# 地方少数民族语言书刊数字化加工规则（2019）

### 规则适用范围：

本技术规则用于2019年地方少数民族语言书刊数字化项目。本规则要求针对少数民族语言书刊规定了加工编号、对象数据、元数据以及命名规则存储规则等应遵循的标准。

# 第一部分 少数民族语言图书标准：

# 1元数据加工标准

## 1.1规范性引用文件

下列文件对于本规则的应用是必不可少的。

《中国文献编目规则》（第二版）

《GB/T25100-2010 信息与文献 都柏林核心元数据》

《信息资源的内容形式和媒体类型标识》（GB/T3469-2013）

## 1.2著录对象

以数字化文献为著录对象。一般以具有独立名称、并可独立使用的一个数字资源为著录单位。

图书的著录单位为一种图书。对于多部分资源，一般可自由选择综合著录或分散著录。

## 1.3著录信息源

著录信息源是资源本身。资源本身是著录的首选来源。资源本身信息不足，可参考其他信息源。

## 1.4著录用文字

题名、责任者等取自资源本身的信息一般按资源本身的文字著录，著录我国各少数民族文献，应遵循其民族文字书写规则。

为方便用户检索，出现在资源中的题名只有少数民族文字时，编目员需在元数据中补充简体中文的题名。

由编目员进行描述或标引的信息均采用简体中文著录。

有规范或规定编码标准的要求按标准著录，如：语种、中图分类。

日期用阿拉伯数字著录。

推荐采用ISBD著录标识符。

一个著录项中多个著录内容之间以半角分号分隔。

## 1.5元数据著录简表

|  |  |  |
| --- | --- | --- |
| **术语** | **必备性** | **著录内容** |
| 加工编号 | 必备 | 著录资源的加工编号。 |
| 正题名 | 必备 | 著录资源的主要名称，根据在资源中出现的形式著录。  少数民族文字著录应遵循其民族文字书写规则。  在题名页或代题名页同时出现少数民族文字和中文题名时，应选择中文题名作为正题名，将少数民族文字题名作为其他题名著录。 |
| 其他题名 | 有则必备 | 除正题名以外的题名信息。包括其他语种的题名或用于限定、补充、解释正题名的题名信息。  当在资源本身未出现中文题名时，应在本字段著录取自资源以外信息源上的中文题名或由编目员翻译的中文题名。 |
| 责任者名称 | 有则必备 | 创建资源内容的主要责任者或对资源内容做出贡献的其他实体。  少数民族责任者应将民族著录于姓名前圆括号内（民族应使用少数民族规范名称，参见《中国文献编目规则》第二版附录3），若同时载有汉译姓名和姓名原文，姓名原文著录于汉译姓名后圆括号内。  例：责任者名称：(维吾尔族)帕尔哈提·艾克木(perhat hékim) |
| 责任方式 | 有则必备 | 著录责任者对资源内容做出贡献的方式。不同责任方式对应的责任者名称应分别著录。  例：责任者名称1:(美)拉斯·克兰德  责任方式1:著  责任者名称2:陈永剑  责任方式2:译 |
| 中图分类 | 必备 | 描述资源主题内容的中图分类法类号。依照CLC（《中国图书馆分类法》第五版）进行标引。 |
| 主题词或关键词 | 必备 | 描述资源主题内容的受控或非受控的词汇。受控主题词依据CCT（中国分类主题词表）进行标引。 |
| 摘要 | 必备 | 用自由行文的方式对资源内容的概括性描述。描述文字不少于100字。 |
| 附注 | 有则必备 | 关于资源内容的其他说明。凡未在其他字段中著录而又有必要补充说明的内容均可著录于本字段。如：作者简介。可采用自由行文方式著录。  应尽可能采用固定导语和规范用语。  不同类型信息应重复本字段。  例：附注：哈米提·阿哈甫(1985~)，哈萨克族，新疆维吾尔自治区民族事务委员会民族宗教政策研究所助理研究员。主要从事民族文化研究。新疆阿勒泰人。 |
| 出版者名称 | 有则必备 | 使资源可以获得和利用的责任实体，如资源的出版者或颁布者。 |
| 出版地 | 有则必备 | 著录出版者所在的城市名称。如具体出版地不详，可著录所在省名或国名。 |
| 出版日期 | 有则必备 | 著录资源的出版日期。具体著录时，通过数字化形式获得的再生数字资源应著录其实体文献的出版日期。 |
| 内容形式 | 必备 | 著录资源的内容形式词及内容限定词。取值参见国标《信息资源的内容形式和媒体类型标识(GB/T 3469-2013)》。  例：内容形式：图像（测绘型）  （注：一本地图集） |
| 媒体类型 | 必备 | 著录资源的媒体类型。取值参见国标《信息资源的内容形式和媒体类型标识(GB/T 3469-2013)》。本项目中缺省取值为“电子”。 |
| 格式 | 必备 | 资源的格式。需用英文大写著录。  例：TIF;PDF |
| 页码 | 必备 | 著录图书的页码。  多卷（册）图书综合著录时，页数连续编码者，先著录总册数，在总册数后的圆括号内著录总页数，页数单独编码者，在总册数后的圆括号内依次著录各册页数，中间用分号分隔。  多卷（册）图书分散著录时，页数连续编码者，著录其起讫页码。  例1：4册(1317页)  例2：3册(80;120;180页)  例3：101-543页 |
| 语种 | 必备 | 著录资源的文字语种。采用3位语种代码，参考《新版中国机读目录格式使用手册》（附录A）。常见代码如下：  汉语：chi  英语：eng  藏语：tib  蒙古语：mon  维吾尔语：uig  朝语：kor |
| 关联 | 有则必备 | 描述与本资源相关的其他资源。这些关系包括包含与被包含、参照与被参照、继承与被继承、替代与被替代、其他版本与原版本、其他格式与原格式等关系。如资源的丛编、合订。相关资源一般独立存在。可著录资源名称或资源标识符。 |
| 时间范围 | 有则必备 | 著录资源内容涉及的时间特征，著录内容包括时代、日期、时间段。  例：正题名：清代锡伯族档案史料选编  时间范围：清代 |
| 空间范围 | 有则必备 | 著录资源内容涉及的空间特征，著录内容包括地点、地理坐标。  例：正题名：云南五州市沙马曲比源流及其分布  空间范围：云南省 |
| 版本 | 有则必备 | 著录资源的版本或版次信息以及其他附加说明。  版本一般是指同一种文献因编辑、传抄、刻板及装订或制作形式的不同而产生的不同本子。如：修订版、增订版、缩印本、影印本等。  对于有关内容特征、体裁特征和适用范围的说明文字一般不著录于本字段。如：缩写本、绘画本、节选本、少年版、科学版等。 |
| 适用对象 | 必备 | 资源适宜使用的人群，取值为：普通用户、少年儿童、残障人士。 |
| ISBN | 有则必备 | 著录在国际标准书号中确认资源的唯一标识。 |
| 统一书号 | 有则必备 | 著录在统一书号中确认资源的唯一标识。 |
| CDOI | 有则必备 | 著录在CDOI中确认资源的唯一标识。 |
| 发布地址 | 有则必备 | 资源在因特网或局域网的发布地址。 |
| 数据提交单位 | 必备 | 著录提交该数据的单位。 |
| 所属任务年份 | 必备 | 著录联建工作的任务年度。 |

