

CY

中华人民共和国新闻出版行业标准

CY/T XXX.8—XXXX

中文古籍数字出版规范
第8部分：古籍数据交换

Standard for digital publishing of Chinese ancient books—Part 8: Data exchange

(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 数据交换类型 1

 4.1 元数据交换 1

 4.2 图像数据交换 1

 4.3 内容数据交换 1

5 数据交换基本要求 1

 5.1 完整性要求 1

 5.2 有效性要求 2

 5.3 使用权限 2

6 数据交换接口要求 2

参考文献 3

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8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 8 部分：古籍数据交换

1 范围

本部分规定了中文古籍数据的交换类型，以及数据完整性、有效性、使用权限及交换接口等方面的要求。

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。

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

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形态。

[来源：CY/T 101.1—2014，7.1.1]

4 数据交换类型

4.1 元数据交换

古籍的描述元数据、管理元数据以及扩展元数据应以XML格式文件进行交换。交换元数据时，应确定元数据包含的内容范围。

4.2 图像数据交换

图像数据宜以TIFF、RAW、JPEG、JPG、GIF等格式文件进行交换，具体有以下2种交换方式：

- 交换插图、表格、生僻字及特殊符号等对应的图片；
- 交换完整版面图像文件。

4.3 内容数据交换

内容数据交换包括以下3种方式：

- 交换含版式信息和全文内容的数据，通常为XML格式；
- 交换PDF版式文件；
- 交换Epub流式文件。

5 数据交换基本要求

5.1 完整性要求

为了保证交换数据的完整性，交换数据应符合以下要求：

- XML文件应完整包含古籍版式信息和全文内容；
- 版式PDF文件应无缺页、漏页等情况；
- Epub文件中的内容应与古籍原版一致，无内容缺失。

5.2 有效性要求

交换数据应能通过相应软件及系统读出，不应出现数据损坏、异常报错、无法打开等错误。读出数据应完整，不允许出现编码混乱、图像失真等无法使用的错误。

5.3 使用权限

根据约定和原则对交换的数据使用权限进行限制，通常包括以下限制：

- a) 是否允许保存或打印；
- b) 是否允许二次传播；
- c) 其他约定。

6 数据交换接口要求

中文古籍数字出版中的数据交换接口要求如下：

- a) 通用性：接口开发应遵循通用的技术规范，确保兼容性和互操作性；
- b) 确定性：应明确交换数据的数据类型、格式、使用权限等信息，确保信息地传递和解析准确；
- c) 安全性：数据传输技术应安全可靠，确保数据不丢失、被篡改或泄漏；
- d) 可管理性：可追踪和管理数据传输过程，确保数据交换一致、准确；
- e) 易用性：接口调用简洁高效，便于应用和集成；
- f) 可扩展性：可支持不同类型和版本数据的扩展。

参 考 文 献

- [1] GB/T 25100—2010 信息与文献 都柏林核心元数据元素集
 - [2] CY/T 96—2013 电子书内容术语 新闻出版行业标准
 - [3] 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
 - [4] 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
 - [5] CY/T 101.6—2014 新闻出版内容资源加工规范 第6部分：数据管理
 - [6] CY/T 102.2—2014 数字对象内容存储、复用与交换规范 第2部分：对象封装、存储与交换
 - [7] CY/T XXX.3—XXXX 中文古籍数字出版规范 第3部分：长期存储
 - [8] WH/T 20—2006 古籍定级标准
 - [9] 胡艳杰. 古籍数字化资源的交换及其意义初探[J]. 图书馆工作与研究, 2014(226):46.
-