

2005 ~ 2017 纸质档案数字化标准对比研究

王芸芸^{1,2,3}, 杨健康^{1,3}, 向禹^{1,3}

(1. 中南大学 信息安全与大数据研究院 湖南 长沙 410083; 2. 长沙医学院档案室 湖南 长沙 410219; 3. 中南大学 档案技术研究所, 湖南 长沙 410083)

摘要: 纸质档案数字化标准关系着档案事业的科学化、规范化发展。DA/T 31 - 2017《纸质档案数字化规范》和 DA/T 31 - 2005《纸质档案数字化技术规范》相比,主要变化体现在数字化前处理部分、数字化过程中元数据采集、档案扫描和图像处理、数据备份和数字化成果管理等方面。2017年新标准的变化,意味着档案领域更加重视纸质档案数字化,对数字化的要求也更加严格。文章以纸质档案数字化的新旧标准为研究对象,通过对新旧数字化标准的对比研究,论述新标准的优点和可行性或是对档案事业发展的作用。

关键词: 纸质档案; 档案数字化; 数字化技术规范; 数字化标准

中图分类号: G270

文献标识码: A

文章编号: 2096 - 2339(2018)04 - 0201 - 03

DOI:10.19534/j.cnki.zyxygc.2018.04.094

档案是国家机构、社会组织或个人在社会活动中直接形成的有保存价值的各种形式的原始记录。随着计算机技术、网络技术以及通信技术的发展和运用,我国档案事业从传统管理逐步迈向档案信息化阶段。纸质档案数字化是档案信息化建设中必不可少的环节,它不仅可以提高办公效率和经济效益,还能更好的保护档案原件并且避免档案受到“孤本”限制。另外,档案数字化也使得远程服务、资源共享得以实现。所以,馆藏资源数字化也是实现档案信息组织与服务网络化、档案实体虚拟化、信息资源共享化、档案利用知识化的基础,更是数字档案馆和数字档案室的核心内容之一。

2005年国家档案局出台了《纸质档案数字化技术规范》,从技术标准方面对纸质档案数字化工作作出了相关要求。但由于信息技术的不断发展、数字化设备等硬件的更新换代以及档案内容形式、管理方法、利用方式等的变化,DA/T 31 - 2005《纸质档案数字化技术规范》(以下简称原标准)已不能完全涵盖当时纸质档案数字化过程中所出现的问题,因此国家档案局于2017年颁布了DA/T 31 - 2017《纸质档案数字化规范》(以下简称新标准),提出了更适用于档案行业的纸质档案数字化工作的规范要求。

1 新原标准对比情况

1.1 标题与范围

新标准将文件标题更改为《纸质档案数字化规范》,与《纸质档案数字化技术规范》相比,扩大了规定的管理范围,从规定纸质档案数字化的主要技术要求变为规定纸质档案数字化技术和管理要求,从适用于扫描仪对各种纸质档案的数字化加工处理到对数字化加工过程的管理。

1.2 规范性引用文件

新标准中增加了引言文件 GB/T 20530 - 2006 文献

档案资料数字化工作导则、DA/T 1 档案工作基本术语、ISO/TR 13028 信息与文献档案数字化实施指南。删除了引用文件 GB/T 17235.1 连续色调静态图像的数字压缩编码、GB/T 17235.2 连续色调静态图像的数字压缩编码、GB/T 18894 - 2002 电子文件归档与管理规范、ITU (CCITT) G3 二值图像压缩算法、ITU (CCITT) G4 二值图像压缩算法。

1.3 术语和定义

新标准将适用于其本身的术语和定义规定为: GB/T 20530 - 2006、DA/T 1、DA/T 18 界定的以及数字化、数字图像、纸质档案数字化、分辨率这四个术语,相对于原标准而言,对几个术语的定义更简洁、准确,而且覆盖范围更广泛。

1.4 调整部分

(1) 新标准将第四点的标题从“数字化流程”改为“总则”,对纸质档案数字化的质量、技术手段提出了相关要求,并强调对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纸质档案数字化作出了要求。

(2) 原标准中第五点仅提到关于案卷整理的相关要求,新标准将第五点的标题改为“组织于管理”,不仅对纸质档案数字化的整理管理作出要求,并增加了对机构及人员、基础设施、工作流程控制、工作文件管理、档案数字化外包这些管理工作的相关要求,同时强调在纸质档案数字化过程中应当制定合理的工作方案,而且要有对应的管理制度。新标准增加的内容更加全面,也能够适应更广范围的数字化工作,同时也对档案外包机构提出了相应要求,使得档案工作朝更加标准化、规范化发展。

