



# 纸质档案数字化 的流程与方法



ZHIZHI DANGAN SHUZHUA  
DE LIUCHENG YU FANGFA

国家档案局组织编写



中国文史出版社

# 纸质档案数字化的流程与方法

国家档案局组织编写

中国文史出版社

## 图书在版编目 (CIP) 数据

纸质档案数字化的流程与方法 / 国家档案局组织编写. —北京: 中国文史出版社, 2018. 5

ISBN 978-7-5205-0323-5

I. ①纸… II. ①国… III. ①档案工作-数字化-研究-中国 IV. ①G270.7

中国版本图书馆 CIP 数据核字 (2018) 第 122202 号

责任编辑: 詹红旗

---

出版发行: 中国文史出版社

网 址: [www.wenshipress.com](http://www.wenshipress.com)

社 址: 北京市西城区太平桥大街 23 号 邮编: 100811

电 话: 010—66173572 66168268 66192736 (发行部)

传 真: 010—66192703

印 装: 廊坊市海涛印刷有限公司

经 销: 全国新华书店

开 本: 787 毫米×990 毫米 1/16

印 张: 6

字 数: 60 千字

版 次: 2018 年 6 月北京第 1 版

印 次: 2018 年 6 月第 1 次印刷

定 价: 18.00 元

---

文史版图书, 版权所有, 侵权必究。

文史版图书, 印装错误可与发行部联系退换。

## 《档案人员培训多媒体教材》编审委员会

主任：李明华

副主任：胡旺林 王绍忠 付 华 刘鲤生

主 编：杜 梅

编审委员会办公室：

杜 梅 刘剑波 刘振华 梁 琨

## 《纸质档案数字化的流程与方法》编写委员会

主 编：孙兆伟

副主编：张大伟

编 委：范 韫 蒋红莉 张 莉 王 玮

## 出版说明

加强档案人员培训是提高档案工作水平的基础。为进一步规范档案人员培训工作，提高档案人员培训质量，国家档案局组织编写制作了《档案人员培训多媒体教材》（以下简称《教材》）。

《教材》以党和国家有关档案工作的政策、法律法规、规定以及档案工作的有关标准、规范为依据，将档案管理工作中的基本原则和技术要求有机结合，充分体现科学严谨、指导规范、立足实际、面向一线。《教材》按照有关档案工作任务的步骤和流程，详细介绍并演示每个工作环节的规范与要求等，指导学习者准确掌握该项工作的基本技能。

《教材》采用多媒体教学演示光盘和纸质载体教材相结合的形式，二者各自独立又互为补充。《教材》以多媒体教学演示光盘为主，重点对实际操作内容进行形象、生动、规范、准确的演示，逐项演示需完成工作任务的具体操作方法、操作要领、操作程序及注意事项等，使学习者直观了解“做什么”“怎么做”“做到什么程度”的问题。纸质载体教材作为演示光盘的辅助形式，进一步说明“为什么这么做”等问题。

《教材》是由国家档案局组织编写的。每个科目采用主编负责制，由制作单位完成编写拍摄，国家档案局《教材》编审委员会对各科教材进行终审。

本教材的使用对象主要是有关档案人员，在编写工作中注重规范性、操作性、实用性，坚持以实际工作需求为导向，以规范和提高岗位技能为重点，使教材易教易学，既规范准确，又通俗易懂。

由于多媒体教材的编写制作时间仓促，教材中难免会有疏漏和不完善之处，敬请各使用单位或个人将使用中发现的问题或建议及时反馈给我们（电话：010—55605241），以便我们修订完善。

国家档案局《教材》编审委员会办公室

2018年6月

# 目 录

第一章 纸质档案数字化概述	1
第一节 概念及意义	1
一、相关概念	1
二、意义	3
第二节 基本要求	5
一、基本原则	5
二、对象确定	6
三、基本环节	7
第三节 组织加工方式	9
一、内部加工	9
二、外包加工	11
三、部分外包加工	15
第四节 组织与管理	15
一、配备机构人员	15
二、配备基础设施	17
三、制订工作方案	19



