

# 纸质档案数字化规范 DA/T31—2017

## 前 言

本标准按照 GB/T 1.1-2009 给出的规则起草。本标准替代 DA/T 31-2005 《纸质档案数字化技术规范》。

本标准与 DA/T 31-2005 相比，主要技术变化如下：

- 标题进行了修改；
- 增强组织与管理部分的内容，完善数字化工作中管理相关要求；
- 增强数字化前处理部分的内容，包括对实体档案保护和档案规范化管理方面的要求；
- 增加数字化过程中元数据采集的要求；
- 修改了档案扫描部分参数要求；
- 修改了图像处理部分内容，更加强调保持档案原貌的要求；
- 细化了数字化成果验收的内容；
- 删除原标准数据备份和数字化成果管理相关内容。 本标准由国家档案局提出并归口。

本标准起草单位：国家档案局档案科学技术研究所、国家档案局信息中心、国家档案局技术部。

本标准主要起草人：王良城、马淑桂、郝晨辉、程春雨、杜琳琳、蔡伟、宋涌、王大众、田军、曹燕、李华峰。 本标准所代替标准的历次版本发布情况为： ——DA/T 31—2005。

# 引 言

DA/T 31-2005 的发布实施，从技术标准方面对纸质档案数字化工作提出了要求，很好的促进了纸质档案数字化工作向科学化、规范化的方向发展，在档案信息化工作中发挥了重要作用。

由于纸质档案数字化工作所依赖的信息技术发展迅速，经过 10 年的时间，数字化设备、存储、网络等技术条件都发生了较大变化，同时，纸质档案数字化工作管理方法和管理理念等都在不断更新、发展。因此，及时对本标准进行修订，适时调整相关要求，从档案行业层面科学规范纸质档案数字化工作，具有重要的现实意义。

标准本次修订更加注重我国纸质档案数字化工作自身的特点，结合目前信息技术发展的水平，提出适用于档案行业的纸质档案数字化工作的规范性要求。

# 纸质档案数字化规范

## 1 范围

本标准规定了纸质档案数字化技术和管理要求。本标准适用于采用扫描设备对纸质档案的数字化加工过程的管理。

## 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 20530-2006 文献档案资料数字化工作导则 DA/T 1 档案工作基本术语 DA/T 18 档案著录规则

ISO/TR 13028 信息与文献 档案数字化实施指南（Information and Documentation — Implementation guidelines for digitization of records）

## 3 术语和定义

GB/T 20530-2006、DA/T 1、DA/T 18 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 各单位应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。纸质档案数字化工作的开展应遵循 ISO/TR 13028 和 GB/T 20530 提出的要求和建议。

4.2 纸质档案数字化的基本环节主要包括：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等。

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。

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. 组织与管理

### 5.1 机构及人员

5.1.1 应建立纸质档案数字化工作组织，对数字化工作进行统筹规划、组织实施、协调管理、安全保障、技术保障、监督检查、成果验收等，确保数字化工作的顺利开展。

5.1.2 应配备具有相应能力的工作人员，包括熟悉档案业务并具有较高的调查研究水平和良好的组织领导能力的管理人员，熟悉相关标准规范并能够为纸质档案数字化工作各环节提供技术支持的技术人员，掌握一定数字化基础知识并熟悉本职工作的操作人员等。应通过科学规范的管理制度，对工作人员进行规范化管理。为强化数字化工作的安全性，应加强对外聘工作人员的审核。

### 5.2 基础设施

5.2.1 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。

5.2.2 加工场地的选择及温湿度等环境的控制不应不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。

5.2.3 应合理规划、配备和管理纸质档案数字化设施设备，确保设施设备安全、先进，能够满足数字化工作的需要。

### 5.3 工作方案

5.3.1 应在充分调研的基础上，制定科学合理的工作方案，确保纸质档案数字化工作达到预期目标。

5.3.2 纸质档案数字化工作方案应包括数字化对象、工作目标、工作内容、成本核算、数字化技术方法和主要技术指标、验收依据、人员安排、责任分工、进度安排、安全管理措施等内容。数字化对象的确定应综合考虑档案的珍贵程度、开放程度、利用率、亟待抢救程度、数字化资金情况等因素。

5.3.3 宜对纸质档案数字化工作方案进行专家论证，确保其科学、规范、合理。

5.3.4 纸质档案数字化工作方案应经审批后严格执行。工作方案审批结果应与数字化工作过程中形成的其他文件一并保存。

### 5.4 管理制度

5.4.1 应制定科学化、规范化的管理制度，并在工作过程中严格执行，以有效保障档案安全和纸质档案数字化成果质量。

5.4.2 纸质档案数字化管理制度应包含岗位管理、人员管理、场地管理、设备管理、数据管理、档案实体管理等方面的制度。

岗位管理制度主要规定数字化工作各岗位的工作目标和职责，形成明确的岗位业务流程规范、考核标准、奖惩办法等；人员管理制度主要对人员的安全责任、日常行为、外聘人员信息审核及管理、非工作人员来访登记等进行规范；场地管理制度主要对人员出入和工作场地内基础设施、环境、网络、监控设施、现场物品、证件等的管理进行规范；设备管理制度主要对数字化工作各环节涉及的全部设备的管理进行规范；数据管理制度主要对数字化各环节所产生的数据的管理进行规范；档案实体管理制度主要对档案实体在数字化过程中的交接、管理、存放等工作进行规范。

