

CY

中华人民共和国新闻出版行业标准

CY/T XXX.4—XXXX

中文古籍数字出版规范
第4部分：版式采集

Standard for digital publishing of Chinese ancient books—Part 4:Layout extract

(点击此处添加与国际标准一致性程度的标识)

(征求意见稿)

(本草案完成时间：2025-11-04)

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX—XX—XX 发布

XXXX—XX—XX 实施

国家新闻出版署 发布

目 次

前言 II

引言 III

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 采集对象 2

5 采集范围 3

 5.1 版面采集 3

 5.2 版式信息采集 3

6 数据规格 3

 6.1 版面采集数据规格 3

 6.2 版式信息采集数据规格 3

7 采集流程 10

 7.1 基本流程 10

 7.2 准备工作 10

 7.3 版面采集 11

 7.4 图像处理 11

 7.5 版式信息采集 11

8 质量要求 11

 8.1 版面采集质量要求 11

 8.2 版式信息采集质量要求 11

附录 A （规范性） 中文古籍数字出版版式信息采集 XML Schema 样例 13

 A.1 中文古籍数字出版版式信息采集 XML Schema 13

附录 B （资料性） 中文古籍数字出版版式信息采集 XML 样例 15

 B.1 中文古籍数字出版版式信息采集 XML 样例 15

参考文献 16

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本文件是CY/T XXX《中文古籍数字出版规范》的第4部分。CY/T XXX已经发布了以下部分：

- 第1部分：术语；
- 第2部分：元数据；
- 第3部分：长期存储；
- 第4部分：版式采集；
- 第5部分：内容采集；
- 第6部分：版式重构；
- 第7部分：古籍数字加工与应用模式；
- 第8部分：古籍数据交换；
- 第9部分：数据加工管理。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由全国新闻出版标准化技术委员会（SAC/TC 527）归口。

本文件起草单位：

本文件主要起草人：

引 言

随着信息技术的进步，数字阅读已经普及，现如今，数字科技不断成熟，许多先进的技术被应用于内容资源数字化中，在新闻出版领域中图书、报纸、期刊、音频、视频的数字化和出版标准都已经发布并实施，而中文古籍数字出版标准尚处于缺失状态。为顺应数字化潮流、推动新时代古籍数字化工作高质量、创新性发展，更有效解决“藏”与“用”的问题，让中文古籍文献焕发新的生命力，使更多中文古籍文献得以深层次的挖掘和现代化的呈现，制定了CY/T XXX—XXXX《中文古籍数字出版规范》。依据中文古籍数字化生产过程，拟由9个部分组成。

——第1部分：术语。目的在于规范与中文古籍数字化相关的术语，统一相关概念，避免由于概念和术语不明确而造成的交流困难、歧义和误解。

——第2部分：元数据。目的在于规范中文古籍数字出版的元数据信息，便于理解数据的含义和用途，有助于提高数据的管理、组织、质量控制、存储、共享和安全保护的效率，为中文古籍元数据应用提供依据和指导。

——第3部分：长期保存。目的在于给出中文古籍长期存储数据的类型、存储原则、存储环境和存储备份策略的相关技术要求，为中文古籍数据长期保存提供依据和指导。

——第4部分：版式采集。目的在于规定中文古籍数字化加工中版式采集对象、采集范围和采集流程并对数据规格和质量要求提出技术要求，为中文古籍数字化版式采集提供依据和指导。

——第5部分：内容采集。目的在于规定中文古籍数字化加工中内容采集目标、采集范围和采集流程，并对文字采集、样式采集、结构采集提出技术要求，为中文古籍数字化内容采集提供依据和指导。

——第6部分：版式重构。目的在于给出中文古籍数字化加工中版式重构的部件元素组成、用字要求、描述文件要求和相应的质量要求，为中文古籍数字化版式重构提供依据和指导。

——第7部分：古籍数字化加工与应用模式。目的在于给出中文古籍数字化加工成品数据类型及规格要求，并描述了长期保存、古籍电子书、古籍资源库应用所需的成品数据类型，为中文古籍数字化加工应用提供依据和指导。

