

纸质档案数字化规范

Specification for digitization of paper-based records

2018 - 11 - 07 发布

2019 - 01 - 01 实施

天津市市场监督管理委员会 发布

目 次

前言	II
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 纸质档案数字化原则	2
5 纸质档案数字化加工组织方式	2
6 纸质档案数字化加工基础设施	2
7 纸质档案数字化加工机构和人员	2
7.1 数字化加工机构	2
7.2 数字化工作人员	2
8 纸质档案数字化外包	2
9 纸质档案数字化项目管理	2
9.1 管理制度	3
9.2 流程控制	3
9.3 文档管理	3
10 纸质档案数字化加工流程	3
10.1 制定纸质档案数字化加工方案	3
10.2 档案出库	3
10.3 数字化前处理	3
10.4 目录数据库建立	4
10.5 元数据采集	4
10.6 数字化扫描	5
10.7 图像处理	5
10.8 纸质档案数字化图像光学字符识别（OCR）	6
10.9 版式转换	6
10.10 数据挂接	6
10.11 数据验收	6
10.12 档案归还入库	7
11 数字化成果移交	7
12 数据备份	7
附录 A（资料性附录） 纸质档案数字化加工流程示例	8
附录 B（资料性附录） 数字化管理登记表示例	9

前 言

本标准按照GB/T 1.1-2009给出的规则起草。

本标准由天津市档案局提出并归口。

本标准起草单位：天津市档案馆。

本标准主要起草人：方昀、仇伟海、王丽春、朱雅晶、陈雪。

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的一般过程与数字化过程中涉及的技术及管理要求等。
本标准适用于本市各级各类档案馆、机关团体企事业单位的纸质档案数字化工作。

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 文献档案数字化工作导则
DA/T1-2000 档案工作基本术语
DA/T 18 档案著录规则
DA/T 22 归档文件整理规则
DA/T31-2017 纸质档案数字化规范
档办发[2014]7号 档案数字化外包安全管理规范

3 术语和定义

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扫描加工,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、元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。

3.5

光学字符识别 optical character recognition

电子设备检查纸上显示的字符,通过检测暗、亮的模式确定其形状,然后用字符识别方法将形状翻译成计算机文字的过程。

3.6

识别准确率 recognition accuracy

通过OCR识别正确文字的比率，识别准确率=（识别正确文字数 / 应识别文字总数）×100%。

4 纸质档案数字化原则

根据本单位档案的珍贵程度、破损情况、利用频率及资金等情况进行统筹安排，分步实施。
加工涉密档案时要按照国家相关保密要求进行实施。

5 纸质档案数字化加工组织方式

根据扫描数量和档案保密程度分别采取本单位自行组织扫描或外包给数字化服务机构的方式。

6 纸质档案数字化加工基础设施

- 6.1 配备专用加工场地并根据实际工作情况合理布局，形成档案存放、数字化前处理、数字化加工、数据验收等区域。
- 6.2 加工场地的选择和温湿度等环境的控制应适合档案实体的保管、存放。
- 6.3 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，保证供电稳定。
- 6.4 配备必要的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。

7 纸质档案数字化加工机构和人员

7.1 数字化加工机构

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、技术支持、监督检查、成果验收等。

7.2 数字化工作人员

数字化工作应配备熟悉档案业务、具有较强的协调沟通和组织领导能力的管理人员，熟悉相关标准规范并能对数字化工作各环节及计算机网络、设备提供技术支持的技术人员，掌握一定数字化工作技术的操作人员。

8 纸质档案数字化外包

8.1 纸质档案数字化工作如需外包，应从数字化服务机构的企业性质、股东组成、企业规模、注册资金、安全保密、工作业绩等方面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，同等条件下优先选用具有档案数字化加工相关涉密资质的服务机构。

8.2 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）要求执行。

8.3 档案部门数字化工作组织的工作人员要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。

9 纸质档案数字化项目管理

9.1 管理制度

9.1.1 纸质档案数字化管理制度包括岗位管理、人员管理、设备管理、场地管理、档案实体管理、数据管理、安全保密制度等。

9.1.2 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、操作规范、工作标准、考核标准、奖惩办法等。

