期防护措施确保档案管理安全合规,以数据治理技术措施充分释放数据价值,依托人才与制度建设夯实转型基础,从而进一步提高大数据环境下的档案管理实效。

[参考文献]

- [1]杨智勇,程佳琪.数字时代电子档案理论与实践双向建构——《电子档案管理理论与实务》书评[J].中国档案研究,2025(1):289-298.
- [2]常消暑.数字化时代人事档案赋能企业人力资源管理的战略价值与实践研究[J].标准生活,2025(S1):4-6.
- [3]刘宏.大数据环境下医疗行业档案安全风险的动态识别与精准防控[N].山西科技报,2025-12-30.
- [4]蒋辰媚.大数据背景下档案分类的实践与探索[J].黑龙江档案,2025(6):97-99.
- [5]王玉龙.大数据视域下地方高校档案信息建设研究[J].黑龙江档案,2025(6):130-132.
- [6]王鹏,吉雅,鄢西.大数据技术在土地出让档案管理中的应用[J].经纬天地,2025(6):96-100.
- [7]朱雪莺.数字时代档案管理智能化转型策略探析[J].参花,2025(36):111-113.

作者简介:

唐燕(1979—),女,汉族,湖南益阳人,本科,政工师,研究方向:档案管理。