ICS 01.140.20
A 14

C A D A L 项 目 标 准

CADAL 10103.1—2019

图书期刊数字对象制作规范

Specification of Book and Periodicals Digitization

第三稿

2019-02-10

2019-02-03 发布 2019-02-10 实施

CADAL 项目管理中心 发布

目次

前言

《CADAL 数字对象加工规范》分成 4 个部分，由 13 个标准组成。

——第 1 部分：CADAL 10101—2019 数字对象采集规范。

——第 2 部分：CADAL 10102—2019 数字对象制作基本流程规范，这部分根据加工对象的不同又分成 8 个子规范。

- 第 1 子规范：CADAL 10103.1—2019 图书期刊数字对象制作规范；
  CADAL 10103.2—2019 Book Digitalization Specification。
- 第 2 子规范：CADAL 10104—2019 报纸数字对象制作规范。
- 第 3 子规范：CADAL 10105—2019 文档数字对象制作规范。
- 第 4 子规范：CADAL 10106—2019 图片数字对象制作规范。
- 第 5 子规范：CADAL 10107—2019 古籍数字对象制作规范。
- 第 6 子规范：CADAL 10109—2019 视频数字对象制作规范。
- 第 7 子规范：CADAL 10110—2019 音频数据加工标准与操作规范。
- 第 8 子规范：CADAL 10227—2019 缩微胶片数字化加工标准与操作规范。

——第 3 部分：CADAL 10111—2019 数字内容编码与内容标记规范。

——第 4 部分：CADAL 10112-2019 数字对象加工与应用等级标准。

本标准为第 2 部分的第 1 子规范之一。

《CADAL 数字对象加工规范》代替 CADAL 项目一期制定的《数字化文本加工规范草案》。

2019 年标准修正主要针对《图书期刊数字对象制作规范》中的目标文档扫描参数及发布级图像格式进行修改。

本标准由大学数字图书馆国际合作计划（CADAL）项目管理中心提出并归口。

本标准起草单位：杭州中元数据科技有限公司、深圳市点通数据有限公司、浙江大学图书馆。

本标准主要起草人：周小芳、郑传双、薛霏。

本标准于 2012 年 05 月发布，于 2013 年做了部分修改，并于 2019 年做了重大的修改，修改详情参见附件。

引　　言

数字对象加工规范是数字图书馆资源建设的基础，制定数字对象加工规范的目的是让数字图书馆资源建设单位，在数字对象采集、加工、封装、存储等环节中有统一的规格和操作方法，保持数字资源的格式与内容形式的一致性。

《CADAL 数字对象加工规范》是 CADAL（China Academic Digital Associative Library）项目关于数字对象加工的规范集，是 CADAL 项目数字对象加工必须遵从的基础性企业标准。

《图书期刊数字对象制作规范》的基本目的是保证 CADAL 项目图书期刊资源采集质量，主要解决：

（1）界定图书期刊资源的加工目标；

（2）规定图书期刊资源的成品数字资源格式、内容、保存方式。

与之相对应的英文规范，请参考 CADAL 10103.2—2019 Book Digitalization Specification。

图书期刊数字对象制作规范

1 范围

本部分规定了图书期刊数字对象制作过程中的原则、采集要素、加工标准、存储格式、目录结构、特例处理等。

本部分适用于图书期刊数字对象加工制作过程管理与质量检测。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅所注日期的版本适用于本文件。凡是不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 3469	文献类型和文献载体代码
GB/T 3792. 1	文献著录第 1 部分总则
ISO10646—1: 2000	信息技术——通用多八位编码字符集
CADAL 10101—2019	数字对象采集规范
CADAL 10102—2019	数字对象制作基本流程规范
CADAL 10111—2019	数字内容编码与内容标记规范
CADAL 10112—2019	数字对象加工与应用等级规范
CADAL 10301—2019	数字对象唯一标识符规范
CADAL 10302—2019	数字对象内部标识与命名规范

3 术语和定义

3.1 数字对象 Digital Object

数字对象指一组通过数字化加工得到的、描述一个特定的实物资源的、可存储于计算机并可利用计算机技术进行再现的数据集合。

3.2 图书期刊数字对象 Book/Issues Digital Object

图书期刊数字对象指从图书或期刊资源（包括原始出版物、缩微或影印复制品）中采集、加工得到的数字对象。

3.3 原始图像 Original Image

原始图像指通过初始扫描、摄影、转换等手段直接获取的图像文件。

3.4 典藏级文件 Archive File

典藏级文件指数字对象采集过程中所获得的原始图像文件、原始音频文件、原始视频文件经过本规范许可的加工方法处理后得到的高精度、无压缩（或高品质压缩）的文件。

3.5 发布应用级文件 Application File

发布应用级文件指典藏级文件经过本规范许可的加工方法处理后得到的用于网上在线浏览的文件或特定应用的各类派生文件。

3.6 双层 PDF Text Hidden PDF

双层 PDF 指通过 OCR 等技术手段，将原文中每行文字内容放在底层，上层放置原始图像，继而形成的 PDF 格式的文件。

3.7 单层 PDF Image Only PDF

单层 PDF 指由原始图像直接转换而成的 PDF 文件。

3.8 DC 元数据 DC Metadata

DC 元数据指 Dublin Core 元数据。

3.9 目录结构 Catalog

目录结构指符合 XML 的 METS 规范的目录结构信息，包括目录节点名称、链接指向的页面文件编号。同时，一册书籍封装的多页 PDF 内部需要以书签形式呈现其目录导航信息。

3.10 资源结构 Guide

资源结构指将资源各部分内容组合成一个整体的内部结构关系，包括各资源片断间的并列、包含、从属、接续、引用关系等。

3.11 资源封装信息 Open Package Format

资源封装信息指数字对象封装成可发布与展示的资源过程中生成的各类信息。

4 原则

4.1 CADAL 项目制作完成的数字对象格式要求

——所有数据应以明码或公开的文件格式保存；
——数字对象能够在浏览器中进行展示；
——数字对象支持跨平台应用。

4.2 数字对象命名要求

所有文件命名应遵守 Q/CADAL 10111—2019《数字内容编码与内容标记规范》。

5. . 采集要素

CADAL 项目对图书数字化加工后形成的数字图书，要求必须包含以下 5 部分内容：
——典藏级图像文件；
——发布应用级图像文件；
——DC 元数据文件；
——目录结构信息文件；
——资源封装信息文件。

6 加工标准

6.1 藏级图像文件的制作要求

典藏级图像文件是对通过扫描，或原生电子数据通过转换采集到的图像文件进行符合 Q/CADAL 10112—2019《数字对象加工与应用等级标准》的加工处理后得到的图像文件，存放于数字对象目录的“otiff”子目录下：

——每一页一个文件，扫描文件从 00000001.JPG 或 00000001.JP2 开始依次命名；

——扫描从书籍封面至封底依次进行，书籍内所有页面（包括封面、封底、书名页、目录页反面的空白页和插页）都需要扫描。无论什么类型书籍，封面封底都必须按原貌采集，如果书籍封面或封底是空白页的请按原貌扫描，如果书籍缺失封面或封底的直接扫缺页。

6.1.1 典藏级图像文件不同的基本要求

CADAL 项目针对不同的扫描方式，分别给出了扫描标准（见表 1）。

表 1 典藏级图像文件扫描标准

页面样式		纯文字黑白页面	配图黑白页面	彩色页面
DPI	传统扫描仪	600	600	600
	拍摄式扫描仪	400	400	400
色阶	传统扫描仪	24位彩色	24位彩色	24位彩色
	拍摄式扫描仪	24位彩色	24位彩色	24位彩色
压缩方式	传统扫描仪	JPEG/JPEG2000	JPEG/JPEG2000	JPEG/JPEG2000
	拍摄式扫描仪	JPEG/JPEG2000	JPEG/JPEG2000	JPEG/JPEG2000

6.1.2 JPEG、JPEG2000 压缩参数设定要求

JPEG 压缩：需要将品质参数选成最高（100% Quality）

JPEG2000 压缩：品质参数不低于 Kakadu 6. 3 中的 Slope Value=51000（或 Kakadu6. 4 中的 Slope Value=42800）。

6.1.3 不完整图像处理要求

针对可能出现的原书缺页情况，CADAL 项目要求有明显标识。制作单位可采取如下两种方案。

方案一：在缺页处插入写有“原书缺页 Page Missed in Original Book”的图像文件（见图 1）。如果缺页涉及到目录导航信息，请在目录导航标题后加上“（缺）”，如“封面（缺）”、“封底（缺）”。正文里的缺页按页码进行扫描，连续缺几个缺页的就连续扫几张缺页。

图 1 缺页替代样例

方案二：提供一个 XML 格式的加工过程记录文件，在其中描述清楚每个页面的情况。

6.1.4 拍摄式扫描的颜色管理

使用拍摄式扫描仪加工图书时，应符合如下要求：

——应建立摄影棚和漫反射光源系统；

——应在封面之前、封底之后各扫描一次标准色卡，用于日后颜色校正处理。

——采用爱色丽 color checker classic 24 色卡 mini 达芬奇色板（X-Rite 24 色迷你色板），此款色卡材质为纸质，尺寸约 10.9cm*6.4cm(约名片大小)与色卡护照内的 class 目标尺寸相同，如图 2 所示：