9.1.3 人员管理制度主要对人员的出入、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。

9.1.4 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，确保设备安全、科学、高效运行。

9.1.5 场地管理制度主要对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。

9.1.6 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。

9.1.7 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。

9.1.8 安全保密制度主要对档案实体的安全保密管理及数字化工作人员在数字化加工过程中的安全保密行为进行规范。

9.2 流程控制

制定纸质档案数字化加工流程并根据工作流程制定各环节的操作规范。对纸质档案数字化全流程的安全、质量、进度要加强管理和控制，在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，确保数字化成果的质量。纸质档案数字化加工流程参见附录A。

9.3 文档管理

纸质档案数字化过程中工作文档包括：纸质档案数字化方案、纸质档案数字化审批书、档案实体出入库登记单、数字化前处理登记单、数字化工作流程单、数字化成果验收单、数字化成果交接单等。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。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。工作单示例参见附录B。

10 纸质档案数字化加工流程

10.1 制定纸质档案数字化加工方案

根据实际情况制定科学合理的纸质档案数字化加工方案，方案中应包括纸质档案数字化对象、数字化技术标准、数字化工作流程、质量控制标准、验收标准、人员分工、进度安排、安全管理等内容。

10.2 档案出库

档案保管部门应根据纸质档案数字化加工方案中确定的数字化对象，对档案实体进行调取、清点、登记等准备工作。并根据库房管理要求经过相关审批程序后，与数字化加工部门共同清点确认，办理相关出库手续，进行档案交接出库。

10.3 数字化前处理

10.3.1 确定扫描页

确定为数字化对象的纸质档案原则上应全部扫描。如有特殊情况，如破损严重、重份文件等确实不能扫描或不需要扫描的页面要在“数字化前处理登记单”中加以标注。

10.3.2 纸质档案整理

10.3.2.1 应以不破坏纸质档案为原则拆除装订物，对于没有页号或页号不正确的档案要重新编制页号。

10.3.2.2 在编制页号过程中应将破损页面、缺页、原件不清等特殊情况进行登记。对于文件顺序混乱的情况要及时进行调整，破损、折皱不平页面要进行必要的技术处理。

10.3.2.3 编写页号时，应以不破坏档案原件为原则，在统一位置书写页号，确保档案页号准确、规范、无遗漏。

10.3.3 目录核校

对于已建立目录数据库的数字化加工对象，应依照目录数据库建立时制定的数据规则，对照档案原件内容，规范目录数据库中目录内容，如发现错误应进行修改。

10.4 目录数据库建立

10.4.1 应按照 DA/T 18 的要求制定目录数据库数据规则，包括字段内容、字段类型、字段长度等。

10.4.2 采用人工校对或软件自动校对的方式，对目录数据库的建库质量进行检查。核对著录项目是否完整、著录内容是否规范、准确，发现不合格的数据应进行修改或重录。

10.5 元数据采集

应同时采集数字化过程中相关元数据，主要包括数字化授权信息、数字化软硬件信息、数字化日期与时间、格式、扫描分辨率、色彩空间、压缩方案等。数字化加工软件应以件为单位自动捕获元数据，若软件无法自动捕获应采取手工著录的方式给元数据赋值。元数据项及著录要求见表1。

表1 元数据项

序号	元数据名称	类型	著录要求
1	数字化时间	日期型	著录纸质档案数字化每份文件的起止时间 格式为：YYYY-MM-DDhh:mm:ss/YYYY-MM-DDhh:mm:ss
2	数字化对象描述	字符型	记录需要特别说明的案卷物理特征等信息。
3	数字化授权信息	字符型	著录纸质档案数字化审批信息。包括《纸质档案数字化审批书》的编号、数字化对象等。
4	色彩空间	字符型	著录扫描图像所应用的色彩空间
5	扫描分辨率	字符型	著录扫描分辨率
6	扫描色彩模式	字符型	著录扫描色彩模式
7	压缩方案	字符型	著录扫描图像的压缩算法
8	压缩率	字符型	著录扫描图像的压缩比率
9	存储格式	字符型	著录扫描图像的存储格式
10	扫描设备类型	字符型	著录纸质档案数字化所用扫描设备的类型
11	扫描设备制造商	字符型	著录纸质档案数字化所用扫描设备的制造商
12	扫描设备型号	字符型	著录纸质档案数字化所用扫描设备的型号
13	数字化软件名称	字符型	著录纸质档案数字化所用软件的名称

14	数字化软件版本	字符型	著录纸质档案数字化所用软件的版本
15	数字化软件生产商	字符型	著录纸质档案数字化所用软件的生产商
16	阅读所需软硬件条件	字符型	著录纸质档案数字化成果阅读所需的软硬件条件
17	数字化成果移交接收信息	字符型	著录纸质档案数字化成果的移交接收信息。包括《纸质档案数字化成果交换单》的批次号、全宗号、内容描述等。

