

河南省档案馆纸质档案数字化技术规范

1 范围

本标准规定了河南省档案馆纸质档案数字化的主要技术要求。

本标准适用于河南省档案馆采用扫描设备对纸质档案数字化加工处理及数字化成果的管理,适用于省直向省档案馆移交纸质档案数字化成果的单位。各市、县级综合档案馆参照执行。

2 规范性引用文件

本规范根据中华人民共和国档案行业标准DA/T 31-2017《纸质档案数字化规范》,结合河南省档案馆纸质档案数字化工作实际情况制定。

3 术语和定义

GB/T 20530—2006、DA/T 1-2000、DA/T 18-1999界定的以及下列术语和定义适用于本文件。

3.1 数字化 digitization

利用计算机技术将模拟信号转换为数字信号的处理过程。

3.2 数字图像 digital image

表示实物图像的整数阵列。一个二维或更高维的采样并量化的函数,由相同维数的连续图像产生。

3.3 纸质档案数字化 digitization of paper-based records

采用扫描仪等设备对纸质档案进行数字化加工,使其转化为存储在磁带、磁盘、光盘等载体上的数字图像,并按照纸质档案的内在联系,建立起目录数据与数字图像关联关系的处理过程。

3.4 分辨率 resolution

单位长度内图像包含的点数或像素数,一般用每英寸点数(dpi)表示。

4 总则

4.1 河南省档案馆根据档案的珍贵程度、开放程度、利用率、亟待抢救程度、数字化资金情况等因素统筹规划、科学开展纸质档案数字化工作。制定数字化规划与年度计划,有目标有计划地开展数字化工作。

4.2 纸质档案数字化的基本环节主要包括：档案出库、数字化前处理、目录数据库建立、档案扫描、图像处理、数据挂接、数字化成果验收与移交、档案归还入库，具体流程见附录A。

4.3 应采取有效的管理和技术手段，确保纸质档案数字化成果质量。纸质档案数字化应遵循档案管理的客观规律，真实反映档案内容，最大程度地展现档案原貌。

4.4 纸质档案数字化过程中，应保存数字化项目信息、技术环境、数字化各类技术参数等方面的元数据。元数据元素的确定应符合ISO/TR 13028提出的要求。

4.5 应加强纸质档案数字化各环节的安全管理，确保档案实体和档案信息的安全。各个环节应该进行详细登记，每个环节都应该通过数字化加工管理表单（见附录B）跟踪记录生产情况、质量情况，每个环节在处理前应将实物与工单进行核查，发现问题及时退回纠正。

4.6 加工涉密档案时，应按照涉密档案相关保密要求开展工作。

5 加工场地规划

河南省档案馆应配备专用加工场地，并进行合理布局，形成档案存放、数字化前处理、档案著录、档案扫描、图像处理、质量检查等工作区域。

加工场地的选择及温湿度等环境的控制应有利于档案实体的保护。场地内应配备可覆盖全部场地的防火、防水、防有害生物、防盗报警、视频监控等安全管理的设施设备。

档案数字化各项工作必须在河南省档案馆指定的场地完成，不得将档案原件、扫描图片等相关文件以任何方式带出场地，整理后剔除的无保存必要的文件材料由河南省档案馆按规定进行处理。

6 档案出库

6.1 河南省档案馆数字化管理人员根据工作进度安排从库房提调档案，档案保管部门工作人员负责将档案移交给档案提调人员。

6.2 数字化管理人员接收档案后，应对照案卷移交目录认真核查档案数量。

6.3 核验完毕后，数字化管理人员需对核验情况进行登记，填写《河南省档案馆纸质档案数字化调卷、还卷登记表》，并提交档案保管部门工作人员签字确

认。

6.4在核验过程中若出现异常情况（如：数量有偏差、案卷密封、案卷严重破损等情况），应立即通知档案保管工作人员进行处理。

7 数字化前处理

7.1 确定扫描页

原则上应将确定为数字化对象的纸质档案全部扫描，不宜进行挑扫。如有不需要扫描的，把同一案卷（卷、件）中的扫描件和非扫描件区分开，不扫描的档案应与扫描档案一同著录案卷目录和卷内文件目录，并做出相应标识，保持在档案中的原有位置，同扫描档案一同编号、排序。

