

# 图书馆特色资源库建设

## ——资源加工与制作

槐燕

2015年9月

1

• 资源库建设背景

2

• 资源库建设标准

3

• 资源库加工与制作



## 国外数字图书馆文献建设

### • 美国

美国数字图书馆启动计划一期、二期工程；美国记忆

### • 英国

elib项目

### • 法国

文化精品数字化项目 (JOUVE)

### • 日本

e-Japan重点计划



## 国内数字图书馆文献建设

- 数字图书馆概念1995年引入我国
- 在我国掀起数字图书馆研究热
- 2000年前后开始数字图书馆建设



# 国内数字图书馆文献建设

- 我国数字图书馆标准与规范建设项目
- 中国高等教育文献保障系统
- 国家数字图书馆工程



## 国家数字图书馆工程

1998年，国家图书馆向文化部提出在全国实施“中国数字图书馆工程”的建议

2001年，国家“十五”重点文化项目“国家图书馆二期工程暨国家数字图书馆工程”正式立项

数图工程建设内容：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

标准规范体系包括：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36项标准规范



## 1

## • 资源库建设背景

## 国家数字图书馆工程

成果类别	规范名称
汉字处理规范	汉字属性字典
	全文版式还原与全文输入XML规范
	古籍用字规范（计算机用字标准）
	计算机中文信息处理规范
生僻字、避讳字处理规范	
唯一标识符	国家图书馆数字资源唯一标识符规范
对象数据	国家图书馆数字资源对象管理规范
	文本数据加工标准与工作规范
	图像数据加工标准与工作规范
	音频数据加工标准与工作规范
	视频数据加工标准与工作规范
元数据总则	国家图书馆元数据应用规范
	国家图书馆元数据置标规范
	国家图书馆核心元数据标准
	国家图书馆专门元数据设计规范
	CNMARC-XML
	CNMARC-DC-国家图书馆核心元数据集的对照转换
	MARC21-DC-国家图书馆核心元数据集的对照转换



## 1

## · 资源库建设背景

## 国家数字图书馆工程

成果类别	规范名称
元数据总则	国家图书馆元数据应用规范
	国家图书馆元数据置标规范
	国家图书馆核心元数据标准
	国家图书馆专门元数据设计规范
	CNMARC-XML
	CNMARC-DC-国家图书馆核心元数据集的对照转换
	MARC21-DC-国家图书馆核心元数据集的对照转换
专门元数据规范——古文献	专门元数据标准与著录规范——拓片
	专门元数据标准与著录规范——舆图
	专门元数据标准与著录规范——甲骨
	专门元数据标准与著录规范——古籍
	专门元数据标准与著录规范——家谱
专门元数据规范——电子书刊	专门元数据标准与著录规范——电子图书
	专门元数据标准与著录规范——电子连续性资源
	专门元数据标准与著录规范——学位论文
	专门元数据标准与著录规范——期刊论文
专门元数据规范——网络及多媒体资源	专门元数据标准与著录规范——网络资源
	专门元数据标准与著录规范——音频
	专门元数据标准与著录规范——视频
	专门元数据标准与著录规范——图像



## 实物文献

包括甲骨、竹简、印章、石碑、木版、铜板

## 纸本文献

包括印刷型：书、刊、报；手工型：手稿、书札等

## 缩微文献

包括缩微胶卷和缩微平片

## 视听文献

包括录音带、唱片、录象带、视盘

## 电子文献

包括网络资源和实体资源（如光盘、磁盘、磁带）



一个数字对象的组成包含三个要素

(1) 对象数据

(2) 元数据

(3) 唯一标识符



## 2.1 对象数据

### 格式体系

- 文本数据

TXT、XML、HTML、PDF

- 图像数据

TIFF、JPEG、JPEG2000、PDF

- 音频数据

MP3、WAV

- 视频数据

MPEG-1、MPEG-2、MPEG-4



## 2.1.1 文本数据

▶ 适用于排版简单，以文字为主要表达形式，可以存在少量图表的普通文献（不包括古籍善本、手稿等特殊文献）的文本数据加工制定。

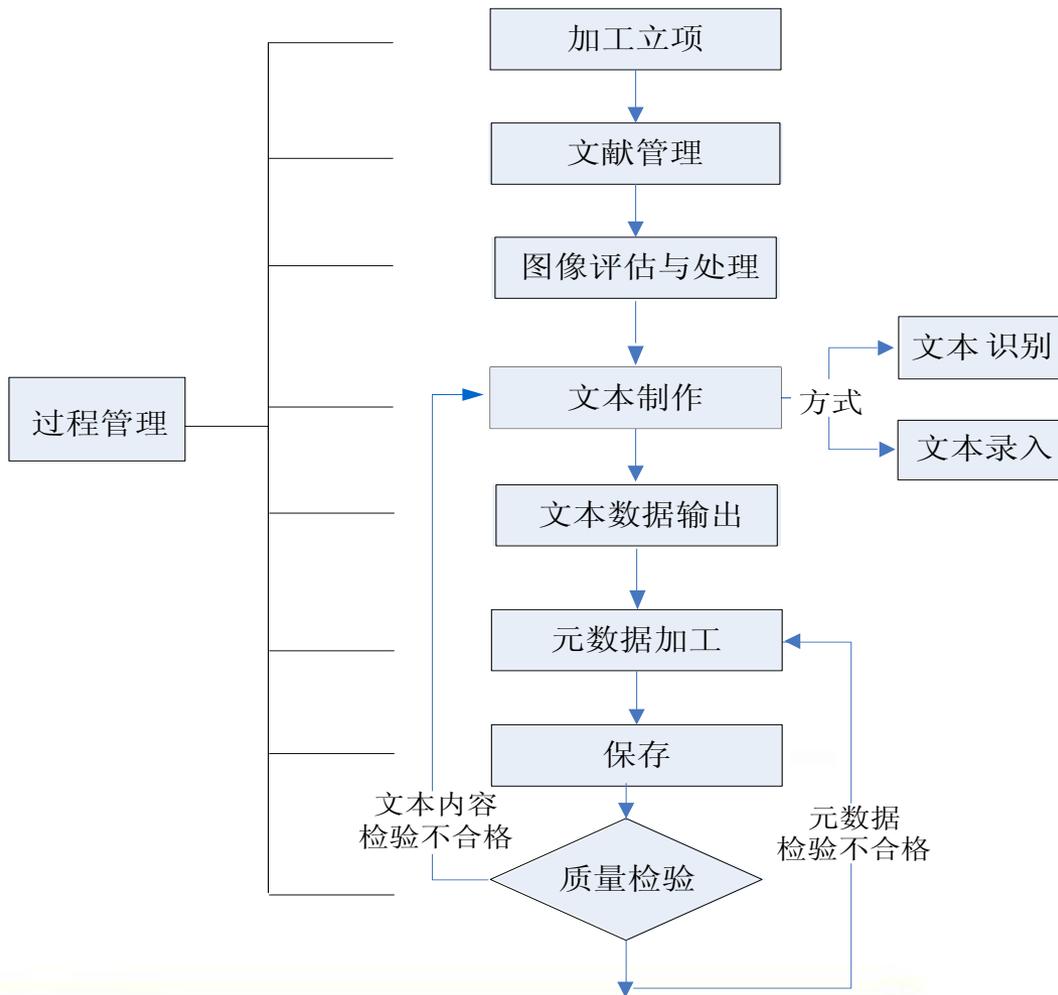
▶ 加工对象：一般纸质文本文献、印刷型文献经过数字转换后的图像文件。



## 2

# 资源库建设标准

### 文本数据加工流程



## 2.1.2 图像数据

### ▶ 资源应用级别

长期保存级 (A)

