

CY

中华人民共和国新闻出版行业标准

CY/T XXX.9—XXXX

中文古籍数字出版规范
第9部分：数据加工管理

Standard for digital publishing of Chinese ancient books—Part 9: Data processing
management

(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 数据加工基本流程 2

5 人员管理 2

 5.1 资质管理 2

 5.2 权限管理 3

 5.3 保密管理 3

 5.4 工作规范管理 3

6 环境管理 3

 6.1 图像获取环境 3

 6.2 古籍存放环境 3

7 资料管理 4

 7.1 资料整理和移交 4

 7.2 资料使用 5

8 数据存储和备份管理 6

 8.1 基础数据存储和备份管理 6

 8.2 过程数据存储和备份管理 6

 8.3 成品数据存储和备份管理 6

 8.4 数据长期保存 6

 8.5 数据核销 6

9 交付管理 6

附录 A（资料性）质量检测报告样例 8

附录 B（资料性）资料清单样例 9

 B.1 资料登记清单 9

 B.2 资料移交清单 9

 B.3 资料使用流转单 10

参考文献 11

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9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9部分：数据加工管理

1 范围

本部分规定了中文古籍数据加工的基本流程，以及人员、环境、资料、数据存储和备份、交付等管理要求。

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

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 18883—2022 室内空气质量标准

GB/T 18204.5—2025 公共场所卫生检验方法 第5部分：集中空调通风系统

3 术语和定义

下列术语和定义适用于本文件。

3.1

资料 materials

中文古籍原件、扫描文件的统称。

[来源：CY/T 101.1—2014，4.1.2]

3.2

资料整理 processing of materials

对中文古籍原件、扫描文件等进行分类、核对、检查、登记的过程。

3.3

资料保管 material custody

维护加工对象资料完整与安全的保存活动。

3.4

损坏 damage

由于外部或内部作用而导致加工原件物理状况有害的变化。

[来源：CY/T 101.1—2014，4.4.8]

3.5

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]

3.6

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]

3.7

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]

3.8

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]

3.9

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]

3.10

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]

3.11

温度日较差 diurnal variation temperature

温度气象要素一昼夜之间最高值与最低值之差。

3.12

相对湿度日较差 diurnal variation of relative humidity

相对湿度气象要素一昼夜之间的最高值与最低值之差。

4 数据加工基本流程

中文古籍数字出版数据加工的基本流程包括以下几个部分。

- a) 古籍接收：待加工古籍的接收、上架和登记或待加工古籍的原始图像文件的接收、存储和登记等。
- b) 资料整理：对中文古籍原件、原始图像文件等进行分类、核对、检查、登记。
- c) 图像获取：包括资料领取和核查、平板扫描/非接触扫描/拍照、完整性检查、图像资源上传、古籍资料归还等流程。
- d) 图像处理：对图像文件进行拼接、纠偏、版心调整、图像裁切、灰阶补偿、去除黑边、图像去噪等处理，使图像满足长期保存和后续版式和内容采集等工作的需要。
- e) 元数据著录：对古籍的描述元数据和管理元数据分别进行著录。
- f) 版式采集：对古籍采集的图像中的天头、地脚、板框、版心、行款、界行、象鼻、鱼尾、印章、书耳等版式内容进行分析、标引并提取相应的布局和坐标等信息。
- g) 内容采集：使用智能版面分析、文字识别和校对等技术对古籍的文字、图表、样式和结构的数字化信息进行采集。
- h) 版式重构：使用版面还原技术将采集的古籍版式信息、文字、图表等元素以矢量化形式进行还原。
- i) 成品数据交付、验收：整合古籍数字化加工后的版式 PDF 文件、结构化 XML 文件、流式 Epub 文件等成品数据，按照质量和检测要求进行数据验收检测，并填写质量检测报告，报告样例见附录 A。