图 2 爱色丽 color checker classic 24 色卡 mini 达芬奇色板

——封面之前的色卡以 00000000.JPG 或 00000000.JP2 命名，封底之后的色卡文件名按封底文件名加 1；

——色卡拍照时要求统一纵向摆放，拍照的底纹用色调均匀的黑色卡纸，拍照的色卡图片保持工整，没有倾斜，色卡基本位于黑色卡纸居中位置，如图 3 所示：

图 3 色卡拍照示例图

——色卡图片直接保留原始扫描图片，不做旋转之外任何图像处理，最终色卡图见图 4，存放到 otiff 文件夹里：

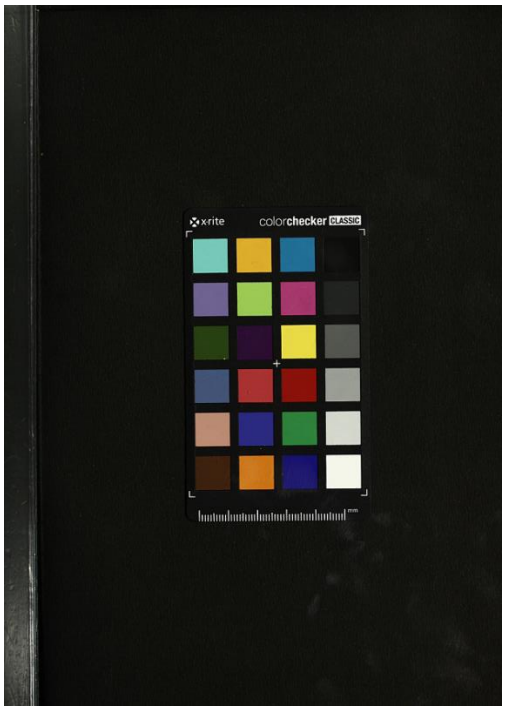


图 4 最终色卡示例图

6.1.5 扫描的一些特殊情况处理

（1）部分书籍本身页面褶皱，扫描后部分页面容易产生细微扭曲或褶皱。这种情况，扫描前请尽量将书本抚摸平整，无法摸平的，首先保证扫描图像可读。另外，在此书的 DC 的 description 字段里予以描述说明“书籍本身页面褶皱，扫描后部分页面产生细微褶皱，特此说明”。

（2）部分书籍装订较紧的（靠近书中缝处有少量文字被装订进去），应尽量扫描出来，确实无法扫描完整导致内容残缺的，在此书的 DC 的 description 里予以描述说明“书籍装订较紧或书籍本身页面内容残缺，部分页面扫描后内容残缺，特此说明”。

（3）原书封面、封底或其他页面破损，扫描时统一在下面垫一张白纸。

6.2 发布应用级图像文件

发布应用级图像文件是普通读者直接看到的页面，应保持基本的整洁。

6.2.1 CADAL 项目发布应用级图像文件的基本要求

——所有发布应用级图像文件应保持页面整洁:图像处理应在遵照书籍原貌的前提下进行，即保留原书籍里所有内容，包括馆藏印章、条形码、馆藏描述性文字、手写批注、各种颜色画线、题词、索书号标签纸等，对扫描带来的黑边须进行裁切处理，确保发布应用级图像文件页面整洁、美观；

——主体文字内容不能出现 90°侧倒或 180°颠倒，当页码和主体文字方向出现不一致时，以遵循文字阅读方向为主；

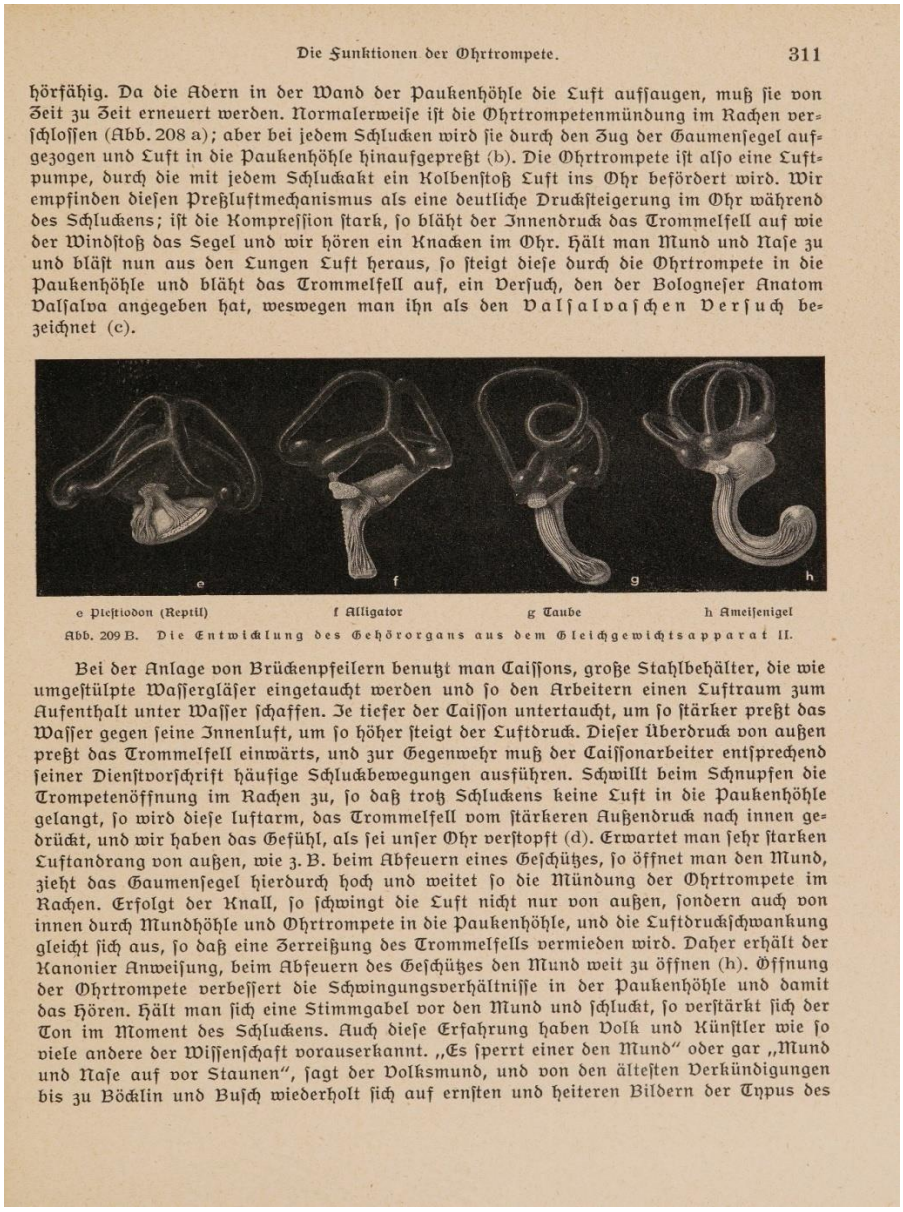
——页面整体倾斜不能超过 3°。

6.2.2 发布应用级图像文件的展现方式

所有发布应用级图像文件以 PDF 的方式展现：

——对手写体文稿、印刷质量较差（如油印、印刷模糊不清或字迹笔画断线）、图册、画册等 OCR 识别结果出现大量乱码不具备检索意义的页面，可直接由图像转换成单层 PDF，如图 5 所示：

Inhalt des vierten Bandes



Die Funktionen der Ohrtrompete. 311

hörfähig. Da die Adern in der Wand der Paukenhöhle die Luft aufsaugen, muß sie von Zeit zu Zeit erneuert werden. Normalerweise ist die Ohrtrompetenmündung im Rachen verschlossen (Abb. 208 a); aber bei jedem Schlucken wird sie durch den Zug der Gaumensegel aufgezogen und Luft in die Paukenhöhle hinaufgepreßt (b). Die Ohrtrompete ist also eine Luftpumpe, durch die mit jedem Schluckakt ein Kolbenstoß Luft ins Ohr befördert wird. Wir empfinden diesen Preßluftmechanismus als eine deutliche Drucksteigerung im Ohr während des Schluckens; ist die Kompression stark, so bläht der Innendruck das Trommelfell auf wie der Windstoß das Segel und wir hören ein Knacken im Ohr. Hält man Mund und Nase zu und bläst nun aus den Lungen Luft heraus, so steigt diese durch die Ohrtrompete in die Paukenhöhle und bläht das Trommelfell auf, ein Versuch, den der Bologneser Anatom Valsalva angegeben hat, weswegen man ihn als den Valsalvaschen Versuch bezeichnet (c).

e Pleftiodon (Reptil) f Alligator g Taube h Ameisenigel

Abb. 209 B. Die Entwicklung des Gehörorgans aus dem Gleichgewichtsapparat II.