(3) 新标准增加了第六点“档案出库”和第七点“数字化前处理”两个章节的内容,对纸质档案进行数字化处

作者简介: 王芸芸(1995 -),女,湖南邵阳人,硕士研究生,图书情报专业,研究方向:档案学方向。

理的前期准备工作作出了相应要求,对纸质档案的保护也提出了相关程序要求。“数字化前处理”这一章节的内容规定了纸质档案进行数字化处理之前的细微工作标准,如:编制页号、目录数据库准备、拆除装订、技术修复等。相对于原标准而言,新标准更加完善、全面。

(4) 原标准中第六点“建目录库”在新标准中对应的内容为第八点“目录数据库建立”,在新标准中要求制定目录数据库规则,并且提出在纸质档案目录准备与目录数据库建立工作中均应严格遵守:此规则的建立应符合 DA/T18 对档案著录的要求。其次,新标准中增加了数据库结构的设计应该注意保持档案的内在联系等要求,并对纸质档案数字化前处理工作的相应环节作出规定。

(5) 关于“档案扫描”,新标准中增加了对档案扫描的基本要求、扫描设备的选择及保护等内容,在这些内容中提到了关于不同尺寸、珍贵程度的档案应当如何扫描,替代了原标准中单独列出的特殊页面的扫描的内容。针对扫描方式,新标准中规定为彩色模式、黑白二值模式和灰度模式,而在原标准中仅有连续色调扫描和黑白二值扫描两种方式。新标准中规定扫描分辨率应不小于 200 dpi,若文字较小时应不小于 300 dpi,并指出目前比较新型的档案利用方式如仿真复制等的扫描分辨率应根据需要进行调整;在原标准中,针对纸张和档案色彩不同规定了不同的分辨率。(如表 1 所示)相比而言,新标准更加细致,两个规范都指出应该根据实际要求进行调整。

表 1 《纸质档案数字化标准》对比情况

项目	DA/T 31 - 2017	DA/T 31 - 2005
扫描方式	宜全部采用彩色模式进行扫描	可采用黑白二值图像和连续色调图像两种方式
页面为单色文字	字迹清晰、不带图片的档案材料	可采用黑白二值模式扫描
	清晰度较差或带有图片的档案材料	可采用灰度模式扫描
页面为多色文字图像的档案材料	采用彩色模式扫描	采用连续色调图像扫描
分辨率	一般情况下	不小于 200 dpi
	特别指出	需要进行 COM 输出的档案,扫描分辨率建议不小于 300 dpi; 需要进行高精度仿真复制的档案,扫描分辨率建议不小于 600 dpi
		单色页面档案文件,扫描分辨率一般建议选择 100 ~ 200 dpi; 彩色页面档案文件,扫描分辨率可选择 100 dpi 以上参数值
		大幅面档案文件,扫描分辨率应选择 100 dpi 以上

(6) 新标准中第十点“图像处理”、第十一点“数据挂接”对应原标准第八点“数据处理”的内容,原标准图像处理 - 纠偏中对偏斜图像要求保证数字图像的偏斜角度小于 1 度,新标准中规定视觉上基本不感觉偏斜就可,灵活性更大。新标准增加了对数字图像进行裁边的要求。新标准中删除了原标准数据备份和数字化成果管理的相关内容,因为本规范是对纸质档案数字化的相关要求,删除这些内容使内容更加简洁。

(7) 关于“存储格式”,新标准中在第九点“档案扫描”中,下设‘存储格式’这个内容,增加 JPEG2000 等数字图像长期保存的格式,同一批档案应采用相同存储格式这些内容,并提出可以根据网络浏览速度、存储空间占用等方面综合考虑,将图像转换为 OFD、PDF 等其他格式。原标准在第九点“信息存储”中规定了数字图像档案的保存格式,并对不同色彩数字图像的保存格式及压缩率进行了不同的规定。相比而言,新标准的灵活性更强,原标准的规定虽然更加细致,但保存格式仅有两种,无法适应当代技术的发展也让档案工作者受到一些局限。

(8) 新标准中单独列出第十二点“数字化成果验收与移交”,规定了验收方式、验收内容、验收指标、验收结

论以及移交方面的工作,原标准中关于数字化成果验收的内容仅在第四点“数字化基本流程”中提到对裸数据进行验收前的相应处理,新标准细化了数字化成果管理相关内容。

(9) 新标准中将进行了数字化处理之后的纸质档案处理办法的相关内容,单列出来并且将这一点放在新标准的最后一点,这样的安排不仅符合纸质档案数字化处理的流程,也能够让档案管理人员更加清晰的完成收尾工作。

