

CY

# 中华人民共和国新闻出版行业标准

CY/T XXX.1—XXXX

## 中文古籍数字出版规范 第1部分：术语

Standard for digital publishing of Chinese ancient books—Part 1: Terminology

(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 通用术语 ..... 1

4 元数据术语 ..... 2

5 数据采集术语 ..... 3

    5.1 版式采集术语 ..... 3

    5.2 内容采集术语 ..... 3

6 长期存储术语 ..... 4

7 数据规格术语 ..... 4

    7.1 数据格式术语 ..... 4

    7.2 质量要求术语 ..... 5

8 数据加工术语 ..... 5

    8.1 资料管理术语 ..... 5

    8.2 版式重构术语 ..... 5

    8.3 存储备份术语 ..... 6

    8.4 数据交付术语 ..... 6

    8.5 数据服务术语 ..... 6

9 数据交换和应用术语 ..... 7

    9.1 数据交换术语 ..... 7

    9.2 数据应用术语 ..... 7

参考文献 ..... 8

索引 ..... 9

## 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1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1部分：术语

### 1 范围

本部分规定了中文古籍数字出版的相关术语和定义，主要包括通用术语、元数据术语、数据采集术语、长期存储术语、数据规格术语、数据加工术语、数据交换和应用术语等。

本部分适用于描述书写或印刷于1912年以前具有中国古典装帧形式的汉文书籍的数字出版。

### 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。

GB/T 5271.4—2000 信息技术词汇 第4部分：数据的组织

CY/T 50—2008 出版术语

CY/T 96—2013 电子书内容术语

CY/T 101.1—2014 新闻出版内容资源加工规范 第1部分：加工专业术语

CY/T 102.2—2014 数字内容对象存储、复用与交换规范 第2部分：对象封装、存储与交换

WH/T 20—2006 古籍定级标准

### 3 通用术语

#### 3.1

**中文古籍** ancient Chinese book

书写或印刷于1912年以前具有中国古典装帧形式的汉文书籍。

#### 3.2

**筒子页** package page

一页书折叠后装订在一起形成的相连接的正反两页。

#### 3.3

**版面** layout

原指按一定行款字数并以边栏、界行、版心、鱼尾等形象为装饰而雕刻或排拣活字形成的平面。后世即指一版一版印纸。

注：参考《古籍基本知识》

#### 3.4

**板框** frame

书页正面图文四边的围栏，一般指印刷的书。也称边栏，上方叫“上栏”，下方叫“下栏”，两旁叫“左右栏”。

#### 3.5

**界行** bounded lattice

**界格**

在版面内分割行字的直线。

#### 3.6

**行款** form and arrangement of lines in calligraphy

**行格**

正文之行数与字数。通常以半页计算，叫做“半页×行×字”，有的直接称“×行×字”，若一行中有两排字（通常是为大字正文做注解），叫做“小字双行×字”。

### 3.7

**天头** page head

图文或板框上方余幅，即版框外空白纸的上方。

### 3.8

**地脚** foundation

**下角**

图文或板框下方余幅，即版框外空白纸的下方。

### 3.9

**版心** type area

**中缝**

通常指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。

### 3.10

**鱼尾** fish tail

版心中间图形，刻于距版心上边大约四分之一处，有时版心上下都有鱼尾。

### 3.11

**象鼻** trunk line

鱼尾上下到板框的一条线。

## 4 元数据术语

### 4.1

**都柏林核心元数据** Dublin Core Metadata; DC

一种根据各类信息资源的共同特点确定的，适用于不同领域间信息资源、特别是网络信息资源的元数据。

### 4.2

**著录规则** description rules

根据一定目的和使用对象，对特定资源形式和内容特征进行描述时所需的具体操作规则。

### 4.3

**版本** version

一书经过抄写或印刷而形成的传本。

[来源：WH/T 20—2006，2.2，有修改]

示例1：书写或印刷的各种形式。

示例2：一书在流传过程中卷帙的存佚。

注：书籍具有的特征。

### 4.4

**丛编** series of books

在内容和形式，或者典藏方式上具有一定联系，并有独立题名的古籍集合。

注：参考《国家图书馆古籍元数据规范与著录规则》

### 4.5

**子目** subheading

古籍丛编的单册。

注：参考《国家图书馆古籍元数据规范与著录规则》

### 4.6

**抄本** transcript

以某一传本为底本，抄写而成的书本。

[来源：WH/T 20—2006，2.2.3]

注：习惯上对元及元以后抄写的书本称为抄本。

## 4.7

**影抄本** facsimile edition

**影写本**

以某一传本为底本，按照底本文字的行款格式、板框大小、文字内容摹抄的与底本版面形象相同的书本。

[来源：WH/T 20—2006，2.2.4]