——第8部分：古籍数据交换。目的在于给出中文古籍数据交换类型、数据交换的基本要求和数据交换的接口要求，为中文古籍数据交换提供指导和帮助。

——第9部分：数据加工管理。目的在于给出中文古籍数字化加工的基本流程以及人员、环境、资料、数据存储、数据备份和数据交付的管理要求，为中文古籍数字化加工管理提供指导和帮助。

中文古籍数字出版规范

第4部分：版式采集

1 范围

本文件规定了中文古籍数字化加工中版式采集对象、采集范围、数据规格、流程和质量要求等内容。本文件适用于中文古籍数字化加工中版式信息的采集工作。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；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。

CY/T XXX. 3—XXXX 中文古籍数字出版规范 第3部分：长期存储

WH/T 22—2006 古籍特藏破损定级标准

WH/T 23—2006 古籍修复技术规范与质量要求

3 术语和定义

下列术语和定义适用于本文件。

3.1

版式采集 layout extract

对中文古籍版面信息及内容位置、布局等版式信息的采集。

3.2

校勘信息 text modify information

对页面内容中校勘的描述。

3.3

书脑 threading part

线装古籍打眼穿线部分。

3.4

板框 frame

边栏

书页正面图文四边的围栏。

[来源：WH/T 23—2006，2.5.1，有修改]

注：上方叫“上栏”，下方叫“下栏”，两旁叫“左右栏”。

3.5

天头 page head

图文或版框上方余幅，即版框外空白纸的上方。

[来源：WH/T 23—2006，2.5.4，有修改]

3.6

天头字 head word

位于天头的汉字。

3.7

地脚 foundation

下角

图文或版框下方余幅，即版框外空白纸的下方。

[来源: WH/T 23—2006, 2.5.5, 有修改]

3.8

书耳 book ears

版框外左上角的空格，里面刻有简单篇名，形状像耳朵。

示例：中文古籍的书耳位置如图 1 所示。

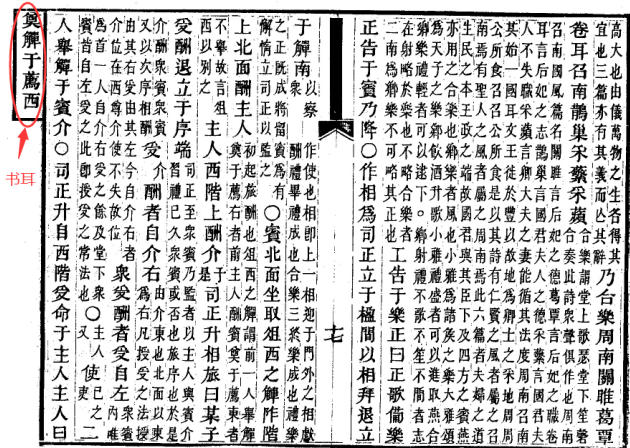


图1 中文古籍书耳位置示例

3.9

界行 bounded lattice

界格

在版面内分割行字的直线。

3.10

鱼尾 fishtail shape

版心中间的图形，用作标示中缝线，作为折叶的标准。

注：有时版心上下都有鱼尾。

3.11

象鼻 trunk line

鱼尾上下到版框的一条线。

3.12

版心 type area

中缝

一版印页的中心比较窄的一行空白行。

3.13

印章 seal

图章

印于古籍上表示鉴定或签署的文具，大体可分为名章、闲章两类。

3.14

书口 book mouth

版口

古籍装订后开合一侧的端面。

4 采集对象

版式采集对象主要包括。

- a) 中文古籍版式信息：记录古籍版式要素组成结构、各要素坐标位置和布局信息以及对应的图像对象数据链接等信息。

b) 图像对象数据：主要包括中文古籍中的插图、生僻字图和表格等。

5 采集范围

5.1 版面采集

采用扫描、拍照等数字化手段对古籍版面信息进行采集，按页生成古籍版面的数字化图像文件。图像文件应包括封面、书牌、封底、书名页、正文页、有板框的空白页、夹签、活页、附件等内容。

5.2 版式信息采集

对中文古籍版面内的天头、地脚、板框、版心、行款、界行、象鼻、鱼尾、印章、书耳等版式内容要素进行逐项分析，提取相应的布局和坐标信息等，标引其属性并生成XML文件。版式信息样式示意图见图2。

