

大学数字图书馆国际合作计划
缩微胶片数字化加工标准与规范

分类：数字对象元数据标准规范集

编号：ysj_zmysjzlgz_wsjp

负责人：曾祥瑞

完成日期：2011年9月9日

文档备忘录

编号	标题	日期	状态
01	缩微胶片数字化加工标准与规范	2011-9-9	初稿 01

(标准规范在正式发布前后, 根据意见会有几次主要修订, 在此留下记录信息)。

版权声明

本报告研究工作属于教育部项目“大学数字图书馆国际合作计划”的一部分。按照有关规定，国家和“大学数字图书馆国际合作计划”项目管理中心拥有本报告的版权，依照《中华人民共和国著作权法》享有著作权。

本报告可以复制、转载、或在电子信息系统上做镜像，但在复制、转载或镜像时须注明真实作者和完整出处，并在明显地方标明“大学数字图书馆国际合作计划项目资助”的字样。

报告版权人不承担用户在使用本作品内容时可能造成的任何实际或预计的损失。

本报告作者谨保证本作品中出现的文字、图片、声音、剪辑和文后参考文献等内容的真实性和可靠性，愿按照《中华人民共和国著作权法》，承担本作品发布过程中的责任和义务。教育部、“大学数字图书馆国际合作计划”项目管理中心等有关管理机构对于本作品内容所引发的版权、署名权的异议、纠纷不承担任何责任。

“大学数字图书馆国际合作计划”项目组网站（<http://www.cadal.cn>）是本报告的第一发表单位。在不发生重复授权的前提下，报告撰写人保留将经过修改的项目成果向正式学术媒体直接投稿的权利。

摘要

对与缩微胶片数字化加工标准和规范相关的术语给出了详细的定义,标准规范内容包含了缩微胶片数字化的基本加工流程,主要包括胶片整理、目录建库、胶片扫描、图像处理和识别转换、目录制作与封装、信息存储、检索利用等工序。

内容目录

1 概述.....	1
2 术语定义.....	1
3 标准规范内容.....	4
4 参考文献.....	9
5 附录.....	9

1 概述

缩微技术是目前信息资料存储最通用最常见的存储技术之一,在图书、档案、金融、卫生、工矿企业等领域得到广泛的应用。特别是《1997-2010年全国公共图书馆文献缩微规划》通过后,公共图书馆大量地开展地方志、民国时期文献、古籍、少数民族文献、外文古旧文献等的缩微工作,制作了数百万的缩微胶片。

随着信息技术和网络技术的飞速发展,缩微胶片查找速度慢、阅读困难、易损坏等缺点严重阻碍了缩微胶片的传播与使用,仅仅作为一种文献存储方式存在。因此,缩微胶片数字化是缩微胶片有效利用的必然趋势。

缩微胶片数字化就需要加工规范的标准化,必须用一系列标准来控制和指导数字化的全过程。由于国内至今无缩微胶片数字化加工标准与规范,只有完善了这方面的工作,才能使缩微文献数字加工规范化。

考虑不同规格、不同材质的缩微胶片实际特点,电子文档国际标准,读者阅读使用习惯,以通用性、经济性、规范性为原则,制订缩微胶片数字化加工标准与规范。

2 术语定义

2.1 缩微胶片

指人类利用胶卷摄影技术,复制书籍、报纸等出版物上的文字和图片,汇集成的微粒高解像力的胶片,包括卷片和平片,本规范特指平片。

2.2 数字化

指利用计算机技术将模拟信号转换为一系列二进制代码的数字信号,引入计算机内部,进行统一处理的过程。

2.3 缩微胶片数字化加工

指采用缩微胶片扫描仪等设备将缩微胶片上的影像转换为存储在光盘、硬盘等载体上并能被计算机识别的数字图像或数字文本的处理过程。

2.4 数字图像

指数码图像或数位图像，它以二进制数字形式进行存储和处理。

2.5 黑白二值图像

指与彩色及灰度相对应的一种图像模式，图像中每个像素的亮度值仅可以取 0 和 1。它在表示图像时只区分黑白二值，把其它色彩根据其特性转换为黑白两种，因此形成只有两种颜色的图像。

2.6 图像压缩

指以较少的比特有损或无损的表示原来的像素矩阵的技术，目的是减少图像中的冗余信息，从而用更加高效的格式存储和传输数据。

2.7 灰度值

指进行灰度扫描时，黑白图像中点的颜色深度，范围一般从 0 到 255，白色为 255，黑色为 0，共 256 级灰度。

2.8 TIFF

指标记图像文件格式，它是一种基于标记的无损(不丢失信息)压缩格式，用于在应用程序之间和计算机平台之间交换文件，它的出现使得图像数据交换变得简单。因为它存储的图像细微层次的信息非常多，图像的质量也得以提高，故而非常有利于原件为黑白胶片的复制储存。

2.9 分辨率

指单位长度内图像包含的点数或像素数，分辨率越高代表图像品质越好，越能表现出更多的细节。单位有 DPI（点每英寸）、LPI（线每英寸）和 PPI（像素每英寸），本规范采用 DPI。

