

CY

中华人民共和国新闻出版行业标准

CY/T XXX. 7—XXXX

中文古籍数字出版规范 第7部分：古籍数字化加工与应用模式

Standard for digital publishing of Chinese ancient books—Part 7: Model about digitalization processing and application of ancient books

(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 中文古籍数字化加工成品数据 2

 4.1 图像文件 2

 4.2 版式文件 2

 4.3 内容文件 2

5 应用模式 3

 5.1 长期保存 3

 5.2 古籍电子书 3

 5.3 古籍资源库 3

6 中文古籍数字化加工成品数据与应用模式的关系 3

参考文献 4

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7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7部分：古籍数字化加工与应用模式

1 范围

本部分规定了中文古籍数字化加工的成品数据类型及应用模式。
本部分适用于中文古籍数字出版的数字化加工及应用。

2 规范性引用文件

本文件没有规范性引用文件。

3 术语和定义

下列术语和定义适用于本文件。

3.1

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。

3.2

单层矢量化版式文件 vectorized fixed-layout document

矢量版式文件

按照原加工对象的文字位置，文字采用矢量字，修饰性图片、艺术字、底纹、线条、图表和公式等以图像形式显示的版式文件。

[来源：CY/T 101.1—2014，6.3.14]

3.3

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]

3.4

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]

3.5

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]

3.6

按需印刷 POD

按照用户的要求，依指定的地点和时间予以提供为目的，直接将所需资料的文件数据进行数码印刷、装订。

3.7

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。

注：电子书包括电子图书、电子期刊和电子报等。

[来源：CY/T 96—2013，4.2.4]

4 中文古籍数字化加工成品数据

4.1 图像文件

中文古籍数字化加工成品数据中图像文件规格要求如下：

- a) 文件格式：采用无损压缩 TIFF 格式；
- b) 分辨率：不低于 300DPI；
- c) 色彩模式：选择适合图像的色彩模式，如 RGB、CMYK 等；
- d) 图像尺寸：保留图像的原始尺寸。

4.2 版式文件

4.2.1 用于长期保存

中文古籍数字化加工成品数据中用于长期保存的PDF文件应由图像文件自动生成，其规格要求如下：

- a) 文件格式：采用 PDF/A 格式；
- b) 版本：采用 PDF1.4 及以上版本；
- c) 图像要求：见本部分 4.1。

4.2.2 用于发布应用

4.2.2.1 图像 PDF

中文古籍数字化加工成品数据中用于发布应用的图片PDF文件规格要求如下：

- a) 文件格式：采用 PDF/A 格式；
- b) 版本：采用 PDF1.4 及以上版本；
- c) 图像要求：采用有损压缩格式 JPEG 或 JPEG 2000 格式，RGB 色彩空间，分辨率不高于 150DPI。

4.2.2.2 版式重构 PDF

中文古籍数字化加工成品数据中用于发布应用的版式重构PDF文件规格要求如下：

- a) 文件格式：采用 PDF/A 格式；
- b) 版本：采用 PDF1.4 及以上版本；
- c) 页面尺寸：与原版古籍尺寸保持一致；
- d) 文字：选取近似字体并遵照古籍原版内容还原文字的字号、相对位置、样式等信息；
- e) 图表：采用图像插入方式，遵照古籍原版还原；
- f) 矢量图形：古籍版式重构 PDF 文件中的版式信息内容要素应使用矢量线条进行绘制还原，还原的矢量图形效果应与原版古籍版式近似一致。

4.3 内容文件

4.3.1 XML 文件

中文古籍数字化加工成品数据中用于发布应用的XML文件内容应包括：描述元数据、管理元数据、目次结构化数据、全文结构化数据等。

4.3.2 Epub 文件

中文古籍数字化加工成品数据中用于发布应用的Epub文件要求如下：

- a) Epub 文件应符合相应技术规范；
- b) Epub 文件中内容应与古籍原版内容一致；
- c) Epub 文件必须拥有清晰有效的导航结构和目录；
- d) Epub 文件中的图像和多媒体元素应该与源文件一致，为保证快速下载和加载应进行必要压缩；

e) Epub 文件应具有良好的可用性，包括字体、字号、行距、段落间距、链接等信息有效、准确。

5 应用模式

5.1 长期保存

中文古籍数字化加工成品数据中的图像文件及用于长期保存的版式文件可支持中文古籍资源的长期保存、索引调阅应用、按需印刷等，还可作为数据基础通过进一步数字化加工处理以满足未来对古籍资源不同的应用需求。

5.2 古籍电子书

中文古籍数字化加工成品数据中的用于发布应用的版式文件或内容文件中的Epub文件可支持以版式呈现或内容浏览为主的古籍电子书应用。

5.3 古籍资源库

中文古籍数字化加工成品数据中的用于发布应用的版式文件和内容文件可支持内容分类、内容检索、原版阅读、全文阅读、对照阅读和知识组织等古籍资源库应用。

6 中文古籍数字化加工成品数据与应用模式的关系

中文古籍数字化加工成品数据与应用模式关系见表1。

表1 中文古籍数字化加工成品数据与应用模式关系对应表

成品数据 \ 应用模式		长期保存	发布应用	
			古籍电子书	古籍资源库
图像文件		▲	-	-
版式文件	用于长期保存	△	-	-
	用于发布服务	-		
内容文件	Epub 文件	-	△	△
	XML 文件	-	-	▲
注：“▲”为必选；“ ”为默认选择；“△”为可选；“-”为无。				

参 考 文 献

- [1] GB/T 5271.4—2000 信息技术词汇 第4部分：数据的组织
 - [2] GB/T 11457—2006 信息技术 软件工程术语
 - [3] GB/T 25100—2010 信息与文献 都柏林核心元数据元素集
 - [4] CY/T 50—2008 出版术语
 - [5] CY/T 96—2013 电子书内容术语
 - [6] 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
 - [7] CY/T 101.2—2014 新闻出版内容资源加工规范 第2部分：数据加工与应用模式
 - [8] 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
 - [9] GC/FH 12—2016 国家数字复合出版系统工程标准 复合文档基础结构
-