

中山市档案局文件

中档〔2016〕16号

关于印发《中山市文件与档案管理服务平台 归档文件采集规范》的通知

各有关单位：

为规范中山市文件与档案管理服务平台对电子文件和数字副本的采集，保证归档文件的真实性、完整性和有效性，根据《电子文件归档与管理规范》（GB/T18894-2002）、《纸质档案数字化技术规范》（DA/T31-2005）和《数字档案室建设指南》（档函〔2015〕114号）等有关规定要求，结合我市实际制定本规范，请遵照执行。



中山市文件与档案管理服务平台 归档文件采集规范

1 适用范围

1.1 本规范规定了中山市文件与档案管理服务平台（以下简称市文档平台）电子文件和数字副本的采集步骤与要求。涉密文件不属市文档平台采集范围。

1.2 本规范适用于使用市文档平台的党政机关、事业单位、国有企业、社会团体、村（社区）及其他单位。

2 定义

2.1 电子文件

电子文件是指通过计算机等电子设备形成、办理、传输和存储的数字格式的信息记录。电子文件包含其形成过程相关信息。

2.2 数字副本

数字副本是指利用计算机等设备，对纸质文档、纸质照片及底片、录音、录像、实物等传统载体文件、档案、资料进行数字化加工，以统一的数字形式存储的信息记录。

2.3 文档

文档是指归档文件最小的文件单元。

3 电子文件采集

3.1 文书类电子文件采集

文书类电子文件是指在本单位电子政务系统中产生的完成办理程序的文件，包括但不限于文件正本、草稿（包含历次修改痕迹）及公文处理单等。

3.1.1 采集方式

根据各单位是否实现文件与档案一体化管理，文书类电子文件采集分为在线和离线两种采集方式。

3.1.1.1 在线采集

市文档平台推荐采用在线采集方式。在线采集要求本单位相应电子政务系统与市文档平台通过 Web Service 对接方式进行数据交换，实现文件与档案一体化管理。在此前提下，只需在市文档平台“文件中心”的“文件导入”模块，对需归档电子文件进行分类、补充著录、归档等操作即可完成在线采集。

3.1.1.2 离线采集

没有实现文件与档案一体化管理的单位采用离线采集方式。离线采集步骤：电子文件收集、本地分类存储、机读目录著录、全文上传及归档。

1) 电子文件收集。将本单位单机或相应电子政务系统中产生的电子文件离线下载，收集到本地存储。

2) 本地分类存储。存储路径按“档案大类名称-年度-部门”分级建立文件夹。在“部门”文件夹下存放电子文件。电子文件命名用8位阿拉伯数字的文件形成日期+文件题名+(稿本)。

例：文书档案\2015\办公室\20150703 中山市公房租赁合同(正本).doc

3) 机读目录著录。在市文档平台“文件著录”模块进行著录。著录方法按《中山市文件与档案管理服务平台著录规则》执行。

4) 全文上传及归档。完成机读目录著录后，将电子全文上传(或批量上传)到市文档平台，一份文件有不同稿本的应分开上传，上传完毕即可进行归档操作。

3.1.2 保存格式要求

上传文档推荐PDF/A、PDF、TIFF、OFD为优先顺序保存格式。鉴于我市电子政务发展现状，对使用Microsoft Office、WPS Office、方正Apabi软件所产生的电子文件也可采集归档。

