

T/JXEA

江西省工程师联合会团体标准

T/JXEA 355—2026

档案数字资源封装与长期保存格式要求

Requirements for encapsulation and long-term preservation formats of archival digital resources

（征求意见稿）

2026—XX—XX 发布

2026 - XX- XX 实施

目次

前 言 II

引 言 III

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 总则 2

5 不同类型数字资源的格式策略 2

5.1 公文文件的格式选择与处理 2

5.2 文献资料的格式要求 2

5.3 影像档案的长期保存格式 3

5.4 音频档案的编码与保存 3

5.5 视频档案的多层保存策略 3

5.6 数据库档案的多源保存 3

6 资源封装与元数据组织 3

7 格式迁移的管理规范 4

8 真实性保护与审计管理 4

9 风险预案与应急处置 4

附 录 A（规范性） 实施细则 5

附 录 B （规范性） 档案数字资源封装与长期保存实施指南 7

前 言

本文件依据GB/T 1.1-2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由***提出。

本文件由江西省工程师联合会归口。

本文件起草单位：

本文件主要起草人：

引 言

随着档案数字化转型不断深化，档案数字资源类型日趋多元，涵盖文本、多媒体、数据库、地理信息等多种形态，其长期保存的可读性、真实性与安全性面临严峻挑战。当前，格式不统一、封装不规范、技术迭代导致的可读性衰减等问题，制约了档案价值的有效传承。为规范档案数字资源管理，依据《档案法》及相关国家、行业标准，特制定本规范。

本规范明确了各类档案数字资源的格式标准化要求、资源封装规则、格式迁移流程及安全保障措施，覆盖各类档案保管单位及资源全生命周期管理，为档案数字资源的规范保存、有效利用提供统一技术指引，保障档案数字资源跨代际安全传承。

档案数字资源封装与长期保存格式要求

1 范围

本规范对档案数字资源的格式标准化和资源封装提出了统一的技术要求，适用范围涵盖国家档案部门、地方各级档案馆、企业档案室、高校档案馆、医疗卫生机构档案部门、司法机关档案部门等各类档案保管单位。

本规范涉及的数字资源类型包括但不限于：数字扫描获得的文本档案、原生数字形式的电子政务文件、多媒体类档案、结构化的数据库档案、地理信息系统数据、遥感影像数据、科研数据档案等。

本规范的适用阶段涵盖档案入库前的格式检查与标准化处理、入库后的保存维护、利用过程中的格式兼容性检测、以及出现技术风险时的应急处理。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本文件。凡是不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 22080 信息安全管理体系要求

GB/T 35273 信息安全技术 个人信息安全规范

DA/T 31 档案管理体系

DA/T 64 数字档案室建设指南

DA/T 65 电子文件管理系统功能要求

《档案法》（2020年12月修订）

3 术语和定义

3.1

档案数字资源形态 archival digital resource form

指档案以二进制数据形式存在的各类表现形式，包括静态图像、动态视频、音频流媒体、文本编码、结构化数据等多种类型。[来源：DA/T 64，3.1]

3.2

格式标准化 format standardization

指按照国际或国家标准将档案数字资源统一为规定的文件格式的过程，目的是确保档案在长期保存中的可读性和可交换性。[来源：DA/T 65，3.2]

3.3

资源封装 resource encapsulation

指将档案内容文件、配套元数据、技术文档、质量验证信息组织成逻辑关联的整体的过程，形成独立的、自描述的档案对象。[来源：DA/T 31，3.3]

3.4

版本管理链 version management chain

指记录档案数字资源从原始采集、格式转换、内容修改到最终保存的完整演变过程的信息追踪体系。[来源：GB/T 22080，3.4]

3.5

可读性风险 readability risk

指因技术过时、软件停维、硬件淘汰等原因导致数字资源无法被访问或理解的风险状态。[来源：GB/T 35273，3.5]

3.6

格式兼容层 format compatibility layer

指为应对某类格式可能面临淘汰而预先创建的中间格式副本或转换规则库。[来源：DA/T 64，

3.6]

3.7

元数据关联度 metadata relevance

指元数据与档案原始内容在逻辑和物理层面的紧密程度指标。

3.8

数据链完整性 data chain integrity

指在档案数字化全过程中保留所有操作记录和中间件，使得档案的来源和演变过程可以完全追溯。[来源：《中华人民共和国档案法》（2020年12月修订），第三条]