## 2 纸质档案数字化的流程与方法

四、落实标准规范 .....	19
五、严格执行安全保密制度 .....	21
六、提供保障机制 .....	25
七、过程管理目标 .....	26
<b>第二章 纸质档案数字化前处理 .....</b>	<b>32</b>
<b>第一节 档案出库 .....</b>	<b>32</b>
一、档案出库前的准备工作 .....	32
二、档案的交接 .....	35
<b>第二节 档案的拆装整理 .....</b>	<b>37</b>
一、确定扫描页 .....	37
二、清查档案内容 .....	38
三、目录数据准备 .....	38
四、拆除装订 .....	39
五、纸张处理 .....	40
六、技术修复 .....	41
七、前处理登记 .....	41
<b>第三章 纸质档案数字化数据采集 .....</b>	<b>44</b>
<b>第一节 设备选择 .....</b>	<b>44</b>
一、设备种类 .....	44
二、综合选择 .....	46
<b>第二节 参数设置 .....</b>	<b>47</b>
一、显示器参数 .....	47
二、扫描图像类型及色彩模式 .....	48

三、扫描分辨率 .....	49
四、亮度、对比度和曝光 .....	50
五、存储格式 .....	50
六、压缩率 .....	52
第三节 数字化操作 .....	52
一、设备操作 .....	52
二、图像命名 .....	53
三、登记备案 .....	54
第四章 纸质档案数字化数据处理 .....	56
第一节 图像处理 .....	56
一、纠偏 .....	56
二、去污 .....	57
三、图像拼接 .....	57
四、裁边处理 .....	57
五、图像优化 .....	58
六、OCR .....	58
第二节 图像质量检查 .....	58
一、质检 .....	58
二、检查记录 .....	60
第五章 纸质档案数字化数据存储 .....	62
第一节 存储格式的选择 .....	62
第二节 文件存储的组织 .....	63
第三节 目录数据库建立 .....	64

## 4 纸质档案数字化的流程与方法

一、制定数据规范 .....	64
二、数据库格式选择 .....	64
三、设计库结构 .....	65
四、档案著录 .....	65
五、数据录入 .....	66
六、检查数据质量 .....	66
第四节 数据挂接 .....	68
一、数据挂接的含义及作用 .....	68
二、数字图像的调阅形式 .....	68
第六章 纸质档案数字化后处理 .....	70
第一节 数字化成果验收 .....	70
一、验收方式 .....	70
二、验收内容 .....	71
三、验收指标 .....	73
四、验收结论 .....	74
第二节 档案装订与归还入库 .....	76
一、档案装订还原 .....	76
二、档案的清点交还 .....	76
第三节 数字化成果移交 .....	77
第四节 数字化成果管理 .....	79
一、数据管理 .....	79
二、数据备份 .....	80
三、软硬件系统维护 .....	82
四、加工文档管理 .....	84

# 第一章 纸质档案数字化概述

## 第一节 概念及意义

### 一、相关概念

数字化是指利用计算机技术将模拟信号转换为数字信号的处理过程。纸质档案数字化是指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。

纸质档案数字化加工方式一般有两种：第一种是将纸质档案文件的每一页看成是一张图片，由扫描仪或数码翻拍仪等设备按页扫入或摄入计算机系统制成数字图像文件；第二种是通过计算机人工录入或通过扫描仪扫描由字符识别软件（OCR）识别，制成文本文件。纸质档案信息

## 2 纸质档案数字化的流程与方法

的记录方式是多种多样的，有的字体是手写体，有的字体是繁体字，有的字迹退化，还有图表、图像等形式，此时采用第二种方式进行数字化加工就会暴露出如下弱点：采用人工录入，需要大量的录入人员，投入大，收效慢。用OCR字符识别年代较久远的档案、手写体的准确率低。而采用第一种方式进行数字化加工制作容易，工作人员只要稍加培训，即可操作设备，差错率相对较低，即便发现错扫、漏扫等现象，也能很快被纠正。另外，因为数字图像保持了档案原貌，可以使利用者在看到文字内容的同时，也能看到纸质档案的其他一些特征，如原件的墨迹、印章、圈点、批注等各种印记，使利用者可以获得更多的线索，开展一些更深入的研究。纸质档案数字化通常的做法是：档案目录采用电子计算机人工录入数据库；档案文件的内容采用扫描或拍摄制作成数字图像；视需要将数字图像经OCR识别转换为双层PDF，上层为图像，下层为文本，为后续信息化利用、数据化共享、全文检索和智能挖掘等技术的实施奠定基础。