# 2图像数字化标准和命名规则

## 2.1图像扫描加工

1. 扫描前根据国际色彩协会（International Color Consortium，简称ICC）标准，做加工设备的基本色彩校正，及针对各类型文献进行色彩校正。
2. 逐页扫描
3. 黑白页和灰度页用灰度方式扫描

色彩位深：8 位

分辨率：300 dpi；小于5号字体用400 dpi

档案典藏级格式：TIFF 不压缩

发布服务级：PDF（经过 JPEG2000 压缩后，再做格式转换）

1. 彩色页用彩色方式扫描

色彩位深：24 位

分辨率：300 dpi；小于5号字体用400 dpi

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）

1. 纠偏处理。对出现偏斜的图像进行纠偏处理，图像歪斜度不可以超过一度，对方向不正确的图像进行旋转还原，以符合阅读习惯。
2. 去污处理。对图像页面中出现的影响图像质量的杂质如黑边等进行去污处理。
3. 图像拼接。若原文献幅面较大，无法整体采集的，可将原件分画幅采集（分画幅采集时，注意各拍摄区域边缘须留有一定的重复采集区域，以保证拼接完成后所有接缝处连贯完整），服务级文件进行拼接处理，合并为一个完整的图像，以保证数字文件的整体性（合并信息应在readme.txt进行说明）。

## 2.2图像文件质量要求

1. 图像文件（各种格式）放大到1：1状态，逐页检查。检查文件是否有彩点、彩线、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。若不符合图像质量要求应进行图像校正或重新扫描；
2. 图像文件真实反映原件，同册图书的图像尺寸相同，不得有失真现象；不能有缺页、错页、数据内容缺失等现象（原件有缺失的除外）；
3. 档案典藏级图像，允许进行适当的纠偏，去污等处理，以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，一页为一个PDF文件，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；
4. 图像符合扫描规格要求和技术参数；
5. 所有文件保存位置以及文件命名正确，可以有效打开和显示；
6. 图像综合错误率不超过1‰。

## 2.3命名规则

### 2.3.1加工编号（book\_id）

文献数字化加工过程中一册文献的唯一标识，它由11位数字和1位下划线组成。

文献基本资料类型（1位）、文献语种（1位）、任务年（2位）、机构代码（4位）、下划线（1位）、单位内部流水号(3位)。

本规则针对地方图书的文献基本资料类型为（0：图书），文献语种为（2：民语），任务年为公元年后两位数字（19年项目统一为19），机构代码见《图书馆机构代码表》（2019），单位内部流水号由各单位自行分配，从1开始，不足3位以0补齐。

例：02190000\_001

### 2.3.2图像文件名

1. 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。

1. 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。

1. 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。

1. 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。当存在大幅图像需分区扫描时其文件命名为Txxxxx\_00zz，其中zz为数字，表示扫描图像分区，按扫描顺序排序，对应的发布服务级进行合并后命名为Txxxxx\_00。例如：该图在原书的第28页，该图分成4个画幅，则4个画幅的扫描文件名分别为T00028\_0001.tif、T00028\_0002.tif、T00028\_0003.tif、T00028\_0004.tif，拼接后的发布级图像文件名为T00028\_00.pdf。