Bei der Anlage von Brückenpfeilern benutzt man Caissons, große Stahlbehälter, die wie umgestülpte Wassergläser eingetaucht werden und so den Arbeitern einen Luftraum zum Aufenthalt unter Wasser schaffen. Je tiefer der Caisson untertaucht, um so stärker preßt das Wasser gegen seine Innenluft, um so höher steigt der Luftdruck. Dieser Überdruck von außen preßt das Trommelfell einwärts, und zur Gegenwehr muß der Caissonarbeiter entsprechend seiner Dienstvorschrift häufige Schluckbewegungen ausführen. Schwillt beim Schnupfen die Trompetenöffnung im Rachen zu, so daß trotz Schluckens keine Luft in die Paukenhöhle gelangt, so wird diese luftarm, das Trommelfell vom stärkeren Außendruck nach innen gedrückt, und wir haben das Gefühl, als sei unser Ohr verstopft (d). Erwartet man sehr starken Luftandrang von außen, wie z. B. beim Abfeuern eines Geschützes, so öffnet man den Mund, zieht das Gaumensegel hierdurch hoch und weitet so die Mündung der Ohrtrompete im Rachen. Erfolgt der Knall, so schwingt die Luft nicht nur von außen, sondern auch von innen durch Mundhöhle und Ohrtrompete in die Paukenhöhle, und die Luftdruckschwankung gleicht sich aus, so daß eine Zerreißung des Trommelfells vermieden wird. Daher erhält der Kanonier Anweisung, beim Abfeuern des Geschützes den Mund weit zu öffnen (h). Öffnung der Ohrtrompete verbessert die Schwingungsverhältnisse in der Paukenhöhle und damit das Hören. Hält man sich eine Stimmgabel vor den Mund und schluckt, so verstärkt sich der Ton im Moment des Schluckens. Auch diese Erfahrung haben Volk und Künstler wie so viele andere der Wissenschaft vorauserkannt. „Es sperrt einer den Mund“ oder gar „Mund und Nase auf vor Staunen“, sagt der Volksmund, und von den ältesten Verkündigungen bis zu Böcklin und Busch wiederholt sich auf ernsten und heiteren Bildern der Typus des

图 5 图像转换成单层 PDF 示例

——对其他类型的文档，要求制作成双层 PDF，其中上层为加工处理后的图像，下层为 OCR 软件自动识别后对应的文本，并且要求双层 PDF 文件中文字的位置与图像能重合（见图 6）。

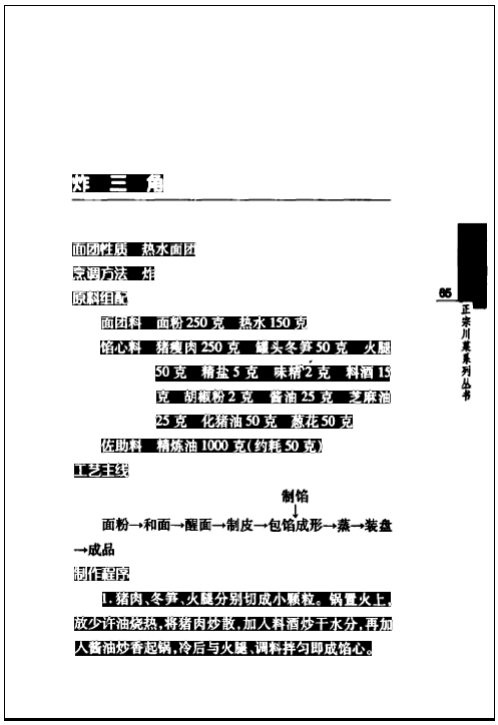


图 6 双层 PDF 效果

所有发布应用级图像文件（单页 PDF 文件，不包括封面前和封底后的色卡文件）置于数字对象目录下“ptiff”目录中，且针对每册书籍，将所有单页 PDF 文件合并到一个多页 PDF 文件(不包括封面前和封底后的色卡文件)，以该数字对象唯一标识符命名，存于数字对象目录的根目录下。

6.2.3 发布应用级图像文件的容量控制

为有效控制发布应用级 PDF 文件容量，允许对扫描图片采用 JPG 或 JP2000 格式进行质量压缩，以 A4 幅面普通书籍为例（推荐采用 JP2000 格式，选择合适压缩比进行压缩）：

——ptiff 目录里单页 PDF 文件容量控制在 200KB 左右（允许正负 20%浮动）；

——每册书籍（按 300 页/册）的总 PDF 文件控制在 58MB 左右（允许正负 20%浮动）；

——确保压缩输出的 PDF 文件在 100%比例浏览下图像清晰可读；

——对于 A3 及其他幅面书籍，按上述规则类推。

对一般普通书籍，原则上不允许缩小图片尺寸，对于一些特殊图片（比如超大幅面地图，颜色信息特别丰富的图片），仅仅压缩图片质量无法控制容量，则允许对图片按比例缩小尺寸（最低允许缩小到 30%）后再封装 PDF，同时允许降低 PDF 分辨率（最低允许降低到 300DPI）。

6.2.4 发布应用级图像文件初始视图控制

为规范发布应用级 PDF 文件，要求对 PDF 文件初始视图属性做统一规定，以 Adobe acrobat pro 软件为例，打开“文件”下的“属性”菜单，弹出文档属性对话框，设置初始视图如图 7 所示：

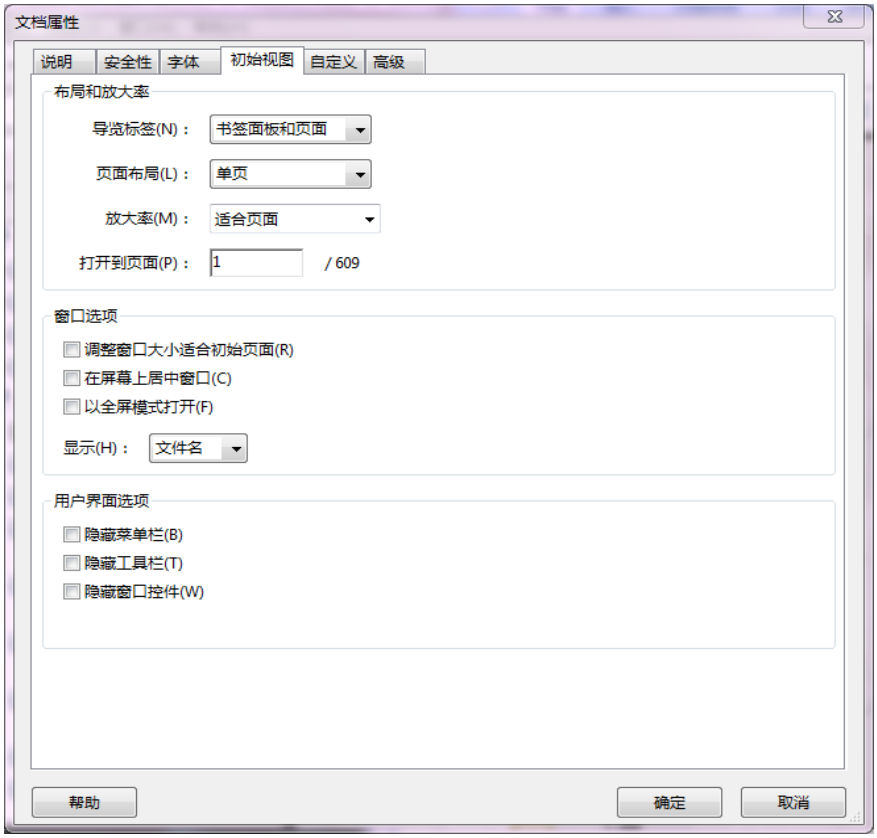


图 7 PDF 文件初始视图属性设置

导览标签：要求书签面板和页面都打开；

页面布局：要求按“单页”显示；

放大率：按“适应页面”；

打开 PDF 默认打开第 1 页。

6.2.5 发布应用级图像文件尺寸统一

为规范发布应用级 PDF 文件，要求合并版的多页 PDF 图像里相同幅面的页面其图像尺寸保持一致，同一幅面的所有页面不允许存在宽度或高度差异。

6.2.6 发布应用级 PDF 版本说明

为规范发布应用级 PDF 文件，所有单页 PDF 和多页 PDF 文件，要求 PDF 版本至少为 1.5（Acrobat 6.x）版本。

6.3 DC 元数据

DC 元数据存于数字对象目录下的“meta”目录中的“dc. xml”文件中，是以 Dublin Core 为核心，加上 CADAL 项目特有的元数据构成，制作要求详见《元数据著录总则》、《中文图书元数据著录规范》和《期刊元数据著录规范》。对于日语类书籍，允许采用中文输入法录入其书籍内出现的中文文字。

6.4 目录结构信息

6.4.1 目录结构信息的基本要求

目录结构信息要求建立每个目录章节信息与发布应用级图像文件的文件名之间的对应关系。

6.4.1.1 Catalog.xml 文件

要求建立符合 XML 的 METS 规范的目录结构信息，包括目录节点名称、链接指向的页面文件编号，目录结构信息存于数字对象目录下的“meta”目录中的“catalo.xml”文件中。如：

示例 1:

```
<METS: div TYPE="Chapter" LABEL="清汤抄手"ORDERLABEL="4"><METS:
fptrFILEID="00000014"/></METS: div>
```

其中：

LABEL="清汤抄手"　　表示章节名

ORDERLABEL="4"　　表示章节编号

METS: fptr FILEID="00000014" 表示发布应用级图像文件的主文件名为"00000014"

