

国家数字图书馆工程标准规范成果

国家图书馆图像数据 加工标准和操作指南

朱 强 张春红 龙 伟 主编



國家圖書館出版社

中国图书馆学会图书馆分会 中国图书馆学会图书馆分会

国家图书馆图像数据 加工标准和操作指南

国家图书馆 编 中国图书馆学会图书馆分会 编



国家图书馆 编 中国图书馆学会图书馆分会 编

国家图书馆图像数据 加工标准和操作指南

朱 强 张春红 龙 伟 主编

國 國家圖書館出版社

图书在版编目(CIP)数据

国家图书馆图像数据加工标准和操作指南/朱强,张春红,龙伟主编. —北京:国家图书馆出版社,2011.12

(国家数字图书馆工程标准规范成果)

ISBN 978 - 7 - 5013 - 4677 - 6

I. ①国… II. ①朱… ②张… ③龙… III. ①视听资料—图书馆工作—指南 IV. ①G255.73 - 62

中国版本图书馆 CIP 数据核字 (2011) 第 196394 号

责任编辑:高爽

书名 国家图书馆图像数据加工标准和操作指南

著者 朱 强 张春红 龙 伟 主编

出版 国家图书馆出版社(原北京图书馆出版社)

(100034 北京市西城区文津街7号)

发行 010 - 66139745 66151313 66175620 66126153

66174391 (传真) 66126156 (门市部)

E-mail btsfxb@nlc.gov.cn (邮购)

Website www.nlcpress.com→投稿中心

经销 新华书店

印刷 北京科信印刷有限公司

开本 787 × 1092 (毫米) 1/16

印张 4.75

版次 2011 年 12 月第 1 版 2011 年 12 月第 1 次印刷

字数 45 (千字)

书号 ISBN 978 - 7 - 5013 - 4677 - 6

定价 58.00 元

丛书编委会

主 编：国家图书馆

编委会：

主 任：周和平

执行副主任：詹福瑞

副主任：陈 力 魏大威

成 员(按姓氏拼音排名)：卜书庆 贺 燕 蒋宇弘

梁蕙玮 龙 伟 吕淑萍 申晓娟 苏品红

汪东波 王文玲 王 洋 杨东波 翟喜奎

赵 悦 周 晨

本书编委会

主 编：朱 强 张春红 龙 伟

编 写：肖 珑 邵 珂 唐 勇 朱本军 宋庆生

王 浩 赵四友 李鹏云 韩 超 李成文

总 序

数字图书馆涵盖多个分布式、超大规模、可互操作的异构多媒体资源库群,面向社会公众提供全方位的知识服务。它既是知识网络,又是知识中心,同时也是一套完整的知识定位系统,并将成为未来社会公共信息的中心和枢纽。数字图书馆建设的最终目标是实现对人类知识的普遍存取,使任何群体、任何个人都能与人类知识宝库近在咫尺,随时随地从中受益,从而最终消除人们在信息获取方面的不平等。“国家图书馆二期工程暨国家数字图书馆工程”是国家“十五”重点文化建设项目,由国家图书馆主持建设,其中国家数字图书馆工程的建设内容主要包括硬件基础平台、数字图书馆应用系统和数字图书馆标准规范体系。

标准规范作为数字图书馆建设的基础,是开发利用与共建共享资源的基本保障,是保证数字图书馆的资源和服务在整个数字信息环境中可利用、可互操作和可持续发展的基础。因此,在数字图书馆建设中,应坚持标准规范建设先行的原则。国家数字图书馆标准规范体系建设围绕数字资源生命周期为主线进行构建,涉及数字图书馆建设过程中所需要的主要标准,涵盖数字内容创建、数字对象描述、数字资源组织管理、数字资源服务、数字资源长期保存五个环节,共计三十余项标准。