### 5.5 工作流程控制

5.5.1 应依据相关的法律法规和各类技术标准，制定相关的工作流程和各环节操作规范等，对纸质档案数字化全过程进行有效的控制，确保数字化成果质量。纸质档

案数字化流程示例参见附录 A。

5.5.2 应加强对纸质档案数字化工作的全流程安全管理。

5.5.3 应建立完善的问题反馈机制，对纸质档案数字化工作过程中后端环节发现前端环节中产生的问题进行及时反馈和修正。

## 5.6 工作文件管理

5.6.1 应根据情况制定符合实际要求的纸质档案数字化工作文件，以此加强对数字化工作的管理。主要包括纸质档案数字化工作方案、纸质档案数字化审批书、纸质档案数字化流程单、数据验收单、项目验收报告、纸质档案数字化成果移交清单等，采取外包方式实施时，还应包括项目招标文件、投标文件、中标通知书、项目合同、保密协议等。部分工作单示例参见附录 B。

5.6.2 应加强对纸质档案数字化工作文件的管理，明确数字化工作过程中形成的工作文件的整理、归档、移交等管理要求。

## 5.7 档案数字化外包

5.7.1 纸质档案数字化工作如需外包，档案部门应从企业性质、股东组成、安全保密、企业规模、注册资金情况等方面严格审查数字化加工企业的相关资质；按照 GB/T 20530-2006 第 5 章的要求评估数字化加工企业的技术能力；从规章制度的建立健全程度等方面考查加工企业的管理能力。

5.7.2 在项目实施过程中，应依据《档案数字化外包安全管理规范》（档办发[2014]7 号），从档案部门、数字化服务机构、数字化场所、数字化加工设备、档案实体、数字化成果移交接收与设备处理等层面执行严格的安全管理要求。

5.7.3 档案部门应指派专门人员参与纸质档案数字化外包业务的监督、指导，完成质量监控、进度监控、投资监控、安全监控和协调沟通等方面的工作。

## 6 档案出库

6.1 档案保管部门应按照纸质档案数字化工作方案确定的数字化对象开展档案调取、清点、登记等前期准备工作，并提交档案出库申请，经相关责任人批准后，严格按照档案库房管理规定为数字化对象办理出库相关手续，并与数字化部门共同清点无误后，对档案进行交接出库。

6.2 纸质档案数字化过程中，应设置距离数字化加工场所较近的保管库用以临时存

放纸质档案，并对纸质档案的领取与归还进行严格管理，认真做好检查、清点、登记等工作，确保纸质档案的安全。

## 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。如有不需要扫描的页面应加以标注。

### 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，且不应压盖档案内容。

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 目录数据准备

7.3.1 按照目录数据库建立时制定的数据规则，对照档案原件内容，规范档案中的目录内容。

7.3.2 对需在目录数据库中进行标记的情况进行标记。

### 7.4 拆除装订

应以对纸质档案的保护为原则确定是否拆除装订。如需拆除装订物，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。

### 7.5 技术修复

7.5.1 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。

7.5.2 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。

## 8 目录数据库建立

8.1 应制定目录数据库数据规则，包括数据字段长度、字段类型、字段内容要求等。目录数据库数据规则的制定应符合 DA/T 18 对档案著录的要求。在纸质档案目录准备与目录数据库建立工作中均应严格遵守。

8.2 数据库选择应考虑可转换为通用数据格式，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，并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 超出所使用扫描仪扫描尺寸的档案可采用更大幅面扫描仪进行扫描，也可以采用小幅面扫描仪分幅扫描后进行图像拼接的方式处理。分幅扫描时，相邻图像之间应留有足够的重叠，并且采用标版等方式明确说明分幅方法；若后期采用软件自动拼接的方式，重叠尺寸建议不小于单幅图像对应原件尺寸的 1/3。

9.2.3 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。

9.3.2 页面中有红头、印章或插有照片、彩色插图、多色彩文字等的档案，应采用彩色模式进行扫描。

9.3.3 页面为黑白两色，并且字迹清晰、不带插图的档案，也可采用黑白二值模式进行扫描。

9.3.4 页面为黑白两色，但字迹清晰度差或带有插图的档案，也可采用灰度模式扫描。

## 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。

9.4.2 扫描分辨率应不小于 200dpi。如文字偏小、密集、清晰度较差时，建议扫描分辨率不小于 300dpi。

9.4.3 如有 COM 输出、仿真复制、印刷出版等其他用途时，可根据需要调整扫描分辨率。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，扫描分辨率建议不小于 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 纸质档案数字图像长期保存格式为 TIFF、JPEG 或 JPEG2000 等通用格式，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 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 应以档号为基础对数字图像命名。图像命名方式的选择应确保图像命名的唯一性。

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

9.6.3 数字图像确需存储为多页文件时，可采用该档案的档号对图像命名。

9.6.4 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 建议档案部门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 验收完成后须经验收组成员签字。验收“通过”的结论，必须经相关领导审核、签字后方有效。

## 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。移交单示例参见附录 B。

## 13 档案归还入库

### 13.1 档案装订

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。

### 13.2 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。