

ICS 35.240.01

CCS L 70

团 体 标 准

T/CPRA 2104—2025

文化数据价值评价指南

Guidelines to the value evaluation of cultural data

2025-12-12 发布

2025-12-12 实施

中国公共关系协会文化大数据产业工作委员会 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
4 评价原则	2
5 文化数据价值评价指标体系	2
5.1 指标体系	2
5.2 评分指南	4
6 评价方法	16
6.1 指标取值过程	16
6.2 等级划分方法	17
6.3 评价过程	17
附录 A	18

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工作委员会提出并归口。

本文件起草单位：北京玖扬科技有限公司、山东数字文化集团有限公司、四川省文化数据服务有限公司、伏羲云（北京）文化科技有限公司、华中文化大数据科技（武汉）有限公司、江苏省文化产权交易所有限公司、北京文创动力信息技术有限公司、广西广电大数据科技有限公司、辽宁北斗云数字科技有限公司。

本文件主要起草人：高凯、陈亮宏、尹晖、宣宏量、宋耀、王伟文、魏传强、幸伍洲、刘春燕、赵鹏举、程玉、崔义娜、夏立雄、陈香竹、唐思、张涵、杨小雪、王鑫、李媛媛、尚乾雄。

文化数据价值评价指南

1 范围

本文件提供了文化数据价值评价方面的指导，给出了文化数据的本体价值、功能价值、情感价值、传播价值、经济价值、市场需求、技术可行性、合作体量等评估指标及其分值确定建议，并给出了综合评分计算方法。

本文件适用于文化机构、企业、科研单位等各类市场主体开展文化数据（包括文化资源数据和文化数字内容）的价值评估，以支撑文化数据交易、投资合作、版权授权等活动的开展。

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 2100.2—2024 中国文化遗产标本库技术要求

T/CPRA 2101.2—2024 中华民族文化基因库技术要求

T/CPRA 2102.2—2024 中华文化素材库技术要求

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

下列术语和定义适用于本文件。

3.1

文化数字内容 **cultural digital content**

是指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]

3.2

文化资源数据 **cultural resources data**

是指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]

3.3

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]

3.4

中华民族文化基因库 gene repository of Chinese culture

中华民族文化基因库的数据主要来自革命文物大数据库和博物馆大数据体系，通常经过高精度数据采集，导入国家文化大数据体系底层关联集成系统，完成数据标注、关联，并提取中华民族文化元素、符号和标识，最终形成中华民族文化基因库。相关数据通过国家文化专网实现共享。

[来源：T/CPRA 301 定义 3.1.3]

3.5

中华文化素材库 material repository of Chinese culture

以文化资源数字化成果为原料，集成运用各种新技术，萃取中华文化之要素，并分门别类标签化，进而形成的可组合使用的素材库。按照文化呈现要素中华文化素材库可划分为中华字库、中华音库、中华像库、中华乐库、中华舞库、中华剧库等。中华文化素材库的来源是多元的，中国文化遗产标本库和中华民族文化基因库的数据是重要来源，文化企事业单位已建成的文化艺术、新闻出版、广播电视、网络视听、电影等数据库，也是重要来源。相关数据通过国家文化专网实现共享。

[来源：T/CPRA 301 定义 3.1.4]

4 评价原则

文化数据价值评价宜遵循以下原则：

- a) 系统性：全面系统收集评价对象的信息，依据准确可靠的数据，采用客观公正的价值评价程序评价文化数据的价值；
- b) 科学性：保持评价指标选取的科学合理性，评价程序的严密有效性；
- c) 可操作性：评价指标切合文化数据的实际情况，评价程序简便可行；
- d) 独立性：评价过程独立自主完成，不受外界因素干扰。

5 文化数据价值评价指标体系

5.1 指标体系

文化数据价值评价体系由本体价值、功能价值、情感价值、传播价值、经济价值、市场需求、技术可行性、合作体量八个一级评价指标和对应的二级评价指标构成，文化数据价值评价具体指标体系见如表 1 所示。