7.2 编制页号

7.2.1 应对没有页号或页号不正确的档案重新编制页号。

7.2.2 重新编制页号时，应在统一位置书写页号，且不压盖档案内容。

7.2.3 书写页号所使用的笔、墨等不应破坏档案原件或对档案长期保存造成影响。

7.2.4 应将破损页面、缺页等特殊情况进行登记。

7.3 拆除装订

应以最小干预原则确定是否拆除装订。如因扫描需要必须拆除装订物，应注意保护档案不受损害，并对排列顺序不准确的档案进行重排。特殊装订且拆除装订后需恢复的档案，在拆除装订物时应采用拍照等方式记录档案原貌，以便于恢复。

7.4 技术修复

7.4.1 破损严重或其他无法直接进行扫描的纸质档案，应先由专业技术人员进行技术修复。

7.4.2 折皱不平影响扫描质量的纸质档案应先进行压平等相应技术处理。

7.4.3 特殊纸张（如：已破碎的文件）将根据实际情况进行必要的处理，如复印、加薄膜等，以防止在扫描工序中对纸张造成二次伤害。

8 目录数据库建立

8.1 按照《河南省档案馆文书档案目录数据库结构规范》，对照档案原件，

对每件档案进行著录，建立起目录数据库。

8.2 目录数据库应选择通用的数据格式，以便于数据交换。

8.3 数据库结构的设计应特别注意保持档案的内在联系，有利于纸质档案数字化成果的管理和利用。

8.4 将纸质档案数字化前处理工作中对纸质档案目录进行修改、补充的结果录入数据库，形成准确、完整的目录数据。

8.5 可采用计算机自动校对与人工校对相结合的方式，对目录数据的质量进行检查，包括著录项目的完整性、著录内容的规范性和准确性等。发现不合格的数据应及时进行修改。

9 档案扫描

9.1 基本要求

档案扫描应根据纸质档案原件实际情况、数字化目的、数字化规模、计算机网络和存储条件等选择相应的扫描设备，和进行相关参数的设置和调整。参数的设置和调整应保证扫描后数字图像清晰、完整、不失真，图像效果最接近档案原貌。

9.2 扫描设备

9.2.1 扫描设备的选择应特别注意对档案实体的保护，尽量采用对档案实体破坏性小的扫描设备进行数字化。

9.2.2 采用A3平板扫描仪或其他适用类型扫描仪按档案编定的次序逐卷逐页扫描。超过A3幅面的档案，必须根据档案页面选择大幅面扫描仪另行扫描。

9.2.3 对于极其珍贵且尺寸不规则的档案，为方便直观显示原件大小，可采用标板、标尺等方式标识原件大小等信息。

9.2.4 应遵循相关设备的使用规律进行定期维护、保养。

9.3 扫描色彩模式

9.3.1 为最大限度保留档案原件信息，便于多种方式的利用，宜全部采用彩色模式进行扫描。

9.4 扫描分辨率

9.4.1 扫描分辨率的选择，应保证扫描后图像清晰、完整，并综合考虑数字图像后期利用方式等因素。扫描分辨率不低于300 dpi。

9.4.3 需要进行COM输出的档案，扫描分辨率建议不小于300 dpi；需要进行高精度仿真复制的档案，扫描分辨率不小于600 dpi；需要进行印刷出版的档案，可结合档案幅面、印刷出版幅面、印刷精度要求等选择合适的分辨率。

9.5 存储格式

9.5.1 图像文件采用单页TIFF或JPEG，图像压缩率的选择可根据实际应用的需求而定。

9.5.2 纸质档案数字图像利用时，也可从网络浏览速度、易操作性、存储空间占用等方面进行综合考虑，将图像转换为OFD、PDF等其他格式。

9.5.3 同一批档案应采用相同的存储格式。

9.6 图像命名

9.6.1 应以原件档号为基础对数字图像命名。确保图像命名的唯一性。

9.6.2 将数字图像存储为单页文件和多页PDF格式，并按档号与图像流水号的组合对图像命名。

9.6.3 应科学建立纸质档案数字图像的存储路径，确保数据挂接的准确性。

10 图像处理

10.1 图像拼接

对分幅扫描形成的多幅数字图像，应进行拼接处理，合并为一个完整的图像，以保证纸质档案数字图像的整体性。拼接时应确保拼接处平滑地融合，拼接后整幅图像无明显拼接痕迹。

10.2 旋转及纠偏

对不符合阅读方向的数字图像应进行旋转还原。对出现偏斜的图像应进行纠偏处理，以达到视觉上基本不感觉偏斜为准。

10.3 裁边

如需对数字图像进行裁边处理，应在距页边最外延至少2至3毫米处裁剪图像。

10.4 去污

如需对数字图像进行去污处理，以去除在扫描过程中产生的污点、污线、黑边等影响图像质量的杂质，应遵循展现档案原貌的原则，处理过程中不得去除档案页面原有的纸张褪变斑点、水渍、污点、装订孔等痕迹。