4 总则

档案机构在进行数字资源格式选择时应遵循实用性、前瞻性、开放性、可审计性综合原则。实用性原则要求选择的格式应能直接满足档案的存储和检索需求；前瞻性原则要求预测相关格式在今后10年至20年的生存周期；开放性原则要求优先选择技术规范公开透明的格式，避免因格式垄断导致的技术绑定；可审计性原则要求所有格式选择和转换过程都应形成可追踪的审计记录。档案机构应成立数字资源格式管理专门小组，由档案业务人员、技术人员、设备厂商代表等组成，定期评估本机构数字资源的格式现状，制定格式优化升级计划。对于新增的数字资源，入库前必须进行格式预审，评估该格式是否符合本规范的要求。档案机构应建立格式转换的标准操作流程，包括转换前的兼容性测试、转换过程的全程监控、转换后的质量验证三个关键环节。所有参与格式管理工作的人员应每年接受不少于40小时的培训，内容包括新兴数字格式知识、格式转换技术工具操作、档案真实性保护等。

5 不同类型数字资源的格式策略

5.1 公文文件的格式选择与处理

公文档案是档案资源的核心组成部分，包括各类行政文件、会议记录、决议决定等。这类档案具有较强的法律效力和参考价值。公文档案原始格式通常为Word、Excel或专有的办公系统格式。为确保长期保存，应采用如下格式策略：对于具有正式印章和签署的公文原件扫描件，应使用TIFF-RGB或TIFF-灰度格式保存，分辨率300DPI，采用LZW无损压缩。对于已电子化的公文文本，应转换为PDF/A-1或PDF/A-2格式，禁止使用嵌入式脚本和外部链接。针对需要长期编辑维护的公文文档（如条例、规程等），应同时保留OpenDocument格式的源文档。公文档案不应采用Microsoft Office原生格式进行保存，因为该格式的版本更新频繁，向下兼容性差。

对于包含表格数据的公文档案，如财务报表、统计汇总等，应将表格部分独立提取为CSV或OpenDocument Spreadsheet格式，确保数据的结构化和可重用性。对于含有签名和印鉴影像的公文，应采用图文混合的处理方式：文字部分按PDF/A处理，印鉴部分保留原始分辨率的TIFF图像，两者通过XML元数据进行关联。

5.2 文献资料的格式要求

文献档案包括图书、期刊、论文、报告等各类学术和研究资料。这类档案的特点是内容复杂多样、格式差异大、参考价值持久。对于已出版的图书和期刊，如果档案机构拥有电子版本，应保留出版社的原始电子格式（通常为PDF或EPUB）作为第一优先级备份。如果仅拥有纸质版本，经扫描后应采用TIFF-RGB格式，分辨率根据文献的重要性分级：核心文献600DPI、重要文献400DPI、一般文献300DPI。

对于手稿、古籍等特殊文献，应采用超高分辨率扫描（1000DPI以上），保留原始色彩信息，并将扫描成果分别保存为原始TIFF和网络传播用的JPEG两个版本。文献中包含的图表、插图应单

独进行高质量扫描，确保细节清晰。对于特殊出版物如地图集、画册等，应采用多页面容器格式如PDF或EPUB，保留其页面间的逻辑关系。

5.3 影像档案的长期保存格式

照片、画像、设计图等影像档案是档案资源中极具历史和艺术价值的部分。这类档案面临的主要挑战是色彩信息的准确保留和细节的长期可读性。原始影像档案（如黑白照片、彩色幻灯片）应采用高保真扫描方式获得数字副本。黑白照片的扫描应采用至少16-bit灰度TIFF格式；彩色照片应采用RGB色彩空间、24-bit或32-bit色深的TIFF格式，并嵌入标准的色彩空间信息（如sRGB或Adobe RGB）。

对于设计类图纸，如建筑图、工程图等，应保留两个版本：一是高分辨率的TIFF栅格化副本用于长期保存；二是矢量格式（如SVG或标准化的CAD格式）副本用于编辑和更新。矢量格式应剥除软件专有扩展，仅保留国际标准定义的元素。对于需要色彩管理的影像档案（如艺术品照片、文物图像），应在TIFF文件中嵌入ICC色彩配置文件，确保在任何设备上的色彩再现一致。