在国家数字图书馆标准规范建设中,国家图书馆本着合作、开放、共建的原则,引入有相关标准研制及实施经验的文献信息机构、科研机构以及企业单位承担标准规范的研制工作,这就使得国家数字图书馆标准规范的研制能够充分依托国家图书馆及各研制单位数字图书馆建设的实践与研究,使国家数字图书馆的标准规范成果具有广泛的开放性与适用性。本次出版的系列成果均经过国家图书馆验收、网上公开质询以及业界专家验收等多个验收环节,确保了标准规范成果的科学性及实用性。

目前,国内数字图书馆标准规范尚处于研究与探索性应用阶段,国家图书馆担

负的职责与任务决定了我们在数字图书馆标准规范建设方面具有的责任。此次将国家数字图书馆工程标准规范研制成果付梓出版,将为其他图书馆、数字图书馆建设及相关行业数字资源建设与服务提供建设规范依据,对于推广国家数字图书馆建设成果,提高我国数字图书馆建设标准化水平,促进数字资源与服务的共建共享具有重要意义。

国家图书馆馆长 周和平
2010 年 8 月

前 言

图像数据加工标准是数字图书馆标准规范体系的重要内容之一。由于图像资源数字化与数字图书馆的资源发布、长期保存和广泛共享密切相关,且具有资金密集和风险性大的特点,因此为了充分获得最佳效益,制定统一、有效、适用的图像数据加工标准具有重要意义。

在过去的几年中,世界各国各地的图书馆及机构如美国国会图书馆(Library of Congress)、英国 AHRC(Arts and Humanities Research Council)、欧洲 TAPE(Training for Audiovisual Preservation in Europe)项目、美国马里兰大学图书馆和哈佛学院图书馆等开展了一系列数字化项目,积累了大量的数字资源和标准规范成果;在国内,国家科技部的“我国数字图书馆标准规范建设(CDLS)”项目和国家“211 工程”的“大学数字图书馆国际合作计划(CADAL)”项目也都在这方面取得了阶段性进展。已有的图像数据加工标准及其实践成果包括数字图像编码规则、元数据规范和著录规则、格式体系、应用级别、加工设备、后期处理、品质检验、命名规则、数据保存和管理等方面,但未能形成面向全流程应用的、完善的标准规范体系和可操作性强的操作指南,导致图像数据资源在保存和利用方面存在很多的问题。

北京大学图书馆先后承建或参加了科技部“中国数字图书馆标准与规范”建设项目的相关子项目——数字资源加工标准、基本元数据规范、数字资源描述标准规范的完善与扩展建设等;参加了 CADLIS“中国高等教育数字图书馆技术标准与规范”的编写;还作为全国文献影像技术标准化技术委员会的成员单位之一共同参与相关国家标准的制修订工作,最新参与的是 2010 年国家标准《文档管理 便携式文档格式 第一部分:PDF 1.7》的制定。设立于北京大学图书馆的北京大学数字加工中心,是数据加工、数字资源存储和发布为一体的公共服务机构,积累了大量的数字资源成果和数字加工实践经验。国家图书馆和北京大学图书馆作为牵头单

位,研制的科技部《数字资源加工标准》《基本元数据规范》等多项成果被广泛应用于北京大学数字加工实践、CALIS 特色库项目和其他数字化项目等方面,受到数字加工业务各委托单位和 CALIS 各成员院校的普遍认可。

2008 年,国家数字图书馆工程标准规范项目“图像数据加工标准与工作规范”研制工作采取竞争性谈判方式向社会各界发出参与研制的邀请。北京大学图书馆在图像数据加工标准规范研制和数字加工实践方面具有深厚的基础,从而得到评选专家首肯而中标,成为该标准的研制单位。

国家图书馆“图像数据加工标准与工作规范”项目,是国家数字图书馆工程标准规范建设项目之一。该项目的主要研究任务是根据国家图书馆图像文献和数字资源内容的情况,参考数字加工领域的最新技术成果,制定图像数据加工标准和操作指南。项目研制成果应该满足国家图书馆数字馆藏建设要求,符合数字对象资源的长期保存和使用目的。