10.6 数字化扫描

10.6.1 数字化扫描基本要求

10.6.1.1 扫描后数字图像要清晰、完整、不失真。

10.6.1.2 若纸张幅面超出扫描仪尺寸，可分幅扫描后再进行拼接，相邻图像之间应留有足够的重叠部分，并且在“纸质档案扫描工作单”中明确说明分幅方法。

10.6.2 数字化扫描技术要求

10.6.2.1 扫描色彩模式

一般情况采用彩色模式进行扫描；页面为黑白两色，字迹清晰并且不带插图的档案，也可采用黑白二值模式进行扫描；页面为黑白两色，字迹清晰度差或带有插图的档案，也可采用灰度模式扫描。

10.6.2.2 扫描分辨率

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。一般文件扫描分辨率建议不小于300dpi。如原件质量较差或字体很小、字迹密集，可适当提高分辨率。

10.6.2.3 存储格式

纸质档案数字图像存储格式可选择TIFF、JPEG、JPEG2000、PDF和OFD，长期保存格式为TIFF、JPEG或JPEG2000等通用格式。存储时压缩率的选择，应在满足实际应用需求的前提下，尽量减小存储容量。同一批档案应采用相同的存储格式。

10.6.3 数字图像命名规则

图像命名方式的选择应确保图像命名的唯一性和存储路径的规律性。如，可以档号为基础对数字图像命名和确定存储路径。

10.7 图像处理

10.7.1 拼接处理

应对分幅扫描的图像进行拼接处理，以保证图像的完整性。拼接处信息要完整，拼接平滑，无明显拼接痕迹。

10.7.2 图像纠偏

数字化图像经纠偏处理后偏斜角度不超过1度，并符合阅读习惯。

10.7.3 裁边处理

数字化图像裁边处理时，应在原始档案的边界最外延2至3毫米处裁剪，保留原始档案的边界。

10.7.4 去污处理

对数字图像进行去污处理，应遵循保持档案原貌的原则，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，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。

10.7.5 图像质量检查

图像质量检查包括数字图像清晰度、失真度、完整性等检查，漏扫、重扫、多扫等情况检查，数字图像拼接、旋转及纠偏、裁边、去污等处理情况检查，数字图像排列顺序检查等。

10.8 纸质档案数字化图像光学字符识别（OCR）

10.8.1 对于数字化图像光学字符识别率高、有实际工作需要的，可在纸质档案数字化的基础上开展此项工作。

10.8.2 要先对图像质量进行总体评估，必要时对图像进行二值化、图像增强、降噪处理等预处理，以提高识别准确率。

10.8.3 可根据实际使用需要选择识别范围。

10.8.4 输出格式可采用双层PDF，命名见10.6.3。

10.9 版式转换

在对纸质档案数字化原始图像进行利用时，可将每一份纸质档案文件扫描所得的一个或多个图像通过版式转换存储为一份文件，并可进行适当的压缩，但要保证图像清晰可打印。一般采用PDF、OFD等通用格式作为利用格式。命名见10.6.3。

10.10 数据挂接

10.10.1 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。

10.10.2 应对挂接结果进行检查，检查内容应包括目录数据与纸质档案数字化图像对应的准确性、挂接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。

10.11 数据验收

10.11.1 验收内容

10.11.1.1 成立专门的验收小组对数字化成果进行验收。

10.11.1.2 目录数据（含元数据项）验收，包括各条目的内容、格式，必注项是否填写齐全，元数据各元素完整性和赋值规范性等。

10.11.1.3 数字图像验收，包括数字化参数、图像的完整性、图像命名的准确性、排列顺序的准确性、图像质量等。

10.11.1.4 数据挂接情况验收，主要是目录数据与其相对应的数字图像的挂接的准确性等。

10.11.1.5 数字化工作文件验收，主要是数字化过程中形成的工作文件的完整性、规范性。

10.11.1.6 存储载体验收，包括载体的安全性、可用性等。

10.11.2 验收指标

采取计算机自动检验与人工抽检相结合的方式验收，可利用计算机检验的项目要100%进行检验，且合格率要达到100%。人工抽检项目的抽检率一般不低于10%，数据库条目与数字图像内容对应的准确性抽检合格率要达到100%，其他内容的抽检合格率要不低于98%。

