

如何建立同时有效管理结构化数据和文档性数据的电子档案系统

李 波

大唐环境产业集团股份有限公司北京新科技产业分公司 北京 100097

【摘 要】随着数字化转型的深入,电子档案系统面临结构化数据与文档性数据并存的“二元结构”挑战。传统档案管理模式将二者分立处理,导致数据孤岛、管理低效与价值埋没。本文基于档案学界最新研究成果与行业实践,提出融合管理的理论框架与实现路径。研究认为,建立同时有效管理两类数据的电子档案系统需要“三层架构、双轨策略”:在战略层面确立融合管理的顶层设计,在平台层面构建支持异构数据的统一底座,在应用层面创新智能化的业务场景。在此基础上,本文系统阐述了制度规范先行、技术平台筑基、业务流程再造、安全保障护航的实施路径,并结合区块链、人工智能等新一代技术应用,展望了从档案数字化向数据智慧化跃迁的未来图景。研究为各类机构建设新一代电子档案系统提供了理论参考与实践指南。

【关键词】结构化数据;非结构化数据;电子档案系统;融合管理;数字化转型

DOI:10.12417/2811-0722.26.09.042

1 引言

数字时代,档案管理的对象正在发生深刻变革。审视当前组织的数字资产构成,不难发现两类形态迥异的数据资源正同步激增:一类是依托关系型数据库严密组织的结构化信息单元,涵盖交易流水、业务表单及各类元数据,其存在形式高度依赖预定义的字段体系与关联规则;另一类则是以松散格式存在的非结构化文档,包括但不限于行政公文、技术图纸、契约合同、往来邮件及多媒体文件,此类数据虽缺乏规整的存储结构,却承载着业务活动的完整语义与深层脉络。这两类数据犹如档案世界的“二进制代码”,共同构成组织数字记忆的完整图景。

然而,将两类数据割裂处置的管理范式,在实践中暴露出三重系统性缺陷:其一,资源耗费居高不下——结构化数据因系统依赖而面临湮没风险,文档性数据则因重复存储造成冗余堆积;其二,检索利用效能折损——用户若需获取跨类型信息,不得不在异构系统间频繁切换、手动拼合,流程割裂严重;其三,知识关联网络断裂——本应相互印证、彼此补充的数据单元被人为阻隔,致使档案作为完整证据链条的固有价值难以充分彰显。

“十四五”全国档案事业发展规划明确提出,要在结构化数据归档上实现突破。这标志着档案管理正从“纸质核心”迈向“数纸协同”,从单一类型管理走向多元融合治理。在此背景下,如何建立一套能够同时有效管理结构化数据和文档性数据的电子档案系统,成为档案学界与实务界共同关注的核心命题。

本文旨在系统回答这一命题。通过整合最新研究进展与典型案例经验,构建融合管理的理论框架,并围绕制度、技术、流程、安全四个维度,提出切实可行的系统建设路径。

2 融合管理的理论基础与核心挑战

2.1 两类数据的本质差异与管理要求

结构化数据与非结构化数据的差异,根植于它们的本质特征。结构化数据以预定义的数据模型为基础,遵循严格的字段类型、长度约束和关系规则,典型代表是关系型数据库中的表格数据。这类数据的优势在于查询高效、统计便捷、事务一致性强,但其劣势同样明显——高度依赖特定的数据库管理系统和数据结构定义,一旦脱离原生环境,便可能变得不可读、不可用。

文档性数据则相反,它以自包含的方式存储信息内容,文本、图像、音视频文件各有其格式规范,但内容解读不依赖于外部数据模型。这类数据的优势在于格式通用、易于交换、人机共读,但检索效率相对较低,内容挖掘需要借助OCR、自然语言处理等智能技术。

从档案管理视角审视,两类数据的合规要求也存在显著差异。结构化数据的归档强调“业务逻辑的完整保存”——不仅要保存数据本身,还要保存数据之间的关联关系、业务规则、时间戳信息,否则数据便失去了凭证价值。文档性数据归档则侧重“原始性与真实性的保障”——电子文件的元数据、固化信息、数字签名需要伴随文件全生命周期。两类数据的管理要求在本质上互补而非冲突,这为融合管理提供了理论基础。