项目由北京大学朱强馆长和肖珑副馆长负责统筹和审核,研制任务的具体承担人如下。

调研报告:国内外数字加工相关的标准规范,邵珂;国外数字加工标准应用项目,唐勇;中国内地和台湾地区数字加工和标准规范建设项目,张春红。

标准规范:技术参数,邵珂;元数据加工,唐勇;流程和质量管理的,张春红。

操作指南:各类型图像资源数字加工流程和操作指南,张春红;流程图和元数据加工指南,唐勇;特殊图像资源加工注意事项,邵珂。

加工标准测试和报告:邵珂、宋庆生、朱本军。

项目研究成果主要参考前述国家科技部相关项目的图像数字对象成果,对国内外主要的数字图书馆资源建设项目和国际通用的数字资源编码、文件格式标准进行较全面的调研。项目研究注重数字对象加工的全流程控制和管理,根据文献载体特点和属性制定加工标准和操作指南,较好地满足了国家图书馆图像数字资源长期保存和应用的消费需求,具有科学性、适用性和可操作性。

本书是以国家图书馆“图像数据加工标准与工作规范”项目的研究成果为基

基础编写而成,主旨是为国家图书馆图像数字对象加工提供原则性的指导,除指导国家图书馆数字加工业务实践,还体现了国际国内数字加工领域技术和标准的最新发展,国内各相关单位在数字加工实践中、各相关领域和行业在进行数字对象加工与管理时,可以参考使用。

编者

2010 年 12 月 31 日

目 录

前 言	(1)
第一部分 国家图书馆图像数据加工标准	(1)
1 范围	(3)
2 规范性引用文件	(3)
3 术语和定义	(4)
4 原则	(5)
5 数字编码标准	(6)
6 内容标记	(6)
7 资源应用级别	(7)
8 加工技术标准和格式体系	(8)
9 数字加工流程	(12)
10 命名规则	(16)
附录 A(资料性附录) 结构代码	(17)
附录 B(资料性附录) 资源级别代码	(18)
第二部分 国家图书馆图像数据加工标准操作指南	(19)
1 各类型图像资源的加工流程说明	(21)
2 各流程工作规范与操作指南	(25)
参考文献	(57)
后 记	(59)

第一部分 国家图书馆图像数据加工标准

1 范围

本标准规范给出了文献资源、原生图像数字资源和派生图像数字资源的加工标准和工作规范。

本标准规范既适用于尚未数字化的、以不同形式和载体存储的信息资源,如书刊、善本,需要运用图像采集与处理,文字、图像的识别以及对数字化初始信息的各种再加工技术,将其转化成能够为计算机处理的数据信息;又适用于最初产生时,即以数字图像形态存在,如图书附带的数字图像资源、源于互联网的自产资源;也适用于文献经过数字化加工后派生出的数字图像资源。

2 规范性引用文件

下列文件中的条款通过本标准的引用而成为本标准的条款。凡是注日期的引用文件,其随后所有的修改单(不包括勘误的内容)或修订版均不适用于本标准,然而,鼓励根据本标准达成协议的各方研究是否可使用这些文件的最新版本。凡是不注日期的引用文件,其最新版本适用于本标准。

GB/T 22113—2008 印刷技术 印前数据交换 用于图像技术的标签图像文件格式 (TIFF/IT)

GB/T 19894—2005 数码照相机光电转换函数(OECF)的测量方法(eqv ISO 14524:1999)

ISO 32000—1:2008 文献管理 便携式文档格式 第一部分:PDF 1.7

ISO 19005—1:2005 文献管理 长期保存用电子文档文件格式 第1部分:1.4版PDF的使用(PDF/A-1)

ISO 4087:2005 缩微摄影技术 在35mm卷片上拍摄永久保存报纸的规定

ISO 21550:2004 摄影 图像用电子扫描仪 动态范围测量

ISO 16067—2:2004 摄影 摄影图像用电子扫描仪 空间分辨率测量 第2部分:胶片扫描仪

ISO 16067—1:2003 摄影 摄影成像用电子扫描仪的空间分辨率测量 第1部分:反射介质用扫描仪

ISO 15739:2003 摄影 电子静止图像成像 噪声测量

ISO/IEC 15444—12:2005 信息技术 JPEG 2000 图像编码系统

ISO/IEC 10918 信息技术 连续色调静止成像的数字压缩和编码

ISO/IEC 14772 信息技术 计算机图形和图像处理 虚拟现实建模语言

ISO/IEC 11544:1993 信息技术 图像和声音信息的代码表示 二级图像的渐进压缩

3 术语和定义

3.1 图像(Image)