建议以TIFF格式存储，不压缩或无损压缩。

复制加工级 (P)

建议以JPG或JPEG2000，有损压缩。

发布服务级 (D)

L (Large, 大图)

M (Medium, 中图)

S (Small, 小图)



## 2

## · 资源库建设标准

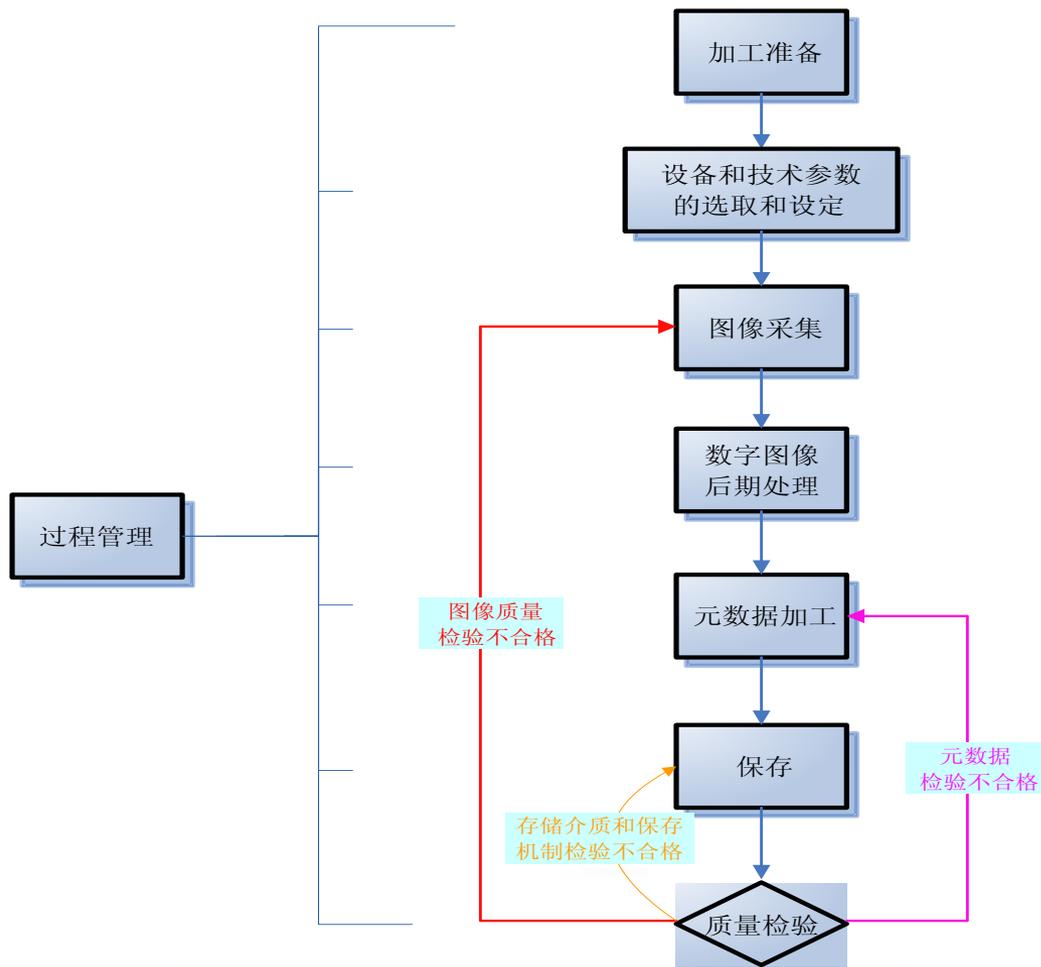
表 图书类图像资源数字化加工标准

文献类型	应用级别	图像分辨率 (dpi)	色彩位深	允许的编辑加工	文件格式 压缩算法		
普通 书刊 报类	A	300-600	黑白 8位, 24 位	↔	TIFF 5 / TIFF 6 不压缩或无损压缩 (如: LZW) ↔ TIFF -G4 ↔		
	D	L			300	锐化, 裁切, 纠偏, 去噪, 色彩管理 ↔	JPG GIF ↔ JPEG2000 或封装到 PDF ↔
		S			72-96		
古籍 手稿 乐谱	A	300-600	8位 24位 更高	↔	TIFF 5 / TIFF 6 不压缩或无损压缩 (如: LZW) ↔		
	P	300-600			JPEG2000 无损压缩 ↔		
	D	L			300	JPG 或封装到 PDF ↔	
		M			150	JPG GIF ↔	
		S			72-96	JPG GIF ↔	
小幅 照片	A	1200	8位 24位 或更高	↔	TIFF 不压缩或无损压缩 (如: LZW) ↔		
	P	原采集像素			JPG2000 无损压缩, ↔		
	D	L	800	24位	锐化, 裁切纠偏, 去噪, 成比例扩展, 最低限度的调整彩色和色调 ↔	JPG ↔ GIF ↔	
		M	300				
S		150					

# 2

## 资源库建设标准

图像资源加工流程图



## 2.1.3 音频数据

资源级别	主要参数			文件格式	后缀名（小写）
	采样率	量化级	通道数		
长期保存级	22/128kHz	16/24bit	由原始资料特性决定	wav	wav
发布服务级	22/44kHz	8/16bit	双声道/单声道	mp3	mp3

## 2.1.4 视频数据

▶ 适用于模拟视频、数字视频的采集、编码转换、保存与编辑

资源级别	主观质量描述	主要参数						
		分辨率	帧数 (帧/秒)	视频速率	音频速率 (kbps)	音频采样	文件格式	后缀名 (小写)
长期保存级	高清质量	1920×1080	25/30/60	固定码率50Mb/s或25Mb/s的可变码率	384	立体声 48kHz	MPEG2编码 AVI格式	mpg avi
	标清质量	720×576	25/30	15Mb/s固定码率或可变码率				
	最低质量	原采集窗口尺寸但不小于352×288	25	5Mb/s固定码率或可变码率	224			
发布服务级	最低质量	不低于352×288	15/25/30	不低于512kb/s	64-384	立体声 44.1kHz	MPEG4 FLV WMV	mpg flv wmv

## 2.2 元数据

- 描述型元数据
- 管理型元数据
- 结构型元数据



## 2.2.1 元数据功能

- **查找**，符合用户检索要求的实体（即利用实体的属性或关系在一个文档或数据库中找到一个或一组实体）；
- **识别**，一个实体（即确认所描述的实体对应于所查找的实体，或者区分具有相似特征的两个或多个实体）；
- **选择**，适合用户需要的一个实体（即选取一个在内容、物理形式等方面能满足用户要求的实体，或放弃一个不适合用户需求的实体）；
- **获取/存取**，所描述的实体（即通过购买、借阅等方式获取一个实体，或者以电子方式通过联机连接远程计算机来检索一个实体）