6.4.1.2 PDF 目录导航信息（Bookmark）

针对每册书籍封装的多页 PDF 文件以书签形式呈现其目录导航信息，以建立每个目录章节信息与发布应用级图像文件的对应关系，包括两大部分：

——目录浏览：即书本的正文章节信息（按照书籍阅读顺序和目录层次结构建立）；

——其他信息：除目录信息之外的所有其他资源结构信息，参考 6.5 资源封装信息。

即所有书籍一级目录都是目录浏览和其他信息，均链接到数字对象的封面。如图 8 所示：

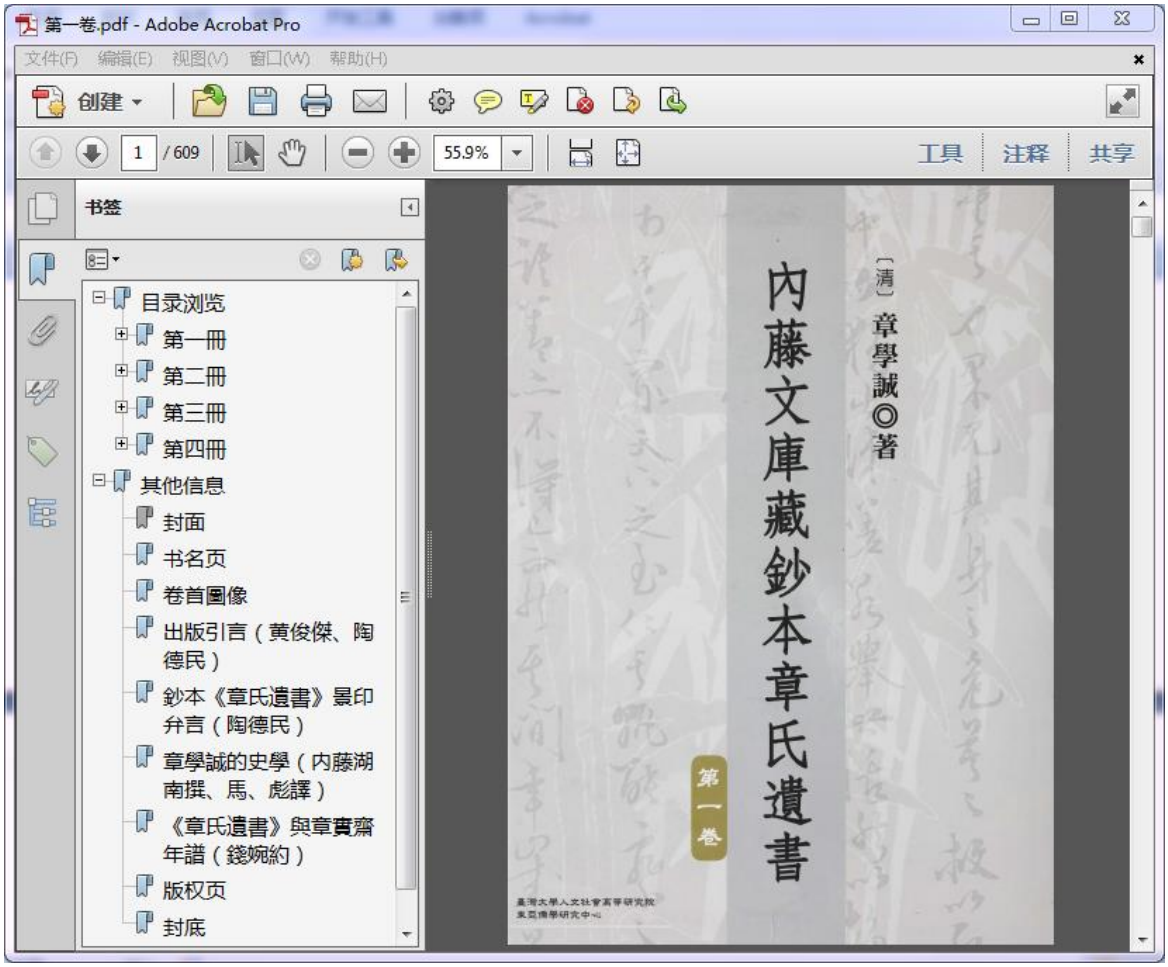


图 8PDF 目录导航信息示例图

6.4.2 目录级别要求

对有目录页的图书，按目录页内容制作前三级内容。对没有目录页的图书，需要由制作人员根据书籍内文的章节信息制作 1～2 级目录。如果书籍正文没有明显标题可以编制目录的，则按以下规则编制目录：

——中文图书

书名（链接书名页）

正文（链接正文第一页）

——外文图书

书名（采用书籍正文对应语种著录，链接书名页）

Text（链接正文第一页）

6.4.3 目录准确率要求

所有目录导航（包括 PDF 书签和 Catalog.xml）的链接要求 100%准确，其文字差错率要求低于 1‰。

6.4.4 目录著录规则

（1）导航信息要求整齐美观，章、节、标题与序号之间必须加一个空格；

（2）文字录入应遵循书籍目录页内文字的简繁体格式，对于 GBK 不包含的汉字，可以录入其对应的全拼拼音，所有英文字符、标点符号、数字字符和一些特殊符号都在全角方式下录入，对于无法录入的特殊符号，可以使用“＃”代替。(注：“＃”必须在全角下录入)。对于日语类书籍，允许采用中文输入法录入其书籍内出现的中文文字；

（3）目录遵照书籍目录页进行著录，一部分书籍出现目录页里的章节标题和正文里标题不一致，但是表达的含义基本相同，主要有以下几种情况：

——简体和繁体差异

——中文和英文翻译（目录是中文，正文是英文，或者反之）

——简写和全称差异

——近义词表达

——近似语义表达

这种情况下，目录著录遵照书本目录客观著录，可以不去对照原文寻找章节标题逐个修改，但是需要在书本的 description 里予以描述“书本目录页上的文章标题和正文标题表述不一致，目录导航遵照目录页客观著录，特此说明”。当然，如果该书的目录编制和书本正文标题大相径庭，就重新按正文重新编制目录后再著录。

（4）特殊情况处理：

——目录中的角标问题：

①目录中包含上、下角标。例如：

X^{2} 可以录入为 X¯2 即(X+上划线+2);

X_{2} 可以录入为 X_2 即(X+下划线＋2);

同时含有上、下角标的先录入上角标，后录入下角标。

②目录中包含繁分式。例如：

$$\frac{\frac{A+B}{C+D}}{\frac{E+F}{G+H}}$$

可以录入为[(A＋B)/(C＋D)]/[(E＋F)/(G＋H)]；

③目录中包含根号。

若根号下为数字，如：“根号 2”，则可以录入为√2。

注意： 3√2(三次根号 2) 与 3*√2(三倍根号 2)的区别！

若根号下为表达式，如：“根号下 A 加 B”，则可以录入为√(A＋B)；

——如果某标题链接的页面是缺页，则在标题后加“（缺）”；

——书本含有多个目录，如有中文目录和外文目录的，则录入书籍正文语种对应的目录；

——对于目录在上册且下册没有目录的图书，应该将在上册中对应于下册的目录录入；

——对于有总目录的图书，应该录入总目录。有简目和详细目录的，录简目；

——对于分册的图书，应该录入本分册对应的目录；

——若书籍目录编制错误，并且正文里标题也有错误（一般指书本标题出现很明显的错别字）时，则改正标题错别字按正确方式著录；

——若目录中的标题为中、英文混合或其它国家文字的混合，则应该将在页码前的文字全部录入；

——若一本图书中有两个或两个以上的目录(但其中一个目录 A 是另一个目录 B 的一部分)，则录入最完整的那个目录。其它的因书的内容制定；

——若目录中字数太多无法标引的，则可以只录入前二十个字，省略部分用“……”表示；

——若书籍目录中页码不是按顺序排列的，比如按照某类规则分类排列的，则按照书籍阅读次序依次著录目录；

——目录页的页码出现其它非阿拉伯数字，全部统一成阿拉伯数字页码。

6.5 资源封装信息

资源封装信息存于数字对象目录下的“meta”目录中的“a.opf”中，包括 4 部分内容，详见《数字内容编码与内容标记规范》。如果书本里所用词条与下面表达不一致的时候，根据书本著录。所有词条统一按首字母大写方式著录。

资源结构信息需要保留：

——封面 Cover；

——书名页 Title；

——CIP 数据页 CIP DataCIP；

——版权页 Copyright；

——插图 Illustration；

——目录 Content；

——摘要 Abstract；

——序言 Foreword；

——前言 Preface；

——感谢 Acknowledge；

——附录 Appendix；

——索引 Index；

——参考文献 Reference；

——后记 Postscript；

——封底 Back Cover。

做结构信息标记时，中文资源用上述汉字词语来标记，非中文资源用相应的语种单词来标记。

示例 2：

```
<guide>
    <reference type= "Other" title= "封面" href= "ptiff/00000001.pdf"/>
    <reference type= "Other" title="书名页" href= "ptiff/00000003.pdf"/>
    <reference type=" Other" title = "版权页" href= "ptiff/00000004.pdf "/>
    <reference type=" Other" title="目录" href= "ptiff/00000005.pdf "/>
    <reference type=" Other" title= "封底" href= "ptiff/00000301.pdf "/>
</guide>
```