5 人员管理

5.1 资质管理

古籍数据加工人员的资质管理包括以下3个方面：

- a) 担任古籍数字加工工作的人员应具有一定古籍、版本学、目录学相关知识和工作经验；
- b) 古籍数字加工人员应在接受专门培训后，方可从事古籍数字加工工作；
- c) 应对数据处理人员的熟练程度、错误类型敏感程度、操作准确性、工作责任心等进行考核。

5.2 权限管理

对古籍数据加工人员的权限管理包括以下2个方面：

- a) 一般不允许加工人员带入、使用移动设备；
- b) 非正常原因需要拷贝数据，需要到管理员处申请备案，由管理员提供移动设备，在管理员监督下完成操作。

5.3 保密管理

古籍数据加工人员对古籍资料和数据负有安全保密义务。

5.4 工作规范管理

对古籍数据加工人员的工作规范管理应包括以下2个方面：

- a) 对各个环节的加工职能都应提供详细的操作任务指导书，保证工作人员的标准化操作；
- b) 相关工作人员需提交工作日志，定期上报工作情况。

6 环境管理

6.1 图像获取环境

古籍数据加工的图像获取环境应符合以下要求：

- a) 在恒温恒湿的环境下进行扫描或拍照，应带手套作业，防止手汗损坏古籍；
- b) 古籍单页下应垫衬纸。

6.2 古籍存放环境

6.2.1 空间要求

古籍的存放空间应符合以下要求：

- a) 古籍不应存放于建筑物顶层和上层；
- b) 存放空间应具有优良的密闭性和保温隔热性能；
- c) 存放空间不应有给、排水和空调、热力水管线通过；
- d) 存放空间入口处应设置缓冲间。

6.2.2 温湿度要求

古籍存放空间的温湿度应符合以下要求：

- a) 加工环境温度应控制在 16℃-22℃；
- b) 加工环境相对湿度应控制在 45%-60%；
- c) 温度日较差不应大于 2℃，相对湿度日较差不应大于 5%；
- d) 应设置温湿度检测仪，检测、记录温湿度变化情况。

6.2.3 空气净化和通风要求

古籍存放空间的空气和通风应符合以下要求：

- a) 灰尘和有害气体浓度参数的检测应符合 GB/T 18883—2022《室内空气质量标准》的规定；
- b) 不应混放缩微胶片等其他可能释放有害气体的物品；
- c) 应保证一定比例的新风量，新风比例应符合 GB/T 18204.5—2025《公共场所卫生检验方法 第5部分：集中空调通风系统》的规定。

6.2.4 照明要求

古籍存放空间的照明条件应符合以下要求：

- a) 照明光源的紫外线含量应小于 75uw/lm；
- b) 应采用防紫外线玻璃和遮阳措施，防止阳光直射；
- c) 人工照明宜选用乳白色灯罩的白炽灯，采用荧光灯时，应有过滤紫外线和防火措施。

6.2.5 消防与安防要求

加工环境应设置水灾、火灾自动报警系统和气体灭火系统。

7 资料管理

7.1 资料整理和移交

7.1.1 资料整理

7.1.1.1 分类

中文古籍数据加工的资料分类是对中文古籍原件、扫描文件的划分。

7.1.1.2 核对

对数据加工的中文古籍原件、原始图像文件等进行核对，核对点应包括：

- a) 核对资料的物理形态是否与原始资源清单一致；
- b) 核对资料出版信息、卷册信息等，确定是否为所加工版本；
- c) 核对原始图像文件是否完整，数据版本是否为最终版。