## 2.2.2 MARC元数据

遵循ISO2709的特定格式，其标准构成为：记录头标区、地址目次区、数据字段区、记录分隔符。

记录结构如下：

记录头标	地址目次区	数据字段区	记录分隔符
------	-------	-------	-------



CNMARC书目格式参照UNIMARC书目格式，并结合我国文献编目的实际情况，将数据字段区划分为以下十大功能块。

0—标识块

1—编码信息块

2—著录信息块

3—附注块

4—款目连接块

5—相关题名块

6—主题分析块

7—知识责任块

8—国际使用块

9—国内使用块



## 2

## • 资源库建设标准

ALEPH 编目系统 - 20.01版 库: NLC01 (NLC01) 服务器: 192.168.116.103:6991(20.01) 用户: WAW

ALEPH(A) 查看(Y) 编目(C) 单册(I) 编辑(E) 编辑文本(x) 记录管理器(M) 远程(R) 服务(S) 帮助(H) ?

BK 系统号 3146902 倾听老舍 专著 又是一年芳草绿 (老舍 (1899~1966) lao she 著) 年: 2006

书目记录号

头标区	记录Id	版本Id	ISBN号	通用数据	语言	出版国别	CDF:专著	CDF:形态	题名与责任	版本项	出版项	载体形态项	从编项	载体附注	连接-总集	其它题名	科学主题	中图法	个人-等同	个人-次要
001	003146902	005	20060714164904.0	100	101	102	105	106	200	205	210	215	225	307	461	517	606	690	701	702
----	oam2-22-----^^^450-			20060426d2006^^^^em^y0chiv50^^^^^^ea	chi	CN	a^^^z^^^000cy	r	倾听老舍 专著 又是一年芳草绿 老舍著 吾人选编	2版 修订本	北京 中国广播电视出版社	229页 图, 照片 23cm 1光盘	倾听文学书系 附光盘: ISBN 7-88719-049-5 ISRC CN-M21-02-0019-0	001012002015977	又是一年芳草绿		散文 中国 现代 选集	I266	4 老舍	4 老舍 (1899~1966) lao she 著 吾人 选编

[R] 编辑记录  
 NLC01-314690  
 NLC01-698172  
 NLC01-695759  
 [I] 输入记录  
 [T] 提示

NLC01 - 314690  
 △ 对象 (没有)  
 NLC60  
 馆藏 - S  
 馆藏 - Z  
 馆藏 - Z  
 馆藏 - Z  
 馆藏 - Z  
 NLC50 - 62:

开始 | ALE... | ALE... | 中... | mar... | 可... | PDF XML... | 2Q... | Mic... | 金... | Mic... | 教... | 11:00 周一 2014-05-05

### 2.2.3 DC元数据

都柏林核心元数据 (Dublin Core Metadata), 1995年由OCLC与国家超级计算机应用中心 (NCSA) 联合发起的, 因创始地在美国俄亥俄府都柏林而得名, 目的是建立一套描述数字资源的标准。



## 15个核心元素

名称	标签	名称	标签
Title	题名	Creator	创建者
Contributor	其他责任者	Publisher	出版者
Subject	主题	Description	描述
Date	日期	Type	类型
Format	格式	Identifier	标识符
Source	来源	Language	语种
Relation	关联	Coverage	时空范围
Right	权限		



### 描述型 元数据元素

Title  
Creator  
Subject  
Description  
Publisher  
Contributor  
Type  
Format  
Identifier  
Language  
Coverage

### 管理型 元数据元素

Date  
Right  
Digitalization  
Explanation  
Contributing  
Institution

### 结构型 元数据元素

Relation  
Source



▶ 特色资源库加工原则

- 特色性
- 通用性
- 规范性
- 关联性



- ▶ 特色资源库建设应具备4个环节：
  - 项目需求调研
  - 项目实施方案的制定
  - 项目实施阶段
  - 结束阶段



## 实施内容

- ▶ 图书整理
- ▶ 扫描制作
- ▶ 数据标引
- ▶ 质量检查
- ▶ 数据验收
- ▶ 成品移交



## 3.1 图书整理

### 交接确认

根据交接清单核对与确认文献实体名称和数量。

### 导入数据

根据条形码从Aleph系统获取MARC元数据，并导入系统。

### 文献查重

根据图书条形码查重，根据ISBN号，系统号确定是否需要重新加工。

### 文献前整理

补充著录文献的缺页、损坏等情况，确定是否加工；确定扫描方式（高速扫描、平板扫描、零边距扫描、大幅面扫描等）；系统指定图书加工编号，生成中文图书数据库之图书基本信息表。



## 3.1 图书整理

### ▶ 图书加工编号

图书数字化加工过程中一册图书的唯一标识。

文献基本资料类型（1位）、文献语种（1位）、加工年（2位）、大流水号（4位）、子流水号（2位）。

注：在加工数据库各个表中，要将文献基本资料类型、文献语种、加工年、大流水号（即前8位）和子流水号（2位）用下划线连接，并作为一个字段使用。如：01090012\_66

例：

		文献语种（中文）		子流水号
文献基本资料类型				
（图书）	0 1	09	0012	66
加工年（2009）			大流水号	

其存储路径：c:\01090012\66\



## 图书基本信息表

实施内容

序号	中文名称	字段名称	对应书目数据（MARC）内容
1	加工编号	book_id	
2	分类	cat_id	第一个690字段\$a
3	书名	book_name	200字段\$a.\$h,\$i,\$e
4	作者	author	200字段\$f
5	出版社	pub_house	210字段\$c
6	出版时间	pub_date	210字段\$d
7	ISBN号	isbn	010字段\$a
8	001	record_id	001字段
9	条码号	barcode	



## 3

## · 资源库加工与制作

book_id	cat_id	book_name	author	instructor	specialty	realm	degree	publish_h	publish_d	page_num
01613042_01	A811.5	马克思的哲学	丁立卿著	胡海波指导	马克思主义理	马克思主义基	博士论文	东北师范大学	2008	125页
01613042_02	A811.6	马克思的理论	韩玲著	袁银传指导	马克思主义理	马克思主义理	博士论文	武汉大学	2007	266页
01613042_03	A811.6 D082	马克思民主思	王国宏著	叶庆丰指导	科学社会主义		博士论文	中共中央党校	2007	100页
01613042_04	A811.6 D616	马克思社会形	黄斌著	荣长海指导	科学社会主义	中国特色社会	博士论文	天津师范大学	2008	175页
01613042_05	A811.6 D095	欧洲近代政治	杨晓东著	阎孟伟指导	马克思主义哲		博士论文	南开大学	2007	304页
01613042_06	A811.6 D990	论马克思恩格	葡运珍著	李爱华指导	马克思主义基		博士论文	山东师范大学	2008	159页
01613042_07	A811.6	自由的意蕴及	曾宇辉著	徐大同指导	政治学理论	中西政治思想	博士论文	天津师范大学	2007	166页
01613042_08	A811.6	马克思阶级意	苗泉竹著	侯惠勤指导	马克思主义哲		博士论文	南京大学	2007	195页
01613042_09	A811.6	马克思恩格斯	李征著	刘建军指导	思想政治教育	思想政治教育	博士论文	中国人民大学	2008	170页
01613042_10	A811.6 B03	马克思的空间	孙江著	任平指导	马克思主义哲	发展哲学	博士论文	苏州大学	2007	143页
01613042_11	A811.6 F319	农村公社与东	熊振均著	盛邦和指导	马克思主义哲		博士论文	上海财经大学	2007	225页
01613042_12	A811.6 B516	马克思和海德	吴书林著	王南?指导	马克思主义哲		博士论文	南开大学	2008	149页
01613042_13	A811.6 B03	从“世界历史	刘招明著	张盾指导	马克思主义哲	马克思主义基	博士论文	吉林大学	2009	101页
01613042_14	A811.6 B03	论马克思恩格	李娟著	刘永富指导	马克思主义哲		博士论文	西安交通大学	2008	111页
01613042_15	A811.6 B0-0	现代性的反思	李重著	张再林指导	马克思主义哲		博士论文	西安交通大学	2008	138页