3.1.3 电子文件稿本

本规范将电子文件稿本分为正本、草稿和其他三种。

3.1.3.1 正本。一般情况下，在本单位电子政务系统中完成处理程序的电子文件作为正本归档。在只有纸质文件或

电子文件与纸质文件内容不一致（如在纸质文件中增加了签名、印章、指纹等内容）时，对纸质文件数字化加工形成的数字副本也可作为正本归档。

3.1.3.2 草稿。除正本之外的历次修改稿作为电子文件草稿归档。

3.1.3.3 其他。除上述两类之外的其他相关归档材料，如文件处理单等过程文档，作为其他稿本归档。

3.1.3.4 稿本标识

稿本	稿本标识	解 释
正本	A	指完成处理程序的正式文件
草稿	B	除正本之外的历次修改稿
其它	C	除上述两类之外的其它相关归档材料，如文件处理单等

3.2 照片类电子文件采集

照片类电子文件是指在工作活动中通过摄影设备形成的静态摄影影像，有完整的 EXIF 信息。市文档平台采集的照片类电子文件主要针对数码照片。

照片类电子文件采集步骤：数码照片收集、本地存储、全文上传、机读目录著录及归档。

3.2.1 数码照片收集

将本单位工作活动中形成的、有保存价值的数码照片收

集到本地存储。

3.2.2 本地存储

本地存储按“档案大类名称-年度”分级建立文件夹（年度一般指拍摄年度）。在“年度”文件夹下建立以“拍摄日期+活动内容+（拍摄者或来源）”命名的照片组文件夹。“拍摄日期”用8位阿拉伯数字表示，“活动内容”应记述清楚活动主题、主要人物的姓名、职务，“拍摄者或来源”用“（）”括起，拍摄者不明的注明照片来源或提供者。将数码照片存储在相应照片组文件夹下，主要人物照片应以“姓名、职务及其在照片的位置”命名。

例：2012年10月15日拍摄的中山市经济工作会议的一组数码照片，该组照片的存储路径为：照片档案\2012\20121015 中山市经济工作会议在中山市会议中心召开，XXX、XXX、XXX等领导参加会议（张三）

3.2.3 全文上传

以组为单位将满足归档质量和保存格式要求的数码照片批量上传到市文档平台。

3.2.3.1 归档质量要求

1) 数码照片画面清晰、完整和准确。对反映同一内容的若干照片，应选择具有主题鲜明、影像清晰、画面完整、人物形象端正、未加修饰剪裁等特点的归档。

- 2) 采集的数码照片应具有符合相关要求的文字说明。
- 3) 数码照片像素数不低于 300 万。
- 4) 常规数码照片除 RAW 格式外存储容量应小于 300MB。

3.2.3.2 保存格式要求

不压缩或无损压缩的主要格式有 RAW、TIFF，有损压缩的主要格式有 JPG、JPEG、PNG 等。本规范推荐以 JPG 格式保存，内容特别重要的宜以 RAW 或 TIFF 格式保存。选择 TIFF 格式时应采用 LZW 压缩方式。

3.2.4 机读目录著录及归档

在市文档平台按照片组及张分别进行机读目录著录，著录完毕即可进行归档操作。

3.3 音频类电子文件采集

音频类电子文件是指在工作活动中通过音频处理设备或 MIDI 电子设备直接形成的以声音为主要表现方式的音频文件。一次工作活动产生的音频为一件。

音频类电子文件采集步骤：音频文件收集、本地存储、全文上传、机读目录著录及归档。

3.3.1 音频文件收集

将本单位工作活动中形成的、有保存价值的音频文件收集到本地存储。

3.3.2 本地存储

本地存储按“档案大类名称-年度”分级建立文件夹，将剪辑后的音频存储在“年度”文件夹下。每段剪辑后的音频文件以“摄录日期+活动内容+（摄录者）”命名。

例：题名为“XXX在中山市档案工作会议上的讲话”的音频文件，摄录日期为2000年11月16日，则电子文件命名为：音频档案\2000\20001116XXX在中山市档案工作会议上的讲话（李四）

3.3.3 全文上传

以件为单位将满足归档质量和保存格式要求的音频文件上传（或批量上传）到市文档平台。

3.3.3.1 归档质量要求

1) 音频文件反映的内容应完整准确，声音清晰，各段句子不短缺也不重复。

2) 声波频率应30Hz~14kHz之间，信噪比应不低于70dB。音频采样率不低于44.1kHz。

3) 若一件音频容量大于2000MB，原文件以离线方式保存，在上传市文档平台前要求进行分割剪辑处理。

3.3.3.2 保存格式要求

推荐以WAV、MP3为优先顺序格式保存。非WAV、MP3格式的音频文件应保留原始格式，根据原始音质等级转换为WAV或MP3格式归档，如是CD音质级及以上的转换为WAV格

式，如是 CD 音质级以下的转换为 MP3 格式。

3.3.4 机读目录著录及归档

在市文档平台按件著录机读目录，著录完毕即可进行归档操作。

3.4 视频类电子文件采集

视频类电子文件是指在工作活动中通过摄影设备直接形成的以视频影像为主要表现方式的视频文件。一次工作活动产生的视频为一件。

视频类电子文件采集步骤：视频文件收集、本地存储、全文上传、机读目录著录及归档。

3.4.1 视频文件收集

将本单位工作活动中形成的、有保存价值的视频文件收集到本地存储。

3.4.2 本地存储

本地存储按“档案大类名称-年度”分级建立文件夹，将剪辑后的视频存储在“年度”文件夹下。每段剪辑后的视频文件以“摄录日期+活动内容+（摄录者）”命名。

例：原文题名是“在改革中创新，在创新中发展：中山市劳动保障工作纪实”，成文日期为 2007 年 3 月 1 日，则电子文件命名为：视频档案\2007\ 20070301 在改革中创新在创新中发展：中山市劳动保障工作纪实（张三）