表 1 文化数据价值评价指标体系

一级指标	二级指标
1 本体价值	
	1.1 独特性
	1.2 历史性
	1.3 艺术性
	1.4 文化符号性
	1.5 文化多样性
2 功能价值	
	2.1 传承性
	2.2 社会性
	2.3 教育性
	2.4 科学性
3 情感价值	
	3.1 故事性
	3.2 情感共鸣性
	3.3 情感主流性
4 传播价值	
	4.1 传播力
	4.2 知名度
	4.3 系列性
5 经济价值	
	5.1 版权与知识产权
	5.2 收入与盈利模式
	5.3 商业应用潜力
6 市场需求	
	6.1 目标受众规模
	6.2 市场热度与趋势
	6.3 竞争态势
7 技术可行性	
	7.1 数据质量
	7.2 数据兼容性
	7.3 数据处理难度
	7.4 存储与传输成本
8 合作体量	
	8.1 数据体量
	8.2 高质量数据占比

在针对具体的文化数据进行价值评价时，可自行设定一级指标与二级指标分值的权重。具体示例参见附录 A。

5.2 评分指南

5.2.1 本体价值

本体价值用于衡量文化数据作为文化载体本身所固有的核心价值，聚焦其在文化内涵、稀缺性、历史关联、艺术特质、符号意义及多样性呈现等方面的本质属性，反映数据自身不可替代的文化内核与独特性。具体指标包括：

- a) 独特性
- b) 历史性
- c) 艺术性
- d) 文化符号性
- e) 文化多样性

在对文化数据的本体价值评价过程中，其各类指标的具体分值可采用表 2 的建议的方式进行确定。

表 2 本体价值评分指南

评分指标	指标释义	分值			
		0~10 分	11~40 分	41~70 分	71~100 分
1.1 独特性	指文化数据在全球、特定领域或地域范围内的稀缺程度及与同类数据的差异程度，反映其不可替代性。	独特性极弱，难以与同类数据区分。	含一定独特元素，但整体与常见文化数据相似（如带地方特色的民间故事集）。	特定领域/地域内具有显著独特性，与同类数据差异明显（如少数民族独有祭祀仪式影像）。	全球范围内独一无二，是特定文化现象的唯一记录（如濒危方言完整语音数据库或仅存的某个失传工艺的制作影像）。
1.2 历史性	衡量文化数据与历史事件、人物的关联程度及作为历史研究资料的核心性，反映其对历史认知的贡献。	历史价值微弱，与历史关联性不紧密。	具备一定历史信息，但重要程度较低（如普通家族近代家谱）。	与历史相关，对某段历史研究有重要参考价值（如地方县志数字化版本）。	直接关联重大历史事件/关键人物，是研究特定历史时期的核心一手资料（如古代皇家密档、重要历史事件的原始影像）。
1.3 艺术性	评估文化数据所承载内容的艺术水准、创意性及对艺术发展的影响，反映其审美价值。	艺术性差，难以体现艺术价值。	达到行业平均水平，但缺乏独特魅力（如普通网络小说）。	艺术水准较高，具有独特审美价值和创意（如优秀独立动画作品）。	在艺术领域具有开创性，对艺术发展产生深远影响（如毕加索早期手稿的数字化作品，或某创新音乐流派的开山之作）。

1.4 文化符号性	衡量文化数据所承载内容成为特定文化、地域或群体标志性符号的程度及象征意义。	无明显象征意义，难以代表特定文化或群体（如普通日常生活场景的随意拍摄）。	在小范围社群中具有象征意义，但未形成广泛认知（如某乡镇的特色建筑记录）。	在区域或民族范围内成为代表性文化符号（如少数民族的标志性服饰数字化档案）。	成为国家或全球性的文化标志性符号，象征意义被广泛认同（如故宫代表中国古代宫廷文化）。
1.5 文化多样性	衡量文化数据所涵盖的不同文化类型、地域文化、族群文化的丰富程度，及对文化差异性的展现能力。	单一文化内容为主，几乎无文化多样性体现。	涉及少数几种文化类型，多样性体现较有限（如某一文化圈内的分支文化记录）。	包含多个不同文化体系的核心内容，能体现显著的文化差异性（如跨洲民间艺术对比数据集）。	全面涵盖全球多区域、多族群的独特文化，系统展现文化多样性的深度与广度（如全球濒危族群文化全景记录数据库）。