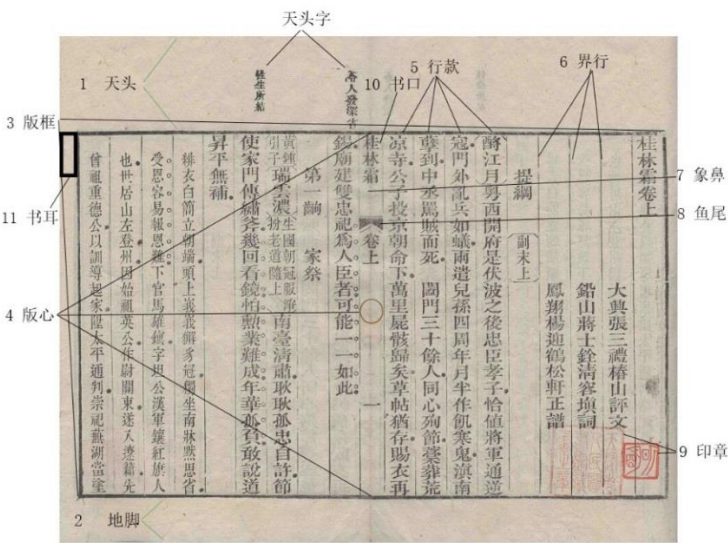


图2 版式信息样式示意图

6 数据规格

6.1 版面采集数据规格

版面采集生成的图像文件应内容完整、清晰。版面采集图像文件成品数据宜采用无损压缩TIFF格式存储，相关技术参数见表1。

表1 版面采集图像文件成品数据的技术参数

参数类型 格式	应用方向	压缩方式	分辨率(DPI)	色彩模式
TIFF	面向长期存储应用或加工应用	LZW	不低于300	彩色图像采用彩色模式、灰度图像采用灰度模式、黑白图像采用二值模式