## 4.8

**影刻本** shadow edition

按照某一底本原样摹刻印制的书本。

[来源：WH/T 20—2006，2.2.6.4]

## 4.9

**影印本** photocopy

以某一版本为底本，用照相的方法制成印版，上机印刷而印成的书本。

[来源：WH/T 20—2006，2.2.13]

## 4.10

**附录** appendix

古籍正文之后的一些附加性文字内容。

## 5 数据采集术语

## 5.1 版式采集术语

## 5.1.1

**版式采集** layout extract

对古籍版面内容的位置布局等信息的采集。

## 5.1.2

**校勘信息** text modify information

对页面内容中校勘的描述。

## 5.1.3

**天头字** head word

位于天头的汉字。

## 5.1.4

**书耳** book ears

板框外左上角的空格，里面刻有简单篇名，形状像耳朵。

## 5.1.5

**文字块** text block

同一页面中，由若干文字行组成的一段文字。

## 5.1.6

**印章** seal

印于古籍上表示鉴定或签署的图像。

## 5.1.7

**书口** book mouth

**版口**

古籍装订后开合一侧的端面。

## 5.2 内容采集术语

## 5.2.1

**外字** character outside the set

在指定集合范围内无法认同的字符。

## 5.2.2

**简洁文字串** simplified text string

对页面中版式特殊的文字内容的一种表述方式，这种内容因版式特殊而难以实现单字采集，仅录入其文字，并以字符串格式处理。

示例：中文古籍中对注文的注释

## 5.2.3

**段落结束标记** paragraph end symbol

添加在行尾字符后的一种标记，用于表示段落的结束。

## 5.2.4

**文字段** text paragraph

一行中版式信息相同的连续字符。版式信息包括文字大小、文字所在行、文字方向等。

## 5.2.5

**文字行** text row

同一行的所有文字段的总和。

## 5.2.6

**小注** small annotation

注释。

## 5.2.7

**残缺文字** incomplete text

因页面残缺而造成的不完整的文字。

## 5.2.8

**模糊文字** fuzzy literal

难以识别的完整文字。

## 5.2.9

**字符集** character set

不同字符的一个有限集合。

[来源：GB/T 5271.4—2000，04.01.02]

示例：GB/T 1988 字符集的国际基准版。

## 6 长期存储术语

## 6.1

**长期存储级数字资源** long-term conservation digital resource

作为档案存储及出版用的数字资源，不用于发布服务，可作为格式转换和复制的母本。

## 6.2

**分辨率** Dots Per Inch; DPI

图像每英寸长度内的像素点数。

## 6.3

**无损压缩** lossless compression

**无失真压缩**

利用数据的冗余进行压缩，可完全恢复原始数据而不引起任何失真。

## 7 数据规格术语

### 7.1 数据格式术语

## 7.1.1

**头文件** head

对一种文献整体描述的XML文件。

## 7.1.2

**版式文件** fixed-layout document



**版式文档**

排版后生成的，包含版面固化呈现需要的全部数据的一种文件。

[来源：CY/T 96—2013，6.1.5]

## 7.1.3

**流式文件** reflowing document

**流式文档**

按照内容逻辑顺序，内容呈现可适应终端设备屏幕或窗口变化的一种文件。

[来源：CY/T 96—2013，6.1.6]

## 7.1.4

**双层矢量化版式文件** vectorized dual-layer fixed-layout document

**双层版式文件**

在单层图像版式文件基础上，同时生成与图像层对应的透明字体模式的文字层，可支持选取、拷贝和查找的版式文件。

[来源：CY/T 101.1—2014，6.3.13]

## 7.2 质量要求术语

## 7.2.1

**差错** error

算出的、观察的或测量的值或状况，与真的、规定的或理论上正确的值或状况之间的差异。

## 8 数据加工术语

## 8.1 资料管理术语

## 8.1.1

**资料** materials

中文古籍原件、扫描文件的统称。

## 8.1.2

**资料整理** processing of materials

对书写或印刷的中文古籍原件、扫描文件等进行分类、核对、检查、登记的过程。

## 8.1.3

**资料保管** material custody

维护加工对象资料完整与安全的保存活动。

## 8.1.4

**附件** annexed matter; accompanying material

内容与出版物主体部分有直接联系，在装订或其他组装方式上与主体部分相分离的附属资料。

[来源：CY/T 50—2008，5.103]

## 8.1.5

**损坏** damage

由于外部或内部作用而导致加工原件物理状况的有害变化。

[来源：CY/T 101.1—2014，4.4.8]

## 8.1.6

**折角** dog-eared

由于故意行为或损耗导致的一页的角被折转。

[来源：CY/T 101.1—2014，4.4.9]