2.2 分立的代价:当前管理模式的问题诊断

当前普遍采用的“结构化数据留库、文档性数据进馆”的分立模式,在实践中暴露出诸多问题。

首先是数据流失风险。结构化数据深度依赖业务系统,而业务系统因技术迭代频繁升级或关停。当系统退役时,大量具有长期保存价值的结构化数据往往随之一同消亡。据嘉兴市档案馆的调研观察,政务业务系统中沉淀的结构化数据普遍存在系统环境依赖度过高的问题,一旦脱离原生业务平台,其可读性与可用性便大幅衰减,由此催生出数据资源割裂、数字记忆

断层的隐患。

其次是管理成本叠加。分立模式下，机构需要同时维护两套技术体系——面向结构化数据的关系型数据库与面向文档性数据的对象存储系统，不仅增加了采购、运维成本，也带来了技术人员双重培养的负担。中国电建的统计数据显示，沿用纸质输出归档的传统做法，单张纸页的综合管理成本高达9.52元，年度累计支出相当可观。

再次是价值挖掘受限。分立管理意味着数据之间无法建立关联，档案的完整叙事被割裂。以建设项目档案为例，设备采购合同（文档性数据）与设备调试参数（结构化数据）分属不同系统，研究人员需要分别查询、手动关联，效率低下且易出错。这种割裂使得档案无法发挥其作为完整证据链的价值。

2.3 融合管理的理论框架

融合管理并非简单地将两类数据堆砌在同一系统中，而是构建一种“形散神聚”的治理体系。基于上述分析，本研究构建了融合管理的三维治理体系。从纵向治理维度审视，该体系包含三个递进层次：位于顶层的战略决策层，主要职责是确立档案数据的资产属性，并统筹制定适用于两类数据的统一管理制度；居于中位的技术赋能层，其核心任务是搭建能够兼容异构数据的底层支撑平台，实现存储资源、索引服务与访问接口的统一调度；处于基座的应用实践层，则聚焦于具体业务场景，开发适配不同数据类型的操作功能模块。

从横向协同维度观察，该体系设计了双轨并行的治理路径：第一轨道可称为“资产化封装路径”，主要面向结构化数据，通过格式封装、环境固化、数据迁移等技术手段，将其从依附于特定业务系统的动态数据转化为独立存续的档案资产；第二轨道可称为“知识化萃取路径”，主要面向文档性数据，借助OCR识别、语义解析、知识图谱构建等智能技术，将其从静态存储的文件对象转化为可供深度挖掘的知识单元。两条路径最终交汇于统一的数据服务中枢，实现两类数据的同步归档与协同治理。

这一框架的核心要义在于：承认差异，但不接受割裂；追求统一，但不抹杀特性。融合管理的目标不是消除两类数据的区别，而是在保持各自优势的基础上，实现“1+1>2”的整体价值。

3 系统建设的核心路径

3.1 制度先行：构建覆盖全生命周期的规范体系

任何技术系统的有效运行，都以制度规范为基石。建立融合管理的电子档案系统，首要任务是构建“顶层设计+操作细则”双层次的制度框架。

顶层设计层面，需要制定涵盖两类数据管理的综合性政策文件。嘉兴市的经验表明，由档案部门主动参与电子文件管理通知的起草修订，可以从制度源头构建“生成—流转—归档”

全链路闭环管理体系。该市通过统一保管期限、简化归档流程、优化管理制度等举措，确保在最小改动范围内实现馆室一体化提升。这一实践启示我们，融合管理不是推倒重来，而是在现有制度框架基础上的整合优化。

操作细则层面，需要针对两类数据分别制定详细规范。对于结构化数据，应明确数据采集范围、封装格式（如XML结构化封装）、元数据标准、保管期限划分规则。对于文档性数据，应规范文件格式要求、元数据著录规则、四性检测指标。中车青岛四方实践显示，通过制定《建设项目电子文件归档和电子档案管理细则》，并建立涵盖900余份标准化表单的模板库，为电子文件归档奠定了坚实基础。国电象山海上风电项目同样构建了包含900余份标准化模板的电子文件模板库，成为试点的突出亮点。

制度建设的难点在于协调各方利益。实践中可以采用“专班指导+定制培训”模式，对业务部门与档案部门开展联合培训，推动形成“全员参与、协同推进”的共识。象山项目的经验表明，当参建单位切身感受到电子文件归档带来的便利——审批时间从数天压缩到数小时、迎检告别“翻箱倒柜”——心态便会发生从“要我做”到“我要做”的根本转变。