3.4.3 全文上传

以件为单位将符合归档质量和保存格式要求的视频文件上传（或批量上传）到市文档平台。

3.4.3.1 归档质量要求

1) 视频文件要求内容准确完整、主题鲜明、影像清晰、画面完整，并具有符合相关规定要求的文字说明。若一件视频容量大于 2000MB，离线保存原文件，在上传市文档平台前应进行分割剪辑处理。

2) 视频画面质量要求参照国家标准 GB/T 7401 -1987。视频幅型比一般为 16:9 或 4:3，首选 16:9 比例；分辨率应不低于 720×576I，宜采用 1280×720I 以上，重要活动的视频宜不低于 1920×1080I。

3.4.3.2 保存格式要求

推荐采用 MPEG-2 或 H.264 的视频编码标准，以 AVI、MP4、MPG 作为归档保存格式，非上述 3 种格式的应保留原始文件格式，并转换成视频编码为 H.264 的 MP4 格式归档。视频编码标准与格式对应表：

视频编码标准	视频文件格式
MPEG-2	AVI、MPG
H.264	MP4

3.4.4 机读目录著录

在市文档平台按件著录机读目录，著录完毕即可进行归档操作。

4 数字副本采集

以下各类数字副本采集，均应在市文档平台完成机读目录著录且生成档号的前提下进行。

4.1 纸质文档数字副本采集

在纸质文档没有相应电子文件和纸质文档与电子文件内容不一致（如纸质文档增加了签名、印章、指纹等内容）时，要对纸质文档进行数字化加工。

纸质文档数字副本采集步骤：数字化加工、本地存储、批量挂接。

4.1.1 数字化加工

4.1.1.1 扫描色彩模式选择

1) 扫描色彩模式一般分类黑白二值、灰度、彩色三种，归档文件通常采用彩色模式扫描。

2) 页面有红头、印章和签字、插图的文档、页面清晰度较差的手写体文档、页面有多种颜色的文档、珍贵的历史文档等，应采用彩色模式扫描。页面字迹清晰、不带图片、只有黑白两种颜色的文档，允许采用黑白二值或灰度模式扫描。

4.1.1.2 扫描分辨率选择

扫描分辨率参数大小的选择，原则上以扫描后的图像清晰、完整、不影响图像的利用效果为准。页面字迹清晰的 A4 幅面的文档一般采用 200DPI 扫描。对于手写的、文字偏小、密集、清晰度较差或幅面较大的可提高到 300DPI 或以上，增加幅度以扫描后的图像按 100%显示是否清晰、是否消除图像噪声为标准。

4.1.1.3 压缩要求

存储时压缩率的选择，应以保证扫描的图像在 100%显示时清晰可读的前提下，尽量减少存储容量为准则。采用黑白色彩模式扫描的文档应采用 TIFF (G4) 压缩格式；采用彩色色彩模式扫描的文档应采用 LZW 压缩算法。

4.1.1.4 图像处理

1) 图像拼接。如若扫描设备不支持大幅面文档扫描时，可分幅扫描形成多幅图像并进行拼接处理。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。应采用有效的技术手段，借助图像拼接软件工具实现无缝拼接，保证拼接后的图像整体亮度、对比度、饱和度等参数一致。

2) 旋转及纠偏。对不符合阅读方向的图像应进行旋转还原。对出现偏斜的图像应进行纠偏处理。

3) 裁边。数字图像裁边要求在距页边最外延至少 2 至 3 毫米处裁剪，不得裁掉纸质文档原有页边，包括破损、歪斜

页边。

4) 去污。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去除。去污处理应严格遵循展现文档原貌原则，不得去除文档页面原有的纸张褪变斑点、水渍、污点、装订孔等。