5.2.2 功能价值

功能价值用于衡量文化数据在实际应用场景中所发挥的功能性作用，聚焦其在文化传承延续、知识教育传递、社会现象反映、科学研究支撑等方面的实用价值，反映数据对文化实践、认知提升及学术研究的支撑能力。具体指标包括：

- a) 传承性
- b) 社会性
- c) 教育性
- d) 科学性

在对文化数据的功能价值评价过程中，其各类指标的具体分值可采用表 3 的建议的方式进行确定。

表 3 功能价值评分指南

评分指标	指标释义	分值			
		0~10 分	11~40 分	41~70 分	71~100 分
2.1 传承性	衡量文化数据对文化传承的必要性及支撑作用，反映其在文化延续中的不可替代性。	对传承作用极小。	对传承有辅助作用，但非不可或缺（如普通民俗活动简单记录）。	对文化传承有重要帮助，是关键资料（如传统节日习俗详细记录）。	是某种文化传承的核心载体，没有该数据文化传承将面临中断风险（如非遗唯一传承人的教学视频）。
2.2 社会性	衡量文化数据与社会现象、社会关系、公共议题的关联程度，及对社会认知、社会反思的推动作用。	与社会议题关联微弱，难以体现社会性。	涉及日常社会现象，能反映部分社会特征，但影响力有限（如普通社区生活记录）。	能体现特定社会群体的生活状态、社会问题，对理解社会某一侧面有重要意义（如弱势群体生存现状的深度调查报告）。	深刻反映重大社会变革、核心社会矛盾或普遍社会心理，对社会研究或公共政策制定有重大参考价值（如某一时代社

					会运动的完整影像记录)。
2.3 教育性	衡量文化数据在知识传递、认知启蒙、技能培养等方面的教育功能,及对不同受众的学习辅助价值。	教育价值微弱,难以满足学习需求。	包含一定教育信息,可作为辅助学习材料,但系统性和深度不足(如科普类短视频合集)。	具有明确的教育目标,能有效传递特定知识或技能,适合特定群体学习使用(如非遗技艺系统教学视频集)。	是某一领域的权威教育资源,能系统构建知识体系,对专业学习或大众启蒙有决定性作用(如高校核心课程的原始教学素材库)。
2.4 科学性	衡量文化数据的科学依据、严谨性,及对科学研究、技术发展的支撑作用,反映其科学价值。	科学性差,存在明显错误或无科学依据。	包含一定科学信息,但缺乏系统性或严谨性不足(如民间经验的简单记录)。	符合科学规范,数据严谨,对特定科学研究或技术应用有重要参考价值(如某类非遗技艺的科学原理分析报告)。	基于重大科学发现或严谨科学实验,是某一科学领域的核心研究数据,对科学发展有突破性贡献(如某一学科开创性理论的原始实验记录)。

5.2.3 情感价值

情感价值用于衡量文化数据通过叙事与情感连接传递文化的效果,聚焦其在故事建构、情感共鸣等方面的感染力,反映数据引发受众情感认同、集体记忆及价值观共鸣的能力,以及通过叙事承载文化内涵的深度。具体指标包括:

- a) 故事性
- b) 情感共鸣性
- c) 情感主流性

在对文化数据的情感价值评价过程中,其各类指标的具体分值可采用表4的建议的方式进行确定。

表4 情感价值评分指南

评分指标	指标释义	分值			
		0~10分	11~40分	41~70分	71~100分
3.1 故事性	衡量文化数据的叙事完整性、情节丰富度及人物塑造能力,反映其通过故事传递文化的效果。	故事性差,情节零散或缺乏故事元素。	故事结构完整,但情节平淡(如普通生活故事集)。	故事结构完整,情节有基本起伏,人物形象较清晰(如地方民间传说合集)。	情节跌宕起伏,人物丰满,能以精彩叙事传递文化内涵(如经典历史演义小说)。