6.2 版式信息采集数据规格

6.2.1 版式内容要素信息采集

版式内容要素信息按照图3所示结构进行采集。

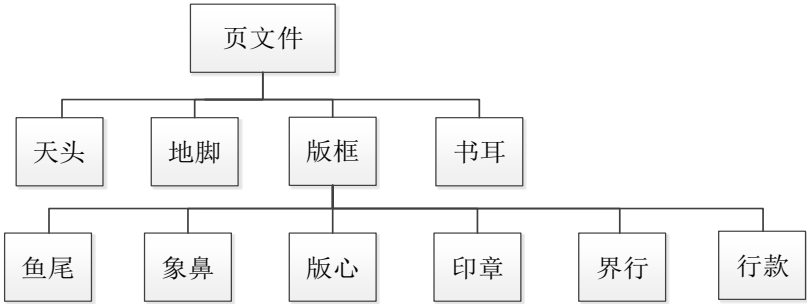


图3 版式内容要素信息结构

6.2.2 页文件

英文标签: <Page>
中文名称: 页文件
说 明: 古籍的独立筒子页。
结 构 图: 页文件元素结构模型见图4。

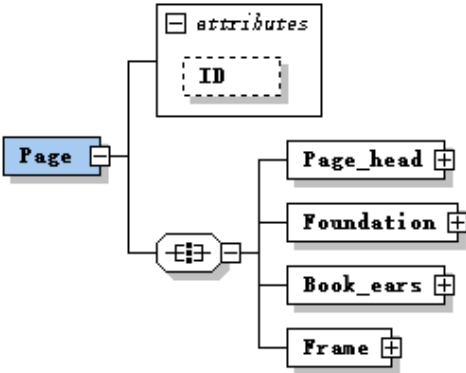


图4 页文件元素结构模型

子 元 素: <Page_head>, <Foundation>, <Book_ears>, <Frame>。
属 性: 页文件元素的属性见表2。

表2 页文件元素属性

中文名称	英文标签	说明	类型	必备性	重复性
标识符	ID	古籍筒子页的唯一标识符	String (255)	必备	不可重复

6.2.3 天头

英文标签: <Page_head>
中文名称: 天头
说 明: 版框外空白纸的上方。
结 构 图: 天头元素结构模型见图5。

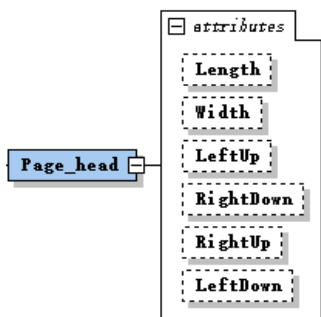


图5 天头元素结构模型

属 性：天头元素属性见表3。

表3 天头元素属性

中文名称	英文标签	说明	类型	必备性	重复性
长度	Length	天头区域长度	Int	可选	不可重复
宽度	Width	天头区域宽度	Int	可选	不可重复
区域左上坐标	LeftUp	古籍筒子页天头区域左上角坐标；(x,y)	Int	必备	不可重复
区域右下坐标	RightDown	古籍筒子页天头区域右下角坐标；(x,y)	Int	必备	不可重复
区域右上坐标	RightUp	古籍筒子页天头区域右上角坐标；(x,y)	Int	必备	不可重复
区域左下坐标	LeftDown	古籍筒子页天头区域左下角坐标；(x,y)	Int	必备	不可重复

6.2.4 地脚

英文标签：<Foundation>
中文名称：地脚
说 明：版框外空白纸的下方。
结 构 图：地脚元素结构模型见图6。

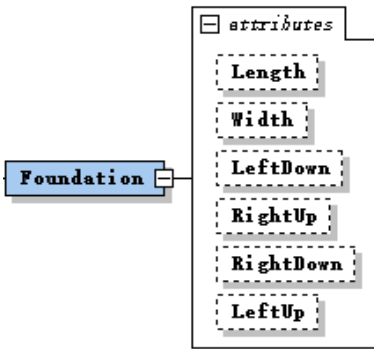


图6 地脚元素结构模型

属 性：地脚元素属性见表4。

表4 地脚元素属性

中文名称	英文标签	说明	类型	必备性	重复性
长度	Length	地脚区域长度	Int	可选	不可重复
宽度	Width	地脚区域宽度	Int	可选	不可重复
区域左上坐标	LeftUp	古籍筒子页地脚区域左上角坐标；(x,y)	Int	必备	不可重复
区域右下坐标	RightDown	古籍筒子页地脚区域右下角坐标；(x,y)	Int	必备	不可重复
区域右上坐标	RightUp	古籍筒子页地脚区域右上角坐标；(x,y)	Int	必备	不可重复
区域左下坐标	LeftDown	古籍筒子页地脚区域左下角坐标；(x,y)	Int	必备	不可重复

6.2.5 书耳

英文标签: <Book_ears>
中文名称: 书耳
说 明: 版框外左上角的空格, 里面刻有简单篇名。
结 构 图: 书耳元素结构模型见图7。

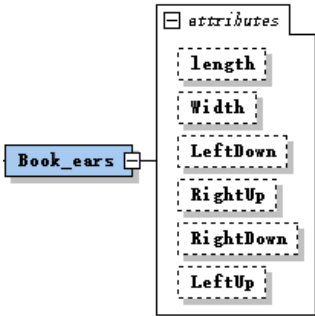


图7 书耳元素结构模型

属 性: 书耳元素属性见表5。

表5 书耳元素属性

中文名称	英文标签	说明	类型	必备性	重复性
长度	Length	书耳区域的长度	Int	可选	不可重复
宽度	Width	书耳区域的宽度	Int	可选	不可重复
区域左上坐标	LeftUp	书耳区域左上角坐标; (x,y)	Int	必备	不可重复
区域右下坐标	RightDown	书耳区域右下角坐标; (x,y)	Int	必备	不可重复
区域右上坐标	RightUp	书耳区域右上角坐标; (x,y)	Int	必备	不可重复
区域左下坐标	LeftDown	书耳区域左下角坐标; (x,y)	Int	必备	不可重复

6.2.6 版框

英文标签: <Frame>
中文名称: 版框
说 明: 书页正面图文四边的围栏。
结 构 图: 版框元素结构模型见图8。

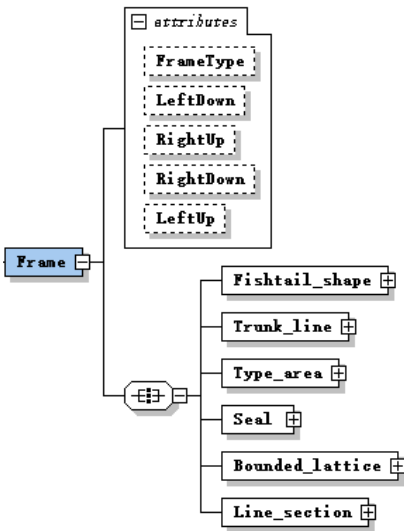


图8 版框元素结构模型

子元素：<Fishtail_shape>，<Trunk_line>，<Type_area>，<Seal>，<Bounded_lattice>，<Line_section>。