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。当插页中存在大幅图像需分区扫描时其命名为Txxxxx\_yyzz，yy为数字，表示插页；zz为数字，表示扫描图像分区，按扫描顺序排序，对应的发布服务级图像进行合并后命名为Txxxxx\_yy。例如：一张图夹在20页、21页之间，且该图分成4个画幅扫描，则4个画幅的扫描文件名分别为T00020\_0101.tif、T00020\_0102.tif、T00020\_0103.tif、T00020\_0104.tif，将4个画幅拼接后，该图的文件名为T00020\_01.pdf。

1. 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。

1. 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。

# 3数据库标准

## 3.1数据库命名

对象数据所对应数据库。数据库名由8位数字组成：

文献基本资料类型（1位）、文献语种（1位）、任务年（2位）、机构代码（4位）。

例：02190000.mdb

## 3.2数据库制作

**3.2.1** 对marc数据进行摘取，制作文献基本信息表（book）。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 中文名称 | 字段名称 | 字符类型 | 必备性 | 备注 |
| 1 | 加工编号 | book\_id | char | 必备 |  |
| 2 | 分类 | cat\_id | char | 必备 | 第一个690字段$a |
| 3 | 书名（民语） | book\_name1 | char | 必备 |  |
| 4 | 书名（汉语） | book\_name2 | char | 必备 |  |
| 5 | 作者（民语） | author1 | char | 有则必备 |  |
| 6 | 作者（汉语） | author2 | char | 有则必备 |  |
| 7 | 出版社 | pub\_house | char | 有则必备 | 210字段$c |
| 8 | 出版时间 | pub\_date | char | 有则必备 | 210字段$d |
| 9 | 001 | record\_id | char | 必备 | 001字段 |

**3.2.2**文献缺页信息著录，生成文献缺页信息表(lostpage)。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 中文名称 | 字段名称 | 字符类型 | 必备性 | 备注 |
| 1 | 加工编号 | book\_id | char | 必备 |  |
| 2 | 缺页前页号 | lost\_text\_page | 数字 | 必备 | 缺页文件名数字部分 |
| 3 | 缺页位置 | lostpage\_place | char | 必备 | 缺页文件名字母部分 |
| 4 | 缺页数量 | lostpage\_num | 数字 | 必备 |  |

注：图书有缺失页时，对该表字段进行标引。

**3.2.3** 记录扫描分辨率、压缩因子、文件数量、存储量等信息，生成文献加工信息表(process)。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 中文名称 | 字段名称 | 字符类型 | 必备性 | 备注 |
| 1 | 加工编号 | book\_id | char | 必备 |  |
| 2 | 书名（汉语） | book\_name | char | 必备 |  |
| 3 | 扫描分辨率 | dpi | 数字 | 必备 |  |
| 4 | 压缩因子 | comp\_factor | char | 必备 |  |
| 5 | 灰度页数量 | grey\_num | 数字 | 必备 |  |
| 6 | 彩色页数量 | col\_num | 数字 | 必备 |  |
| 7 | TIFF数量 | tiff\_num | 数字 | 必备 |  |
| 8 | PDF数量 | pdf\_num | 数字 | 必备 |  |
| 9 | TIFF存储量 | tiff\_mb | 数字 | 必备 | 存储单位：MB |
| 10 | PDF存储量 | pdf\_mb | 数字 | 必备 | 存储单位：MB |
| 11 | TIFF硬盘位置 | hdA\_place | char | 必备 | 硬盘号 |
| 12 | PDF硬盘位置 | hdB\_place | char | 必备 | 硬盘号 |

## 3.3 数据库制作要求

数据库的制作符合下列要求。

（1）文件格式为mdb格式。

（2）使用字符集为UTF-8。

（3）著录信息应严格按照文献实际内容进行描述，所著内容与对象文件应正确链接。

（4）无法录入的生僻字等用“〓”表示。

（5）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。

（6）各种著录、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‰。

# 4说明文件

## 4.1 制作文献的说明文件

说明文件格式：Microsoft Excel

说明文件命名：由8位组成：文献基本资料类型（1位）、文献语种（1位）、加工年（2位）、机构代码（4位）。如：02190000.xls

说明文件命名：地方文献数据说明（说明文件的4张表需保存在同一个Excel文件的不同工作表中，每个工作表按下列表命名）。

### 4.1.1数据总体说明表

|  |  |
| --- | --- |
| **名称** | **说明** |
| 项目名称 | 填写项目名称，如2019年地方少数民族语言图书数字化 |
| 承建馆 | 承建馆名称全名 |
| 质检单位 | 第三方质检单位全名 |
| 项目时间 | 填写项目所属年份 |
| 数据加工编号 | 以起始～终止表示，如02190000\_001～02190000\_040 |
| 完成文献总册书 | 这批数据图书总册数 |
| 完成文献总页数 | 这批数据图书总页数 |
| 序号 | 以1开始，顺次编号 |
| 存储内容 | 如：扫描图像，PDF文件,第三方质检报告、版权证明等 |
| 数据格式 | TIFF，PDF等 |
| 存储介质 | 硬盘 |
| 介质编号 | 硬盘编号，如02190000 |
| 介质数量 | 存储硬盘的数量 |
| 存储量 | 以GB为单位表示 |
| 备注 | 获得版权证明的图书册数 |