## 3.2 扫描制作

### 色彩校正

扫描前根据国际色彩协会ICC标准，做基本的色彩校正，及针对各类型图书进行色彩校正。

### 文献扫描

根据文献整理中确定的扫描方式采用相应的扫描仪进行扫描。图书为全书逐页扫描方式，依照扫描规格进行图书数字加工；按照命名规则命名图像文件。



## 3.2 扫描制作

### ▶ 灰度方式扫描

色彩位深：8 位

分辨率：300 dpi；小于5号字体用400 dpi

JPEG2000压缩因子：20（根据图像规格、颜色、数据量适当调整）

档案典藏级格式：TIFF 不压缩

发布服务级格式：PDF（经过JPEG2000压缩后，再做格式转换）



## 3.2 扫描制作

### ▶ 彩色方式扫描

色彩位深：24 位

分辨率：300 dpi；小于5号字体用400 dpi

JPEG2000压缩因子：120（根据图像规格、颜色、数据量适当调整）

档案典藏级格式：TIFF 不压缩

发布服务级格式：PDF（经过JPEG2000压缩后，再做格式转换）



## 3.2 扫描制作

### ▶ 图像命名规则

#### 前封（含封一、封二）

扫描文件名为Axxxxx\_00，其中xxxxx为5位数字，按原书顺序依次排序。

#### 前附页

目录页之前的前附页扫描文件名为Bxxxxx\_00，其中xxxxx为5位数字，按原书顺序依次排序。

目录页之后的前附页扫描文件名为Dxxxxx\_00，其中xxxxx为5位数字，按原书顺序依次排序。

#### 目录页

扫描文件名为Cxxxxx\_00，其中xxxxx为5位数字，按原书顺序依次排序。



## 3.2 扫描制作

### ▶ 图像命名规则

#### 正文

有页码的正文扫描文件名为Txxxxx\_00,其中xxxxx为5位数字,与原书页号一致,按原书顺序依次排序。

正文中插页扫描文件名为Txxxxx\_yy,其中xxxxx为5位数字,表示插页的前一页顺序号,yy为数字,表示插页,并按原书顺序依次排序。

#### 后附页

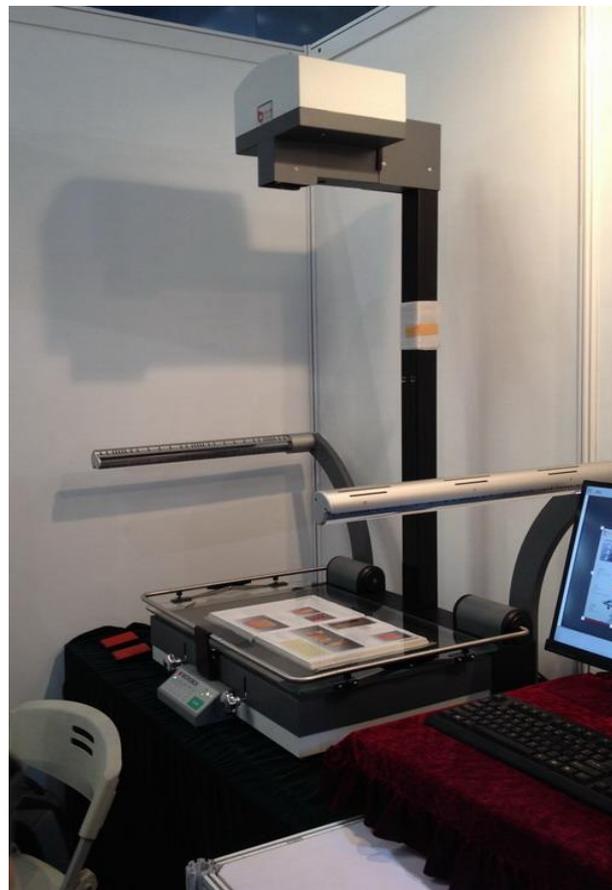
扫描文件名为Yxxxxx\_00,其中xxxxx为5位数字,按原书顺序依次排序。

#### 后封(含封三、封四)

扫描文件名为Zxxxxx\_00,其中xxxxx为5位数字,按原书顺序依次排序。



# 高清扫描仪





高速扫描仪

零边距扫描仪



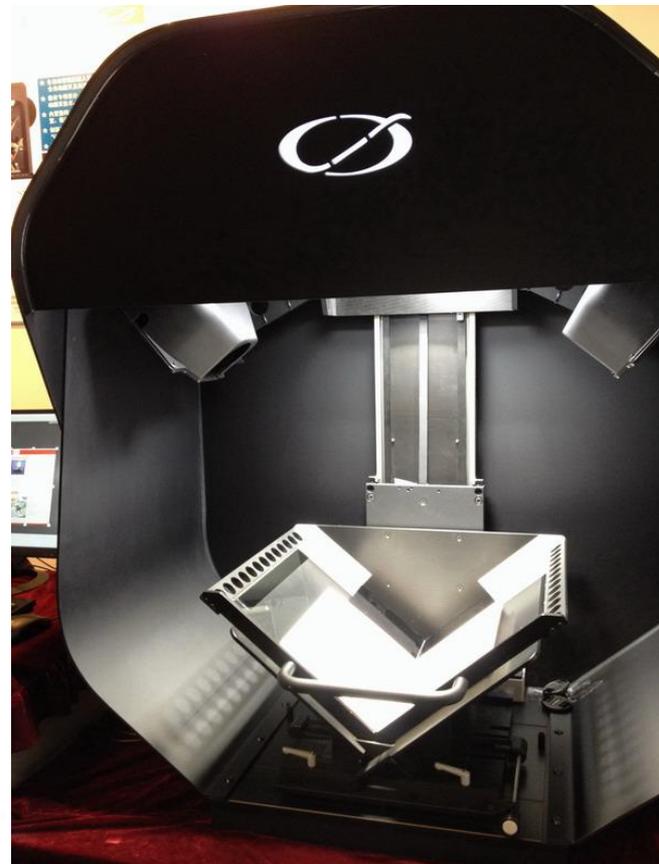


缩微胶卷高速扫描仪





非接触式扫描仪



## 3.3 数据标引

- 目录信息标引
- 摘要及内容简介标引
- 图书版权页标引
- 插页信息标引
- 缺页信息标引
- 图书结构信息标引
- 图书加工信息标引



## 3.3.1 目录信息标引

记录图书目录页内容及图像文件信息。

图书目录信息表 (catalog表)