父元素：<Page>

属性：版框元素属性见表6。

表6 版框元素属性

中文名称	英文标签	说明	类型	必备性	重复性
版框类型	FrameType	用来描述版框的四周边类型，数值可选“四周单边 四周双边 左右双边”；三种类型的样式应单独定义	String	必备	不可重复
区域左上坐标	LeftUp	版框区域左上角坐标；(x,y)	Int	必备	不可重复
区域右下坐标	RightDown	板框区域右下角坐标；(x,y)	Int	必备	不可重复
区域右上坐标	RightUp	板框区域右上角坐标；(x,y)	Int	必备	不可重复
区域左下坐标	LeftDown	板框区域左下角坐标；(x,y)	Int	必备	不可重复

6.2.7 鱼尾

英文标签：<Fishtail_shape>

中文名称：鱼尾

说明：版心中间的图形“𩚑”，用作标示中缝线，作为折叶的标准。

结构图：鱼尾元素结构模型见图9。

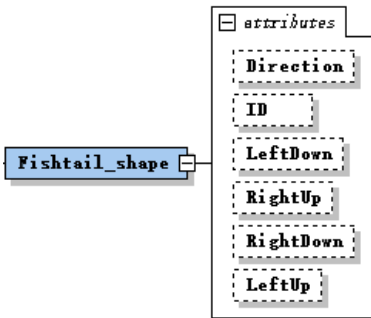


图9 鱼尾元素结构模型

属性：鱼尾元素属性见表7。

表7 鱼尾元素属性

中文名称	英文标签	说明	类型	必备性	重复性
方向	Direction	鱼尾的方向	String	必备	不可重复
编码	ID	鱼尾的页内编码	Int	必备	不可重复
区域左上坐标	LeftUp	鱼尾左上角坐标；(x,y)	Int	必备	不可重复
区域右下坐标	RightDown	鱼尾右下角坐标；(x,y)	Int	必备	不可重复
区域右上坐标	RightUp	鱼尾右上角坐标；(x,y)	Int	必备	不可重复
区域左下坐标	LeftDown	鱼尾左下角坐标；(x,y)	Int	必备	不可重复

6.2.8 象鼻

英文标签：<Trunk_line>

中文名称：象鼻

说明：鱼尾上下到板框的一条线。

结构图：象鼻元素结构模型见图10。

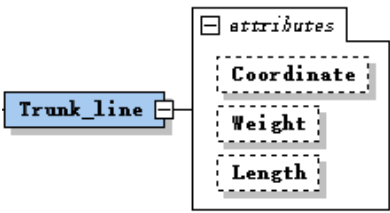


图10 象鼻元素结构模型

属性：象鼻元素属性见表8。

表8 象鼻元素属性

中文名称	英文标签	说明	类型	必备性	重复性
线宽	Weight	象鼻的线条宽度	Int	必备	不可重复
线长	Length	象鼻的线条长度	Int	必备	不可重复
象鼻一端坐标	Coordinate	象鼻一端的坐标；(x,y)	Int	必备	可重复

6.2.9 版心

英文标签：<Type_area>
中文名称：版心
说明：一版印页的中心比较窄的一行空白行，是书页对折时的标准。
结构图：版心元素结构模型见图11。

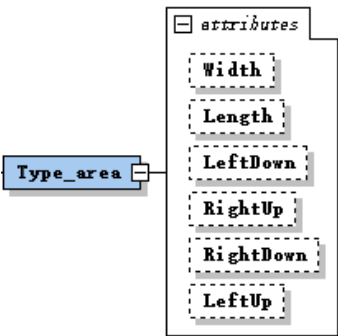


图11 版心元素结构模型

属性：版心元素属性见表9。

表9 版心元素属性

中文名称	英文标签	说明	类型	必备性	重复性
宽度	Width	版心区域的宽度	Int	可选	不可重复
长度	Length	版心区域的长度	Int	可选	不可重复
区域左上坐标	LeftUp	版心区域左上角坐标；(x,y)	Int	必备	不可重复
区域右下坐标	RightDown	版心区域右下角坐标；(x,y)	Int	必备	不可重复
区域右上坐标	RightUp	版心区域右上角坐标；(x,y)	Int	必备	不可重复
区域左下坐标	LeftDown	版心区域左下角坐标；(x,y)	Int	必备	不可重复