10.5 图像质量检查

10.5.1 数字图像不完整、无法清晰识别或图像失真度较大时，应重新扫描。

10.5.2 对于漏扫、重扫、多扫等情况，应及时改正。

10.5.3 数字图像的排列顺序与档案原件不一致时，应及时进行调整。

10.5.4 对数字图像拼接、旋转及纠偏、裁边、去污等处理情况进行检查，发现不符合图像质量要求时，应重新进行图像处理。

11 数据挂接

11.1 应借助相关软件对数据库中的目录数据与其对应的纸质档案数字图像进行挂接，以实现目录数据与数字图像的关联。

11.2 逐条对挂接结果进行检查，包括目录数据与纸质档案数字图像对应的准确性、已挂接数字图像与实际扫描数量的一致性、数字图像是否能正常打开等，发现错误及时进行纠正。

12 数字化成果验收与移交

12.1 验收方式

12.1.1 河南省档案馆成立专门的验收组对纸质档案数字化成果进行验收。

12.1.2 应采用计算机自动检验与人工检验相结合的方式对纸质档案数字化成果进行验收检验。

12.2 验收内容

12.2.1 纸质档案数字化成果包括数字图像、档案目录数据、元数据、数字化工作中产生的工作文件、存储载体等。

12.2.2 应对目录数据进行验收，主要包括数据库中各条目的内容、格式等的准确程度、必填项是否填写等。

12.2.3 应对元数据进行验收，主要包括元数据元素的完整性和赋值规范性等。

12.2.4 应对数字图像进行验收，主要包括数字化参数、存储路径、命名的准确性、图像的完整性、排列顺序的准确性、图像质量等。

12.2.5 应对数据挂接进行验收，主要包括目录数据与其对应的数字图像的挂接的准确性等。

12.2.6 应对工作文件进行验收，主要包括工作文件的完整性、规范性等。

12.2.7 应对存储载体进行验收，主要包括载体的可用性、有无病毒等。

12.3 验收指标

能够采用计算机自动检验的项目应采用计算机自动检验的方式进行100%检验，检验合格率应为100%。对于无法用计算机自动检验的项目，可根据情况以件或卷为单位采用抽检的方式进行人工检验。抽检比率不得低于5%，对于数据库条目与数字图像内容对应的准确性，抽检合格率应为100%，其他内容的抽检合格率应不低于95%。