序号	中文名称	字段名称	备注
1	加工编号	book_id	
2	标引序号	serial_num	数值型
3	章节号	chapter_num	
4	章节名	chapter_name	
5	作者	author	
6	页码	page_num	客观著录, 如实反映目录页原貌 (可为空)
7	绝对页码	ppage_num	数值型, 文件名数字部分
8	页位置	page_place	文件名字母部分
9	属性	page_prop	1) “目录”属性为“1”; 2) “无目录”属性为“2”; 3) 每册书除第一条目录外, 其余记录的属性默认为“0” 数值型

## 图书目录信息表 (catalog表)

book_id	serial_	chapter_num	chapter_name	autho_	page_	ppage_	page_	page_I_
01613042_01	1	目录						1
01613042_01	2		中文摘要		I		2 B	0
01613042_01	3		英文摘要		III		4 B	0
01613042_01	4		目录		V		1 C	0
01613042_01	5		引言		1		1 D	0
01613042_01	6 (一)		为什么研究《1844年手稿》?		1		1 D	0
01613042_01	7 1.		我们是否“读懂”了《1844年手稿》?		1		1 D	0
01613042_01	8 2.		从当代中国的理论需求来研究《1844年手稿》		4		4 D	0
01613042_01	9 (二)		转变研究《1844年手稿》的观念与方式		5		5 D	0
01613042_01	10 1.		回到《1844年手稿》本身		5		5 D	0
01613042_01	11 2.		研究《1844年手稿》的“经济学”与“哲学”双重维度		7		7 D	0
01613042_01	12 一、		《1844年手稿》是马克思主义哲学的真正诞生地和秘密		10		10 T	0
01613042_01	13 (一)		马克思的第一部“经济学——哲学”批判文本		10		10 T	0
01613042_01	14 1.		第一部“经济学——哲学”批判文本的著作特征		10		10 T	0
01613042_01	15 2.		马克思“新哲学”的研究方式——“经济学——哲学”批判		13		13 T	0
01613042_01	16 (二)		马克思对“理论之谜”与“历史之谜”的最初“科学”解答		15		15 T	0

## 3.3.2 摘要及内容简介标引

记录图书摘要及内容简介图像文件信息。

### 摘要及内容信息表

序号	中文名称	字段名称	备注
1	加工编号	book_id	
2	书名	book_name	200字段\$a,\$h,\$i,\$e
3	作者	author	200字段\$f
4	中文摘要位置	chiabst_place	记录中文摘要首页文件名
5	中文摘要页数	chiabst_num	
6	英文摘要位置	engabst_place	记录英文摘要首页文件名
7	英文摘要页数	engabst_num	

## 摘要及内容简介信息表

book_id	book_name	author	chiabst_p	chiabst_n	engabst_p	engabst_n
01613042_01	马克思的哲学革命	丁立卿著	B00002_00	2	B00004_00	2
01613042_02	马克思的理论教育思想研究	韩玲著	B00003_00	3	B00006_00	4
01613042_03	马克思民主思想研究	王国宏著	B00003_00	1	B00004_00	2
01613042_04	马克思社会形态理论的当代阐释与	黄斌著	B00002_00	1	B00003_00	2
01613042_05	欧洲近代政治哲学的发展与马克思	杨晓东著	B00003_00	2	B00005_00	2
01613042_06	论马克思恩格斯的国际法思想	蔺运珍著	D00001_00	3	D00004_00	5
01613042_07	自由的意蕴及其当代价值	曾宇辉著	B00001_00	2	B00003_00	2
01613042_08	马克思阶级意识理论及其当代发展	苗泉竹著	D00001_00	2	D00003_00	2
01613042_09	马克思恩格斯对思想政治教育的贡	李征著	B00002_00	4	B00006_00	6
01613042_10	马克思的空间生产理论及其当代展	孙江著	B00002_00	3	B00005_00	3
01613042_11	农村公社与东方社会结构	熊振均著	B00001_00	2	B00003_00	2
01613042_12	马克思和海德格尔技术思想比较研	吴书林著	B00003_00	1	B00005_00	2
01613042_13	从“世界历史”到“全球化”	刘招明著	B00004_00	1	Y00003_00	2
01613042_14	论马克思恩格斯历史唯物主义的‘	李娟著	B00001_00	2	B00003_00	2
01613042_15	现代性的反思与超越	李重著	B00001_00	2	B00003_00	2
01613042_16	论马克思主义新人道主义的科学向	张少军著	B00002_00	2	B00004_00	2
01613042_17	列宁农民问题理论研究	许蓉著	D00001_00	1	D00002_00	1
01613042_18	列宁司法思想研究	王建国著	D00001_00	1	D00002_00	2
01613042_19	毛泽东中国社会主义发展道路思路	牛田盛著	B00003_00	2	B00005_00	2
01613042_20	国际政治与中国	张菲著	B00002_00	3	B00005_00	3
01613042_21	毛泽东的人生观研究	刘慧著	B00003_00	2	B00005_00	2
01613042_22	毛泽东的社会主义文化观研究	陈桂芝著	B00002_00	1	B00003_00	2



### 3.3.3 图书版权页标引

记录图书版权页图像文件信息。

图书版权信息表（copyright表）

序号	中文名称	字段名称	备注
1	加工编号	book_id	
2	书名	book_name	
3	作者	author	
4	001	record_id	
5	版权页位置	copyright_place	记录版权页文件名

book_id	book_name	author	record_id	copyright
00100201312_01	2011年国家司法考试	杨艳霞, 徐金桂, 杨雄编著	005273254	B00002_00



## 3.3.4 插页信息标引

记录图书插页信息。

图书插页信息表 (inset表)

序号	中文名称	字段名称	备注
1	加工编号	book_id	
2	插页前正文页号	prior_text_page	图书印刷页码
3	插页数量	inset_num	数值型

book_id	prior_text_page	page_place	inset_num
01613046_52	208	T	1

## 3.3.5 缺页信息标引

记录图书缺页信息。

图书缺页信息表 (lostpage表)

序号	中文名称	字段名称	备注
1	加工编号	book_id	
2	缺页前正文页号	start_text_page	图书印刷页码
3	缺页数	lostpage__num	数值型

book_id	start_text	page_plac	lostpage_
01613043_16	205	T	1
01613043_16	139	T	1
01613043_16	111	T	1
01613043_16	71	T	1
01613043_16	13	T	1



