

《纸质档案数字化规范》解读

文/杜琳琳

近年来,在纸质档案数字化工作迅速推进的同时,纸质档案数字化对象、管理与利用方法以及技术条件等都在不断发展。因此,及时修订DA/T 31—2005《纸质档案数字化技术规范》,适时调整相关要求,对科学规范纸质档案数字化工作具有重要的现实意义。2017年8月2日,DA/T 31—2017《纸质档案数字化规范》正式发布,并于2018年1月1日起正式实施。

修订思路

1.与有关标准相协调

在标准修订过程中,课题组特别注意与相关标准的协调,主要体现在以下方面:

首先,按照GB/T 1.1—2009《标准化工作导则 第1部分:标准的结构和编写原则》的规定和要求编写,在修订过程中,遵循GB/T 20000《标准化工作指南》、GB/T 20001《标准编写规则》、GB/T 20002《标准中特定内容的起草》等基础性标准的要求。其次,保持与DA/T 31—2005相关要求的连贯性,汲取其精华内容。但同时也与时俱进,紧跟技术发展趋势,对相关管理、技术提出恰当要求。再次,既保持与国际标准ISO/TR 13028《信息与文献档案数字化实施指南》和国家标准GB/T 20530—2006《文献档案资料数字化工作导则》等标准的一致性,又突出我国现阶段及未来档案工作发展的特色。

2.增加管理要求

DA/T 31—2005主要从技术角度对纸质档案数字化工作提出要求,对管理方面的内容涉及较少。随着纸质档案数字化工作的推进,实际工作中对纸质档案数字化工作标准化管理需求愈加急迫。因此,在修订过程中,课题组特别对纸质档案数字化工作的组织和管理进行了详细设计,并提出了明确要求。为此,此次修订将标准名称由《纸质档案数字化技术规范》修改为《纸质档案数字化规范》。

3.细化和调整技术要求

纸质档案数字化工作在很大程度上依赖于信息技术,而信息技术的迅速发展使相关技术条件也发生了很大变化。及时对技术要求进行细化和调整,使之符合当前纸质档案数字化工作的实际需求是此次修订的一项重要任务。

4.强化安全要求

安全是档案工作的重中之重,随着纸质档案数字化工作的广泛开展,特别是在任务量巨大的情况下,很多单位采取了外包方式开展纸质档案数字化工作,其安全问题不容忽视。此次修订特别注重对档案安全(包括档案实体安全和档案信息安全)的保障,从管理和技术等不同角度对其进行规范。

部分内容解读

1.组织与管理

在纸质档案数字化工作广泛开展的

今天,科学规范的管理工作对于确保数字化成果质量和数字化工作效率具有十分重要的意义。课题组希望通过对相关章节内容的完善,使纸质档案数字化管理工作更加系统化、科学化,从而便于纸质档案数字化管理工作的规范化开展。具体内容集中体现在标准第5章,内容包括机构及人员、基础设施、工作方案、管理制度、工作流程控制、工作文件管理、档案数字化外包等,基本涵盖了纸质档案数字化管理工作的全部内容。管理工作的实施,应达到确保档案安全、提高档案数字化成果质量、提高工作效率的目标。

2.数字化前处理

此次修订将DA/T 31—2005的“档案整理”更名为“数字化前处理”,并对其内容进行了完善,主要包括确定扫描页、编制页号、目录数据准备、拆除装订、技术修复等。这一章的修订重点突出了两方面的思想:一是建议纸质档案数字化全部扫描,不建议进行挑扫;二是保护档案原件。“全部扫描”主要体现在“确定扫描页”部分,主要目的是确保档案信息完整,同时避免对纸质档案进行二次扫描。“档案保护”思想则贯穿整章内容。例如,针对档案数字化过程中是否拆除装订物这一问题,DA/T 31—2005要求:“在不去除装订物情况下,影响扫描工作进行的档案,应拆除装订物”。此次修订,要求从对档案保护的角度出发,如

钢铤装订的档案,随着时间的推移,钢铤生锈,会破坏档案,对于这种情况,就应该建议拆除钢铤装订,改用线装。如果装订本身具有历史特色,就要慎重考虑是否需要恢复原装订和能否恢复原装订。为此,DA/T 31—2017规定:“应以对纸质档案的保护为原则确定是否拆除装订”“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”。

3. 技术参数

随着时代的发展,纸质档案数字化成果需求逐渐趋于多样化,同时,由于技术条件发展迅速,使存储成本等方面的限制逐渐淡化。因此,纸质档案数字化技术参数的选择需多元化、细化,才能满足不同需求。同时,也要遵循一个原则,即对纸质档案原件的保护,尽量只动一次纸质档案,就满足多重利用需求。