10.11.3 验收结论

10.11.3.1 符合本标准 10.11.1 和 10.11.2 要求的数据给予验收“通过”，没有达到要求的数据要退回数字化相应加工环节进行返工或修改，重新验收，直到验收“通过”。

10.11.3.2 验收通过后经验收组签字确认后方可进行数据移交。

10.12 档案归还入库

10.12.1 档案装订

数据验收合格后对于拆除装订的纸质档案要重新装订，并在备考表中标注数字化相关信息，如数字化年度、外包公司、特殊说明等。重新装订要保持档案原貌。

10.12.2 纸质档案验收

纸质档案重新装订后要进行逐一验收，主要包括文件装订顺序、是否存在漏页、压字、页面颠倒等情况。对于验收合格的档案给予验收“通过”，验收不合格的纸质档案要重新整理装订直到验收通过。验收通过后，经验收组成员签字确认方可归还库房。

10.12.3 档案入库

纸质档案验收合格后，按照库房管理要求，与库房管理人员对纸质档案进行清点，办理相关入库手续，归还入库。

11 数字化成果移交

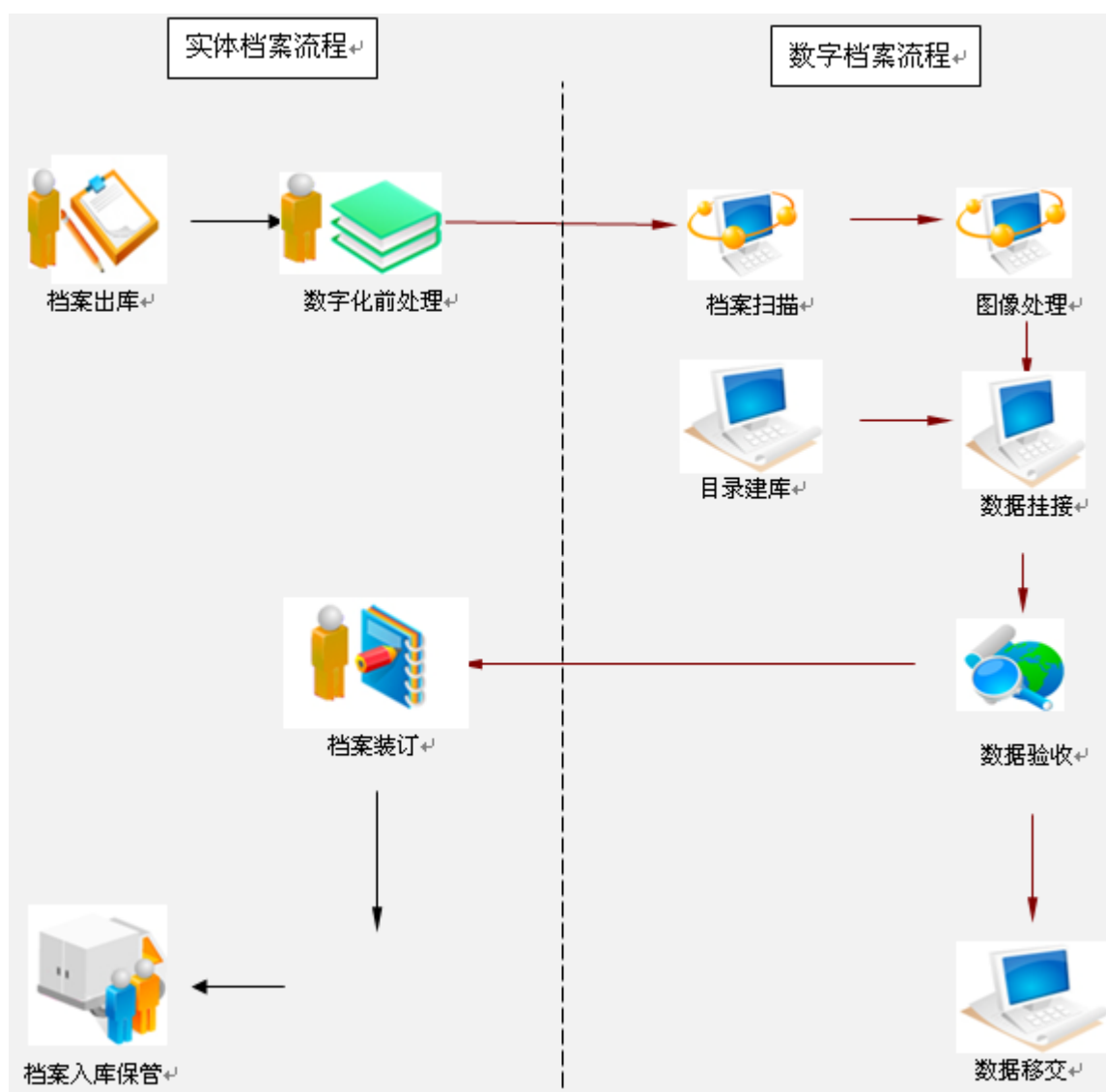
11.1 目录数据文件（含元数据项）以 DBF、XML、XLS 等常用格式进行移交，数字图像以 TIFF、JPEG、JPEG2000 或 PDF、OFD 等格式进行移交，版式转换及 OCR 识别后的文件以 PDF、OFD、双层 PDF 等通用格式进行移交。数据移交单示例参见附录 B。