### 4.1.2保存级对象数据硬盘存储清单

|  |  |
| --- | --- |
| **名称** | **说明** |
| 数据内容 | 存储的数据内容，如2019年少数民族语言图书数字化 |
| 数据格式 | 以大写英文字母表示，如TIFF |
| 序号 | 以1开始，顺次编号 |
| 硬盘编号 | 本硬盘的硬盘编号 |
| 图书加工编号 | 本块硬盘存储的图书加工编号起止号，  如02190000\_001～02190000\_040 |
| 图书册数 | 本块硬盘存储的图书总册数 |
| 图书页数 | 本块硬盘存储的图书总页数 |
| 存储量 | 以GB为单位表示 |
| 备注 | 需要说明的特殊情况 |
| 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 |

注：硬盘编号按递增顺序排列

### 4.1.3发布级对象数据硬盘存储清单

|  |  |
| --- | --- |
| **名称** | **说明** |
| 数据内容 | 存储的数据内容，如2019年少数民族语言图书数字化 |
| 数据格式 | 以大写英文字母表示，如PDF |
| 序号 | 以1开始，顺次编号 |
| 硬盘编号 | 本硬盘的编号 |
| 图书加工编号 | 本块硬盘存储的图书加工编号起止号，  如02190000\_001～02190000\_040 |
| 图书册数 | 本硬盘存储的图书总册数 |
| 图书页数 | 本硬盘存储的图书总页数 |
| 存储量 | 以GB为单位表示 |
| 备注 | 需要说明的特殊情况 |
| 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 |

注：硬盘编号按递增顺序排列

### 4.1.4图书单册数据量统计表

|  |  |
| --- | --- |
| **名称** | **说明** |
| 序号 | 以1开始，顺次编号 |
| 加工编号 | 图书加工编号 |
| 灰度页 | 本册书中按灰度方式扫描的图像数量 |
| 彩色页 | 本册书中按彩色方式扫描的图像数量 |
| 图像合计 | 灰度页与彩色页数量之和 |
| 目录条数 | 本册文献目录信息表条目数 |
| 合计 | 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 |

## 4.2 制作每册文献和保存数据介质的说明文件

**每册文献的说明文件 （bookinfo.txt）**

book\_id=02190101\_001

book\_name=xxxxxx

author=xxxxxx

file\_number=xxxxxx

说明：

book\_id：文献加工编号

book\_name：书名

author：作者

file\_number：总文件数

**存储介质的说明文件（readme.txt）**

一、文献类型：例如，少数民族语言图书

二、制作时间：2019年 月

三、承建馆：

四、加工单位：

五、存储介质信息

1、介质名称：

2、加工起止号：例如，02190101\_001 ～ 02190101\_099

3、文献数量：

4、文件数量：

5、存储容量：

六、技术参数

1、存储格式：例如，不压缩tif格式

2、加工设备：

3、加工软件：

4、参数指标

扫描分辨率：例如，300dpi

扫描方式：

备注：如：02192100\_002中T00006\_0101.tif、T00006\_0102.tif合并为T00006\_01.pdf

# 5存储规则

## 5.1存储介质及命名

存储介质为硬盘，硬盘命名由8位组成：

文献基本资料类型（1位）、文献语种（1位）、加工年（2位）、机构代码（4位）。

## 5.2数据存储规则

存储内容包括：TIFF图像、PDF、对应数据库、说明文件、介质说明文件（readme.txt）、书目数据文件、版权证明、第三方质检报告，各一份。

各类型对象数据按加工编号递增顺序依次放置在硬盘内，保存目录及文件名由四级组成：

第一级目录为加工编号的前8位，如02190000，同时将介质说明文件(readme.txt) 及第三方质检报告电子版并行保存；

第三方质检报告扫描时采用300DPI，以PDF格式保存，图像需进行拼接处理，合并为一个完整的图像，命名规则为“加工编号前8位+T”,如：02190000T.pdf

第二级目录为TIFF,PDF,同时将对应数据库（02190000.mdb）、说明文件（02190000.xls）、书目数据文件（02190000.iso）及版权证明电子版并行保存；版权证明扫描时采用300DPI，以PDF格式保存，图像需进行拼接处理，合并为一个完整的图像。文件夹命名为CERTIFICATE。

第三级目录TIFF,PDF为加工编号，CERTIFICATE文件夹内版权证明分以下两种情况命名：一是本年度加工图书按册出版权证明的，命名规则为“加工编号+C”,如02190000\_001C.pdf；二是本年度加工图书总共出一份版权证明的，命名规则为“加工编号前8位+C”，如02190000C.pdf（下图所示为第一种情况）。