6.2.10 印章

英文标签：<Seal>
中文名称：印章
说明：印于古籍上表示鉴定或签署的图像。
结构图：印章元素结构模型见图12。

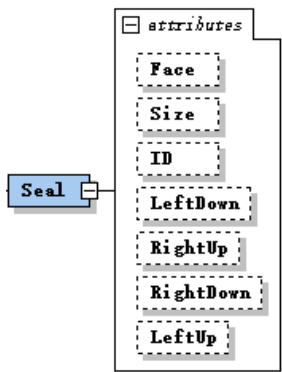


图12 印章元素结构模型

属 性：印章元素属性见表10。

表10 印章元素属性

中文名称	英文标签	说明	类型	必备性	重复性
字体	Face	印章内的字体	String	可选	不可重复
字号	Size	印章内字体的大小	Int	可选	不可重复
编码	ID	印章的页内编码	Int	必备	不可重复
区域左上坐标	LeftUp	印章区域左上角坐标；(x,y)	Int	必备	不可重复
区域右下坐标	RightDown	印章区域右下角坐标；(x,y)	Int	必备	不可重复
区域右上坐标	RightUp	印章区域右上角坐标；(x,y)	Int	必备	不可重复
区域左下坐标	LeftDown	印章区域左下角坐标；(x,y)	Int	必备	不可重复

6.2.11 界行

英文标签：<Bounded_lattice>
中文名称：界行
说 明：在版面内分割行字的直线。
结 构 图：界行元素结构模型见图13。

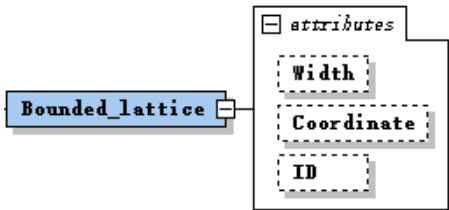


图13 界行元素结构模型

属 性：界行元素属性见表11。

表11 界行元素属性

中文名称	英文标签	说明	类型	必备性	重复性
界行一端坐标	Coordinate	界行一端的坐标位置	Int	必备	可重复
宽度	Width	界行的宽度	Int	必备	不可重复
编码	ID	界行的页内编码	Int	必备	不可重复

6.2.12 行款

英文标签：<Line_section>

中文名称：行款
说明：版面内筒子页半页的行数及字数。
结构图：行款元素结构模型见图14。

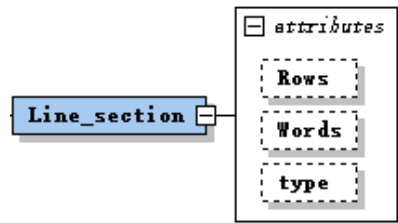


图14 行款元素结构模型

属性：行款元素属性见表12。

表12 行款元素属性

中文名称	英文标签	说明	类型	必备性	重复性
行类	type	筒子页的行的类别；单行或双行	String	必备	不可重复
行数	Rows	筒子页半页的行数	Int	必备	不可重复
字数	Words	筒子页半页的字数	Int	必备	不可重复

7 采集流程

7.1 基本流程

版式采集流程如图15所示。

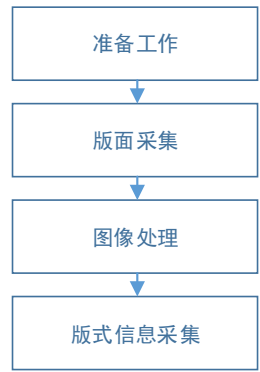


图15 版式采集流程

7.2 准备工作

- 准备工作包括。
- a) 中文古籍破损定级：依据标准 WH/T 22—2006 的规定对中文古籍进行破损定级。
 - b) 中文古籍修复：参考中文古籍破损等级要求，依据标准 WH/T 23—2006 的规定对中文古籍进行修复。
 - c) 书脑保护：在采集前将书的封面、封底用纸张进行包裹，然后用 3 个夹子夹住书脑，减少页面损伤。
 - d) 选择采集区域：在封面、封底与正文页面分别选择采集区域，且采集区域要略大于古籍的开本。