2.10 失真度

指胶片原件在进行数字化处理后，在同等测试环境下，数字图像与胶片原件在色彩、几何、压缩算法等方面的差值。

2.11 可懂度

指数字图像向人或机器提供信息的能力。

2.12 对比度

指画面黑与白的比值，也就是从黑到白的渐变层次。比值越大，图像越清晰醒目，色彩也越鲜明艳丽。高对比度可以使数字图像在黑白之间具有较少的灰色层次并且显示出较少的细节，低对比度具有更多的灰度层次及细节。

2.13 曝光亮度

指决定数字图像敏感强度的指数。高曝光亮度可以使数字图像的色度加深并显示出较多的细节，突出浅色信息的痕迹，同时黑白之间出现的底灰加大。

2.14 密度

指在缩微摄影技术中衡量感光材料曝光和显影后变黑程度的物理量，也可用透射密度 (D_T) 来表示，定义为不透明度的基 10 对数。即：

$$D_T = \log_{10} (1 / \tau) = \log_{10} (I_i / I_t)$$

D_T ——透射密度

τ ——透射率， $\tau = (I_t / I_i)$

I_i ——光能

I_t ——透射光

2.15 解像力

指衡量胶片影像细部能力的物理量，解像力的高低通常以在每毫米胶片上所能分辨的最大线条对数来表示。

3 标准规范内容

3.1 基本内容

缩微胶片数字化的基本加工流程主要包括胶片整理、目录建库、胶片扫描、图像处理和识别转换、目录制作与封装、信息存储、检索利用等工序。

3.2 胶片整理

3.2.1 分类

根据胶片所属报告的类别以及编号对胶片进行分类，根据拟数字化的情况，挑出图像模糊、曝光过度、破损的胶片，并做出标识。估算数字化后所需要的存储空间。

3.2.2 分本

对批量扫描前的胶片进行胶片数量、胶片号的标注。如在标注中与原胶片中的胶片数量、胶片号不一致，以此为准。

3.2.3 拆分

去除胶片中原来的封装物，保证扫描的顺序，确保数字化后正确归卷并不损坏胶片。

3.2.4 建立作业单

制作并填写数字化加工作业单，全程监控扫描（转换）过程，并记录整理数字化过程中各种相关情况。

3.2.5 封装

图像质检完成后，拆除过封装物的胶片按胶片原貌及保管的要求重新封装，做到安全、准确、无遗漏。

3.3 建目录库

3.3.1 目录著录

根据 DC 元数据确定胶片著录项，利用 CADAL OEB editor 工具进行元数据著

录，包括确定胶片目录的著录项、字段长度和内容要求。如有错误或不规范的题名、责任者、时间、张号（数）、等，要进行修改和补充，使其规范化。

3.3.2 数据格式选择

所选定的数据格式通用，应能直接或间接与 DBF 文件格式或通过 XML 文档进行数据交换。

3.3.3 目录输入

将著录好的目录输入计算机内，建立机读目录数据库。

3.4 胶片扫描

3.4.1 扫描方式

根据胶片幅面的大小在扫描时选用不同的焦距。对于较清晰的图像，可以采用 FS300 高速扫描仪进行扫描，以提高工作效率。对于清晰度较差的图像，用 MS7000 手动扫描仪比较容易调整黑白度以及对比度。对于特别模糊的图像，用 MS7000 手动扫描仪扫描时将黑度调整到 100，对比度调整到 1。对于一些特殊的胶片扫描成负片。

3.4.2 扫描色彩模式和分辨率

扫描色彩模式一般有黑白二值、灰度、彩色等，在实际操作中，选择黑白二值模式，分辨率选择 600dpi。

3.4.3 文件命名方式

扫描时文件夹的建立按胶片编号进行建立。文件夹命名方法，以科技报告的命名为例，一般为 10 位数，如 PB84127431，文件夹下图像文件命名以编号 00000001 开始顺序递增。

3.4.4 数据上传

将扫描的图像上传到 FTP 服务器，进行下一步的图像处理。

3.5 图像处理和识别转换

3.5.1 图像处理

3.5.1.1 纠偏

对扫描过程中出现的偏斜图像进行整体纠正，保证数字图像的偏斜角度小于1度。{图像偏斜不超过页面内半个文字}。

3.5.1.2 去污

去除数字图像中影响可懂度的杂质。胶片数字化图像的去污，遵循在不影响可懂度的前提下展现原貌的原则。具体工作分为局部去污，如去除黑边、污点。整体去污，可一次性去除页面上的污渍。

3.5.1.3 校对

3.5.1.3.1 一次校对，检查扫描完的图像质量情况，确保图像完整无误，并视需要对有问题的扫描图像进行取版心、去黑边、纠偏操作，对不合格的图像进行标记，退回重新扫描。