7.1.1.3 检查

对中文古籍原件的外观、污损、缺失等品相要素进行全面检查，对原始图像文件的有效性和完整性进行检查，检查点应包括：

- a) 是否存在缺页和少页等；
- b) 是否存在纸张缺角、残边和折角等损坏；
- c) 资料附带的附件是否完整；
- d) 原始图像文件的完整性和有效性等。

7.1.1.4 登记

登记、核对、检查过程中发现的各类问题，应生成完整的资料登记清单，清单样例见附录B.1。资料登记清单应包括基本信息和检查结果信息，基本信息登记项目见表1。

表1 基本信息登记项目

序号	登记项	资源类型		说明
		古籍纸质原件	原始图像文件	
1	ID 号	✓	✓	序号
2	题名	✓	✓	待加工资源的名称
3	卷册	✓	✓	卷册信息
4	从套书	✓	✓	系列图书相关信息
5	作者	✓	✓	作者信息
6	出版单位	✓	✓	出版单位信息
7	出版时间	✓	✓	出版时间信息
8	页数	✓	✓	总页数
9	字数	✓	✓	总字数
10	电子数据介质	—	✓	硬盘、磁带、光盘
11	分类	✓	✓	纸质、电子数据
12	可否拆分	✓	✓	原始古籍是否可以拆分为单页资源
13	归还要求	✓	✓	明确资料归还要求
14	其他	✓	✓	其他需要登记的信息

7.1.2 资料移交

在委托方无法提供场所直接进行图像采集的情况下，委托方需要将中文古籍数据加工资料登记清单所列资料移交加工方，加工方对资料拥有临时支配权。加工方清点、核对并接收资料，双方签署资料移交清单，移交清单样例见附录B.2，移交清单主要内容应包括：

- a) 项目名称：项目代称或资料统称；
- b) 交接时间：完成移交的时间；
- c) 资料数量：批次数，单批次数，总量；
- d) 种类数量：需加工种类数量；
- e) 加工批次：如资料分批加工，需填写；
- f) 介质类型：如纸质或移动硬盘等；
- g) 签名签章：双方负责交接的管理人员签字或签章确认；
- h) 移交清单附件：附上资料整理详细清单。

7.2 资料使用

7.2.1 使用原则

为避免中文古籍原件在加工过程中出现损坏或遗失，资料的使用应遵循以下基本原则：

- a) 指定专人管理；
- b) 使用扫描图像替代物理件；
- c) 在资料归还上架后，才能被再调用，禁止私下转调；
- d) 无关人员禁止翻阅或使用资料。

7.2.2 使用要求

7.2.2.1 资料借出

使用者应填写资料使用流转单，流转单样例见附录B.3，然后由资料管理员提取资源，在双方确认资源状况后，使用者签字领取。资料使用流转单主要内容应包括：

- a) 资料编号：资料的流水号；
- b) 书架编号：资料存放书架号；
- c) 名称信息：资料的名称或期次信息等；
- d) 借出时间：资料出库时间；
- e) 归还时间：资料入库时间；
- f) 交接签字：使用者借出或交还的签名确认。

7.2.2.2 资料使用

资料使用的基本要求应包括以下5个方面：

- a) 做好资料保管、保护和保养；
- b) 不应损坏、勾划、删改、加注、折页等；
- c) 不应转借他人、移作他用、私自带出，若发生丢失应及时登记；
- d) 如需拆分、裁切等，应有委托方的许可；
- e) 使用保密资料时，应做好资料的保密工作。

7.2.2.3 资料交还

资料交还的基本要求应包括以下3方面：

- a) 资料使用完毕后使用者应及时交还；
- b) 资料交还时，双方在确认没有新的污损后由资料管理员做交还登记；
- c) 资料交还时，若有新的资料污损应及时填写问题反馈登记表。