3.2 情感共鸣性	衡量文化数据引发受众情感认同、集体记忆或价值观共鸣的范围与强度。	难以引发受众情感共鸣。	在小范围内（如单一社区、家族）引起一定情感共鸣，但强度较弱（如地方怀旧老照片集）。	在特定群体中引起较强烈情感共鸣，形成群体记忆点（如某代人的童年动画主题曲）。	引发跨文化、地域、阶层的广泛强烈共鸣（如反映时代变迁的经典纪录片）。
3.3 情感主流性	衡量文化数据与主流价值观的匹配度，以及对主流价值观的传播推进程度。	不符合主流价值观，无法体现对主流价值观的引导作用。	与主流价值观匹配程度不高，难于对主流价值观形成积极正面的引导。	较为符合主流价值观，能够一定程度上帮助主流价值观的宣传引导。	完全符合主流价值观，对主流价值观的引导和传播有积极推进作用。

5.2.4 传播价值

传播价值用于衡量文化数据在不同圈层、时代背景下的扩散潜力与影响力延伸能力，聚焦其传播广度、认知度、内容延展性及与当代语境的适配性，反映文化数据实现文化扩散、持续影响受众的综合潜力。具体指标包括：

- a) 传播力
- b) 知名度
- c) 系列性

在对文化数据的传播价值评价过程中，其各类指标的具体分值可采用表 5 的建议的方式进行确定。

表 5 传播价值评分指南

评分指标	指标释义	分值			
		0~10 分	11~40 分	41~70 分	71~100 分
4.1 传播力	衡量文化数据在不同文化圈层的自发传播能力及文化扩散效应。	传播能力弱，依赖强制推送，受众主动传播意愿极低。	在一定范围内传播，但范围有限（如小众文化爱好者创作的同人作品）。	在特定文化圈层内广泛传播，形成区域或领域热点（如某省份流行的非遗技艺短视频）。	在多文化圈层中具有强大自发传播力，形成广泛扩散（如全球走红的短视频文化现象）。
4.2 知名度	衡量文化数据在目标受众中的认知度、影响力及品牌效应。	知名度极低，鲜为人知。	知名度较低，仅小部分爱好者或研究者知晓（如地方小众非遗项目的记录资料）。	在特定领域/行业内知名度较高，为专业群体广泛知晓（如某艺术流派的代表性作品集）。	在全球或全国范围内的目标群体中认知度极高，形成强品牌效应（如知名 IP 作品）。
4.3 系列性	衡量文化数据的延展性及形成系列化产品、传播内容的潜力。	基本不具备系列化特征。	内容关联性较弱，仅能勉强扩展为简单系列（如同一作者	有明确主题关联，可形 2-3 个连贯系列（如同一地域不同季节的民俗	内容体系完整，可衍生出多维度系列产品（如经典 IP 的小说、影视、