示例 3：

```
<guide>
    <reference type= "Other" title= "Cover" href= "ptiff/00000001.pdf " />
    <reference type= "Other" title= "Title" href="ptiff/00000005.pdf " />
    <reference type= "Other" title= "Copyright" href= "ptiff/00000006.pdf " />
    <reference type="Other" title= "Foreword" href= "ptiff/00000007.pdf " />
    <reference type="Other" title= "Content" href= "ptiff/00000013.pdf " />
    <reference type= "Other" title= "Preface" href="ptiff/00000019.pdf " />
    <reference type="Other" title="Appendex" href="ptiff/00000383.pdf " />
    <reference type="Other" title="Index" href="ptiff/00000391.pdf " />
    <reference type="Other" title= "Back Cover" href="ptiff/00000399.pdf " />
</guide>
```

7 数字对象文件目录结构

存放内容指定如图 9 所示。

```
└──33039778……数字对象唯一标识符
├─meta
          ||__catalog.xml……目录数据
           ||__dc.xml……DC 元数据
            ||__a.opf…… 资源封装文件
           └──otiff……典藏级图像文件
          └──ptiff…… ……发布应用级图像文件(单页 PDF 文件)
           └──33039778.pdf……发布应用级图像文件(多页含书签PDF文件)
```

图 9 单个数字对象文件存放目录结构

附录

（规范性附录）

《CADAL 数字对象制作规范修正说明（2019）》

本次标准修改主要针对《图书期刊数字对象制作规范》中的目标文档扫描参数、发布级图像格式及目录要求进行修正。修改了以下三个方面及相对应的条款，特此说明：

A.1 修改了目标文件扫描参数：考虑还原图书原貌，CADAL 资源一律采用彩色扫描模式

相应规范修改条款如下：

A.1.1 “条款 6.1”中“子目录下：”后的文字修改为：

——每一页一个文件，扫描文件从 00000001.JPG 或 00000001.JP2 开始依次命名；

——扫描从书籍封面至封底依次进行，书籍内所有页面（包括封面、封底、书名页、目录页反面的空白页和插页）都需要扫描。无论什么类型书籍，封面封底都必须按原貌采集，如果书籍封面或封底是空白页的请按原貌扫描，如果书籍缺失封面或封底的直接扫缺页。

A.1.2 “条款 6.1.1 ”的表 1 修改为：

表 1　典藏级图像文件扫描标准

页面样式		纯文字黑白页面	配图黑白页面	彩色页面
DPI	传统扫描仪	600	600	600
	拍摄式扫描仪	400	400	400
色阶	传统扫描仪	24位彩色	24位彩色	24位彩色
	拍摄式扫描仪	24位彩色	24位彩色	24位彩色
压缩方式	传统扫描仪	JPEG/JPEG2000	JPEG/JPEG2000	JPEG/JPEG2000
	拍摄式扫描仪	JPEG/JPEG2000	JPEG/JPEG2000	JPEG/JPEG2000

A.1.3 “条款 6.1.3”的“方案一”修改为：

方案一：在缺页处插入写有“原书缺页 Page Missed in Original Book”的图像文件（见图 1）。如果缺页涉及到目录导航信息，请在目录导航标题后加上“（缺）”，如“封面（缺）”、“封底（缺）”。正文里的缺页按页码进行扫描，连续缺几个缺页的就连续扫几张缺页。

图 1 缺页替代样例

A.1.4 “条款 6.1.4”中“应符合如下要求：”后的内容修改为：

——应建立摄影棚和漫反射光源系统；

——应在封面之前、封底之后各扫描一次标准色卡，用于日后颜色校正处理。

——采用爱色丽 color checker classic 24 色卡 mini 达芬奇色板（X-Rite 24 色迷你色板），此款色卡材质为纸质，尺寸约 10.9cm*6.4cm(约名片大小)与色卡护照内的 class 目标尺寸相同，如图 2 所示：

图 2 爱色丽 color checker classic 24 色卡 mini 达芬奇色板

——封面之前的色卡以 00000000.JPG 或 00000000.JP2 命名，封底之后的色卡文件名按封底文件名加 1；

——色卡拍照时要求统一纵向摆放，拍照的底纹用色调均匀的黑色卡纸，拍照的色卡图片保持工整，没有倾斜，色卡基本位于黑色卡纸居中位置，如图 3 所示：

图 3 色卡拍照示例图

——色卡图片直接保留原始扫描图片，不做旋转之外任何图像处理，最终色卡图见图 4，存放到 otiff 文件夹里：

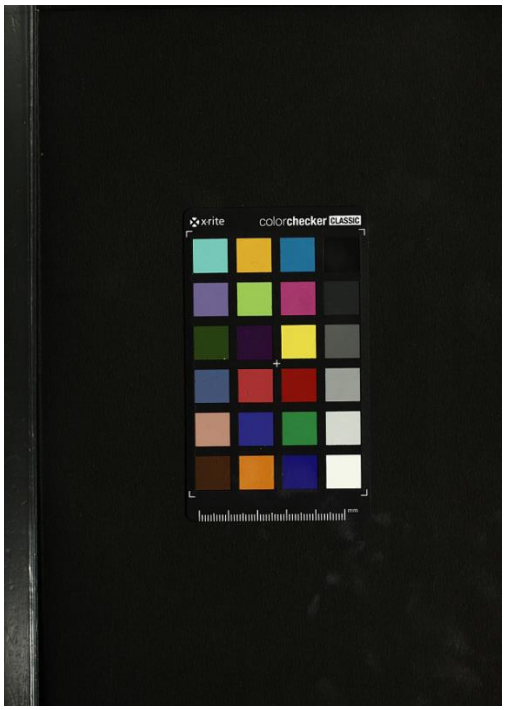


图 4 最终色卡示例图

A.1.6 增加“条款 6.1.5”，标明了扫描过程中一些特殊情况处理：

（1）部分书籍本身页面褶皱，扫描后部分页面容易产生细微扭曲或褶皱。这种情况，扫描前请尽量将书本抚摸平整，无法摸平的，首先保证扫描图像可读。另外，在此书的 DC 的 description 字段里予以描述说明“书籍本身页面褶皱，扫描后部分页面产生细微褶皱，特此说明”。

（2）部分书籍装订较紧的（靠近书中缝处有少量文字被装订进去），应尽量扫描出来，确实无法扫描完整导致内容残缺的，在此书的 DC 的 description 里予以描述说明“书籍装订较紧或书籍本身页面内容残缺，部分页面扫描后内容残缺，特此说明”。

（3）原书封面、封底或其他页面破损，扫描时统一在下面垫一张白纸。”

A.2 修改了发布级图像格式：数字对象加工的目标文件由原有的.djvu 格式改成.pdf 格式

PDF 格式与操作系统平台无关性使其几乎成为数字化信息行业的一个工业标准。考虑目前 PDF 文档的通用性及兼容性，CADAL 所有加工的目标文件采用 pdf 格式。相应规范修改文档如下：

A.2.1 修改了“条款 3.6”为：

“双层 PDF Text Hidden PDF

双层 PDF 指通过 OCR 等技术手段，将原文中每行文字内容放在底层，上层放置原始图像，继而形成的 PDF 格式的文件。”

A.2.2 修改了“条款 3.7”为：

“单层 PDF Image Only PDF

单层 PDF 指由原始图像直接转换而成的 PDF 文件。

A.2.3 修改了“条款 6.2.1”的内容为：

——所有发布应用级图像文件应保持页面整洁:图像处理应在遵照书籍原貌的前提下进行，即保留原书籍里所有内容，包括馆藏印章、条形码、馆藏描述性文字、手写批注、各种颜色画线、题词、索书号标签纸等，对扫描带来的黑边须进行裁切处理，确保发布应用级图像文件页面整洁、美观；