## 8.2 版式重构术语

## 8.2.1

**版式重构** layout rebuild

按照原版面结构、内容和排版信息，以数字化和结构化的形式描述并呈现原版式。

## 8.2.2

**页文件** page

由筒子页扫描后形成的两个图像拼接而成的文件。

## 8.2.3

**文本行** text\_line

**文本区**

行文本在原图像中对应的矩形区域。

## 8.2.4

**点校** punctuate and proofread

**校点**

对古籍进行标点、校订。

## 8.2.5

**注释** annotation

古籍中解释词义、语法、修辞及典故，表达文意及思想，标注与分析音读，考证人名与地名，以及校正文字等部分内容。

## 8.2.6

**内文** text page

内文是去除标题、注释、图题、表题等的剩余正文内容。

## 8.3 存储备份术语

## 8.3.1

**在线存储** online storage

存储设备和所存储的数据保持“在线”状态，可随时读取、修改。

[来源：CY/T 101.1—2014，6.4.5]

## 8.3.2

**离线存储** offline storage

**备份级存储**

对在线存储数据进行备份。

[来源：CY/T 101.1—2014，6.4.6]

## 8.4 数据交付术语

## 8.4.1

**数据文件** data file

包含资源（如数字内容）数字信息的文件。

[来源：CY/T 102.2—2014，3.4]

## 8.4.2

**数据交付** data delivery

古籍数字加工后的成品数据和相关清单文件的提交。

## 8.5 数据服务术语

## 8.5.1

**委托方** distributor

委托数据加工处理服务的单位或个人。

[来源：CY/T 101.1—2014，7.2.1]

## 8.5.2

**加工方** processing vendor

承接数据加工处理服务的单位或个人。

[来源：CY/T 101.1—2014，7.2.2]

## 9 数据交换和应用术语

### 9.1 数据交换术语

#### 9.1.1

##### 成品数据 finished data

已经完成全部数据处理过程且符合预先设定的规格和质量要求，可实现交付的最终数据。

#### 9.1.2

##### 古籍数据交换 data exchange

图书馆、博物馆、研究机构、企业等单位之间，根据自身数字资源规划需要，利用资源交换的方式，实现古籍数字资源的共建共享。

### 9.2 数据应用术语

#### 9.2.1

##### 内容加工 content processing

将大量分散、零乱、无序的内容资源进行整理、提炼，并按照一定的逻辑顺序和科学体系加以编排使之系统化的活动。

[来源：CY/T 101.1—2014，5.1.3]

#### 9.2.2

##### 数字化加工 digital processing

##### 数字化处理

针对纸质出版物、表单、公文、档案文献等，通过专业化的解决方案，加工成PDF、HTML、TXT、Doc、Xls等多种可编辑的电子文件格式的活动。

[来源：CY/T 96—2013，5.2.21]

#### 9.2.3

##### 电子书 ebook

通过相关设备直接呈现文字、图像、音频、视频等内容的数字出版产品。

[来源：CY/T 96—2013，4.2.4]

注：电子书包括电子图书、电子期刊和电子报等。

### 参 考 文 献

- [1] 刘俊熙. 传统文献编目·元数据·都柏林核心集理论探讨[J]. 现代图书情报技术, 2001, 1:10
- [2] 国家图书馆古籍元数据规范与著录规则

索 引

板框 ..... 1

版本 ..... 2

版面 ..... 1

版式采集 ..... 3

版式文件 ..... 4

版式重构 ..... 5

版心 ..... 2

残缺文字 ..... 4

差错 ..... 5

抄本 ..... 2

成品数据 ..... 7

丛编 ..... 2

地脚 ..... 2

点校 ..... 6

电子书..... 7

都柏林核心元数据..... 2

段落结束标记 ..... 4

分辨率..... 4

附件 ..... 5

附录 ..... 3

古籍数据交换..... 7

加工方..... 6

简洁文字串..... 4

界行 ..... 1

离线存储 ..... 6

流式文件 ..... 5

模糊文字 ..... 4

内容加工 ..... 7

内文 ..... 6

书耳 ..... 3

书口 .....	3
数据交付 .....	6
数据文件 .....	6
数字化加工 .....	7
双层矢量化版式文件 .....	5
损坏 .....	5
天头 .....	2
天头字 .....	3
筒子页 .....	1
头文件 .....	4
外字 .....	3
委托方 .....	6
文本行 .....	6
文字段 .....	4
文字块 .....	3
文字行 .....	4
无损压缩 .....	4
象鼻 .....	2
小注 .....	4
校勘信息 .....	3
行款 .....	1
页文件 .....	6
印章 .....	3
影抄本 .....	3
影刻本 .....	3
影印本 .....	3
鱼尾 .....	2
在线存储 .....	6
长期存储级数字资源 .....	4
折角 .....	5
中文古籍 .....	1
注释 .....	6
著录规则 .....	2

资料 ..... 5

资料保管 ..... 5

资料整理 ..... 5

子目 ..... 2

字符集..... 4

\_\_\_\_\_