7.3 版面采集

中文古籍版面采集的图像以面为单位，宜在干燥、通风且恒温的环境中进行扫描或拍照，采集过程中应尽量使页面平整，确保页面端正、完整、字迹清晰不扭曲。

7.4 图像处理

图像处理包括以下几方面。

- a) 图像调整：包括去除图像噪点和黑边、调整图像倾斜度、调整页面顺序、删除无效页面文件等。
- b) 纠偏：纠正采集过程中歪斜的图像。
- c) 拼接：当中文古籍原始文献为筒子页时，需将采集的页面图像拼接合并为筒子页图像。
- d) 版心调整：调整页边距确保每面采集内容版心一致。
- e) 图像裁切：调整图像画布大小确保采集图像页面尺寸一致、大小准确。

7.5 版式信息采集

中文古籍数字出版版式信息采集XML Schema文件样例见附录A，中文古籍数字出版版式信息采集XML文件样例见附录B。

8 质量要求

8.1 版面采集质量要求

8.1.1 图像清晰度

图像清晰，版心居中，倾斜度控制在1度以内，清除图像噪点、黑边。

8.1.2 图像完整性

图像完整，文件页码连续，没有重页、缺页、错页等情况（原书缺页、错页除外）。

8.1.3 图像准确性

8.1.3.1 图像准确性要求

图像质量评定标准单位为1000采集图像页，图像差错率要求在千分之三以下。

8.1.3.2 图像差错统计方法

以采集图像页面为基本检测单位，页面中出现任何一种或几种错误均按1个差错计数，差错类型包括：

- a) 文件格式错误、图像分辨率错误、色彩模式错误、压缩算法错误等规格类错误；
- b) 图像尺寸与原版不一致；
- c) 颜色失真、图像过浓或过淡；
- d) 水平倾斜大于1度。

8.2 版式信息采集质量要求

8.2.1 完整性

8.2.1.1 类型完整

采集数据的类型应与版式信息采集的目标要求保持一致，不应出现缺漏和错误。

8.2.1.2 内容完整

采集数据的内容范围与数量应与版式信息采集的目标要求保持一致，不应出现缺漏和乱序等错误。

8.2.1.3 质量管理文档完整

完整的版式信息采集质量管理文档应包括：

- a) 质量检测方案；

- b) 质量检测报告。

8.2.2 规范性

8.2.2.1 数据格式

采集的数据格式应与版式信息采集目标保持一致，主要包括两类数据：

- a) 版面采集生成的图像数据，应使用无损压缩 TIFF 格式；
- b) 版式信息采集生成的中文古籍数字出版版式结构化信息，应采用 XML1.0 及以上版本文件格式，结构化规范描述文件应采用 XSD1.0 及以上版本。

8.2.2.2 数据文件命名

采集的数据文件命名应与版式信息采集目标保持一致，遵循CY/T XXX. 3—XXXX的规定。

8.2.2.3 数据存储

采集的数据文件存储应与版式信息采集目标保持一致，通常以古籍采集基本单位（卷）为存储文件夹，其下分类存储XML文件及对应的图像数据。

8.2.3 有效性

采集的数据应能通过相关软件及系统读出，不应出现数据损坏、异常报错、无法打开等错误。读出的数据应完整，不应出现编码混乱、图像失真等无法使用的错误。

8.2.4 准确性

8.2.4.1 准确性要求

采集的数据文件质量评定标准单位为1000采集图像页，差错率要求在千分之三以下。

8.2.4.2 差错统计方法

采集的数据与原版排版方式不一致，每出现一处按1个差错计数，差错主要为排版格式采集错误：天头、地脚、书耳、版框等版式图片大小、位置等信息采集错误。

附 录 A

(规范性)