——主体文字内容不能出现 90°侧倒或 180°颠倒，当页码和主体文字方向出现不一致时，以遵循文字阅读方向为主；

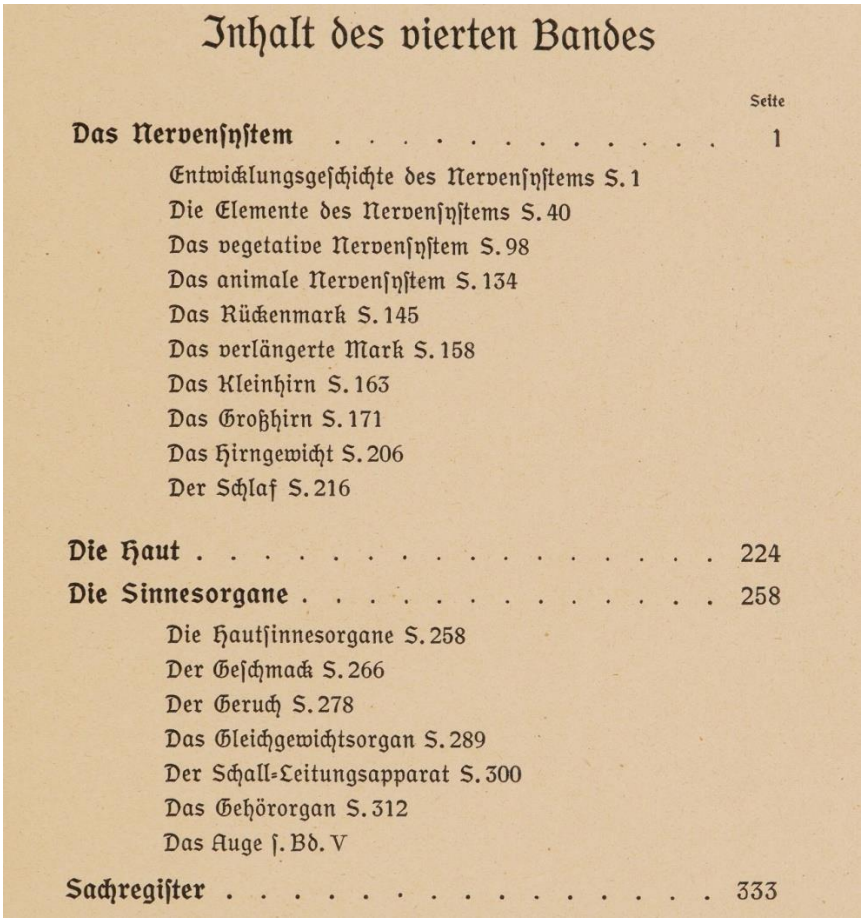
——页面整体倾斜不能超过 3°。

A.2.4 修改了“条款 6.2.2”的内容为：

所有发布应用级图像文件以 PDF 的方式展现：

——对手写体文稿、印刷质量较差（如油印、印刷模糊不清或字迹笔画断线）、

图册、画册等 OCR 识别结果出现大量乱码不具备检索意义的页面，可直接由图像转换成单层 PDF，如图 5 所示：



Inhalt des vierten Bandes

Seite

Das Nervensystem 1

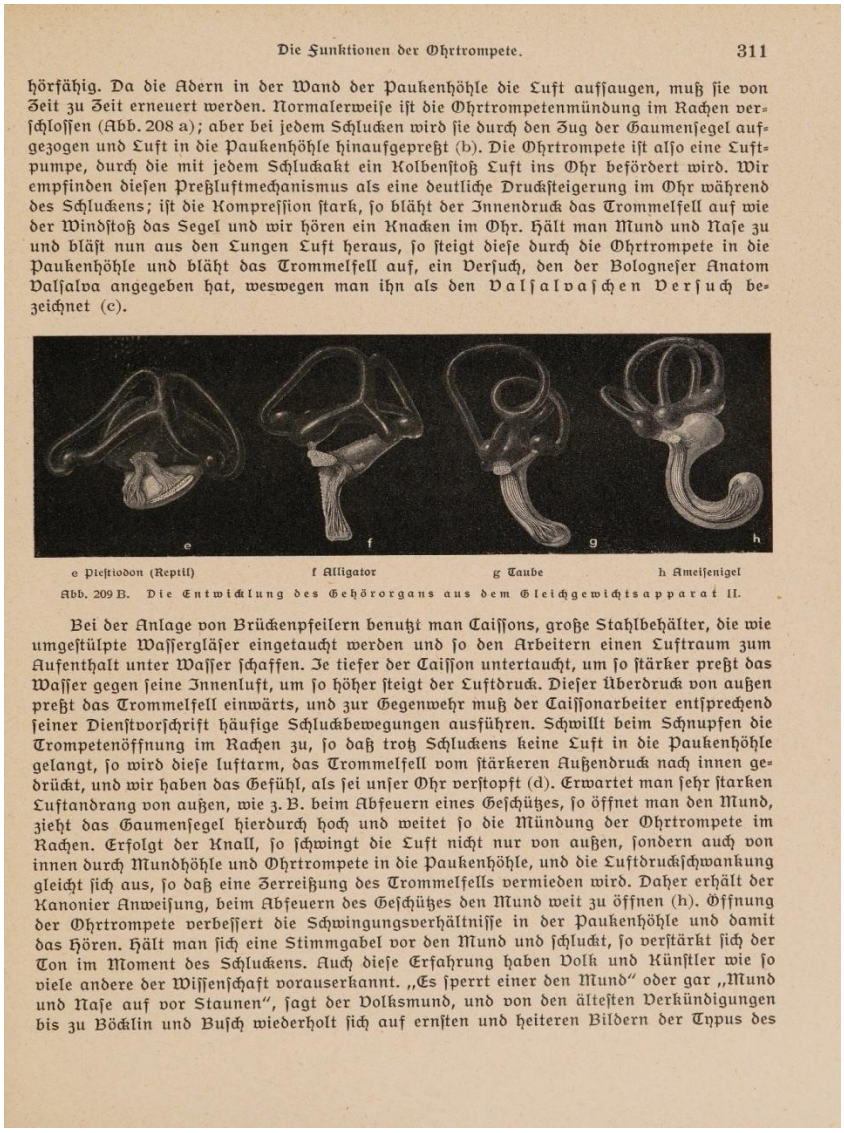
Entwicklungsgeschichte des Nervensystems S. 1
Die Elemente des Nervensystems S. 40
Das vegetative Nervensystem S. 98
Das animale Nervensystem S. 134
Das Rückenmark S. 145
Das verlängerte Mark S. 158
Das Kleinhirn S. 163
Das Großhirn S. 171
Das Hirngewicht S. 206
Der Schlaf S. 216

Die Haut 224

Die Sinnesorgane 258

Die Hautsinnesorgane S. 258
Der Geschmack S. 266
Der Geruch S. 278
Das Gleichgewichtsorgan S. 289
Der Schall-Leitungsapparat S. 300
Das Gehörorgan S. 312
Das Auge s. Bd. V

Sachregister 333



Die Funktionen der Ohrtrompete. 311

hörfähig. Da die Adern in der Wand der Paukenhöhle die Luft aufsaugen, muß sie von Zeit zu Zeit erneuert werden. Normalerweise ist die Ohrtrompetenmündung im Rachen verschlossen (Abb. 208 a); aber bei jedem Schlucken wird sie durch den Zug der Gaumensegel aufgezogen und Luft in die Paukenhöhle hinaufgepreßt (b). Die Ohrtrompete ist also eine Luftpumpe, durch die mit jedem Schluckakt ein Kolbenstoß Luft ins Ohr befördert wird. Wir empfinden diesen Preßluftmechanismus als eine deutliche Druck-steigerung im Ohr während des Schluckens; ist die Kompression stark, so bläht der Innendruck das Trommelfell auf wie der Windstoß das Segel und wir hören ein Knacken im Ohr. Hält man Mund und Nase zu und bläst nun aus den Lungen Luft heraus, so steigt diese durch die Ohrtrompete in die Paukenhöhle und bläht das Trommelfell auf, ein Versuch, den der Bologneser Anatom Valsalva angegeben hat, weswegen man ihn als den Valsalvaschen Versuch bezeichnet (c).

e Plestiodon (Reptil) f Alligator g Taube h Ameisenigel

Abb. 209 B. Die Entwicklung des Gehörorgans aus dem Gleichgewichtsapparat II.

Bei der Anlage von Brückenpfeilern benutzt man Caissons, große Stahlbehälter, die wie umgestülpte Wassergläser eingetaucht werden und so den Arbeitern einen Luftraum zum Aufenthalt unter Wasser schaffen. Je tiefer der Caisson untertaucht, um so stärker preßt das Wasser gegen seine Innenluft, um so höher steigt der Luftdruck. Dieser Überdruck von außen preßt das Trommelfell einwärts, und zur Gegenwehr muß der Caissonarbeiter entsprechend seiner Dienstvorschrift häufige Schluckbewegungen ausführen. Schwillt beim Schnupfen die Trompetenöffnung im Rachen zu, so daß trotz Schluckens keine Luft in die Paukenhöhle gelangt, so wird diese luftarm, das Trommelfell vom stärkeren Außendruck nach innen gedrückt, und wir haben das Gefühl, als sei unser Ohr verstopft (d). Erwartet man sehr starken Luftandrang von außen, wie z. B. beim Abfeuern eines Geschützes, so öffnet man den Mund, zieht das Gaumensegel hierdurch hoch und weitet so die Mündung der Ohrtrompete im Rachen. Erfolgt der Knall, so schwingt die Luft nicht nur von außen, sondern auch von innen durch Mundhöhle und Ohrtrompete in die Paukenhöhle, und die Luftdruckschwankung gleicht sich aus, so daß eine Zerreißung des Trommelfells vermieden wird. Daher erhält der Kanonier Anweisung, beim Abfeuern des Geschützes den Mund weit zu öffnen (h). Öffnung der Ohrtrompete verbessert die Schwingungsverhältnisse in der Paukenhöhle und damit das Hören. Hält man sich eine Stimmgabel vor den Mund und schluckt, so verstärkt sich der Ton im Moment des Schluckens. Auch diese Erfahrung haben Volk und Künstler wie so viele andere der Wissenschaft vorauserkannt. „Es sperrt einer den Mund“ oder gar „Mund und Nase auf vor Staunen“, sagt der Volksmund, und von den ältesten Verkündigungen bis zu Böcklin und Busch wiederholt sich auf ernsten und heiteren Bildern der Typus des

图 5 图像转换成单层 PDF 示例

——对其他类型的文档，要求制作成双层 PDF，其中上层为加工处理后的图像，下层为 OCR 软件自动识别后对应的文本，并且要求双层 PDF 文件中文字的位置与图像能重合（见图 6）。

