

ICS 35.240.01

CCS L 70

# 团 体 标 准

T/CPRA 200.3—2025

## 文化资源数据标注系统

Cultural resources data annotation system

2025-01-09 发布

2025-01-09 实施

中国公共关系协会文化大数据产业委员会

发布

## 目 次

前 言 .....	II
1 范围 .....	1
2 规范性引用文件 .....	1
3 术语和定义 .....	1
3.1 术语和定义 .....	1
4 系统总体操作流程 .....	2
5 技术要求 .....	3
5.1 登录注册模块技术要求 .....	3
5.2 图片输入模块技术要求 .....	3
5.3 多标签标注模块技术要求 .....	3
5.4 大众打标模块技术要求 .....	3
5.5 专家资料库模块技术要求 .....	4
5.6 数据校验模块技术要求 .....	4
6 系统性能指标 .....	4
7 网络接口与通信协议 .....	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。

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。

本文件由中国公共关系协会文化大数据产业委员会提出并归口。

本文件起草单位：北京邮电大学、北京玖扬科技有限公司、伏羲云（北京）文化科技有限公司。

本文件主要起草人：徐坤、高凯、赵海英、侯小刚、薛晓鹏、周月、徐鹏举、尹晖、张炼、曹明炜、崔义娜、王琰、徐娜、陈磊、许培、李媛媛。

# 文化资源数据标注系统

## 1 范围

本文件规定了文化资源数据标注系统的总体操作流程、技术要求、系统性能指标和网络接口与通信协议。

本文件适用于国家文化大数据体系文化遗产标本库中文化资源数据标注系统的设计、开发及运行。

## 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。

T/CPRA 1 国家文化大数据标准体系

T/CPRA 300 文化数字内容分类与代码

T/CPRA 301 文化资源数据分类与代码

## 3 术语和定义

### 3.1 术语和定义

下列术语和定义适用于本文件。

#### 3.1.1 文化数字内容 cultural digital content

以数字形式存在的文化产品，一般以文字、图片、音频、视频、多媒体和其他形式表现。通常包括传统文化产品的数字化以及以数字形式存在的文化产品。国家文化大数据体系下文化数字内容主要涵盖中华优秀传统文化、革命文化和社会主义先进文化等。

文化数字内容作为数字化文化生产线的产出，由文化大数据服务平台管理与分发，并通过文化体验设施和设备消费。通常情况下，文化数字内容又称为文化数字产品、文化数字内容产品等。

[来源：T/CPRA 300 定义 3.1.1]

#### 3.1.2 文化资源数据 cultural resources data

对人类文化中传承下来并可以传播利用的文化（包括物质的和非物质的）进行数字化采集后，所得到的用于识别和展现文化的图像、文字、声音、动画、影片、三维全景、三维模型等数据。国家文化大数据体系下文化资源数据主要包括中国文化遗产标本库、中华民族文化基因库、中华文化素材库中的数据。

[来源：T/CPRA 301 定义 3.1.1]

#### 3.1.3 中国文化遗产标本库 specimen repository of Chinese cultural heritage

中国文化遗产标本库是基于历次全国文化资源普查的数据（包括古籍、文物、美术馆藏品、地方

戏曲、非物质文化遗产等普查已经获得的数据），按照国家文化大数据标准，结构化存储于服务器，并通过国家文化专网实现共享的数据。

[来源：T/CPRA 301 定义 3.1.2]

#### 4 系统总体操作流程

文化资源数据标注系统主要由登录注册模块、图片输入模块、多标签标注模块、大众打标模块、专家资料库模块以及数据校验模块组成，各模块间通过有序的流程进行数据交互与协同，系统操作流程如图 1 所示。