8 数据存储和备份管理

8.1 基础数据存储和备份管理

8.1.1 基础数据存储

中文古籍数据加工的基础数据在线存储，供后续数据加工环节随时调用，使用权限均设置为只读。

8.1.2 基础数据备份

中文古籍数据加工的基础数据离线存储1~2份备份并长期保存，也可用于覆盖恢复出现问题的在线存储基础数据。应确保离线存储备份数据的有效性和安全性。

8.2 过程数据存储和备份管理

8.2.1 过程数据存储

中文古籍数据加工的过程数据按项目、环节和日期等信息进行区分并集中在线存储。

8.2.2 过程数据备份

中文古籍数据加工的过程数据除集中在线存储，还应在相应的数据加工设备内保留备份。修改过程数据时，应同时更新覆盖2份备份。

8.3 成品数据存储和备份管理

8.3.1 成品数据存储

中文古籍数据加工的成品数据需在线存储，以便随时检查、核对或修改。

8.3.2 成品数据备份

中文古籍数据加工的成品数据离线存储1~2份备份并长期保存。应确保离线存储备份数据的有效性和安全性。

8.4 数据长期保存

中文古籍的长期保存数据应达到以下要求：

- a) 数据的存储应至少有2份独立的物理备份；
- b) 数据的存储应优先使用磁介质和光介质；
- c) 存储刷新应在介质理论寿命的一半时进行；
- d) 当出现先进、成熟的存储介质和方法时，应及时迁移到新介质上。

8.5 数据核销

中文古籍成品数据验收合格后，加工方保留时间不超过3个月；如委托方有特殊要求，可按委托方要求时间存储，一般不超过1年。存储到期后应进行数据核销，核销数据应包括以下5种类型：

- a) 基础数据；
- b) 过程数据；
- c) 成品数据；
- d) 其他数据；
- e) 各种数据的备份。

数据核销后，委托方向加工方出具数据核销证明文档。

9 交付管理

古籍数字加工完成后，相关文档和数据交付物应完备，应包括：

- a) TIFF 文件：经过图像处理存储的无压缩图像文件；
- b) PDF 文件：版式风格完全尊重原件的版式还原 PDF 文件；

- c) XML 文件：按照相应的 XML Schema 或 DTD 定义，输出要求定制的可接入后续应用系统或全文检索系统的交换文件，以及体现版式与文章对应关系的文件；
- d) 其他文件：EXCEL 格式的相关说明文件和清单文件。

附 录 A
(资料性)
质量检测报告样例

质量检测报告样例如表A. 1所示。

表A. 1 质量检测报告样例

质量检测报告						
项目编号			项目名称			
提交日期			检测日期			
加工方			检测方			
检测情况	检测项	规格要求	抽检率	检测数据	检测工具	检测结果
数据及检测 明细清单						
检测结论						
问题描述及 错误修改 说明						
检测人						

附录 B
(资料性)
资料清单样例

B.1 资料登记清单

资料登记清单样例如表B. 1所示。

表B. 1 资料登记清单样例

资料登记清单											
项目名称				整理人员				整理时间		年 月 日	
序号	资源信息										
	ID	题名	卷册	丛套书	作者	出版单位	出版时间	页数	字数	损坏情况	备注
合计											

B.2 资料移交清单

资料移交清单样例如表B. 2所示。

表B. 2 资料移交清单样例

资料移交清单			
项目名称			
提交单位		接收单位	
提交时间		加工批次	
提交数据方式	<input type="checkbox"/> 光盘 <input type="checkbox"/> 硬盘 <input type="checkbox"/> 其他		
移交资料的类型和数量			
提交方签字（盖章）		接收方签字（盖章）	
移交资料见《附件清单》			

B.3 资料使用流转单

资料使用流转单样例如表B.3所示。

表B.3 资料使用流转单样例

资料使用流转单						
项目名称:					管理人员:	
序号	书架 编号	批次	版数	ID 号	资源名称	流转类型
1						纸质
2						扫描件
3						
借出时间:	年 月 日 时				归还时间	年 月 日 时
签字:					签字	
归还资源 情况:						

参 考 文 献

- [1] 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
 - [2] CY/T XXX. 7—XXXX 中文古籍数字出版规范 第7部分：古籍数字加工与应用模式
 - [3] 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
 - [4] WH/T 24—2006 图书馆古籍特藏书库基本要求
-