

文化行业标准《汉文古籍版式描述规范》编制说明

一、工作简况，包括任务来源、协作单位、主要工作过程、国家标准主要起草人及其所做的工作等

2017年根据原文化部科技司批复《汉文古籍文本数据规范》正式立项为文化行业标准制定计划项目，项目计划号WH2017-04，该项目由国家图书馆牵头联合北京汉王数字科技有限公司和天津图书馆共同参与起草，项目归口单位是全国图书馆标准化技术委员会（以下简称“图标委”）。2019年11月底，在图标委组织的专家会上，该项目未能通过专家审查。根据专家意见，原项目调整为《汉文古籍版式描述规范》。起草组根据专家意见，向图标委提交了项目变更申请。调整后，项目名称变更为《汉文古籍版式描述规范》，项目计划号、牵头单位和参与单位与原项目保持一致。

2020年10月，标准起草组完成了标准征求意见稿及编制说明的撰写工作，提交图标委秘书处审核。经多次修改，于2021年3月形成标准征求意见稿。

二、标准编制原则和确定标准主要内容（如技术指标、参数、公式、性能要求、试验方法、检验规则等）的论据（包括试验、统计数据），修订标准时，应增列新旧标准水平的对比

本标准从古籍文本化工程的角度出发，规定了汉文古籍版式数据的结构和具体描述，符合本规范的文本数据可以用来还原汉文古籍原貌，适合汉文古籍文献长期存储与使用。本标准对汉文古籍的版式进行了规范，适用于汉文古籍文本化加工结果的描述。

本标准可以客观描述汉文古籍的版式特点（如：有无版框、四周单边、四周

双边、左右双边、大小字等），文本内容和位置，插图大小和位置等内容。每个书叶应该只有一种描述方式；每个描述数据只有一种解释方式，没有歧义。数据格式简单，保证了XML文件容易加工和解释，保证汉文古籍文本后续更深层次加工使用。拥有统一版式的书叶，版式相关数据保持严格一致，如版框位置、版心位置、文本列数、文字大小等。可以对XML进行扩展以适应更多的汉文古籍样式，例如新的字符修饰样式等。扩展部分必须是新增加的内容，不得使用已有的内容代替，不得与已有内容发生逻辑冲突。

三、主要试验（或验证）的分析、综述报告，技术经济论证，预期的经济效果

目前我国现存的汉文古籍超过20万种、50万个版本，已文本化的古籍不超过五分之一。作为重要的学术资源，古籍数字化的数量与质量都不能满足学界的要求。本标准适用于古籍文本化及应用的全流程，能够解决数据质量、资源整合、集成发布等一系列问题。更为重要的是通过采用本标准，目前市场价格在15至50元每筒子页（按照数据批量、加工具体要求、文字正确率等核定价格），在相同条件下，数据加工费用可降低30%至60%。

四、采用国际标准和国外先进标准的程度，以及与国际、国外同类标准水平的对比情况，或与测试的国外样品、样机的有关数据对比情况

相关领域暂时没有国际标准和国外先进标准。

五、与有关的现行法律、法规和强制性国家标准的关系

符合现行法律、法规，相关领域暂时没有强制性国家标准。

六、重大分歧意见的处理经过和依据

无。

七、标准作为强制性标准或推荐性标准的建议

建议作为推荐性行业标准。

八、贯彻标准的要求和措施建议（包括组织措施、技术措施、过渡办法等内容）

标准公布后，将为古籍数字化企业开展相关资源加工提供规范依据，逐步将已有数据转换为标准格式。在标准制定过程中，已经进行了数据格式转换方面的测试，可以实现大规模自动批处理。

九、废止现行有关标准的建议

无。

十、其他应予说明的事项。

无。