## 3.3.6 图书结构信息标引

记录图书封面、前附页、目录、正文等结构信息。

图书结构信息表（struct表）

序号	中文名称	字段名称	备注
1	加工编号	book_id	
2	封面页数	fore_cover_num	数值型
3	目录前，前附页数	preface1_num	数值型
4	目录前，前附页起始页号	preface1_start_page	
5	目录页数	content_num	数值型
6	目录起始页号	content_start_page	
7	目录后，前附页数	preface2_num	数值型
8	目录后，前附页起始页号	preface2_start_page	
9	正文页数	text_num	数值型
10	正文起始页号	text_start_page	
11	后附页数	appendix_num	数值型
12	后附页起始页号	appendix_start_page	
13	封底页数	back_cover_num	数值型

## 图书结构信息表 ( struct表)

book_id	fore_cove	preface1	prefa	content	conter	preface2	preface2	text_num	text_star	appendix	appendix	back_cove
01613042_01	2	5		2	V	9	1	116	10	0		1
01613042_02	2	9		3		21	1	245	22	0		1
01613042_03	2	5		2		6	1	94	7	0		1
01613042_04	2	4		3	I	10	1	165	11	0		1
01613042_05	1	6		3	V	30	1	274	31	0		1
01613042_06	2	1	1	2		18	I	149	11	0		1
01613042_07	2	4	I	2	i	9	1	157	10	0		1
01613042_08	1	0		3	1	4	1	191	5	1		1
01613042_09	2	11		4	1	0		170	1	0		1
01613042_10	1	7		1		3	1	140	4	0		1
01613042_11	3	4	2	2	6	8	8	210	16	0		1
01613042_12	1	6		4	V	7	1	142	8	0		1
01613042_13	2	4		2	I	0		101	1	4	1	1
01613042_14	3	4	I	4	V	0		111	1	1		1
01613042_15	3	4	I	4	V	0		138	1	1		1
01613042_16	2	5		2	V	13	1	80	14	0		1
01613042_17	1	1		3		9	I	109	8	0		1
01613042_18	1	1		3	I	3	I	214	1	0		1
01613042_19	1	6		3	5	9	1	164	10	0		1
01613042_20	2	7		4	1	0		297	1	0		1
01613042_21	2	6		2	V	9	1	71	10	0		1
01613042_22	2	4		2	IV	2	1	104	3	0		1
01613042_23	2	8		4	i	33	1	219	34	0		1
01613042_24	1	6		4	V	16	1	111	17	0		1

## 3.3.7 图书结构信息标引

记录扫描分辨率、压缩因子、文件数量、存储量等信息。

图书加工信息表 (process表)

序号	中文名称	字段名称	备注
1	加工编号	book_id	
2	书名	book_name	
3	扫描分辨率	dpi	数值型
4	压缩因子	comp_factor	
5	灰度页数量	grey_num	数值型
6	彩色页数量	col_num	数值型
7	TIFF数量	tiff_num	数值型
8	PDF数量	pdf_num	数值型
9	TIFF存储量	tiff_mb	数值型, 存储单位: MB
10	PDF存储量	pdf_mb	数值型, 存储单位: MB
11	典藏级光盘位置	cdA_place	第一套典藏级光盘编号
12	服务级光盘位置	cdB_place	第一套服务级光盘编号
13	典藏级硬盘位置	hdA_place	典藏级硬盘财产号
14	服务级硬盘位置	hdB_place	服务级硬盘财产号

## 图书加工信息表 (process表)

book_id	book_name	dpi	comp	grey	col_r	tiff_l	pdf_l	tiff_mt	pdf_m	cdA_place	cdB_r	hdA_place	hdB_place
01613042_01	马克思的哲学	300		133	2	135		987.71		01613A1013		200801135	
01613042_02	马克思的理论	300		279	2	281		1986.88		01613A1013		200801135	
01613042_03	马克思民主思	300		108	2	110		794.63		01613A1013		200801135	
01613042_04	马克思社会形	300		183	2	185		1339.01		01613A1013		200801135	
01613042_05	欧洲近代政治	300		313	2	315		2255.96		01613A1013		200801135	
01613042_06	论马克思恩格	300		171	2	173		1242.94		01613A1013		200801135	
01613042_07	自由的意蕴及	300		173	2	175		1317.21		01613A1013		200801135	
01613042_08	马克思阶级意	300		199	2	201		1540.77		01613A1013		200801135	
01613042_09	马克思恩格斯	300		186	2	188		1447.49		01613A1013		200801135	
01613042_10	马克思的空间	300		151	2	153		1158.37		01613A1013		200801135	
01613042_11	农村公社与东	300		225	3	228		1756.48		01613A1013		200801135	
01613042_12	马克思和海德	300		159	2	161		1241.27		01613A1013		200801135	
01613042_13	从“世界历史	300		112	2	114		846.91		01613A1013		200801135	
01613042_14	论马克思恩格	300		122	2	124		992.09		01613A1013		200801135	
01613042_15	现代性的反思	300		149	2	151		1201.35		01613A1013		200801135	
01613042_16	论马克思主义	300		101	2	103		778.8		01613A1013		200801135	
01613042_17	列宁农民问题	300		122	2	124		975.69		01613A1013		200801135	
01613042_18	列宁司法思想	300		221	2	223		1793.03		01613A1013		200801135	
01613042_19	毛泽东中国社	300		181	3	184		1455.77		01613A1013		200801135	



## ▶ 数据库命名规则

### 数据库命名

#### (1) 硬盘中数据库文件

硬盘中的对象数据所对应数据库。数据库名由 4 位组成：

文献基本资料类型（1 位）、文献语种（1 位）、加工年（2 位）。

本规范针对中文图书其文献基本资料类型为 0，文献语种为 1，加工年为公元年后 2 位数字。

例：0109.mdb

#### (2) 光盘中数据库文件

以 10000 册为单位建立数据库，按批次集中提供的数据库。数据库名由 6 位组成：

文献基本资料类型（1 位）、文献语种（1 位）、加工年（2 位）、流水号（2 位）。

本规范针对中文图书其文献基本资料类型为 0，文献语种为 1，加工年为公元年后两位数字（如 09），流水号为 2 位数字 01-99。

例：010901.mdb



## 3.4 质量检查

### ▶ 质量要求

1. 中文图书均采用ASCII码进行标引，无法录入的生僻字、公式、符号等内容用“■”表示。同时将“■”所对应图像文件保存在档案典藏级数据内，以“■”命名的文件夹内。
2. 标引信息应严格按照原书实际内容进行描述，真实反映图书原貌。各类链接准确无误。
3. 图书封面和各种内页的扫描方式正确，不得随意改变。



## 3.4 质量检查

4. 每本图书相同扫描方式生成的图像保持相同的清晰度，不得有失真现象。

5. 图像歪斜度不可以超过一度；去除与文字、图片、版式无关的杂点、黑边、污迹等信息。

6. 拼接图像接缝处无错位、无缝吻合，不应出现白边和内容缺失，没有明显的歪斜。

7. 长期保存级（档案典藏级）图像，无需进行转换，数字文件保持原采集信息，以无损压缩和不压缩标准格式存档。发布服务级的图像，为有损压缩图像格式，在转换工作中应在图像轮廓清晰可读的前提下（可放大到实际尺寸检查判定），尽量减小数据量。