3.2 平台筑基：混合存储架构与统一数据底座

技术平台是融合管理的物理承载。面对两类数据的异构特征，单一存储技术难以两全。学术研究与行业实践共同指向“混合存储架构+统一数据底座”的解决方案。

从技术适配性角度考量，面向采用严整模式定义的结构化信息资产，优先选择具备强事务保障能力的分布式关系型数据库，此类系统能够有效维系数据操作的原子性与一致性，同时保障复杂查询场景下的响应效率；而针对形态多样、规范各异的非结构化数字对象，则宜采用基于对象存储技术构建的存储体系，凭借其扁平的命名空间设计与近乎无限的横向扩展能力，满足海量文件长期保存与高频读取的双重需求。铁路电子档案领域的研究表明，这种混合存储架构能够有效兼顾两类数据的管理需求，同时通过统一的元数据管理实现二者的业务协同。

统一数据底座的核心功能是建立两类数据之间的“语义桥梁”。具体包括三个层面：一是统一元数据管理，建立涵盖两类数据的元数据标准，记录档案的属性信息和关联关系；二是统一索引服务，建立跨数据类型的全局索引，支持用户无需感知数据物理位置即可检索；三是统一访问接口，提供标准化的数据上传、下载、检索、更新API，屏蔽底层存储差异。

分布式架构是平台建设的关键技术选择。某智能化档案数字化平台采用分布式架构，存储系统基于对象存储技术将数据分散存储在多个节点，通过数据分片和多副本机制保证可靠性；计算资源采用集群部署，支持弹性伸缩，根据业务负载动

态调整计算节点。测试表明,该平台在1000用户并发访问下保持99.9%的事务处理成功率,每小时可处理100GB级档案数据。

成本控制是平台建设必须考虑的现实问题。中车青岛四方的实践显示,通过高效整合各类存储资源构建统一的非结构化存储底座,结合数据特点与利用频次实现冷、温、热三层架构动态流转,存储成本相较传统模式下降低约80%。

3.3 流程再造:从线性管理向闭环治理演进

传统档案管理遵循“形成—归档—保管—利用”的线性流程。融合管理要求打破这种单向流动,构建“业务生成—数据归档—知识反哺”的闭环治理体系。

前端控制是闭环治理的起点。传统模式中,档案管理往往处于业务链条的末端,被动接收业务部门移交的数据。融合管理要求档案管理前移,在业务系统设计阶段就介入数据规范制定。嘉兴市的“馆室一体化”实践实现了从“末端管理”向“全程管理”的转变。象山项目更进一步,将“档案管理人员前置审批”环节嵌入文控系统,使档案管理作为合规性审查的关键一环被正式纳入项目管理核心循环。

自动化归档是流程再造的技术支撑。通过构建统一的标准归档接口,采用可复用的业务系统适配器,实现业务系统电子文件和元数据的在线自动归档。中车青岛四方研发了结构化数据可视化归档技术,可将结构化数据自动转换为非结构化电子文件,实现业务系统结构化数据以“文件化+数据化”两种方式的自动归档。嘉兴市采用基于分布式消息中间件和列式数据库的海量数据处理技术,支持单日峰值300万条数据的实时采集与高效存储。

模型驱动是实现自动化归档的智能引擎。中车青岛四方的实践显示,通过配置归档活动、元数据规则、文件权限,提供快速归档的标准模板,实现对同一类业务分类下档案的集中统一管理。该企业根据业务范围创建了设备类、基建类、项目管理类、设备调试类四类归档模型,以归档模型为载体创建归档计划,进行计划驱动式归档,业务人员仅通过鼠标“拖、拉、拽”即可完成档案移交。

闭环治理的最终目标是实现“档案服务业务”的价值反哺。传统观念认为档案管理是消耗资源的“后台”工作,闭环治理则彰显档案对业务的反向赋能。通过将大量利用需求配置到系统后台,业务人员在权限范围内可直接在线利用档案,大幅简化利用手续。通过数据挖掘,系统可以为员工精准推荐个性化内容——相似设计图纸推荐、新员工规章制度推荐等,多倍幅提升档案数据访问量。工程项目中,电子化审批流转将效率提升数倍至数十倍,电子组卷使每卷案卷整理审批时间平均缩短6小时,项目完工到档案专项验收时间较以往提前近1年。