5.4 音频档案的编码与保存

语音录音、音乐表演、口述访谈等音频档案是重要的民族文化和历史资料。音频格式的选择需要在音质保留和存储成本之间找到平衡。对于重要的语音档案和学术研究用录音，应采用WAV或FLAC格式保存，其中FLAC格式可将文件体积压缩至原始大小的60-70%，同时保证零信息丧失。录音的采样率应不低于44.1kHz，位深不低于16-bit。对于音乐档案，特别是民族音乐和非遗音乐，建议采用高保真规格：采样率48kHz或96kHz，位深24-bit，保存为FLAC或其他高保真无损格式。

对于日常听写、会议录音一般性音频档案，采用高比特率MP3（320kbps）或AAC（256kbps）格式作为主存版本，同时保留一份低质量的用于快速预览。对于存在版权保护需求的音乐档案，可考虑采用带有水印和权限信息的受保护格式，但必须同时保留一份无保护的版本用于长期保存。所有音频档案应在文件元数据中记录录音时间、地点、设备、采样参数等信息。

5.5 视频档案的多层保存策略

视频档案包括新闻素材、历史影像、教学视频、会议记录等，是重要的动态信息载体。鉴于视频文件体积大、编码复杂、依赖特定解码器的特点，应采用多层保存策略。第一层为高保真源文件，采用H.264编码、H.265编码或Apple ProRes编码，码率不低于15Mbps，分辨率与原素材保持一致。第二层为标准流通版本，使用H.264编码、6-10Mbps码率，便于网络传播和预览。第三层为低分辨率版本，用于快速检索和目录管理。

对于具有历史价值的视频档案，如果原始格式已过时（如Motion JPEG、Sorenson等），应进行格式迁移至H.264或H.265。迁移过程中应采用逐帧检验方式，对比原始版本和迁移版本的关键帧，确保无视觉差异。视频文件应采用MP4或MKV容器，后者可提供更丰富的元数据支持。对于包含字幕、评论或附加信息的视频档案，应将这些信息以独立的XML或SRT文件形式保存，并通过容器格式的多轨道功能与视频关联。

5.6 数据库档案的多源保存

经营数据、业务数据、统计数据等结构化档案通常存储于关系型或非关系型数据库中。这类档案的保存难点在于数据的复杂关联关系和对特定数据库引擎的依赖。为确保数据库档案的长期可用，应采用“原始格式+导出格式+文档化”三层策略。原始层保留数据库的原生备份文件和日志，确保数据的完整恢复能力。导出层将数据导出为开放格式：关系型数据库导出标准SQL脚本文件；非关系型数据库如MongoDB导出为JSON格式，文档库导出为XML格式。

文档化层应包含完整的数据库架构文档：包括表结构定义、字段类型说明、索引设置、约束条件、存储过程说明等。这些文档应采用Markdown或HTML格式，便于长期阅读和理解。对于涉及隐私信息的数据库档案，在导出开放格式时应进行适当的脱敏或匿名处理，在元数据中记录处理方式和脱敏规则。定期（至少每年一次）对数据库备份进行恢复测试，验证备份的可用性和完整性。

6 资源封装与元数据组织

资源封装是指将档案文件、描述信息、技术文档、质量证明等相关信息按照标准结构组织在一起的过程。档案机构应采用标准化的包装结构来管理数字资源。在实际操作中，可采用ZIP容器或

TAR等实现方式。结构应包括档案文件目录、manifest文件（校验码列表）、metadata.xml（包的元数据）等关键组件。

元数据应使用标准化结构，采用XML或JSON格式。最低元数据集应包括档案标题、创建时间、创建者、内容摘要、档案类型、保存级别、访问限制等描述性元数据；页面顺序、文件关系、目录结构等结构性元数据；数字签名、校验码、格式信息、转换历史等保存性元数据。元数据应与档案文件一起保存，而不是单独存放在档案管理系统中。对于重要档案，应在元数据中添加质量验证信息：扫描参数、转换工具版本、验证人员署名、验证时间等。

7 格式迁移的管理规范

档案数字资源保存期间必然会面临格式过时的问题。格式迁移应遵循预防性原则，即在格式风险出现之前主动进行迁移，而非被迫应急。档案机构应建立格式生命周期评估机制，对现有格式的技术活力进行评估：技术规范是否仍在更新、相关软件是否还在维护、硬件支持是否广泛等。根据评估结果制定迁移计划。通常，广泛应用的开放标准格式（如PDF、PNG、TIFF）的预期生命周期为15-20年，应在第10-12年进行迁移计划评估。