(1) 色彩模式

色彩模式是数字世界中表示颜色的一种算法。扫描仪的色彩模式一般分为彩色、灰度、黑白3种色彩模式。色彩模式选择,应考虑多种因素:①原件情况。对于影像清晰、档案原件色彩简单、色彩对比强烈的原件,如白纸黑字的普通打印文件,在其他条件均达到最佳的情况下,采取3种色彩模式扫描的图像清晰程度差异不明显。而对于影像不够清晰,以及色彩对比不够强烈的原件,采取彩色模式扫描的图像,在展现档案原件细节、清晰程度等方面要明显优于灰度模式或黑白模式扫描的图像。②硬件条件。过去,由于硬件条件的限制,一般会综合考虑图像质量、访问速度、存储成本等方面的因素选择色彩模式。而随着技术的进步,硬件考量因素将越来越被弱化。③后期应用。彩色模式可以转换为灰度模式或黑白模式,所以,为了尽可能多地采集信息,彩色模式无疑是最理想的选择,对后期应用也是较为灵活的选择。但是,也有例外情况。对于需要进行COM输出的档案,色彩模式对输出效果的影响与缩微胶片输出设备可接受的图像色彩模式有很大关系。例如,国内使用的一部分缩微胶片输出设

备只能接受黑白色彩模式的图像,如扫描为彩色模式的图像需转换为黑白模式方可进行输出,这与直接扫描成黑白模式的图像输出效果相比,后者的效果更佳。所以,在制定纸质档案数字化方案时,在条件允许的情况下,应尽量避免格式转化,以减少信息损失。如果无法避免格式转化,则应根据原件情况、软件情况等制定合理的方案,将信息损失降到最低。

结合课题组的调研结果,从最大程度保留原件信息的角度考虑,标准建议在进行批量数字化的过程中,应尽量将其扫描成彩色模式。同时此次修订提出必须扫描成彩色模式的情况,即“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。”此外,考虑到我国各级各类档案部门差异,要求全部扫描成彩色模式不够现实。因此,此次修订提出了可扫描成黑白模式和灰度模式的情况。

(2) 分辨率

分辨率是指扫描设备对扫描对象每英寸的采样点数。分辨率与扫描图像大小和清晰程度以及扫描时长都有直接关系。扫描图像分辨率越高,所占用的扫描时间越长,扫描形成文件也会越大,图像清晰程度相对越好。但是,通过采用技术标板测试研究可知,当分辨率的设置达到一定程度时,分辨率的增加不会明显改善清晰程度,但是却会大大增加存储空间。所以,在实际工作中,应根据可接受的图像大小和清晰程度的实际情况来选择合适的分辨率。

在现有技术条件下,综合考虑不同档案馆基础条件的差异,标准提出:“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”的一般原则。当然,如果扫描后的档案还需进行后续处理和利用的话,分辨率的选择则应考虑利用需求因素。例如,进行COM输出的图像分辨率不应低于300DPI,进行仿真复制的数字图像分辨率不应低于600DPI。

(3) 存储格式

存储格式即图像文件存放在计算机

存储设备上的格式。目前,常用的图像存储格式主要有TIFF、JPEG、JPEG2000、PDF、OFD等,不同格式适用的应用场景不同。TIFF存储信息量大,适用于图像的长期存储;JPEG具有调节图像质量的功能,允许用不同的压缩比例对文件进行压缩,方便在图像质量和文件大小之间找到平衡点;而提供网络应用时,则建议存储为PDF、OFD格式。值得一提的是,在“对纸质档案只动一次”的原则下,应首先满足档案长期安全保存的需要。

4. 图像处理

随着信息技术的发展,存储设备成本将呈下降趋势。而纸质档案数字化成果的利用则是一个长期过程,应用方式也多种多样。因此,需要对图像质量进行严格控制。当图像质量与存储空间存在矛盾时,课题组始终坚持将数字化图像质量作为主要考量因素,将图像占用的存储空间作为次要考量因素的原则。因此,此次修订去掉了一些节省存储空间的处理原则。在执行该标准开展图像处理工作时,也应时刻坚持上述原则。

5. 成果验收

与DA/T 31—2005相比,DA/T 31—2017对纸质档案数字化成果验收环节的要求进行了较为详细的划分。验收内容主要包括: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、存储载体等指标。验收方式的选择对数字化成果质量的保障也具有关键作用。随着信息技术的发展,计算机自动检测的方法更为多样,范围更为广泛。目前,纸质档案数字化成果质量检查软件正日趋成熟,在标准中提出并提倡采用计算机自动检测的要求具备相应的技术条件。为此,课题组针对档案数字化成果验收指标进行细分,对于能够通过计算机程序自动化方式验收的指标,提出100%验收比率和100%合格率的要求。对于无法采用计算机程序自动化验收的指标,根据具体情况提出不同的合格率要求。

作者单位:国家档案局科研所

责任编辑:黄佳音