3.4 安全护航:可信保障与全程溯源

融合管理扩大数据汇聚范围的同时,也放大了安全风险。结构化数据与文档性数据的融合,意味着敏感信息更集中、攻击面更宽广、泄露后果更严重。因此,安全体系必须从“外围防御”升级为“内生安全”。

区块链技术为电子档案真实性保障提供了新的解决方案。中国电建开发了自主可控的区块链基础平台,选取7个节点搭建电建联盟链。其核心设计思路是:从电子文件自形成起触发区块链技术,将对电子文件进行修改的一切行为作为交易进行存证,建立以哈希值识别、以管理环节上链信息为基础的电子档案溯源多重认证体系。这一机制打通了电子文件从形成系统到档案系统管理的全过程,从源头上为电子文件真实性及全生命周期可追溯提供保障。

嘉兴市档案馆同样运用区块链全程保真技术,通过逐条哈希计算并将哈希值存储于多节点联盟链,构建了从数据生成到归档的全流程可信溯源与防篡改机制。海南省电子文件档案单套制应用云平台则整合了多租户数据隔离、区块链可信认证与OCR文本识别等关键技术,实现电子文件归档、管理、移交、利用全流程数字化规范管理。

档案行业通行的质量验证框架——涵盖真实性核验、完整性检查、可用性测试与安全性评估四大维度——构成了安全保障体系的操作化抓手。中车青岛四方的实践显示,通过应用40余项安全保障技术,实现了电子档案在移交与接收、归档、长期保存三大环节45个检测项全覆盖。象山项目针对不同来源的数据文件,研发了离线接收、在线归档、长期保存等不同场景下的电子文件四性检测功能。安全保障体系贯穿档案采集、整理、传输、保存四大环节,维护档案数据全生命周期安全性。

备份策略是安全体系的最后防线。重要电子档案应实行异地异质备份,中车青岛四方在公司(青岛)与中车集团(北京)之间定期实行异地异质备份。数据分级存储策略则根据数据访问频率和重要程度自动调整存储层级,提高热点数据访问速度,降低冷数据存储成本。

4 关键技术应用与创新趋势

4.1 区块链:可信存证与溯源

区块链技术在电子档案管理中的应用日益深入。从中国电建的实践来看,区块链不仅解决了电子文件真实性保障难题,还开拓了跨单位、跨系统应用区块链交互的场景。在建设项目中,区块链系统帮助建立一个多方参与的电子文件协作管理可信环境,可实现设计、施工、监理、运营文件无缝流转和文档一体化管理。

区块链应用的关键在于“上链数据”的设计。并非所有数据都需要上链,而是将电子文件办理完毕、修改、归档、迁移等关键行为作为交易上链存证。嘉兴市的实践进一步验证了区

区块链在结构化数据归档中的应用价值——通过逐条哈希计算，构建数据全流程可信溯源机制。

4.2 人工智能：智能鉴定与知识挖掘

当前，人工智能技术正通过两条路径深度介入电子档案管理变革：其一是推动业务流程自动化，其二是促进档案内容知识化。

在流程自动化维度，AI 技术显著提升了归档鉴定与检索效率。中国电建基于档案鉴定要求和大数据思想设计计算机程序，形成一套适用于计算机自动判定保管期限的规则和逻辑，打通了由计算机自动处理电子文件归档的“最后一公里”。江西省档案馆的实践显示，人工智能辅助开放审核实现了 25% 的一审替代率、20% 的二审替代率，非直接替代的人工智能审核总准确率为 70.68%，大大提高了档案开放审核工作效率。

在内容知识化维度，AI 技术将静态档案转化为动态知识库。江西省档案馆通过构建人物知识图谱、关键词知识图谱和专题知识图谱，实现档案资源的语义关联和可视化展示。某智能化档案数字化平台整合知识图谱技术，构建档案领域知识库，实现档案资源的语义关联和智能推荐。OCR 识别技术的进步同样显著——基于深度学习的 PaddleOCR 识别工具持续优化，文字识别率、准确率和耗时等指标不断提升，某平台的档案文本识别准确率已提升至 98%。

4.3 数据封装：长期保存的关键技术

审视结构化数据的长期保存实践，最突出的障碍莫过于其对原生系统的深度依赖——当数据脱离既有的业务运行环境，其原本清晰的语义内涵与逻辑关联可能随之湮没。XML 结构化封装技术提供了解耦方案：将数据库表结构、业务逻辑与数据实体打包成自描述、自解释的标准化信息包，使数据彻底脱离原业务系统环境，确保其长期可读可用。