第四级目录为相应格式的对象数据文件，需要相应的单册文献的说明文件。

├─02190000、readme.txt、02190000T.pdf

│ ├──TIFF

│ │ ├─加工编号

│ │ │ ├─A00001\_00.tif

│ │ │ ├─A00002\_00.tif……

│ │ │ ├─bookinfo.txt……

│ │ ├─加工编号

│ │ │ ├─A00001\_00.tif

│ │ │ └─A00002\_00.tif……

│ │ └─……

│ ├──PDF

│ │ ├─加工编号

│ │ └─加工编号……

│ ├──CERTIFICATE

│ │ ├─加工编号C.pdf

│ │ ├─加工编号C.pdf

│ ├──02190000.iso

│ ├──02190000.mdb

│ └──02190000.xls

# 6.修订说明

6.1本年度新增第三方质检报告及版权证明电子版的提交工作，具体修订见5.2数据存储规则。

# 第二部分 少数民族语言期刊标准：

# 1元数据加工标准

## 1.1规范性引用文件

下列文件对于本规则的应用是必不可少的。

《中国文献编目规则》（第二版）

《GB/T25100-2010 信息与文献 都柏林核心元数据》

《信息资源的内容形式和媒体类型标识》（GB/T3469-2013）

## 1.2著录对象

以数字化文献为著录对象。一般以具有独立名称、并可独立使用的一个数字资源为著录单位。

期刊的著录单位为一种期刊。

对于连续性资源的著录对象是否发生改变，可依据题名、发行者、版本等的变化情况进行判断，具体规则参照《中国文献编目规则》（第二版）。

## 1.3著录信息源

著录信息源是资源本身。资源本身是著录的首选来源。资源本身信息不足，可参考其他信息源。

期刊的著录信息源是本题名下第一期或第一部分，若无法获得本题名下第一期或第一部分，应以可获得的最早发行的一期或部分著录。首选来源是期刊的题名页或代题名页（无题名页时，以封面、目次页、编辑页、版权页作为代题名页）。

## 1.4著录用文字

题名、责任者等取自资源本身的信息一般按资源本身的文字著录，著录我国各少数民族文献，应遵循其民族文字书写规则。

为方便用户检索，出现在资源中的题名只有少数民族文字时，编目员需在元数据中补充简体中文的题名。

由编目员进行描述或标引的信息均采用简体中文著录。

有规范或规定编码标准的要求按标准著录，如：语种、中图分类。

日期用阿拉伯数字著录。

推荐采用ISBD著录标识符。

一个著录项中多个著录内容之间以半角分号分隔。

## 1.5元数据著录简表

|  |  |  |
| --- | --- | --- |
| **术语** | **必备性** | **著录内容** |
| 加工编号 | 必备 | 著录资源的加工编号。 |
| 正题名 | 必备 | 著录资源的主要名称，根据在资源中出现的形式著录。  少数民族文字著录应遵循其民族文字书写规则。  在题名页或代题名页同时出现少数民族文字和中文题名时，应选择中文题名作为正题名，将少数民族文字题名作为其他题名著录。  以特定名称单独出版的分辑或副刊（或附刊、增刊、补编等），可单独制作元数据。当该题名能够单独存在且与共同题名或主刊题名比较具有更为重要的作用时，以分辑题名或副刊（或附刊、增刊、补编等）题名作为正题名，并将共同题名或主刊题名著录在关联字段。 |
| 其他题名 | 有则必备 | 除正题名以外的题名信息。包括其他语种的题名或用于限定、补充、解释正题名的题名信息。  当在资源本身未出现中文题名时，应在本字段著录取自资源以外信息源上的中文题名或由编目员翻译的中文题名。 |
| 中图分类 | 必备 | 描述资源主题内容的中图分类法类号。依照CLC（《中国图书馆分类法》第五版）进行标引。 |
| 主题词或关键词 | 必备 | 描述资源主题内容的受控或非受控的词汇。受控主题词依据CCT（中国分类主题词表）进行标引。 |
| 附注 | 有则必备 | 关于资源内容的其他说明。凡未在其他字段中著录而又有必要补充说明的内容均可著录于本字段。如：正题名的变化、书目沿革等。可采用自由行文方式著录。  应尽可能采用固定导语和规范用语。  不同类型信息应重复本字段。  例：正题名：齐齐哈尔师院  附注：1976年复刊时改为：《齐齐哈尔师院学报》 |
| 出版者名称 | 有则必备 | 使资源可以获得和利用的责任实体，如资源的出版者或颁布者。  如果连续出版过程中出版者发生变化，若认为这一变化重要，可在附注中注明改变后的出版者。 |
| 出版地 | 有则必备 | 著录出版者所在的城市名称。如具体出版地不详，可著录所在省名或国名。 |
| 开始日期 | 有则必备 | 著录期刊最早一期的正式出版日期。著录内容是该刊出版情况的客观反映，与编目机构的收藏情况及数字化情况无任何联系。  非公元纪年应如实著录，其后再著录公元纪年，并置于方括号内。  例：开始日期：民国15年[1926] |
| 终止日期 | 有则必备 | 著录期刊结束出版的日期。著录内容是该刊出版情况的客观反映，与编目机构的收藏情况及数字化情况无任何联系。  经休刊而又复刊的出版物，不在本字段进行著录，休刊复刊情况在附注项说明。 |
| 内容形式 | 必备 | 著录资源的内容形式词及内容限定词。取值参见国标《信息资源的内容形式和媒体类型标识(GB/T 3469-2013)》。  例：正题名：漫画会  内容形式：文本;图像（静态） |
| 媒体类型 | 必备 | 著录资源的媒体类型。取值参见国标《信息资源的内容形式和媒体类型标识(GB/T 3469-2013)》。本项目中缺省取值为“电子”。 |
| 格式 | 必备 | 资源的格式。需用英文大写著录。  例：TIF;PDF |
| 语种 | 必备 | 著录资源的文字语种。采用3位语种代码，参考《新版中国机读目录格式使用手册》（附录A）。常见代码如下：  汉语：chi  英语：eng  藏语：tib  蒙古语：mon  维吾尔语：uig  朝语：kor |
| 关联 | 有则必备 | 描述与本资源相关的其他资源。这些关系包括丛编、副刊（附刊、特刊）、继承、替代、合并、分自、吸收、其他版本与原版本、其他格式与原格式等。相关资源一般独立存在。可著录资源名称或资源标识符。 |
| 时间范围 | 有则必备 | 著录资源内容涉及的时间特征，著录内容包括时代、日期、时间段。  例：正题名：中国古代史. 二, 宋辽金元明清  时间范围：宋代;辽金时代;元代;明代; 清代 |
| 空间范围 | 有则必备 | 著录资源内容涉及的空间特征，著录内容包括地点、地理坐标。  例：正题名：澳门历史研究  空间范围：澳门 |
| 适用对象 | 必备 | 资源适宜使用的人群，取值为：普通用户、少年儿童、残障人士。  例：正题名：时事画刊 : 《时事报告》小学生版  适用对象：少年儿童 |
| 出版频率 | 必备 | 著录期刊出版的时间间隔。如：月刊、每年3期、季刊、年刊、不定期出版、不详。 |
| 馆藏范围 | 必备 | 著录馆藏电子期刊的收录范围。  如果某一时间段内的电子资源全部存在，按时间段著录；如果存在缺失情况，则需分段著录，中间用半角分号分隔。 |
| ISSN | 有则必备 | 著录在国际标准连续出版物编号中确认资源的唯一标识。 |
| 统一刊号 | 有则必备 | 著录在国内统一刊号中确认资源的唯一标识。 |
| CDOI | 有则必备 | 著录在CDOI中确认资源的唯一标识。 |
| 发布地址 | 有则必备 | 资源在因特网或局域网的发布地址。 |
| 数据提交单位 | 必备 | 著录提交该数据的单位。 |
| 所属任务年份 | 必备 | 著录联建工作的任务年度。 |