格式迁移前应进行充分的试验工作：选择具有代表性的样本档案进行小规模迁移试验，验证目标格式的可行性和转换工具的质量。转换应采用自动化工具，但关键环节需人工检验。对于大规模档案的迁移，应分批进行，优先迁移重要档案，其次迁移易损坏的档案，最后迁移一般档案。迁移过程应保留完整的日志记录：迁移时间、涉及档案数量、转换工具及版本、转换参数、质量检验结果等。迁移后应长期保留原始格式的至少一份副本作为参考。

8 真实性保护与审计管理

档案的真实性是其最核心的价值属性。数字保存过程中，应采取多种措施防止档案被篡改。对于重要档案和涉及法律效力的档案，应在入库时计算并保存密码学校验码（使用SHA-256或更强的算法）。后续任何对档案的访问或修改都应重新计算校验码，如发现校验码不匹配，应立即启动调查程序。对于需要法律认可的档案，应采用数字签名技术，由档案部门管理员或授权人员对档案进行签名，使用符合国家标准数字证书。

档案机构应建立全面的审计日志系统，记录所有档案的生命周期事件：何人何时访问了哪些档案、进行了何种操作、结果如何。审计日志本身也应受保护，防止被篡改。对于涉及国家秘密或商业秘密的档案，审计日志应采用加密存储。定期（至少每月）对审计日志进行审查，检查是否有异常访问行为。审计日志保存期限应与相应档案相同。

9 风险预案与应急处置

档案数字保存中可能面临多种风险：硬件故障、软件缺陷、格式过时、安全事件、自然灾害等。档案机构应制定详尽的风险预案。针对硬件故障风险，应采用冗余存储策略，关键数据应至少存储在两个独立的物理设备上，且地理位置相距足够远。针对格式过时风险，建立格式监测机制，定期评估和更新；同时为可能过时的格式预先研究替代方案。针对安全事件风险，采用访问控制、加密存储、定期备份等措施。

档案机构应每年进行至少一次的灾难恢复演练，验证备份的可用性和恢复流程的有效性。演练应包括全量恢复和部分恢复两种场景。演练过程应形成详尽的总结报告，发现问题应及时整改。如果在演练中发现数据损坏或恢复失败的情况，应立即进行原因分析和技术改进。

附录 A

(规范性)

档案数字资源封装与长期保存格式要求现场考核
实施细则

A.1 考核环境要求

- A.1.1 现场考核应在联合会考核工作委员会指定的考核场所进行。
- A.1.2 考试场所的建筑、安全、电力、照明、消防等设施须符合国家有关标准、规定。
- A.1.3 考核场所应具备应急安全疏散条件，具有由有关部门鉴定合格的安全区域和应急疏散通道。
- A.1.4 考试场所能与非考试场所分开，考试期间能够实行封闭管理。
- A.1.5 考核场所应具有专业工程能力考核所需的电脑、网络、显示设备、测试仪器等考试业务系统。

A.2 考核流程

- A.2.1 考核方案由联合会考核工作委员会事先审批通过，考官由考核工作委员会从考官专家库中选任。
- A.2.2 由考官现场核查申请人员的身份信息是否与报名信息相符，申请人员与考官一起签字确认。
- A.2.3 根据考核类别和级别，对同批次考核人员采用统一题、自动随机、现场抽题等多种考核模式。
- A.2.4 由考官分配考核位，申请人就位后，考官验证考核初始状态后，宣布考核开始并启动计时器。
- A.2.5 考核时间到，考官即时确认申请人已脱离对考核设备的操控，并确认考核记录保存正确。
- A.2.6 申请人离场，考官复核考试成绩，在评定单上签字确认。
- A.2.7 联合会根据考核评定单，以及申请人员的申请材料初审情况，公布考核结果。