(10) 新标准增加了元数据采集要求。在数字化过程中应当保存各类技术参数方面元数据这一项内容,提出元数据应符合 ISO/TR13028 中的要求。新标准还规定,纸质档案数字化成果包括数字图像、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,验收时还必须对元数据进行验收,主要包括元数据元素的完整性和赋值规范性等。元数据是描述文件背景、内容、结构及管理过程的数据,它使得记录档案特征、背景,以及档案管理工作更加便利。国家档案局于 2009 年、2014 年、2017 年先后出台了《文书类电子文件元数据方案》、《照片类电子档案元数据方案》、《录音录像类电子档案

元数据方案》其中都提到了相应档案元数据的设计、捕获、著录的一般要求,这些方案在纸质档案数字化新标准发布之前起到了指导作用。旧标准在纸质档案数字化过程中并未提及对元数据的要求,新标准增加了对元数据的要求,使得纸质档案数字化更加规范,同时也很好的契合了国家档案局出台的相应规范,能够更好地满足档案长期保存和利用的需求。

2 新旧标准的应用现状

2.1 学术研究现状

分别以“纸质档案数字化规范”、“纸质档案数字化技术规范”为篇名,设置截止时间为2018年4月30日在知网中检索文献,得到以新标准为题的文献共1篇,以原标准为题的文献共5篇,由结果可知,新旧标准的相关学术研究并不多。

2.2 实际应用现状

2016年国家档案局印发《全国档案事业发展“十三五”规划纲要》要求加快档案管理信息化进程,全国省级、地市级和县级国家综合档案馆馆藏永久档案数字化的比例,分别达到30%~60%、40%~75%和25%~50%。自此多省发布纸质档案数字化工作实施方案,从《纸质档案数字化规范》(DA/T 31-2017)2017年发布,2018年1月1日实施以来,大部分省、市、区还是以《纸质档案数字化技术规范》(DA/T 31-2005)为标准来建立纸质档案数字化工作细则,如湖北省于2018年4月2日颁布的《湖北省纸质档案数字化操作规程》、酒泉市肃州区档案局于2018年3月28日出台的《全区档案数字化工作实施方案》、上海市松江区于2018年1月18日发布的《松江区档案馆纸质档案数字化加工质量标准》、甘肃省于2018年1月9日发布的《甘肃省纸质档案数字化操作规范(试行)》、江西省于2017年10月31日发布的《江西省纸质档案数字化管理与技术规范》等都是根据旧标准来制定的。根据新标准来制定纸质档案数字化工作标准的只有较少的省市,如重庆市于2018年4月27日发布的《重庆市纸质档案数字化实施细则》、北京市朝阳区档案局发布正式启动2018年馆藏纸质档案数字化加工项目的通知,要求该项目严格按照国家档案局新颁布的《纸质档案数字化规范》(DA/T 31-2017)来执行。

2.3 综合应用现状

新标准中提到元数据元素的确定应符合ISO/TR 13028提出的要求。中南大学是湖南省首个贯彻实施档案元数据、电子签名及新标准的档案机构。其新标准档案数字化项目验收会议于2018年3月4日下午召开,并成功通过验收。中南大学在进行纸质档案数字化实施过程中,除遵循新标准的要求外,还增加了双层PDF和全文检索质量检查;元数据采集与电子文件封装、阅读等;电子签名与验证等三方面的内容。中南大学档案数字化

项目建设贯彻实施了国家新标准,在高校档案工作中具有一定的示范性。目前,中南大学档案数字化加工是多个国家标准的综合应用,如DA/T 31-2017《纸质档案数字化规范》、DA/T 46-2009《文书类电子文件元数据方案》、DA/T 47-2009《版式电子文件长期保存格式需求》、DA/T 48-2009《基于XML的电子文件封装规范》、DA/T 54-2014《照片类电子档案元数据方案》、GB/T 33480-2016《党政机关电子公文元数据规范》等等。

