

《录音录像类电子档案元数据方案》解读

文/毛海帆 李鹏达 田丹华 傅培超

根据档案馆(室)录音录像类电子档案管理要求,经国家档案局批准,江西省档案局于2011年着手开展录音录像类电子档案元数据标准化研究工作,2014年开始文本编制工作,2017年8月最终成果《录音录像类电子档案元数据方案》(DA/T 63—2017)正式发布,并于2018年1月1日起实施。

编制背景

在电子档案生命周期全程,依托应用系统持续捕获、著录元数据是保证电子档案真实性、可靠性、完整性和可用性(以下简称“四性”)的重要措施之一。为此,国家档案局先后组织编制并发布了《文书类电子文件元数据方案》(DA/T 46—2009)、《照片类电子档案元数据方案》(DA/T 54—2014)等多项元数据标准。由于录音录像类电子档案物理结构复杂、著录与管理要求不同等自身特性,上述档案行业元数据标准均不能完全适应其管理要求。比如,数字音视频编码标准有MPEG-X、H.2XX等10余种,一种编码标准下具体技术参数亦可能不同,如色彩空间、视频比特率、色度采样率等。任何格式的计算机文件的写入、存储、读取和呈现都是依据编码标准进行编码和解码的过程。在电子档案体外采用结构元数据记录编码标准及关键技术参数,就是保持录音录像类电子档案解码播放、格式转换等功

能,以此为电子档案的长期可读和可用提供保障。录音录像类电子档案获取方式还包括网上下载、购买、收录等多种途径,如果涉及摄录者个人,还需要进行知识产权管理。此外,录音录像类电子档案的可读性要依靠对主题内容、责任者、业务背景等进行较为全面的著录方可实现,这些都需要通过元数据给予记录。

根据上述情况,有必要编制专门的元数据标准用于规范、指导录音录像类电子档案元数据的捕获、著录等管理活动。

编制依据与适用范围

1. 编制依据

编制过程中,该标准主要依据《标准化工作导则 第1部分:标准的结构和编写原则》(GB/T 1.1—2009)设置框架及文本内容,依据《电子档案管理基本术语》(DA/T 58—2014)等相关标准定义术语。同时,依据或参照《信息与文献 文件元数据管理 第2部分 概念与实施》(ISO 23081—2009)、《档号编制规则》(DA/T 13—1994)、《档案著录规则》(DA/T 18—1999)等标准,并结合我国档案管理实际设置录音录像类电子档案元数据集框架及其具体元数据。

2. 适用范围

该标准主要适用于各级各类档案馆及其立档单位,可用来描述、管理以卷、件为单元的录音录像类电子档案。基于该标准

对各个元数据应用层次属性的描述可组合形成附录B所列的案卷级、文件级元数据集,将其实现于相关应用系统后即可支持档案馆(室)开展案卷级、文件级录音录像类电子档案著录工作,实现元数据全自动和半自动化捕获、著录,并为录音录像类电子文件归档与电子档案管理提供支撑。

为实现录音录像档案数字副本的集成管理和长期保存,该标准所设元数据均适用于录音录像档案数字副本,其中,数字化责任信息、数字化时间、原始载体、生成方式等元数据具有明确的指向性,视频参数、音频参数与《录音录像档案数字化规范》(DA/T 62—2017)第9章所列技术参数完全兼容。因此,录音录像档案数字副本及其元数据管理可参照该标准执行。

主要内容

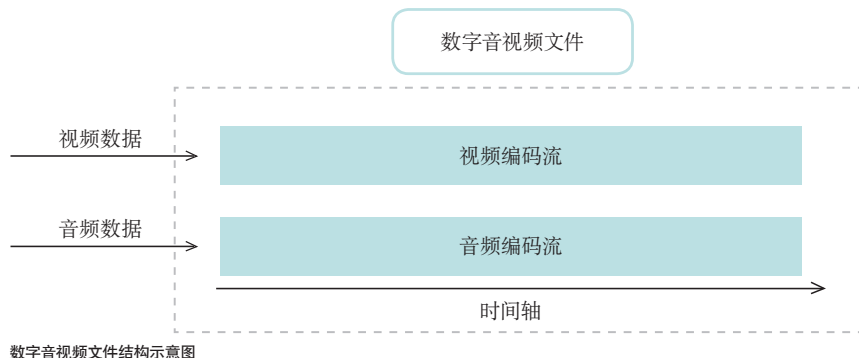
1. 标准结构

该标准共9部分,第1~4章分别为范围、规范性引用文件、术语和定义、元数据实体及元数据描述方法,第5~8章则分别对元数据集的4大类元数据作出逐一描述,附录部分给出了5个实用性工具,均为资料性附录。元数据集共计96个元数据,其中,必选元数据18个,可选元数据45个,条件选元数据33个,覆盖了录音录像类电子档案全生命周期。元数据捕获方式主要有全自动、半自动和手工著录3种,有12个容器型元数据无需著录,47个元数据应由