纸质档案数字化过程中还将出现如下概念：

### 1. 数字图像

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。

### 2. 交接核查

对交接档案的完整性进行检查，记录并反馈缺页、重页、破损，以及案卷（文件）目录、备考表、封面上的错

误等现象。

### 3. 档案修整

根据检查及反馈的情况对档案实体错误进行纠正处理,包括重编页码、补编页码、纠正案卷(文件)目录、备考表等。

### 4. 数据导入

将案卷目录、文件目录及数字图像等正确导入到档案管理系统,确保条目(目录)与图像(全文)完全对应。

## 二、意义

目前大部分档案保管机构库藏档案载体仍然是以纸质为主,传统载体的档案占用的存储空间大,保管成本高,使用过程中易磨损,并且检索效率低,难以满足人们日渐增强对信息资源的需求。纸质档案数字化的目的是为了能够更好地保护档案原件,提高档案服务效率,方便利用者使用档案。与传统的档案服务模式相比,在档案检索、提供利用、保护原件方面,数字化档案都有着得天独厚的条件。具体来讲,纸质档案数字化有如下好处:

### 1. 有利于档案信息的传输,实现档案信息共享

数字化档案信息可以通过网络等现代技术实现异地的远程利用,为需求者提供方便快捷的档案信息服务。档案数字化能改善档案信息的传输方式,使档案的利用不再受“孤本”的限制,一份文件可以同时提供给所有需要它的人共

#### 4 纸质档案数字化的流程与方法

享。档案信息的利用时效性和档案内在价值可以得到充分的释放和体现，可以实现档案工作价值的增值。

##### 2. 有利于保护档案原件

档案的唯一性、原始性和凭证性决定了档案是真实记录历史的重要载体，因此保护原始档案安全和可用性是档案工作的重要内容之一，档案数字化之后，档案利用方式从利用原件变成主要使用档案数字副本，档案原件从而得到了更加充分和有效的保护。档案数字副本可以异地保存，这样可使这些档案在遇到突发性灾害时避免遭到毁灭性破坏。

##### 3. 有利于档案利用者方便快捷地查阅档案

档案工作的最终目的除保存历史记录以外主要是为了利用。这是档案工作生存发展的基础。传统档案利用工作是档案工作者根据利用需求，结合档案整理的相关规则和检索方法，以及档案排架规则调出原始档案提供利用，数字化后档案信息的利用工作转变成由授权档案利用者，根据自身的调档需求，组建检索关键词在网络上查阅档案信息，减轻了档案工作者的劳动强度，方便了档案利用者的查阅，提高了利用工作的效率。

##### 4. 有利于档案信息资源的开发利用

利用档案数字化信息可以高效开展档案信息资源的开发利用工作。通过网络终端随时查阅、整合、编辑丰富的档案信息可以大大提高档案信息利用效率。数字化信息，不仅为档案管理人员开展编研工作提供了方便，还为档案利用者提供了根据业务和专业需要开发利用档案信息的途径。

## 5. 有利于电子政务和无纸化办公

随着国家电子政务的不断发展和推广，无纸化办公正被普遍应用，根据国家对档案信息化建设的进一步规范，在相关法律允许的范围内，纸制档案的数字副本可替代原件直接用于电子政务，与电子档案共同组成数字档案信息库，为电子政务和无纸化办公创造条件。

## 第二节 基本要求

### 一、基本原则

纸质档案数字化目的是使档案信息资源准确、方便、快捷地提供利用，使可以公开的档案信息资源得到共享，以满足社会对档案利用的需求。为此应坚持以下基本原则：一是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。二是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。三是纸质档案数字化过程中，应保存数字化项目信息、技术环境信息和数字化各类技术参



## 6 纸质档案数字化的流程与方法

数。四是应加强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。五是加工涉密档案时，应按照涉密档案相关保密要求开展工作。

### 二、对象确定

纸质档案数字化对象的确定原则是应当对所要进行数字化的对象按照一定的原则和方法进行确认，只有符合一定要求的纸质档案文献才能进行数字化。

其中“一定的原则”是指：

1. 符合国家法律法规的原则：纸质档案的数字化，必须符合国家档案开放、知识产权保护以及信息安全保密等相关规定。

2. 价值性原则：属于归档范围且应永久或长期保存的、社会利用价值高的档案可列入数字化加工的范围。

各档案保管机构库藏纸质档案数量浩繁，而资金和人力资源相对有限，将库藏档案全部进行数字化加工，既不现实，也不必要。无论从社会需求还是数字化成本考虑，都应该有选择的实施档案数字化。因此，各档案保管机构应该从实际出发，在符合国家档案开放、知识产权保护以及信息安全保密等相关规定的前提下，采取优先选择原则进行档案数字化。

一是库藏珍贵档案应优先数字化。各级档案保管机构都有其最为珍贵的“镇库之宝”，从保护人类历史遗产和抢救