5) 存储文件分割处理。若一份文件超过 300MB，需将文件进行切割存储，在文件名中增加切割流水号，用下横线符号“-”与档号隔开。

4.1.1.5 保存格式要求

推荐以 PDF/A 格式保存，允许以 PDF、TIFF 为保存格式，选择 TIFF 保存格式时，应采用 LZW 压缩方式保存。

4.1.2 本地存储

纸质文档数字副本在本地存储路径按“档案大类名称-年度”分级建立文件夹。在“年度”文件夹下存放数字副本文件，数字副本文件按“稿本标识+档号”命名。档号各要素（全宗号、分类号、保管期限代码、年度、件号等）之间用半角减号“-”符号隔开，档号排列顺序号用阿拉伯数字标识，前面不用补“0”。若一份文件（含正本、草稿、其他稿本）是分割存储的，则按“稿本标识-档号-切割流水顺序号”格式命名，稿本标识、档号、切割流水顺序号之间用半角下横线“-”隔开。

例：档号为 999-A-4-2009-23 的文件，数字副本有两份正本、两份草稿，则正本命名为：A_999-A-4-2009-23_1, A_999-A-4-2009-23_2，草稿命名为：B_999-A-4-2009-23_1, B_999-A-4-2009-23_2

4.1.3 批量挂接

将完成命名的数字副本文件在市文档平台“文件中心”的“数字化加工”模块进行批量挂接。每次批量上传的文件不超过 2000MB。

4.2 照片类数字副本采集

对早期形成的没有电子文件的纸质照片或底片，应将其转化为数字形式。本规范所指的底片是指 120 或 135 的正负片。

照片类数字副本采集步骤：数字化加工、本地存储、批量挂接。

4.2.1 数字化加工

4.2.1.1 扫描色彩模式选择

彩色照片或底片采用彩色方式扫描，黑白照片或底片采用灰度方式扫描，扫描色彩模式宜采用 24 位以上 RGB 色彩模式。

4.2.1.2 扫描分辨率选择

纸质照片扫描分辨率应 ≥ 300 DPI，对于珍贵历史照片或

破损照片应适当提高扫描分辨率。底片扫描分辨率应 ≥ 2500DPI，对于珍贵底片可提高到 4000DPI。

4.2.1.3 压缩要求

黑白二值模式和灰度模式采用 CCITT G4 压缩标准，彩色模式采用 LZW 压缩标准。

4.2.1.4 图像处理

参照 4.1.1.4 纸质文档数字副本图像处理要求。

4.2.1.5 保存格式

参照 3.2.3.2 数码照片保存格式要求。

4.2.2 本地存储

照片数字副本在本地存储路径按“档案大类名称-年度-组号”分级建立文件夹，“组号”文件夹下存放数字副本照片。数字副本照片文件名按已生成的档号命名。档号各要素（全宗号、分类号、年度、组号、张号等）之间用半角“-”符号隔开，档号排列顺序号用阿拉伯数字标识，前面不用补“0”。照片档案分类号为 F3。相同画面的纸质照片或底片与数码照片编同一档号。

例：1998 年 10 月 15 日拍摄的一张中山市经济工作会议照片，归档时编的档号为 999-F3-1998-2-23。2015 年扫描该组照片的存储路径和命名为： 照片档案\1998\2\999-F3-1998-2-23.jpg

4.2.3 批量挂接

参照 4.1.3 纸质文档数字副本批量挂接要求。

4.3 音频类数字副本采集

若原始音频文件不是 WAV 或 MP3 格式的，在保留原始音频文件的同时，要求将其数字化加工转换成 WAV 或 MP3 格式数字副本。

音频类数字副本采集步骤：数字化加工、本地存储、批量挂接。

4.3.1 数字化加工

4.3.1.1 参数选择

采样频率选择 44.1kHz 或更高的 48kHz 两个等级。采样位数采用 16 位量化位数，并保持原始轨道数。压缩编码标准：对于 CD 音质级及以上的音频文件采用 WAVE 使用编码压缩标准，保留原始文件的同时保存为 WAV 格式音频文件；对于 CD 音质级以下的音频文件采用 MPEG-2 Audio 压缩标准，保留原始文件的同时保存为 MP3 格式音频文件，注意采集过程中避免强磁场等外部干扰。

4.3.1.2 质量要求

参照 3.3.3.1 音频类电子文件归档质量要求。

4.3.1.3 格式要求

参照 3.3.3.2 音频类电子文件保存格式要求。

4.3.2 本地存储

音频数字副本在本地存储路径按“档案大类名称-年度”分级建立文件夹。在“年度”文件夹下存放音频数字副本文件，音频数字副本文件按档号命名。档号各要素（全宗号、分类号、年度、件号）之间用半角“-”符号隔开，档号排列顺序号用阿拉伯数字标识，前面不用补“0”。音频分类号为“F2”。若一件音频是分割存储的，则按“档号-切割流水顺序号”命名，档号、切割流水顺序号之间用半角“-”隔开。