12.4 验收结论

12.4.1 每批纸质档案数字化成果质量检验达到本标准12.2和12.3的要求，予以验收“通过”。

验收未通过应视情况进行返工或修改后，重新进行验收。

12.4.2 验收完成后，验收组成员认真填写《纸质档案数字化验收登记表》，并签字确认。

12.5 移交

验收合格的数据应按照纸质档案数字化工作方案及时移交，并履行交接手续。

13 档案归还入库

13.1 档案装订还原

纸质档案数字化工作完成后，拆除过装订物的档案如需装订，应注意保持档案原貌，做到安全、准确、无遗漏。

13.1.1 档案装订应按照历史档案原有的装订方向进行，不可更换装订的位置。

13.1.2 对于装订时某些页面必须进行折叠的情况，应以尽可能地减少折痕数为原则。

13.1.3 档案装订必须采用专业装订线进行装订，不可使用金属装订物，如原始档案中发现金属装订物，应予以剔除。

13.1.4 档案装订，应遵循“两对齐”要求，即装订线一侧边缘对齐，档案内页下边缘对齐。

13.1.5 档案装订应尽可能地按照原来的装订孔位进行穿线装订，尽量不要新打孔装订，力求保护原件。

13.1.6 恢复装订时，要保持档案的排列顺序不变，不得漏页、错页，不压字，装订牢固、不掉页，做到安全、准确、无遗漏。

13.1.7 在还原过程中，如果发生问题，应进行异常情况登记。

13.2 档案归还入库

按照档案入库相关要求对纸质档案进行处理和清点，并履行档案入库手续。

附录 A 纸质档案数字化工作流程

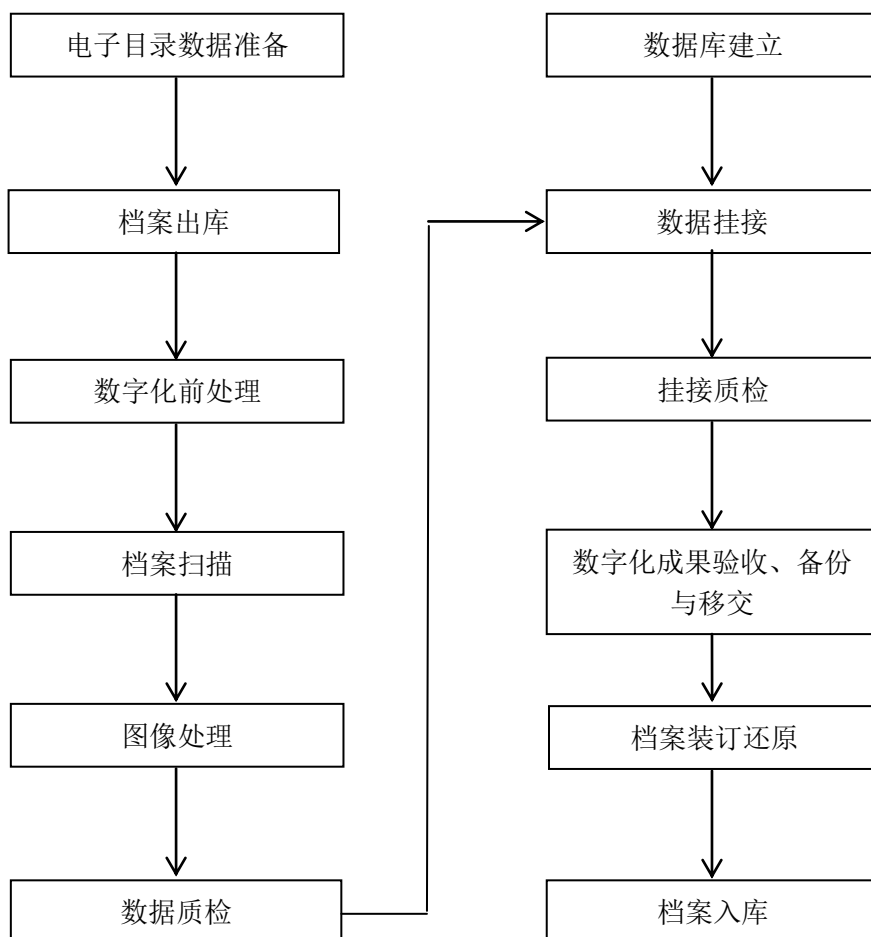


表 B.2 调卷、还卷登记表

河南省档案馆纸质档案数字化调卷、还卷登记表

调卷日期	全宗号	案卷号	还卷日期	暂未还档案	未还原因	调/还卷人	备注

表 B.6 数据质检登记表

河南省档案馆纸质档案数字化数据质检登记表

检验时间： 年 月 日

文件级档号	图 像				目 录		备 注
	裁切比例	不清晰	歪 斜	拼接不平	著录项齐全	准确性	

检验人：

审核人：

验收人：

表 B.8 数字化验收登记表

河南省档案馆纸质档案数字化验收登记表

批次： _____ 验收人： _____ 验收时间： _____ 年 _____ 月 _____ 日

全宗号	总卷数	图像数据						目录数据					元数据			数据挂接				工作文件				载体				验收意见			
		图像总数	计算机全检项	全检结果	抽检数	抽检项	抽检结果	条目总数	计算机全检项	全检结果	抽检条目数	抽检项	抽检结果	计算机全检项	全检结果	抽检项	抽检结果	计算机全检项	全检结果	抽检条数	抽检项	抽检结果	总册数	抽检条数	抽检项	抽检结果	载体类型		载体数量	检验项	检验结果
合计																															