用各种观测系统以不同形式和手段观测客观世界而获得的,可以直接或间接作用于人眼并进而产生视知觉的实体。

3.2 数字图像(Digital Image)

是由扫描仪、摄像机等输入设备捕捉实际的画面产生的,是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。

3.3 缩微胶卷(Microfilm)

条或卷形状的缩微品。

注:来源于 GB/T 4894—2009。

3.4 像素(Pixel)

构成数字图像的最小单位,通常用若干不同色彩像素排列而组成的矩阵来表示数字光栅图。

3.5 图像分辨率(Image Resolution)

图像中存储的信息量,可以有多种衡量法。本标准规范中以每英寸的像素 ppi(pixels per inch)来衡量。

注1:图像分辨率和图像尺寸一起决定文件的大小和输出质量。该值越大,图像文件所占用的磁盘空间也越大,打印或修改图像等操作所需时间也就越多。

注2:图像分辨率以比例关系影响着文件的大小,即文件大小与其图像分辨率的平方成正比关系,图像分辨率也影响到图像在屏幕上显示的大小。

注3:本书“第一部分 国家图书馆图像数据加工标准”中凡是涉及各类图像资源的图像分辨率标准的,一律以 ppi 表示;“第二部分 国家图书馆图像数据加工标准操作指南”中涉及设备指标时使用的设备分辨率描述则用 dpi 表示。

3.6 动态范围(Dynamic Range)

最暗区域到最亮区域的密度差值。动态范围对扫描设备来说它指的是能记录原稿的色调范围。对感光元件和胶片来说,它表示图像所包含的从最暗到最亮的范围。其范围越大,所能记录的暗部和亮部细节越多,所能表现层次也越丰富。

3.7 色调(Tonality)

色彩外观的基本倾向。通常可以从色相、明度、冷暖、纯度四个方面来定义一幅作品的色调,在上述四个要素中,某种因素起主导作用,可以称之为某种色调。

3.8 色度(Colorfulness)

色调和饱和度的合称。色调决定色彩的本质类别,饱和度决定颜色的深浅。

3.9 小型实物(Physical Object)

图书馆所藏甲骨、织帛、竹简、陶瓷、雕塑等早期实体文物。

3.10 对比度(Contrast Ratio)

一幅图像的最高密度与最低密度的差异。

3.11 无损压缩(Lossless Compression)

去掉或减少数据中的冗余,但这些冗余信息可以重新插入到数据中。无损冗余压缩是可逆的过程,也称无失真压缩。

3.12 有损压缩(Lossy Compression)

允许一定程度的失真,可用于图像、声音、动态视频等数据的压缩,压缩比可达到几十倍甚至上百倍。有损压缩是不可逆的过程,损失的信息不能再恢复。

3.13 图像拼接(Image Stitching)

将数张有重叠部分的图像(可能是不同时间、不同视角或者不同传感器获得的)拼成一幅大型的无缝高分辨率图像。

注:图像拼接技术在与图像数字加工相关的档案数字化保存、遥感图像处理、医学图像分析等领域有广泛的应用。图像拼接技术大致可以分为图像预处理、图像配准和图像缝合三个基本步骤,各步骤中均有所需要的各类经典算法。

4 原则

前瞻性:尽量提高数字化产品在将来使用的可能性。

适用性:充分考虑不同类型资源的特点和对其进行数字加工的特殊性因素。

科学性:在标准制定过程中,尽可能直接引用和采纳国际上已经普遍应用的技术标准和规范。

兼容性:尽量保持与“中国数字图书馆标准与规范建设”项目(CDLS)、中国高等教育数字