# 2图像数字化标准和命名规则

## 2.1图像扫描加工

1. 扫描前根据国际色彩协会（International Color Consortium，简称ICC）标准，做加工设备的基本色彩校正，及针对各类型文献进行色彩校正；
2. 逐页扫描
3. 黑白页和灰度页用灰度方式扫描

色彩位深：8 位

分辨率：300 dpi；小于5号字体用400 dpi

档案典藏级格式：TIFF 不压缩

发布服务级：PDF（经过 JPEG2000 压缩后，再做格式转换）

1. 彩色页用彩色方式扫描

色彩位深：24 位

分辨率：300 dpi；小于5号字体用400 dpi

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）

1. 纠偏处理。对出现偏斜的图像进行纠偏处理，图像歪斜度不可以超过一度，对方向不正确的图像进行旋转还原，以符合阅读习惯；
2. 去污处理。对图像页面中出现的影响图像质量的杂质如黑边等进行去污处理；
3. 图像拼接。若原文献幅面较大，无法整体采集的，可将原件分画幅采集（分画幅采集时，注意各拍摄区域边缘须留有一定的重复采集区域，以保证拼接完成后所有接缝处连贯完整），服务级文件进行拼接处理，合并为一个完整的图像，以保证数字文件的整体性（合并信息应在readme.txt进行说明）。

## 2.2图像文件质量要求

1. 图像文件（各种格式）放大到1：1状态，逐页检查。检查文件是否有彩点、彩线、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。若不符合图像质量要求应进行图像校正或重新扫描；
2. 图像文件真实反映原件，同册期刊的图像尺寸相同，不得有失真现象；不能有缺页、错页、数据内容缺失等现象（原件有缺失的除外）；
3. 档案典藏级图像，允许进行适当的纠偏，去污等处理，以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，一页为一个PDF文件，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；
4. 图像符合扫描规格要求和技术参数；
5. 所有文件保存位置以及文件命名正确，可以有效打开和显示；
6. 图像扫描错误率不超过1‰。

## 2.3命名规则

### 2.3.1加工编号（book\_id）

文献数字化加工过程中一种文献的唯一标识，它由11位数字和1位下划线组成。

文献基本资料类型（1位）、文献语种（1位）、任务年（2位）、机构代码（4位）、下划线（1位）、单位内部流水号(3位)。

文献基本资料类型（1：期刊），文献语种（2：民语），任务年为公元年后两位数字（19年项目统一为19），机构代码见《图书馆机构代码表》（2019），单位内部流水号由各单位自行分配，从1开始，不足3位以0补齐。

如：加工对象为某一种期刊时其加工编号为：12190000\_001;

### 2.3.2图像文件名

同种期刊下的每一期期刊由4位数字的流水号表示，流水号各单位自行分配，从1开始，不足4位以0补齐。每期期刊中的图像文件名为3位序号，后缀为tif，不足3位以0补齐。如：《民国》期刊第6卷第1期，其加工编号为12190000\_001，同种内期刊流水号为0001，则其存储为:

\12180000\_001\0001\001.tif、002.tif、……

# 3数据库标准

## 3.1数据库命名

对象数据所对应数据库。数据库名由8位数字组成：

文献基本资料类型（1位）、文献语种（1位）、任务年（2位）、机构代码（4位）。

如：12190000.mdb

## 3.2数据库制作

**3.2.1** 对marc数据进行加工处理，制作文献基本信息表（book）。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 中文名称 | 字段名称 | 类型及长度 | 必备性 | 备注 |
| 1 | 加工编号 | book\_id | char |  | 12位加工编号 |
| 2 | 分类 | cat\_id | char | 必备 | 中图分类 |
| 3 | 书名（民语） | book\_name1 | char | 必备 |  |
| 4 | 书名（汉语） | book\_name2 | char | 必备 |  |
| 5 | 出版社 | pub\_house | char | 有则必备 |  |
| 6 | 出版时间 | pub\_date | char | 有则必备 |  |
| 7 | 001字段 | record\_id | char | 必备 |  |

**3.2.2**卷期信息表(volpis)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 中文名称 | 字段名称 | 类型及长度 | 必备性 | 备注 |
| 1 | 加工编号 | book\_id | char | 必备 |  |
| 2 | 期刊序号 | pserial\_num | char | 必备 | 期刊流水号 |
| 3 | 年 | pyear | char | 有则必备 | 期刊年信息，如1956 |
| 4 | 卷序号 | pvol | char | 有则必备 | 期刊卷信息。如第6卷则著录6 |
| 5 | 期 | pissue | char | 有则必备 | 期刊期信息。如第4期则著录4 |
| 6 | 期刊代码 | pclass | char | 必备 | 各种代码附后 |

其中期刊代码著录内容如下：n:普通刊；z：增刊；a：专刊；b：专辑；c：专集；d：专号；t：特刊；u：特辑；w：特集；s：试刊；h：合刊；f：副刊（复刊）；q：其它。

**3.2.3**文献缺页信息著录，生成文献缺页信息表(lostpage)。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 中文名称 | 字段名称 | 字符类型 | 必备性 | 备注 |
| 1 | 加工编号 | book\_id | char | 必备 |  |
| 2 | 期刊序号 | pserial\_num | char | 必备 | 期刊流水号 |
| 3 | 缺页前页号 | lost\_text\_page | 数字 | 必备 | 缺页前文件名数字部分 |
| 4 | 缺页数量 | lostpage\_num | 数字 | 必备 |  |

注：期刊有缺失页时，对该表字段进行标引。

**3.2.4**记录扫描分辨率、压缩因子、文件数量、存储量等信息，生成文献加工信息表(process)。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 中文名称 | 字段名称 | 字符类型 | 必备性 | 备注 |
| 1 | 加工编号 | book\_id | char | 必备 |  |
| 2 | 书名（汉语） | book\_name | char | 必备 |  |
| 3 | 扫描分辨率 | dpi | 数字 | 必备 |  |
| 4 | 压缩因子 | comp\_factor | char | 必备 |  |
| 5 | 灰度页数量 | grey\_num | 数字 | 必备 |  |
| 6 | 彩色页数量 | col\_num | 数字 | 必备 |  |
| 7 | TIFF数量 | tiff\_num | 数字 | 必备 |  |
| 8 | PDF数量 | pdf\_num | 数字 | 必备 |  |
| 9 | TIFF存储量 | tiff\_mb | 数字 | 必备 | 存储单位：MB |
| 10 | PDF存储量 | pdf\_mb | 数字 | 必备 | 存储单位：MB |
| 11 | TIFF硬盘位置 | hdA\_place | 数字 | 必备 | 硬盘号 |
| 12 | PDF硬盘位置 | hdB\_place | 数字 | 必备 | 硬盘号 |

**注：在制作数据库表时，数据库表使用英文名称，同时将以上4张表格中“字段名称”一栏内容作为著录字段，横向排列。**

## 2.3 数据库制作要求

数据库的制作符合下列要求。

（1）文件格式为mdb格式。

（2）使用字符集为UTF-8。

（3）著录信息应严格按照文献实际内容进行描述，所著内容与对象文件应正确链接。

（4）无法录入的生僻字等用“〓”表示。

（5）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。

（6）各种著录、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‰。

# 4说明文件

## 4.1制作文献的说明文件

说明文件格式：Microsoft Excel

说明文件命名：由8位组成：文献基本资料类型（1位）、文献语种（1位）、加工年（2位）、机构代码（4位）。如：12190000.xls

（说明文件的4张表需保存在同一个Excel文件的不同工作表中，每个工作表按下列表命名）。

### 4.1.1 数据总体说明表

|  |  |
| --- | --- |
| **名称** | **说明** |
| 项目名称 | 填写项目名称，如2019年地方少数民族语言期刊数字化 |
| 承建馆 | 填写承建馆名称全名 |
| 质检单位 | 填写项目第三方质检单位全名 |
| 项目时间 | 填写项目所属年份 |
| 数据加工编号 | 以每种期刊的起始～终止表示，如（12190000\_001）0001～0050 |
| 完成期刊文献种数 | 这批数据期刊的总种数 |
| 完成期刊文献期数 | 这批数据期刊的总期数 |
| 完成期刊文献页数 | 这批数据期刊的总页数 |
| 序号 | 以1开始，顺次编号 |
| 存储内容 | 如：扫描图像，PDF文件，第三方质检报告，版权证明等 |
| 数据格式 | TIFF，PDF等 |
| 存储介质 | 硬盘 |
| 介质编号 | 填写硬盘编号，如12190000 |
| 介质数量 | 填写存储硬盘的数量 |
| 存储量 | 以GB为单位表示 |
| 备注 | 获得版权证明的期刊种数 |