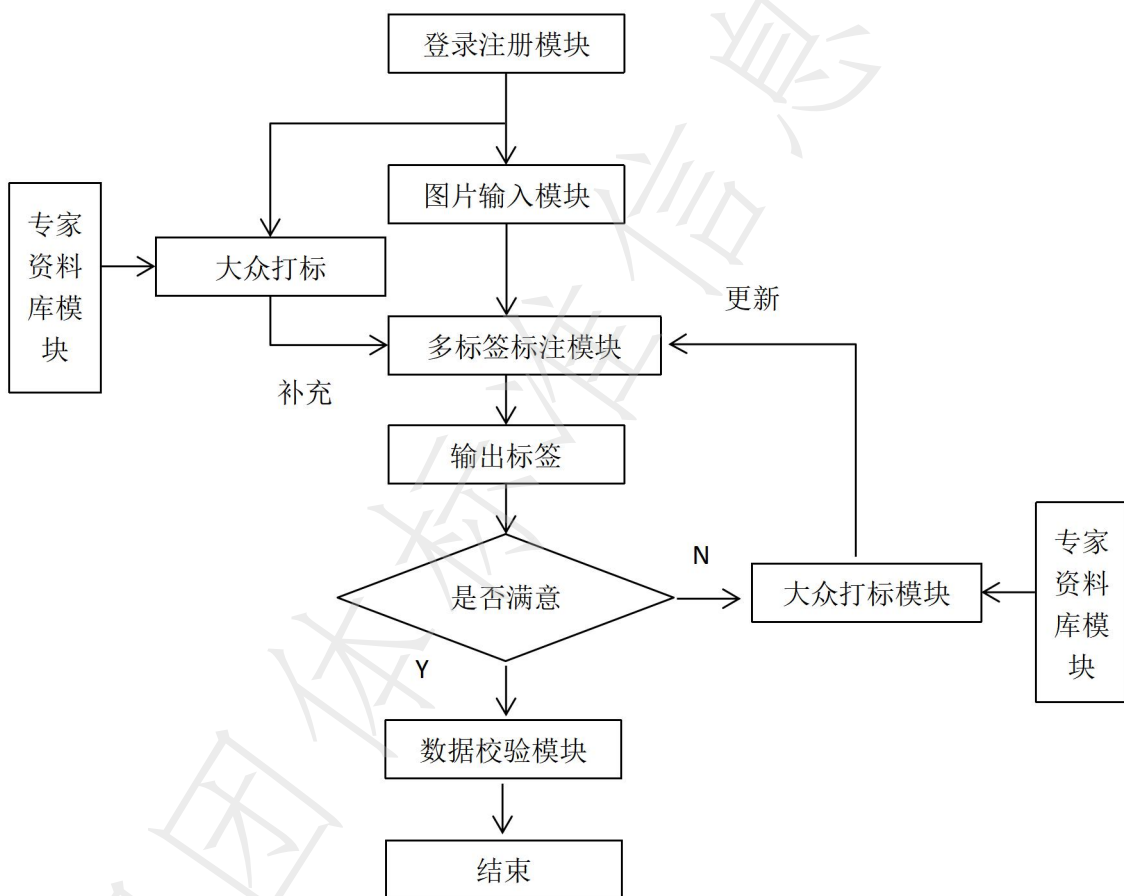


图 1 系统操作流程

内部模块间关系如下：

- 用户通过登录注册模块验证身份后进入图片输入模块，上传的图像文件进入标注任务流程；
- 上传的图像文件通过图片输入模块传递至多标签标注模块，多标签标注模块对图像文件进行智能化处理，生成初步标注结果；
- 多标签标注模块生成的初步结果将传递至大众打标模块，供用户进一步标注优化。大众用户通过直观的界面对标注结果进行补充或校正；
- 标注任务完成后，大众打标模块将不确定或需要校验的标注结果提交至专家资料库，由专家进行深度审核与最终确认；

- e) 数据校验模块负责从标注任务的标注结果进行收集和反馈，检查标注结果的一致性和准确性。如标注结果不符合要求，数据将返回至相应模块进行调整；
- f) 将最终确认的标注数据进行持久化存储，并通过标准化接口提供给外部系统，以支持文化资源的管理与应用。

## 5 技术要求

### 5.1 登录注册模块技术要求

登录注册模块应满足如下要求：

- a) 用户注册与权限管理：用户可通过填写基本信息完成注册，用户角色分为普通用户和专家用户，用户权限需根据邀请码进行分配。支持管理员界面对用户的权限进行审查和调整，记录所有权限更改的操作日志；
- b) 登录验证：支持用户名和密码，或邮箱和密码的传统登录方式；
- c) 安全性要求：对用户密码使用 MD5 加密算法进行存储，以增强系统安全性。

### 5.2 图片输入模块技术要求

图片输入模块应满足如下要求：

- a) 文件上传支持：系统需兼容多种图片格式（JPEG、PNG、TIFF 等），确保适配多样化的文化资源图像，最大单文件大小可达 50 MB，上传界面应提供友好的用户体验，包括拖拽式上传、多文件批量选择以及清晰的上传状态提示；
- b) 图像预览：上传的图片需实时生成缩略图，并可点击放大按钮查看原始图片的细节。

### 5.3 多标签标注模块技术要求

多标签标注模块应满足如下要求：

- a) 自动标注：系统需集成算法（如字典学习算法或者深度学习算法），对文化资源图像进行自动多标签标注，自动标注的标签需包括多种标签（如纹样，材质，颜色等），需确保模型标注的准确率不低于 80%；
- b) 人工协同标注：对于系统调用算法自动生成的标注结果，如果标注结果不准确，可以进一步采用人工协同标注方式对结果进行修改并上传反馈。

### 5.4 大众打标模块技术要求

大众打标模块应满足如下要求：

- a) 图像预览：大众打标图像展示支持多条件搜索（如关键词、类别标签等），根据分类标签展示不同类别图像，并提供排序和筛选功能，并支持分页功能高效显示图像列表和直接跳转到指定页功能，为后续大众打标选取图像提供便捷；
- b) 目标图像选择：可以对数据库图像进行人工打标，然后系统对其提交的标注内容进行内部审核，支持标注结果反馈至专家资料库进行进一步审核，支持任务完成进度可视化与结果反馈。

### 5.5 专家资料库模块技术要求

专家资料库模块应满足如下要求：

- a) 专家领域知识：提供领域专家标注模板支持，支持专业词库动态更新，保证与最新研究同步；
- b) 图像展示：支持分类标签和关键词的复合查询，展示专家已标注的高质量图像。

### 5.6 数据校验模块技术要求

数据校验模块应满足如下要求：

- a) 数据校验：提供领域专家标注模板支持，支持专业词库动态更新，保证与最新研究同步；
- b) 反馈机制：对于数据校验结果需要进行记录存档，并同时反馈到系统上，确保数据质量的提升可持续进行。

## 6 系统性能指标

- a) 响应时间：主功能操作平均响应时间不超过 1 秒；
- b) 数据处理能力：每日数据处理量不低于 10MB；
- c) 网络条件：系统运行需具备稳定的百兆以上带宽支持。

## 7 网络接口与通信协议

系统采用标准化网络接口与通信协议，支持以下功能：

a) 数据接口：基于RESTful API进行开发，提供系统内部的资源访问，并同时提供多标签标注、大众打标、专家资料库任务分发服务功能；

b) 通信协议：支持HTTP/HTTPS协议，数据加密方式采用TLS 1.2；

c) 兼容性：支持主流数据库格式与云平台对接。

---