## 3.4 质量检查

8. 图像名称必需正确，同一数据流水号不得有跳号情况，按顺序排列命名；图像文件的排列顺序应与原书一致。

9. 数据库字段、说明文件、各类表格等内容严格按照附件规定和样例版式，加工方不得擅自更改。

10. 介质中不得存放与备份内容无关的文件，严禁携带病毒，严禁浪费介质空间。



## 3.4 质量检查

### ▶ 数据检查

1. 图像文件（各种格式）放大到1:1状态，逐页检查。检查文件是否有太淡、太浓、黑边、污点、歪斜、模糊（马赛克等）或图像内容不完整等现象，不符合图像质量要求应进行图像校正或重新扫描。
2. 发现文件漏扫时，应及时补扫并正确插入图像。
3. 检查是否符合扫描规格要求。
4. 所有文件保存位置正确，可以正常地打开、显示。



## 3.4 质量检查

5. 检查图像页码是否连续，不得跳页。
6. 文献以册/件为单位检查标引（描述和管理）数据是否完整、准确。
7. 按照命名规则检查目录、文件、数据库、文档、介质等名称是否正确。
8. 检查各类说明、统计、验收等文档是否齐全。



## 3.5 数据验收

1. 数据验收将采取抽样检验，验收人员随机抽选。
2. 各种标引、说明文件的文字、符号、版式、位置和文件名称准确，其综合错误率不超过0.3%。
3. 图像数据扫描方式、扫描规格、文件格式、文件命名、图像处理、压缩方式等符合要求，其综合错误率不超过1%。
4. 成品数据备份数量、保存介质命名、数据存放方式、数据内容符合规范要求，且各类型保存介质内无坏死文件、不准携带病毒，错误率为0。



## 3.5 数据验收

5. 装订后图书保持与扫描前图书的封面、内页等书页的位置和数量前后一致；保证书本规格相差无几。若原书有磁条、RFID标签，加工单位应在本馆购买后贴放在书中。

6. 达到验收标准的数据视为合格，合格范围内检查出的问题进行修正；未达到验收标准的数据由加工单位重新对数据进行检查、修改、重扫等返工工作。

7. 本馆验收人员撰写《数字资源验收记录单》，将验收结果及时通知加工单位，并监督处理过程和结果。



## 3.6 成品移交

### ▶ 文献移交

装订文献，归还文献实体。

### ▶ 数据移交

元数据（书目数据）

数据库文件

数据说明文件

### ▶ 介质移交



## 数据库文件

实施内容

所有 Access 对象	book_id	book_name	author	chiabst_p	chiabst_n	engabst_p	engabst_n
表	01613042_01	马克思的哲学	丁立卿著	B00002_00		2 B00004_00	2
abstract	01613042_02	马克思的理论	韩玲著	B00003_00		3 B00006_00	4
book	01613042_03	马克思民主思	工国宏著	B00003_00		1 B00004_00	2
catalog	01613042_04	马克思社会形	黄斌著	B00002_00		1 B00003_00	2
inset	01613042_05	欧洲近代政治	杨晓东著	B00003_00		2 B00005_00	2
lostpage	01613042_06	论马克思恩格	蔺运珍著	D00001_00		3 D00004_00	5
process	01613042_07	自由的意蕴及	曾宇辉著	B00001_00		2 B00003_00	2
struct	01613042_08	马克思阶级意	苗泉竹著	D00001_00		2 D00003_00	2
查询	01613042_09	马克思恩格斯	李征著	B00002_00		4 B00006_00	6
	01613042_10	马克思的空间	孙江著	B00002_00		3 B00005_00	3
	01613042_11	农村公社与东	熊振均著	B00001_00		2 B00003_00	2
	01613042_12	马克思和海德	吴书林著	B00003_00		1 B00005_00	2
	01613042_13	从“世界历史	刘招明著	B00004_00		1 Y00003_00	2
	01613042_14	论马克思恩格	李娟著	B00001_00		2 B00003_00	2
	01613042_15	现代性的反思	李重著	B00001_00		2 B00003_00	2
	01613042_16	论马克思主义	张少军著	B00002_00		2 B00004_00	2
	01613042_17	列宁农民问题	许蓉著	D00001_00		1 D00002_00	1
	01613042_18	列宁司法思想	王建国著	D00001_00		1 D00002_00	2
	01613042_19	毛泽东中国社	牛田盛著	B00003_00		2 B00005_00	2

## • 数据说明文件

### 1. 数据总体说明

序号	名称	说明
1	项目名称	项目名称
2	加工单位	项目加工单位全名
3	项目时间	项目完成年份
4	数据加工编号	以起始~终止表示, 如 01090001_01~01090213_20
5	完成图书总册数	项目实际完成图书册数
6	完成图书总页数	项目实际完成图像页数
7	序号	以 1 开始, 顺次编号
8	存储内容	存储的数据内容, 如扫描图像、数据库、说明文件等
9	数据格式	以大写英文字母表示, 如 TIFF、PDF、EXCEL、ACCESS
10	存储介质	数据存储的介质, 如硬盘、DVD 光盘
11	介质编号	以起始~终止表示, 光盘填写光盘编号, 硬盘填写硬盘财产编号, 如 0109-1-A100001~0109-1-A100050、200801161
12	介质数量	存储介质的数量
13	存储量	以 GB 为单位表示
14	备注	需要说明的特殊情况, 如“第一套”“第二套”



## • 数据说明文件

### 2. 保存级对象数据硬盘存储清单<sup>4</sup>

序号 <sup>4</sup>	名称 <sup>4</sup>	说明 <sup>4</sup>
1 <sup>4</sup>	数据内容 <sup>4</sup>	存储的数据内容，如中文图书扫描图像、对应数据库 <sup>4</sup>
2 <sup>4</sup>	数据格式 <sup>4</sup>	以大写英文字母表示，如 TIFF、ACCESS <sup>4</sup>
3 <sup>4</sup>	序号 <sup>4</sup>	以 1 开始，顺次编号 <sup>4</sup>
4 <sup>4</sup>	硬盘编号 <sup>4</sup>	本硬盘的财产编号，如 200802162 <sup>4</sup>
5 <sup>4</sup>	图书加工编号 <sup>4</sup>	本块硬盘存储的图书加工编号起止号， <sup>4</sup> 如 01090001_01~01090050_99 <sup>4</sup>
6 <sup>4</sup>	图书册数 <sup>4</sup>	本块硬盘存储的中文图书总册数 <sup>4</sup>
7 <sup>4</sup>	图书页数 <sup>4</sup>	本块硬盘存储的中文图书总页数 <sup>4</sup>
8 <sup>4</sup>	存储量 <sup>4</sup>	以 GB 为单位表示 <sup>4</sup>
9 <sup>4</sup>	备注 <sup>4</sup>	需要说明的特殊情况 <sup>4</sup>
10 <sup>4</sup>	合计 <sup>4</sup>	图书加工编号以起止号表示，合计图书册数、图书页数、存储总量 <sup>4</sup>