### 4.1.2保存级对象数据硬盘存储清单

|  |  |
| --- | --- |
| **名称** | **说明** |
| 数据内容 | 存储的数据内容，如2019年地方少数民族语言期刊数字化 |
| 数据格式 | 以大写英文字母表示，如TIFF |
| 序号 | 以1开始，顺次编号 |
| 硬盘编号 | 本硬盘编号 |
| 期刊加工编号（起始/终止卷期） | （12位加工编号）起始/终止卷期流水号  （12190000\_001）0001～0050 |
| 种数 | 本块硬盘存储的期刊总种数 |
| 期数 | 本块硬盘存储的期刊总期数 |
| 页数 | 期刊的总页数 |
| 存储量 | 以GB为单位表示 |
| 备注 |  |
| 合计 | 合计这批数据种数、期数、页数以及存储量 |

注：硬盘编号按递增顺序排列

### 4.1.3.发布级对象数据硬盘存储清单

|  |  |
| --- | --- |
| **名称** | **说明** |
| 数据内容 | 存储的数据内容，如2019年地方少数民族语言期刊数字化 |
| 数据格式 | 以大写英文字母表示，如PDF |
| 序号 | 以1开始，顺次编号 |
| 硬盘编号 | 本硬盘编号 |
| 期刊加工编号（起始/终止卷期） | （12位加工编号）起始/终止卷期流水号  （12190000\_001）0001～0050 |
| 种数 | 本块硬盘存储的期刊总种数 |
| 期数 | 本块硬盘存储的期刊总期数 |
| 页数 | 期刊的总页数 |
| 存储量 | 以GB为单位表示 |
| 备注 |  |
| 合计 | 合计这批数据种数、期数、页数以及存储量 |

注：硬盘编号按递增顺序排列

### 4.1.4每种文献数据量统计表

|  |  |
| --- | --- |
| **名称** | **说明** |
| 序号 | 以1开始，顺次编号 |
| 加工编号 | 12位加工编号 |
| 题名（民语） | 期刊名称（民语） |
| 题名（汉语） | 期刊名称（汉语） |
| 总期数 | 该种期刊的总期数 |
| 总页数 | 该种期刊的总页数 |
| 备注 |  |
| 合计 | 合计这批数据的题名数、总期数、总页数 |

## 4.2存储介质的说明文件（readme.txt）

一、文献类型：例如，少数民族语言期刊

二、制作时间：2019年 月

三、承建馆：

四、加工单位：

五、存储介质信息

1、介质名称：

2、加工起止号：

3、文献数量：

4、文件数量：

5、存储容量：

六、技术参数

1、存储格式：例如，不压缩TIFF格式

2、加工设备：

3、加工软件：

4、参数指标

扫描分辨率：例如，300dpi

扫描方式：

# 5存储规则

## 5.1存储介质及命名

存储介质为硬盘，硬盘命名由8位组成：

文献基本资料类型（1位）、文献语种（1位）、加工年（2位）、机构代码（4位）。

## 5.2 数据存储规则

存储内容包括：TIFF图像、PDF、对应数据库、说明文件、介质说明文件、文献的书目数据、版权证明、第三方质检报告各一份。

各类型对象数据按加工编号递增顺序依次放置在硬盘内，保存目录如下：

第一级目录为加工编号的前8位，如12190000，同时将介质说明文件(readme.txt)及第三方质检报告电子版并行保存；

第三方质检报告扫描时采用300DPI，以PDF格式保存，图像需进行拼接处理，合并为一个完整的图像，命名规则为“加工编号前8位+T”,如：12190000T.pdf

第二级目录为TIFF，PDF，同时将对应数据库（12190000.mdb）、说明文件（12190000.xls）、文献的书目数据(12190000.iso) 及版权证明电子版并行保存；版权证明扫描时采用300DPI，以PDF格式保存，图像需进行拼接处理，合并为一个完整的图像。文件夹命名为CERTIFICATE。

第三级目录TIFF，PDF为加工编号；CERTIFICATE文件夹内版权证明分以下两种情况命名：一是本年度加工期刊按种类出版权证明的，命名规则为“加工编号+C”,如12190000\_001C.pdf；二是本年度加工期刊总共出一份版权证明的，命名规则为“加工编号前8位+C”，如12190000C.pdf（下图所示为第一种情况）

第四级目录为4位期刊流水号；

第五级目录为对应的对象数据。

文件结构目录如下：

├─12190000、readme.txt、12190000T.pdf

│ ├──TIFF

│ │ ├─加工编号

│ │ │ ├─期刊流水号

│ │ │ │ ├001.tif

│ │ │ │ ├002.tif……

│ │ │ ├─期刊流水号……

│ │ ├─加工编号

│ │ │ ├─期刊流水号

│ │ │ ├─期刊流水号……

│ │ └─……

│ ├──PDF

│ │ ├─加工编号

│ │ └─加工编号……

│ ├──CERTIFICATE

│ │ ├─加工编号C.pdf

│ │ └─加工编号C.pdf

│ ├──12190000.iso

│ ├──12190000.mdb

│ └──12190000.xls

# 6.修订说明

6.1本年度新增第三方质检报告及版权证明电子版的提交工作，具体修订见5.2数据存储规则。