11.2 数字化过程中的工作文档应作为数字化工作卷与验收合格后的数据一并移交给接收部门。数字化工作卷包括的材料详见 9.3。

12 数据备份

数据保管部门要对接收的数据及时进行备份，备份的载体应多样化，可采用在线、离线相结合的方式实现多套备份和异地备份等，以保证数据的安全。

附录 A
 (资料性附录)
 纸质档案数字化加工流程示例



图A.1 纸质档案数字化加工流程

纸质档案数字化流程单见图B.2。

纸质档案数字化流程单

全宗号 保管期限 目录号 案卷号

进度	<input type="checkbox"/> 数字化前处理 <input type="checkbox"/> 前处理质检 <input type="checkbox"/> 数据库建立 <input type="checkbox"/> 目录质检 <input type="checkbox"/> 档案扫描 <input type="checkbox"/> 图像处理 <input type="checkbox"/> 图像质检 <input type="checkbox"/> OCR 识别 <input type="checkbox"/> 版式转换 <input type="checkbox"/> 数据挂接 <input type="checkbox"/> 挂接质检 <input type="checkbox"/> 数据验收 <input type="checkbox"/> 装订还原															
	数字化前处理			前处理 质检	数据库建 立（数据 录入含元 数据项）	目录 质检	档案 扫描	图像 处理	图像 质检	OCR 识别	版式 转换	数据 挂接	挂接 质检	数据 验收	装订 还原	纸质档 案验收
整理 编页	目录核 校	技术 修复														
完成 人员																
完成 时间																

图B.2 纸质档案数字化流程单

数字化前处理登记单见图B. 4。

数字化前处理登记单

档号	编页起止号	需扫描页数	是否存在不需扫描页	特殊情况记录	待修复页	修复完成情况	质检情况说明
档案整理人： 完成时间： 档案修复人： 完成时间： 档案检验人： 完成时间：							

图B. 4 数字化前处理登记单

版式转换、数据挂接工作单见图B.9。

版式转换、数据挂接工作单

全宗号 保管期限 目录号

案卷起止号	案卷总数	文件总数	版式转换开始时间	版式转换完成时间	数据挂接完成时间	特殊情况记录	质检情况说明
版式转换、数据挂接人： 质量检验人：			完成时间： 完成时间：				

图B.9 版式转换、数据挂接工作单

纸质档案数字化验收登记表见图B.10。

纸质档案数字化验收登记表

批次：

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				验收人	验收时间	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	完整性	规范性	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	抽检册数	抽检项				抽检结果						
合计			—	—		—	—		—	—		—	—																					
<p>该批次数据已通过验收，可以进行数据交换。</p> <p style="text-align: right;">负责人签字：</p> <p style="text-align: right;">年 月 日</p>																																		

图B.10 纸质档案数字化验收登记表

纸质档案数字化成果移交单见图B. 12。

纸质档案数字化成果移交单

批次				
全宗号				
内容描述				
移交数字图像数量 (页)		移交条目数 (条)		数字化工作文档 (件、册)
载体起止顺序号		移交载体类型、规格		
检验内容	单位名称			
	移交单位:		接收单位:	
准确性检验				
完整性检验				
可用性检验				
安全性检验				
载体外观检验				
填表人(签名)	年 月 日		年 月 日	
审核人(签名)	年 月 日		年 月 日	
单位(印章)	年 月 日		年 月 日	

图B. 12 纸质档案数字化成果移交单