相关应用系统在不同业务节点和时间点全自动捕获,33个元数据可以半自动方式捕获,只有题名、责任者、摄录者、附注4个元数据需手工著录,最大限度地实现了元数据标准化与著录自动化。

2.元数据集的一体化设置

录音录像类电子档案是指录音类电子档案和录像类电子档案。经对比研究,二者的元数据存在较大范围的重合,不同之处在于责任者的称呼,录像类电子档案技术参数涉及视频编码部分。在最简单的数字音视频文件结构中,沿着时间轴的伸展,至少由一个音频编码数据流和与其并行的一个视频编码数据流组成,比数字音频文件多了视频编码数据(如图所示)。为此,该标准提供了录音录像类电子档案一体化元数据集,通过适用门类属性和摄录者元数据的设置予以区分。



数字音视频文件结构示意图

3.元数据集构成及其作用

录音录像类电子档案元数据集由档案实体、业务实体、机构人员实体和授权实体元数据4个部分组成。档案实体元数据用于描述录音录像类电子档案本身,为必选元数据实体,共75个元数据,涵盖了电子档案标识、内容、结构、背景和管理权限等多个方面。档号、题名、责任者、摄录者、时间、保管期限等18个必选元数据都位列其中;业务实体、机构人员实体和授权实体元数据为管理过程元数据,均为可选元数据实体,该标准推荐采用多元数据实体模式,此时,业务实体、机构人员实体为必选元数据实体。将业务实体的管理活动元数据与另一个或两个元数据实体组配后即可描述一系列管理活动,使后人能够追溯任一电

子档案管理过程,并与档号、原始载体、捕获设备等元数据共同解答“它从哪里来”等重要问题,为录音录像类电子档案的真实性和可靠性提供证据链。

4.元数据描述

该标准第4章第4节规定了元数据描述方法,描述内容包括23个属性。除对注册版本、注册机构、字符集、语言4个共同属性进行集中描述外,还需从编号、中文名称、英文名称、定义等19个属性对每个元数据进行逐一描述。

下面只对约束性等4个元数据属性作出进一步说明:

约束性属性有3个值域,包括必选、条件选和可选。该标准明确的18个必选元数据是最小元数据集,同时适用于档案室和档案馆。条件选有2种情况:其一,当该元数据所属实体或其上位元数据被选用时,

该元数据必选,如业务实体的管理活动元数据和档案实体的主题、来源、存储、原始载体元数据的子元数据;其二,格式名称、格式版本元数据的约束性为条件选,当其上位复合型元数据格式信息仅作为容器型元数据时,这两个元数据必选。

可重复性属性用于说明该元数据或元数据组是否可以循环著录新的元数据值。不可重复元数据仅记录最后一次著录信息。可重复元数据有主题、捕获设备、计算机文件大小等11个(组)元数据。主题元数据用于一件录音录像电子档案的多个片断内容的深度著录,比如,描述某次重大活动的各项议程信息。其他可重复元数据则可以反映录音录像类电子档案形成、变化或管理过程,如再次执行格式转换操

作,则电子档案编码标准、计算机文件格式等必然变化,应通过相互关联的元数据捕获、著录新的元数据值。

信息来源属性从捕获节点和捕获方式2个紧密关联的方面对元数据进行描述。捕获节点主要有登记、挂接、手工著录、触发或完成管理行为等业务节点。捕获方式主要有5种:从操作系统和应用系统参数自动捕获、从电子档案内部编码数据中自动提取、基于应用系统数据字典半自动捕获、由外部数据导入以及手工著录。

著录说明属性则明确了复合型和简单型元数据值著录的规范性要求,并提供了具体的著录示例,附录D、附录E分别提供了整体著录实例,为标准实施提供具体参考。

元数据集的实施应用

该标准推荐全面采用所设元数据集,或由省级档案行政管理部门统一编制录音录像类电子档案元数据集及其XML Schema,统一实施于相关应用系统。编制本地区元数据集时,需考虑区域性电子档案全程管理、集成管理和异构应用系统间的数据共享、交换等因素,有助于优化应用系统功能设计、开发和运维升级;本地区档案馆(室)需要管理录音录像档案数字副本时,应将数字化责任信息等元数据一并纳入元数据集,以便实现二者的集成管理;使用商业化媒资管理系统的档案馆(室),应要求中标供应商完成系统功能对标工作。

在技术层面,任何门类电子档案元数据集的实施应用都是通过嵌入相关应用系统而实现的,录音录像类电子档案元数据集也不例外。这个“嵌入”包含数据库与一系列管理功能的设计与开发。因此,在实施元数据集之前,档案局(馆)和应用软件开发商等相关责任人要全面读懂每个元数据的23个属性,在此基础上方能编制本地区录音录像类电子档案元数据集,开展相关应用系统功能需求分析、采购需求撰写、软件设计开发等工作。

作者单位:江西省档案局

责任编辑:黄佳音