3.5.1.3.2 纠错处理，根据一次校对中提出的削口污渍、黑边、偏斜、图像质量等问题，对每份图像进行相应的处理。

3.5.1.3.3 二次校对，对一次校对及数据处理后的数字图像再次进行检查，并标记图像处理过程中不清晰的页面。不合格的退回重新处理。

3.5.2 图像质量检查

对扫描后形成的图像文件进行清晰度、污渍、黑边、偏斜等问题的控制，以达到要求的图像质量。

保证扫描后的数字图像清晰，便于清楚阅读。

3.5.3 数据挂接

胶片数据与图像文件的对应准确程度的控制。在生成裸数据光盘之前，根据资料整理信息、扫描前的标引信息及胶片文件信息核查图像文件总数是否与实际文件数相等，如不相等则不能生成裸数据光盘，可打印出清单，回退给图像处理

人员补扫。

3.5.4 数据检验

3.5.4.1 将裸数据进行验收前的相应处理,包括文件的格式转换、逻辑分盘处理、添加说明性文件。

3.5.4.2 对数据进行检验,给出合格或不合格结论。

3.5.4.3 将数字加工后的数据转换为裸数据光盘格式,拷贝到活动硬盘上。

3.5.4.4 系统要能自动记录检验进度。

3.5.5 数据验收

3.5.5.1 数据抽检

3.5.5.1.1 以抽检方式检查目录数据、图像数据的质量。

3.5.5.1.2 一个全宗的档案,数据抽检的比率不得低于5%。

3.5.5.2 验收指标

3.5.5.2.1 目录数据、图像数据有不完整、不清晰等质量问题时,抽检标记为“不合格”。

3.5.5.2.2 一个全宗的档案,数据质量抽检的合格率达到95%以上(含95%)时,验收予以“通过”。

3.5.5.3 合格率计算方法

1、统计抽检标记为“不合格”的文件数。

2、抽检合格的文件数 = 抽检文件总数 - 抽检不合格的文件数

3、合格率 = 抽检合格的文件数 / 抽检文件总数 × 100%

3.5.5.4 验收登记

填写数据验收登记表,记录验收结果。

3.5.5.5 验收审核

验收结论必须经有关领导审核、签字后方有效。

3.5.6 数据上载

胶片数字化各工序的数据通过网络及时上载到数据服务器端汇总,其中数字图像自动搜索对应的目录数据,加入对应的电子地址数字图像文件名,建立起一一对应的关系。

3.5.7 数据备份

定期对服务器的各类数据进行备份，防止数据丢失。

3.6 目录制作与封装

对处理完的图像，封装成册，生成文件、制作胶片目录、添加页面链接。

3.7 信息存储

以 JPEG 数字压缩编码或 TIFF 国际通用标准格式压缩后，按胶片编号存储。

数字图像的存储管理保持原胶片的保管模式对应存储。

3.7.1 压缩存储格式

胶片数字图像格式采用 TIFF 和 JPEG 数字压缩编码。

3.7.1.1 黑白二值图像

黑白二值扫描图像文件采用 TIFF (CCITTG3) 格式二值图像压缩算法，压缩串(Cr)为 15: 1。也可以采用 TIFF (CCITTG4) 格式二值图像压缩算法，压缩率(Cr)为 30: 1。

3.7.1.2 连续色调静态图像

连续色调静态图像采用 JPEG 数字压缩编码，平均压缩率(Cr)为 15: 1。

3.7.2 编码方式

采用国际通用编解码算法。

3.7.3 存储方式与载体

可选用在线和离线、不同载体进行存储。

3.8 检索利用

3.8.1 检索利用方式

胶片数字图像检索利用可采用单机、局域网和互联网三种方式。

3.8.2 检索软件配置

检索软件需符合 CADAL 制定的各项要求,具备目录检索与胶片数字图像方便调阅的基本功能。

3.9 加工设备与技术规范

3.9.1 设备类型:所采用的缩微胶片扫描仪(MS7000)和高速缩微胶片扫描仪(FS300)。

3.9.2 转换方式:缩微胶片的构成方式采用“高对比”,压缩方式采用“G4-2D”,保存方式采用“黑白”方式。

3.9.3 分辨率:分辨率采用 600dpi。

3.9.4 文件存储类型:转换的图像存为 TIF 格式。

3.9.5 文件存储方式:胶片按报告号编号、分级建文件夹存储,其中的每一张胶片单独建文件夹;按照资料编号建文件夹存储。

3.9.6 转换采用母片来进行。

3.10 补充说明

3.10.1 不同的缩微胶片扫描仪可能参数设置会有不同。

3.10.2 转换后的图像按照胶片的不同而分别存储。

3.10.3 按照以上办法转换后的图像容量大小跟图像的清晰程度有很大关系,高质量的胶片转换后不仅图像的清晰度更高,而且形成的文件容量也更小。因此在转换时,采用高清晰度的母片。

4 参考文献

- 1 GB/T 18894-2002 《电子文件归档与管理规范》
- 2 DA/T 31-2005 《纸质档案数字化技术规范》
- 3 DA/T 43-2009 《缩微胶片档案数字化技术规范》
- 4 辽阳市档案数字化技术标准 2009年6月9日
- 5 郑州市档案数字化技术标准

5 附录