例：档号为 999-F2-2009-3 的音频文件，存储路径和命名为：音频档案\2009\999-F2-2009-3.mp3

4.3.3 批量挂接

参照 4.1.3 纸质文档数字副本批量挂接要求。

4.4 视频类数字副本采集

若视频文件不是 MP4、MPG、AVI 格式的，要求离线保存原视频文件，在上传至市文档平台前通过视频处理转换工具，将其转换为视频编码为 H.264 的 MP4 格式归档。

视频类数字副本采集步骤：数字化加工、本地存储、批量挂接。

4.4.1 数字化加工

4.4.1.1 参数选择

视频文件数字化主要技术参数见下表:

文件格式	MP4
视频编码格式	H. 264
制式	与原件相同
帧率	与原件相同
画面宽高比	与原件相同
视频分辨率	标清不低于 720×576 , 高清不低于 1280×720
视频采样率 (YUV)	4: 2: 2
视频量化位数	不低于 8bit
视频码率	标清不低于 8Mbps, 高清不低于 16Mbps
音频编码标准、采样率、音频、量化位数、声道等	参见 3.3 音频类电子文件采集

4.4.1.2 视频文件的首尾空白无内容的部分如果时间过长, 应适当剪切, 在画面开始前和结束后各保留 5 秒左右的空白。

4.4.1.3 视频文件在提供利用前, 可采用压缩比更高的编码格式(如 f1v)进行文件转换, 可进行适当的降噪、振幅标准化等处理以抑制和去除噪音、爆音, 可对影像画面进行

去蒙尘、去划痕、校色、画面稳定处理等。处理前的视频文件原始拷贝应进行保存。

4.4.1.4 质量要求

参照 3.4.3.1 视频类电子文件归档质量要求。

4.4.1.5 保存格式要求

参照 3.4.3.2 视频类电子文件保存格式要求。

4.4.3 本地存储

视频数字副本文件在本地存储路径按“档案大类名称-年度”分级建立文件夹。在“年度”文件夹下存放视频数字副本文件，视频数字副本文件按档号命名。档号各要素（全宗号、分类号、年度、件号）之间用半角“-”符号隔开，档号排列顺序号用阿拉伯数字标识，前面不用补“0”。视频文件分类号为“F1”。若一件视频是分割存储的，则按“档号-切割流水顺序号”命名，档号、切割流水顺序号之间用半角“-”隔开。

例：某视频文件档号要素为全宗号、分类号、年度、件号，其档号为 999-F1-2009-3，则命名为：视频档案\2009\999-F1-2009-3

4.4.4 批量挂接

参照 4.1.3 纸质文档数字副本批量挂接要求。

4.5 实物类数字副本采集

实物类档案是指以特定有型的物质实体为载体，能够反映本单位职能活动和历史真实面貌的具有保存价值的档案，一般包括本单位公务活动中获得的奖杯、奖牌、奖章、锦旗、荣誉证书或获赠的纪念品；上级领导、知名人士的题词字画；本单位使用过的牌匾、印章及重大活动中形成的纪念品等。没有电子文件的实物要进行数字化加工采集。

实物类数字副本采集步骤：数字化加工、本地存储、批量挂接。

4.5.1 数字化加工

4.5.1.1 数字化方法

一般实物档案通过数码相机拍摄进行数字化，字画、证书等通过扫描进行数字化。

4.5.1.2 质量要求

用数码相机对实物进行拍摄时，考虑大多实物属于有光泽的反光体，不适合使用点光源，最好以打光面积大的面光源为基础布光，光源的面积越大越好，尽量使环境光线统一、均匀，有条件的可使用柔光箱或柔光屏加强基础布光，以使实物色调更真实，有质感，要求拍摄出来的实物照片上没有明显的光斑或黑斑。此外，要根据实物个体形状选择不同的角度进行拍摄，要求最大限度的展示实物本身的样貌。

4.5.2 本地存储

实物数字副本文件本地存储按“档案大类名称-年度”分级建立文件夹，在“年度”文件夹下存放实物数字副本文件。实物数字副本文件可按“拍摄日期+实物名称+(拍摄者)”命名。

4.5.3 批量挂接

参照 4.1.3 纸质文档数字副本批量挂接要求。

信息公开选项： 主动公开

抄送： 广东省档案局。

中山市档案局办公室

2016年5月11日印发

(共印2份)