## • 数据说明文件

### 3. 保存级对象数据光盘存储清单<sup>4</sup>

序号 <sup>4</sup>	名称 <sup>4</sup>	说明 <sup>4</sup>
1 <sup>4</sup>	数据内容 <sup>4</sup>	存储的数据内容，如中文图书扫描图像 <sup>4</sup>
2 <sup>4</sup>	数据格式 <sup>4</sup>	以大写英文字母表示，如 TIFF <sup>4</sup>
3 <sup>4</sup>	序号 <sup>4</sup>	以 1 开始，顺次编号 <sup>4</sup>
4 <sup>4</sup>	光盘编号 <sup>4</sup>	本张光盘的光盘编号，如 0109-1-A100001 <sup>4</sup>
5 <sup>4</sup>	图书加工编号 <sup>4</sup>	本张光盘存储的图书加工编号起止号， <sup>4</sup> 如 01090001_01~01090001_05 <sup>4</sup>
6 <sup>4</sup>	图书册数 <sup>4</sup>	本张光盘存储的中文图书册数， <u>同册图书图像跨盘存放</u> ，册数计入刻录的首张光盘内 <sup>4</sup>
7 <sup>4</sup>	图书页数 <sup>4</sup>	本张光盘存储的中文图书总页数 <sup>4</sup>
8 <sup>4</sup>	存储量 <sup>4</sup>	以 GB 为单位表示 <sup>4</sup>
9 <sup>4</sup>	备注 <sup>4</sup>	需要说明的特殊情况 <sup>4</sup>
10 <sup>4</sup>	合计 <sup>4</sup>	光盘编号及图书加工编号以 <u>起止号</u> 表示，合计图书册数、图书页数、存储总量 <sup>4</sup>



## • 数据说明文件

┆

### 4. 发布级对象数据硬盘存储清单

序号	名称	说明
1	数据内容	存储的数据内容，如中文图书扫描图像、对应数据库
2	数据格式	以大写英文字母表示，如 PDF、ACCESS
3	序号	以 1 开始，顺次编号
4	硬盘编号	本硬盘的财产编号，如 200802162
5	图书加工编号	本块硬盘存储的图书加工编号起止号， 如 01090001_01~01090050_99
6	图书册数	本硬盘存储的中文图书总册数
7	图书页数	本硬盘存储的中文图书总页数
8	存储量	以 GB 为单位表示
9	备注	需要说明的特殊情况
10	合计	图书加工编号以起止号表示，合计图书册数、图书页数、存储总量



## • 数据说明文件

### 5. 发布级对象数据光盘存储清单<sup>↵</sup>

序号 <sup>↵</sup>	名称 <sup>↵</sup>	说明 <sup>↵</sup>
1 <sup>↵</sup>	数据内容 <sup>↵</sup>	存储的数据内容，如中文图书扫描图像 <sup>↵</sup>
2 <sup>↵</sup>	数据格式 <sup>↵</sup>	以大写英文字母表示，如 PDF <sup>↵</sup>
3 <sup>↵</sup>	序号 <sup>↵</sup>	以 1 开始，顺次编号 <sup>↵</sup>
4 <sup>↵</sup>	光盘编号 <sup>↵</sup>	本张光盘的光盘编号，如 0109-1-B100001 <sup>↵</sup>
5 <sup>↵</sup>	图书加工编号 <sup>↵</sup>	本张光盘存储的图书加工编号起止号， <sup>↵</sup> 如 01090001_01~01090001_05 <sup>↵</sup>
6 <sup>↵</sup>	图书册数 <sup>↵</sup>	本张光盘存储的中文图书册数 <sup>↵</sup>
7 <sup>↵</sup>	图书页数 <sup>↵</sup>	本张光盘存储的中文图书页数 <sup>↵</sup>
8 <sup>↵</sup>	存储量 <sup>↵</sup>	以 GB 为单位表示 <sup>↵</sup>
9 <sup>↵</sup>	备注 <sup>↵</sup>	需要说明的特殊情况 <sup>↵</sup>
10 <sup>↵</sup>	合计 <sup>↵</sup>	光盘编号及图书加工编号以 <u>起止号</u> 表示，合计图书册数、图书页数、存储总量 <sup>↵</sup>



## • 数据说明文件

6. 图书单册数据量统计表<sup>↵</sup>

序号 <sup>↵</sup>	名称 <sup>↵</sup>	说明 <sup>↵</sup>
1 <sup>↵</sup>	序号 <sup>↵</sup>	以 1 开始, 顺次编号 <sup>↵</sup>
2 <sup>↵</sup>	加工编号 <sup>↵</sup>	图书加工编号 <sup>↵</sup>
3 <sup>↵</sup>	灰度页 <sup>↵</sup>	本册书中按灰度方式扫描的图像数量 <sup>↵</sup>
4 <sup>↵</sup>	彩色页 <sup>↵</sup>	本册书中按彩色方式扫描的图像数量 <sup>↵</sup>
5 <sup>↵</sup>	图像合计 <sup>↵</sup>	灰度页与彩色页数量之和 <sup>↵</sup>
6 <sup>↵</sup>	目录条数 <sup>↵</sup>	本册书目录信息表 (catalog 表) 条目数 <sup>↵</sup>
7 <sup>↵</sup>	合计 <sup>↵</sup>	图书加工编号以起止号表示, 合计灰度页数、彩色页数、图像总数、目录标引总条数 <sup>↵</sup>



## • 光盘命名规则

- (1) 对象数据光盘名  
由 12 位组成：  
文献基本资料类型（1 位）、文献语种（1 位）、加工年（2 位）、加工单位代码（1 位）对象数据加工级别（1 位）、光盘复本号（1 位）、流水号（5 位）。  
本规范针对中文图书其文献基本资料类型为 0，文献语种为 1，加工年为公元年后两位数字（如 09），加工单位代码（本馆分配），加工级别为档案典藏级（A）或发布服务级（B），光盘复本号为第一套（1）或第二套（2），流水号为 5 位数字 00001-99999。  
例：档案典藏级光盘第一套：0109-1-A101234  
档案典藏级光盘第二套：0109-1-A201234  
发布服务级光盘第一套：0109-1-B101234  
发布服务级光盘第二套：0109-1-B201234
- (2) 数据库光盘名  
由 9 位组成：  
文献基本资料类型（1 位）、文献语种（1 位）、加工年（2 位）、加工单位代码（1 位）、光盘内容（1 位）、光盘复本号（1 位）、流水号（2 位）。  
本规范针对中文图书其文献基本资料类型为 0，文献语种为 1，加工年为公元年后两位数字（如 09），光盘内容为数据库（C），光盘复本号为第一套（1）或第二套（2），流水号为 2 位数字 01-99。  
例：数据库光盘第一套：0109-1-C101  
数据库光盘第二套：0109-1-C201
- (3) 说明文件光盘名  
由 7 位组成：  
文献基本资料类型（1 位）、文献语种（1 位）、加工年（2 位）、加工单位代码（1 位）、光盘内容（1 位）、光盘复本号（1 位）。  
本规范针对中文图书其文献基本资料类型为 0，文献语种为 1，加工年为公元年后两位数字（如 09），光盘内容为光盘说明文件（D），光盘复本号为第一套（1）或第二套（2）。  
例：说明文件光盘第一套：0109-1-D1  
说明文件光盘第二套：0109-1-D2





谢谢!  
THANK YOU!