中文古籍数字出版版式信息采集 XML Schema 样例

A.1 中文古籍数字出版版式信息采集 XML Schema

```

<?xml version="1.0" encoding="UTF-8"?>
<!-- edited with XMLSpy v2013 (http://www.altova.com) by () -->
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
xmlns:mml="http://www.w3.org/1998/Math/MathML"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xlink="http://www.w3.org/1999/xlink" elementFormDefault="qualified">
  <xs:element name="Page">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="Page_head">
          <xs:complexType>
            <xs:attribute name="Length"/>
            <xs:attribute name="Width"/>
          </xs:complexType>
        </xs:element>
        <xs:element name="Head_word">
          <xs:complexType>
            <xs:attribute name="Position"/>
            <xs:attribute name="Face"/>
            <xs:attribute name="Size"/>
            <xs:attribute name="ID"/>
          </xs:complexType>
        </xs:element>
        <xs:element name="Foundation">
          <xs:complexType>
            <xs:attribute name="Length"/>
            <xs:attribute name="Width"/>
          </xs:complexType>
        </xs:element>
        <xs:element name="Book_ears">
          <xs:complexType>
            <xs:attribute name="length"/>
            <xs:attribute name="Width"/>
            <xs:attribute name="Position"/>
            <xs:attribute name="Weight"/>
          </xs:complexType>
        </xs:element>
        <xs:element name="Frame">
          <xs:complexType>
            <xs:sequence>
              <xs:element name="Fishtail_shape">
                <xs:complexType>

```

```

        <xs:attribute name="Position"/>
        <xs:attribute name="Weight"/>
        <xs:attribute name="Direction"/>
        <xs:attribute name="Number"/>
        <xs:attribute name="ID"/>
    </xs:complexType>
</xs:element>
<xs:element name="Trunk_line">
    <xs:complexType>
        <xs:attribute name="Position"/>
        <xs:attribute name="Weight"/>
        <xs:attribute name="Length"/>
    </xs:complexType>
</xs:element>
<xs:element name="Type_area">
    <xs:complexType>
        <xs:attribute name="Position"/>
        <xs:attribute name="Width"/>
        <xs:attribute name="Length"/>
    </xs:complexType>
</xs:element>
<xs:element name="Seal">
    <xs:complexType>
        <xs:attribute name="Position"/>
        <xs:attribute name="Face"/>
        <xs:attribute name="Size"/>
        <xs:attribute name="ID"/>
    </xs:complexType>
</xs:element>
<xs:element name="Bounded_lattice">
    <xs:complexType>
        <xs:attribute name="Position"/>
        <xs:attribute name="Width"/>
    </xs:complexType>
</xs:element>
</xs:sequence>
<xs:attribute name="Position"/>
<xs:attribute name="Weight"/>
</xs:complexType>
</xs:element>
</xs:sequence>
</xs:complexType>
</xs:element>
</xs:schema>

```

附 录 B

(资料性)

中文古籍数字出版版式信息采集 XML 样例

B.1 中文古籍数字出版版式信息采集 XML 样例

```

<?xml version="1.0" encoding="UTF-8"?>
<!--Sample XML file generated by XMLSpy v2013 (http://www.altova.com)-->
<Page xsi:noNamespaceSchemaLocation="%e7%89%88%e5%bc%8f%e9%87%87%e9%9b%86schema.xsd"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <Page_head Length="长度" Width="宽度"/>
  <Head_word Position="位置" Face="字体" Size="字号" ID="编码">楫</Head_word>
  <Foundation Length="长度" Width="宽度"/>
  <Book_ears length="长度" Width="宽度" Position="位置" Weight="线宽"/>
  <Frame Position="位置" Weight="线宽">
    <Fishtail_shape Position="位置" Weight="线宽" Direction="方向" Number="数量"
ID="编码"/>
    <Trunk_line Position="位置" Weight="线宽" Length="线长"/>
    <Type_area Position="位置" Width="宽度" Length="长度"/>
    <Seal Position="位置" Face="字体" Size="字号" ID="编码"/>
    <Bounded_lattice Position="位置" Width="宽度" />
  </Frame>
</Page>

```

参 考 文 献

- [1] GB/T 38548.2—2020 内容资源数字化加工 第2部分：采集方法
 - [2] CY/T 101.4—2014 新闻出版内容资源加工规范 第4部分：数据加工质量
 - [3] CY/T 101.8—2014 新闻出版内容资源加工规范 第8部分：图书加工
-