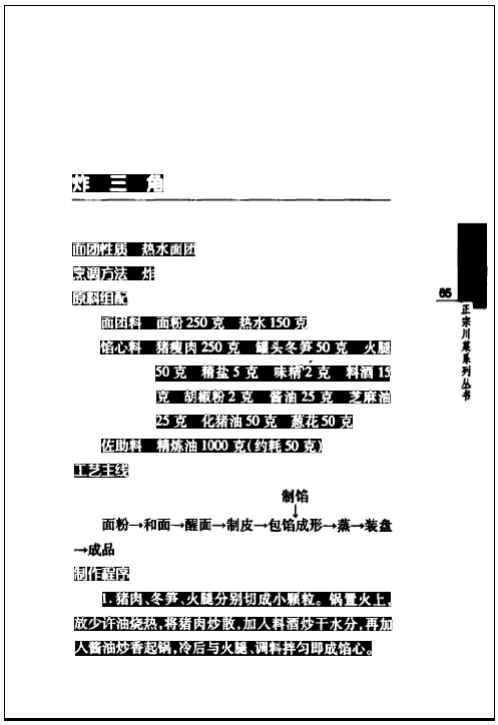


图 6 双层 PDF 效果

所有发布应用级图像文件（单页 PDF 文件，不包括封面前和封底后的色卡文件）置于数字对象目录下“ptiff”目录中，且针对每册书籍，将所有单页 PDF 文件合并到一个多页 PDF 文件(不包括封面前和封底后的色卡文件)，以该数字对象唯一标识符命名，存于数字对象目录的根目录下。

A.2.4 增加“条款 6.2.3”

发布应用级图像文件的容量控制

为有效控制发布应用级 PDF 文件容量，允许对扫描图片采用 JPG 或 JP2000 格式进行质量压缩，以 A4 幅面普通书籍为例（推荐采用 JP2000 格式，选择合适压缩比进行压缩）：

——ptiff 目录里单页 PDF 文件容量控制在 200KB 左右（允许正负 20%浮动）；

——每册书籍（按 300 页/册）的总 PDF 文件控制在 58MB 左右（允许正负 20%浮动）；

——确保压缩输出的 PDF 文件在 100%比例浏览下图像清晰可读；

对于 A3 及其他幅面书籍，按上述规则类推。

对一般普通书籍，原则上不允许缩小图片尺寸，对于一些特殊图片（比如超大幅面地图，颜色信息特别丰富的图片），仅仅压缩图片质量无法控制容量，则允许对图片按比例缩小尺寸（最低允许缩小到 30%）后再封装 PDF，同时允许降

低 PDF 分辨率（最低允许降低到 300DPI）。”

A.2.5 增加“条款 6.2.4”

“发布应用级图像文件初始视图控制

为规范发布应用级 PDF 文件，要求对 PDF 文件初始视图属性做统一规定，以 Adobe acrobat pro 软件为例，打开“文件”下的“属性”菜单，弹出文档属性对话框，设置初始视图如图 7 所示：

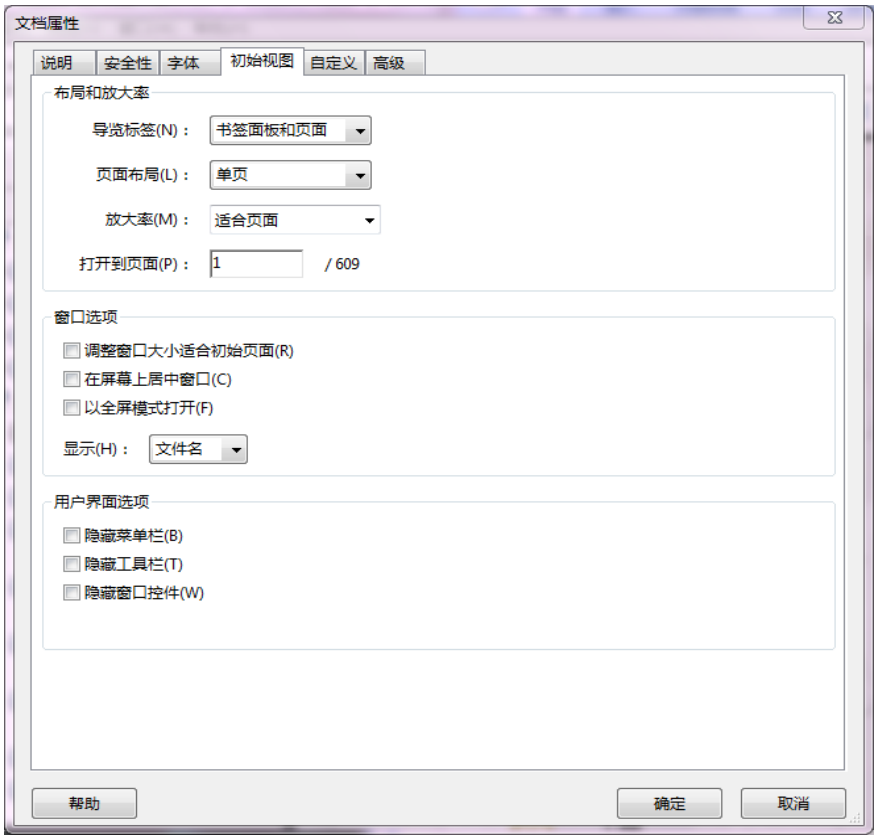


图 7 PDF 文件初始视图属性设置

导览标签：要求书签面板和页面都打开；

页面布局：要求按“单页”显示

放大率：按“适应页面”

打开 PDF 默认打开第 1 页。”

A.2.6 增加“条款 6.2.5”

“发布应用级图像文件尺寸统一

为规范发布应用级 PDF 文件，要求合并版的多页 PDF 图像里相同幅面的页面其图像尺寸保持一致，同一幅面的所有页面不允许存在宽度或高度差异。”

A.2.7 增加“条款 6.2.6”

“发布应用级 PDF 版本说明

为规范发布应用级 PDF 文件，所有单页 PDF 和多页 PDF 文件，要求 PDF 版本至少为 1.5（Acrobat 6.x）版本。”

A.2.8 修改“条款 6.5”示例 2 和示例 3 中的“djvu”为“pdf”

示例 2：

```
<guide>
    <reference type= "Other" title= "封面" href= "ptiff/00000001.pdf"/>
    <reference type= "Other" title="书名页" href= "ptiff/00000003.pdf"/>
    <reference type=" Other" title = "版权页" href= "ptiff/00000004.pdf "/>
    <reference type=" Other" title="目录" href= "ptiff/00000005.pdf "/>
    <reference type=" Other" title= "封底" href= "ptiff/00000301.pdf "/>
</guide>
```

示例 3：

```
<guide>
    <reference type= "Other" title= "Cover" href= "ptiff/00000001.pdf " />
    <reference type= "Other" title= "Title" href="ptiff/00000005.pdf " />
    <reference type= "Other" title= "Copyright" href= "ptiff/00000006.pdf " />
    <reference type="Other" title= "Foreword" href= "ptiff/00000007.pdf " />
    <reference type="Other" title= "Content" href= "ptiff/00000013.pdf " />
    <reference type= "Other" title= "Preface" href="ptiff/00000019.pdf " />
    <reference type="Other" title="Appendex" href="ptiff/00000383.pdf " />
    <reference type="Other" title="Index" href="ptiff/00000391.pdf " />
    <reference type="Other" title= "Back Cover" href="ptiff/00000399.pdf " />
</guide>
```

A.2.9 修改“条款 7”中的图 9 为：

```
└──33039778……数字对象唯一标识符
    |──meta
    |  |__catalog.xml……目录数据
    |  |__dc.xml……DC 元数据
    |  |__a.opf …… 资源封装文件
    └──otiff …… ……典藏级图像文件
    └──ptiff …… ……发布应用级图像文件(单页 PDF 文件)
    └──33039778.pdf  ……发布应用级图像文件(多页含书签PDF文件)
```

图 9 单个数字对象文件存放目录结构

A.3 修改了目录结构信息的基本要求和目录级别要求，增加了准确率要求和目录著录规则

A.3.1 修改“条款 6.4”的标题为：

“目录结构信息”

A.3.2 修改了“条款 6.4.1”

“目录结构信息的基本要求”

目录结构信息要求建立每个目录章节信息与发布应用级图像文件的文件名之间的对应关系。

6.4.1.1Catalog.xml 文件

要求建立符合 XML 的 METS 规范的目录结构信息，包括目录节点名称、链接指向的页面文件编号，目录结构信息存于数字对象目录下的“meta”目录中的“catalo.xml”文件中。如：

示例 1:

```
<METS: div TYPE="Chapter" LABEL="清汤抄手"ORDERLABEL="4"><METS:
fptr FILEID="00000014"/></METS: div>
```

其中：

LABEL="清汤抄手"　　表示章节名

ORDERLABEL="4"　　表示章节编号

METS: fptr FILEID="00000014" 表示发布应用级图像文件的主文件名为"00000014"

A.3.3 增加了“条款 6.4.1.2”