由于新标准并非强制性标准,还存在应用较少、普及度不高等现象,但在多方力量的推动下,档案行业会越来越规范和透明,档案工作也会更加完善。

3 新标准的积极影响

新标准的实施对档案数字化加工需求及档案工作的产生了积极的影响。目前,我国档案数字化加工的确定性需求大致可分为四个层次:①符合数字化档案长久保存的基本要求;②实现数字化档案检索与利用;③实现基于全文检索的档案数据化;④展现档案信息的关联性。原标准中提到信息应以JPEG数字压缩编码或TIFF国际通用的标准格式进行存储,这仅能满足档案数字化加工需求的第一个层次,即长久保存的需求。新标准中提出纸质档案数字图像长期保存格式为TIFF、JPEG或JPEG2000,同时还可以考虑网络浏览速度、存储空间等因素,将图像转换为OFD、PDF等格式。JPEG和JPEG2000相比而言,后者压缩率比前者高出30%左右,还同时支持有损、无损压缩,而且JPEG2000不仅适用于扫描仪、数码相机等传统的数字化加工领域,还可以用于网络传输、无线通讯等新领域。此外,OFD、PDF等存储格式,可以在一定程度上实现OCR识别,这满足了档案数字化过程中第二层次的需求。需要提出的是,新标准要求纸质档案数据化过程中,保存数字化项目信息等各类技术参数方面的元数据,以及在验收环节中要求对元数据元素的完整性和赋值规范性等进行验收,这些规定与档案数字化第二层次的需求相契合,并且为第三、四层次的需求打下了坚实的基础。毫无疑问,新标准的出台将推动档案数字化需求层次实现的进程,它使得资源共享更加方便快捷,并将对档案大数据的研究、档案信息资源开发以及整个档案工作产生深远影响。

4 结语

综上所述,DA/T 31-2017《纸质档案数字化规范》在章节设置、内容组成等方面的安排更科学、细致,并且更好的契合了档案数字化需求层次中的要求。档案数字化工作是档案信息化建设中的重要内容,档案行业想要改善传统的被动服务状态,提高社会地位,就需要树立主动服务观念。通过档案数字化需求层次的实现,能使用户更加方便快捷的查询到所需信息,通过档案数据化技术,

(下转第206页)

正确处理好这种职业伦理关系,是现代图书馆职业伦理的要求。

3 结论

图书馆的开放与资源共享,要在保证图书馆读者实现其求知需求和学习的同时,切实保护好图书馆读者个人信息与隐私。这就要求图书馆工作人员,必须具有较高的职业素养;这就要求图书馆工作人员务必提高法律意识,尊重隐私权;这就要求图书馆工作人员务必做好精益服务,实现人文关怀;这就要求图书馆工作人员务必遵守图书馆职业伦理守则。在不断开展图书馆信息资源信息服务的同时,使图书馆更好地发挥其文化教育功能。

参考文献:

[1] 巴山. 图书馆服务的伦理内涵及调控[J]. 图书馆, 2016

(4):75.

[2] 周淑云. 地方性图书馆法规相关信息伦理调查与分析[J]. 图书馆, 2016(4):45.
[3] 张康之, 李传军. 行政伦理学教程[M]. 长沙: 中国人民大学出版社, 2012.
[4] 谢小红. 民生档案伦理利用分析[J]. 兰台世界, 2015(12): 116.
[5] 赵钊. 浅谈档案职业伦理问题与对策[J]. 兰台世界, 2015(12):117.
[6] 唐敏. 土建类高职院校图书馆与学生信息素养教育实践研究[J]. 兰台世界, 2015(12):159.
[7] 陈林. 弘扬图书馆精神, 推进图书馆事业发展[J]. 兰台世界, 2015(12):163.
[8] 谢小红. 民生档案伦理利用分析[J]. 兰台世界, 2015(12): 115.

(上接第203页)

可以提升档案远程服务能力,实现档案资讯化应用,最大限度地发挥档案和档案信息的价值;档案数字化需求的第四层次,可以使得档案信息关联性的展示成为现实,使档案不再是单一的个体,成为一个有机的联系体,将彻底改变传统档案死板的印象。新标准的出台,将使得档案数字化需求得到更好的实现,加快档案信息化建设。

《纸质档案数字化规范》自发布以来,虽还较少的应用到实际工作中,但将会更有效的督促纸质档案数字化工作,促进各领域档案数字化工作的标准化发展。同时,新标准也在一定程度上对档案外包企业的技术提出了要求,新标准的出台对提高档案工作质量、促进档案事业发展的目标产生了积极影响。另外,新标准在一定程度上还能促使档案行业的相关人员和机构提升自身技术水

平,有利于档案事业的发展。

参考文献:

[1] 郝晨辉. 档案行业标准《纸质档案数字化技术规范》的修订[J]. 数字与缩微影像, 2016(6):9-12.
[2] GB/T 26163.1-2010. 信息与文献 文件管理过程 文件元数据[S].
[3] 国家档案局. 全国档案事业发展“十三五”规划纲要[J]. 中国档案, 2016(5):14-17.
[4] 吕玉洁, 向禹. 确定性需求驱动的档案数字化加工监理模式构建与适用性分析[J]. 档案学研究, 2018(1):107-112.
[5] 尹显东, 李在铭, 姚军, 等. 图像压缩标准研究的发展与前景[J]. 信息与电子工程, 2003(12):326-331.
[6] DA/T 31-2017. 纸质档案数字化规范[S].