A.3 评分规则

- A.3.1 现场考核每个项目的分数由考核方案规定，各个项目总分数应为100分。
- A.3.2 现场考核项目为实际操作或演示，每个项目均有明确评价指标，达成为满分，未达成为零分。
- A.3.3 现场考核项目均规定有时间限制，申请人超时后仍未脱离对考核设备的操控，视为零分。
- A.3.4 现场考核的项目可以采用图文答题方式，但此类项目的分数占比不能超过50%。
- A.3.5 现场考核的所有项目总评分超过80分以上（含80分），视为考核合格。
- A.3.6 原则上现场考核的每个项目只有零分和满分两个结果，但如果现场两位考官均同意给予特殊评分，则由现场两位考官和申请人一同写明原因并签字确认，封存所有现场记录，上报联合会的考核工作委员会复核，复核通过后评分成绩有效。

A.4 可复核性要求

- A.4.1 申请人不允许携带任何外接的存储介质、计算设备等进入考场。
- A.4.2 考前，考官应复核考核设备和系统，确保无连接外网、预存答案、特殊工具等情况，并签字确认。
- A.4.3 考核过程中，不允许申请人将考核设备与任何非指定设备进行数据交互。
- A.4.4 考核过程由现场数字摄像头、自动录屏软件等进行考场记录，该记录应保留半年。
- A.4.5 所有申请人现场编辑的工程代码、电路光路设计、美术工程等数字内容，均应保存2个月，必要时录制现场运行视频或拍照存证。
- A.4.6 现场考核的争议处理

- A. 4. 7 考核场所应张贴考核工作委员会指定的考场纪律，并由考官现场宣读。
- A. 4. 8 违反考场纪律者，考官可即时终止其考核资格，并记录在考核评定单上。
- A. 4. 9 现场考核期间，因考场故障导致考核中断，考官可延长考核时间，并记录在考核评定单上。
- A. 4. 10 考核出现重大异常，考官可立即上报考核工作委员会备案，申请更换在线考核的日程、场所、试卷等，并将考核工作委员会的决定通知申请人。
- A. 4. 11 考核争议和处置，均以考核场所的录屏、录像、录音等获取的数据信息作为证据，考官的书面、口头申明为重要参考资料。
- A. 4. 12 考核工作委员会对所有的考核争议具有最终决定权。

A. 5 成绩公布

- A. 5. 1 现场考核成绩应在联合会考核工作委员会指定的登录网址进行公布和查询。
- A. 5. 2 考核成绩，联合会根据现场考核评分数据汇总以及考核评定单进行成绩审核并公布。
- A. 5. 3 申请人对现场考核成绩有异议，可在成绩公布后15天内向联合会考核工作委员会提请复核，联合会考核工作委员会应在10个工作日内完成复核，并公布复核结论。

附 录 B

(规范性)

档案数字资源封装与长期保存实施指南

B.1 范围

本附录提供了档案数字资源封装与长期保存的参考性实施指南，包括常用保存格式推荐列表、资源封装结构示例、元数据模板示例、格式迁移计划模板、常见格式过时风险评估表以及相关参考文献。本附录旨在帮助档案保管机构更好地理解和应用本规范的主体要求，不作为强制性规定。

B.2 常用长期保存格式推荐列表

表 B.1 长期保存格式推荐列表

资源类型	推荐保存格式	备选/过渡格式	说明
文本公文	PDF/A-1、PDF/A-2	TIFF（扫描件）	PDF/A是ISO标准化的长期保存格式，禁止依赖外部内容。
扫描图像	TIFF（未压缩或LZW压缩）	JPEG2000（无损）	保持原始分辨率（300-600DPI），嵌入ICC色彩配置文件。
数字照片	TIFF、DNG（数字负片）	JPEG（高质量备份）	原始相机RAW格式应转换为DNG或TIFF保存。
矢量图形	SVG、PDF/A	DXF（标准交换格式）	避免使用专有格式（如AI、CDR）。
音频	WAV、FLAC	MP3（320kbps）	重要录音采用48kHz/24bit规格。
视频	MXF、MKV（含FFV1编码）	MP4（H.264）	采用无压缩或帧内压缩编码，保留原始帧率。
数据库	SQL（标准脚本）、XML、CSV	原生备份+文档化	同时保存数据库结构和数据，以及完整的文档说明。
电子邮件	EML、MBOX	PDF/A（单封打印版）	保留原始邮件头、附件和元数据。
网页	WARC（Web ARChive）	PDF/A（截图备份）	保存完整的网页资源（HTML、图片、脚本）及其关系。
地理信息	GeoJSON、GeoTIFF、GML	Shapefile（需文档化）	确保坐标系信息完整保存。