“PDF 目录导航信息（Bookmark）”：

针对每册书籍封装的多页 PDF 文件以书签形式呈现其目录导航信息，以建立每个目录章节信息与发布应用级图像文件的对应关系，包括两大部分：

——目录浏览：即书本的正文章节信息（按照书籍阅读顺序和目录层次结构建立）；

——其他信息：除目录信息之外的所有其他资源结构信息，参考 6.5 资源封装信息。

即所有书籍一级目录都是目录信息和其他信息，均链接到数字对象的封面。如图 8 所示：

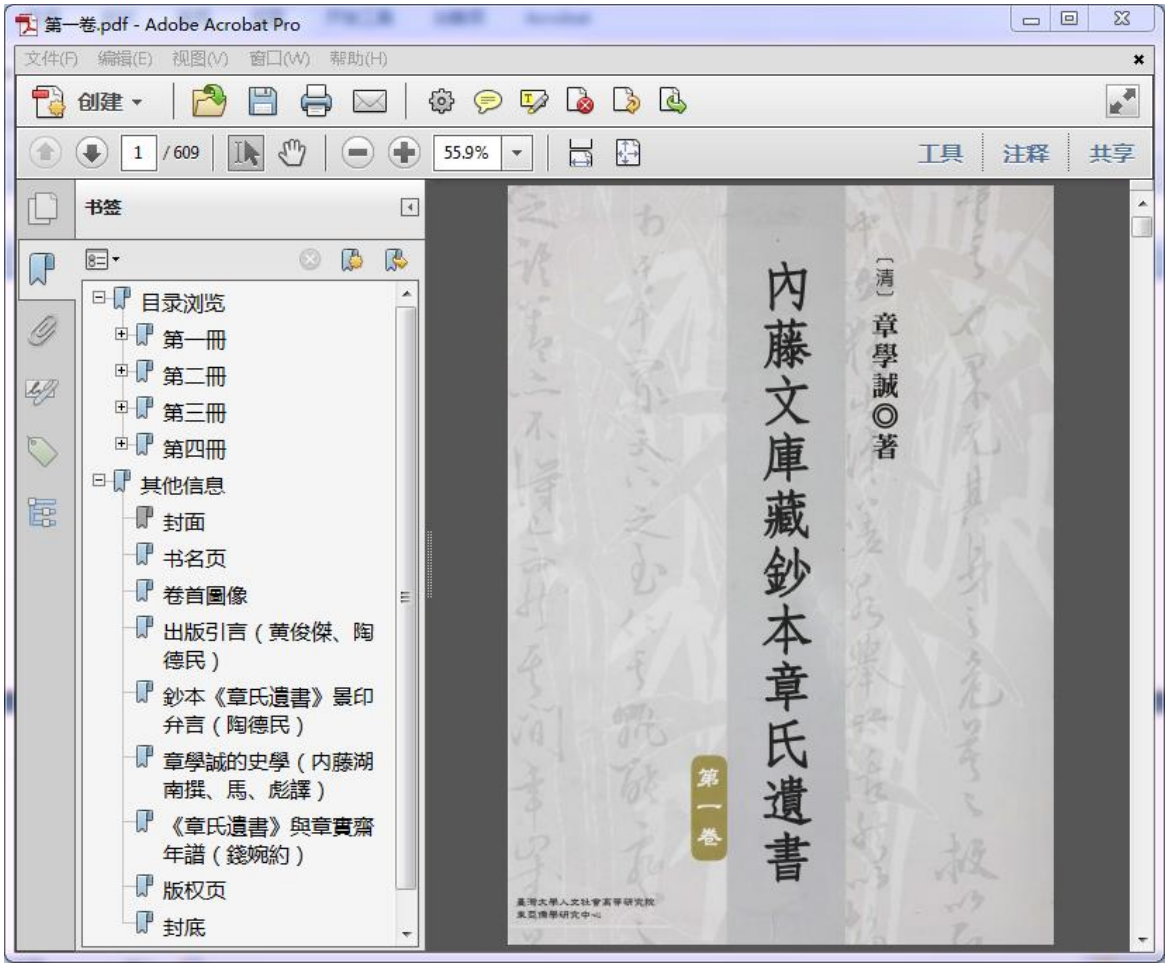


图 8PDF 目录导航信息示例图

A.3.4 补充了“条款 6.4.2”中“制作 1~2 级目录。”后面的内容为：

如果书籍正文没有明显标题可以编制目录的，则按以下规则编制目录：

——中文图书

书名（链接书名页）

正文（链接正文第一页）

——外文图书

书名（采用书籍正文对应语种著录，链接书名页）

Text（链接正文第一页）

A.3.5 增加了“条款 6.4.3”

“目录准确率要求：

所有目录导航（包括 PDF 书签和 Catalog.xml）的链接要求 100%准确，其文字差错率要求低于 1‰。”

A.3.6 增加了“条款 6.4.4”

“目录著录规则：

（1）导航信息要求整齐美观，章、节、标题与序号之间必须加一个空格；

（2）文字录入应遵循书籍目录页内文字的简繁体格式，对于 GBK 不包含的汉字，可以录入其对应的全拼拼音，所有英文字符、标点符号、数字字符和一些特殊符号都在全角方式下录入，对于无法录入的特殊符号，可以使用“＃”代替。(注：“＃”必须在全角下录入)。对于日语类书籍，允许采用中文输入法录入其书籍内出现的中文文字；

（3）目录遵照书籍目录页进行著录，一部分书籍出现目录页里的章节标题和正文里标题不一致，但是表达的含义基本相同，主要有以下几种情况：

——简体和繁体差异

——中文和英文翻译（目录是中文，正文是英文，或者反之）

——简写和全称差异

——近义词表达

——近似语义表达

（4）特殊情况处理：

——目录中的角标问题：

①目录中包含上、下角标。例如：

X^2 可以录入为 X¯2 即(X+上划线+2);

X_2 可以录入为 X_2 即(X+下划线＋2);

同时含有上、下角标的先录入上角标，后录入下角标。

②目录中包含繁分式。例如：

$$\frac{\frac{A+B}{C+D}}{\frac{E+F}{G+H}}$$

可以录入为[(A＋B)/(C＋D)]/[(E＋F)/(G＋H)]；

③目录中包含根号。

若根号下为数字，如：“根号 2”，则可以录入为√2。

注意： 3√2(三次根号 2) 与 3*√2(三倍根号 2)的区别！

若根号下为表达式，如：“根号下 A 加 B”，则可以录入为√(A＋B)；

——如果某标题链接的页面是缺页，则在标题后加“（缺）”；

——书本含有多个目录，如有中文目录和外文目录的，则录入书籍正文语种对应的目录；

——对于目录在上册且下册没有目录的图书，应该将在上册中对应于下册的目录录入；

——对于有总目录的图书，应该录入总目录。有简目和详细目录的，录简目；

——对于分册的图书，应该录入本分册对应的目录；

——若书籍目录编制错误，并且正文里标题也有错误（一般指书本标题出现很明显的错别字）时，则改正标题错别字按正确方式著录；

——若目录中的标题为中、英文混合或其它国家文字的混合，则应该将在页码前的文字全部录入；

——若一本图书中有两个或两个以上的目录(但其中一个目录 A 是另一个目录 B 的一部分)，则录入最完整的那个目录。其它的因书的内容制定；

——若目录中字数太多无法标引的，则可以只录入前二十个字，省略部分用“……”表示；

——若书籍目录中页码不是按顺序排列的，比如按照某类规则分类排列的，则按照书籍阅读次序依次著录目录；

——目录页的页码出现其它非阿拉伯数字，全部统一成阿拉伯数字页码。

A.4 其他

A.4.1 明确标准修订历史：

（1）本标准于 2012 年 05 月发布，于 2013 年做了部分修改，并于 2019 年做了重大的修改。
（2）2019 年标准修正主要针对《图书期刊数字对象制作规范》中的目标文档扫描参数及发布级图像格式进行修改。

A.4.2 内容新增

（1）“条款 3.9”句末新增：“同时，一册书籍封装的多页 PDF 内部需要以书签形式呈现其目录导航信息。”

（2）“条款 6.3”中“制作要求详见”后面的内容修改为：“《元数据著录总则》、《中文图书元数据著录规范》和《期刊元数据著录规范》。对于日语类书籍，允许采用中文输入法录入其书籍内出现的中文文字。”

（3）参考文献新增“[7] 龙伟 肖禹 梁爱民 向辉 鲍国强 王浩 包菊香 杨照坤, 国家古籍保护中心编制《古籍数字化工作手册》（试用本）2014.5”

参考文献

[1] 孙一钢，龙伟，赵四友．数字资源加工标准研究报告[成果]．项目年度编号：2002DEA20018．完成单位：国家图书馆．成果编号：CDLS-S03-008．成果公布日期：2006-06.

[2] International Digital Publishing Forum. Open Packaging Format （OPF） 2. 0. 1 v1. 0. 1. 2010-09-04. [2013-10-15]. http://www.idpf.orgepub20specOPF_2.0_ latest.htm.

[3] The Library of Congress Standards, Metadata Encoding &. Transmission Standard.2012-03-22.[2013-10-15].http://www.loc.gov/standards/mets/mets-schemadocs.html.

[4] LIZARD TECH, INC. DjVu Technology Primer. 2004-11. [2013-10-15]. http: //djvu.org/docs/DjVu_Tech_Primer. djvu.

[5] LIZARD TECH, INC. LizardtechDjVu Reference v3. 2005-11. [2013-10-15]. http: //djvu.org/docs/DjVu3Spec.djvu.

[6] 牛筱桔，冯春术，金赛英．美术图像数字化元数据标准规范．中国美术学院图书馆．http: //www.cadal.cn/bzgf/

[7] 龙伟，肖禹，梁爱民，向辉，鲍国强，王浩，包菊香，杨照坤. 国家古籍保护中心编制《古籍数字化工作手册》（试用本）2014.5