这一技术的本质是“数据+元数据”的联合封装。封装包不仅包含原始数据，还包含描述数据含义、结构、上下文的元数据，以及数据之间的关联关系。即使原业务系统早已关停，封装包依然可以被解析、理解和利用。嘉兴市的实践表明，这一技术对于关停系统数据抢救作用突出，已为 6025 人次提供精准查询服务，维护了 480 人的养老保险权益。

4.4 趋势展望：从数字化向智慧化跃迁

当前档案管理正站在从数字化向智慧化跃迁的关口。江西

省档案馆的研究者指出，档案数据作为新生产要素，只有依靠信息技术才能充分挖掘其内在价值。在当前物联网、云计算、大数据、人工智能、深度学习等技术普遍应用的新形势下，档案智慧化是必然选择。

智慧档案的核心特征在于“数据驱动决策”。传统的档案管理以“保管”为中心，智慧档案则以“知识服务”为中心。通过大数据分析预判利用需求，通过知识图谱揭示隐性关联，通过智能推荐推送个性化内容，档案从“死”数据变为“活”知识。象山项目的经验表明，当档案管理与工程建设深度融合、双向赋能，档案人员与工程人员名字并列出现在创新成果奖状上时，档案的价值得到了最有力的彰显。

展望未来，随着大语言模型的爆炸性发展，档案智慧化面临新的机遇与挑战。如何在保证档案数据绝对安全的前提下，将新技术导入档案工作领域，探索智能辅助编研、智慧检索、自动摘要等全新应用场景，是全体档案工作者需要共同回答的时代命题。

5 结语

构建能够统筹管理结构化数据与文档性数据的电子档案系统，其复杂性远超单纯的技术选型范畴，而是交织着技术架构创新、管理模式变革与认知理念跃迁的多重挑战。这意味着档案工作者需要完成三重转变：在观念层面，将认知焦点从“文件实体保管”转向“数据资产运营”；在方法层面，以融合治理思维替代传统的条块分割模式；在职能定位层面，从被动响应的档案保管者转型为主动赋能的智慧服务提供者。

从浙江嘉兴的馆室一体化实践到中国电建的区块链应用，从江西档案馆的智慧数据探索到中车青岛四方的模型驱动归档，从国电象山的海上风电试点到海南全省的单套制云平台，档案行业正在以多元路径回答同一命题：如何让档案从数字世界的“记忆”升华为“智慧”。这些实践共同揭示：融合管理不是技术堆砌，而是制度、平台、流程、安全的系统重构；不是一蹴而就的工程，而是持续迭代的旅程；不是档案部门的独角戏，而是全员参与的协奏曲。

“十四五”时期是档案事业数字化转型的关键阶段。构建融合管理的电子档案系统，既是落实国家战略的必然要求，也是释放档案价值的根本途径。期待更多机构在实践中探索创新，共同书写档案管理现代化的新篇章。

参考文献：

- [1] 冯佳,徐潇潇,莫思敏.浙江嘉兴馆室一体化再升级[N].中国档案报,2025-09-29.
- [2] 李燕明,王成海,刘向阳,等.集成创新,融合应用新一代信息技术提升电子档案管理效率[EB/OL].国家档案局,2024-07-04.
- [3] 陈泽玺.某智能化档案数字化平台的架构与性能分析[J].机电兵船档案,2025(4):32-34.
- [4] 江西省档案馆.档案大数据分析应用[J].中国档案,2023(8).

- [5] 李永妮,王西山,韩彩夏.如何实现模型驱动式电子档案智能收集与管理——以中车青岛四方机车车辆股份有限公司的实践为例[N].中国档案报,2025-08-04.
- [6] 嘉兴市档案馆.嘉兴市档案馆构建原生数据归档管理新机制[EB/OL].中国档案网,2025-12-25.
- [7] 陈娉婷,裘蓉.当“数字化种子”播撒在这片海上——全国首个海上风电建设项目电子文件归档和电子档案管理试点建设纪实[N].中国档案报,2025-09-08.
- [8] 俞佳,江若飞.铁路电子档案海量非结构化数据存储方案[J].兰台世界,2025(2).
- [9] 海南省档案局.省档案局电子文件档案单套制应用云平台入选国家档案局档案信息化创新案例[EB/OL].海南省人民政府,2025-11-28.