			的非连贯短篇故事集)。	记录系列)。	游戏等跨媒介系列)。
--	--	--	-------------	--------	------------

5.2.5 经济价值

经济价值用于衡量文化数据在商业开发、版权交易等方面的经济潜力及收益稳定性。具体指标包括：

- a) 版权与知识产权
- b) 收入与盈利模式
- c) 商业应用潜力

在对文化数据的经济价值评价过程中，其各类指标的具体分值可采用表 6 的建议的方式进行确定。

表 6 经济价值评分指南

评分指标	指标释义	分值			
		0~10分	11~40分	41~70分	71~100分
5.1 版权与知识产权	衡量文化数据版权归属的清晰度、知识产权保护完善度及法律风险。	版权无明确归属，存在重大法律纠纷（如盗版数据）。	版权归属模糊，存在较多知识产权纠纷隐患。	版权基本明确，存在少量潜在争议风险，法律文件基本齐全。	版权归属清晰无争议，保护体系完善，授权链条及法律文件完整。
5.2 收入与盈利模式	衡量文化数据已形成的收入来源、盈利模式的成熟度及收益稳定性。	盈利模式不清晰，难以实现稳定收入。（如无固定交易渠道，文化数据多疑等以免费分享为主，收入难以覆盖基础运营成本）。	盈利模式初步建立，但尚未完全稳定，有一定收入，但不稳定（如形成专属图片库，与2-3家地方文旅平台达成合作，收入能覆盖文物数据采集、生产、运营成本等）。	盈利模式较为成熟，收入稳定且有一定规模（如与多个主流文化创意平台等建立合作，服务影视剧组、文创企业等B端客户，形成一定收入规模）。	盈利模式成熟稳定，收入可观（如为核心独家文化数据供应商，票房、文创周边、文化活动或文化数据授权等形成稳定收益）。
5.3 商业应用潜力	衡量文化数据在文化核心领域及文化相关领域的应用覆盖广度、适配深度及市场前景。	仅在极小范围场景（如单一地方小众文化平台）有应用可能，市场需求微弱，商业转化难度极高。	在特定小众领域（如地方文旅内容分发、小型文创设计）有一定适配性，存在基础商业需求，但盈利空间有限，需依赖特定合作渠道。	在主流文化领域（如影视内容开发、文创产品设计、文化消费电子适配）有明确适配场景，市场需求稳定，具备成熟的商业转化路径（如与中型文创企业、区域文化平台合作），盈利潜力中等偏上。	在全产业链场景（从内容创作、设计开发到传播分发、终端消费）均具备高适配性，是头部文化企业、跨领域平台的核心需求资源，商业转化路径多元（如IP授权、跨媒介开发、全国性文化项目合作），盈利潜力极高且可持续。

5.2.6 市场需求

市场需求用于衡量文化数据的目标受众规模、市场热度及竞争态势，反映其市场接受度和竞争力。

具体指标包括：

- a) 目标受众规模
- b) 市场热度与趋势
- c) 竞争态势

在对文化数据的市场需求评价过程中，其各类指标的具体分值可采用表7的建议的方式进行确定。

表7 市场需求评分指南

评分指标	指标释义	分值			
		0~10分	10~40分	40~70分	70~100分
6.1 目标受众规模	衡量文化数据覆盖的X/Y/Z世代男性、女性群体范围，及覆盖群体的消费潜力，反映数据的受众基础广度与商业转化潜力。	仅覆盖1个低消费潜力群体，且为单一性别世代，受众基础极窄，消费潜力弱。	覆盖1~2个中低消费潜力群体，受众基础较窄，仅具备基础消费潜力。	覆盖3~4个中高消费潜力群体，或包含1~2个核心高消费潜力群体，受众基础较广，具备中等商业转化潜力。	覆盖5~6个全消费潜力群体，受众基础极广，核心高价值群体集中，商业转化潜力极强。
6.2 市场热度与趋势	衡量文化数据所属文化主题的市场关注度及与未来发展趋势的契合度。	低热度，弱趋势关联。所属文化主题的市场关注度较低或局限于小众群体，商业活跃度不高；与未来发展趋势关联较弱，缺乏明显的契合点，趋势性不突出，发展前景较模糊（例如：传统手工艺的纯记录式内容（无创新或传播亮点）、特定年代的小众娱乐形式回顾）。	中等热度，趋势适配。一定市场关注度，受众群体明确但规模中等，商业表现平稳；与未来发展趋势有较好适配性，虽非核心热门，但符合趋势延伸方向，发展空间可预期（例如：经典文学作品的现代化改编、区域性非遗文化的创新传播）。	高热度，高趋势契合。当前市场热门领域，关注度高，有稳定的受众群体和商业活跃度；与未来发展趋势高度相关，具备持续增长潜力，属于市场重点布局的方向（例如新兴的国风数字艺术、基于大数据的文化消费行为分析内容）。	顶尖热度，强趋势契合。当前市场绝对热门领域，受到全行业及大众的高度关注（如现象级IP、全民讨论的文化潮流）；且与未来3-5年的核心发展趋势（如人工智能、元宇宙、绿色文化等）深度契合，具有前瞻性和引领性，商业价值与社会关注度均处于峰值。

6.3 竞争态势	衡量文化数据在市场中的竞争优势及与同类数据的差异化程度。	无竞争优势，数据同质化严重，用户选择依赖价格、渠道，议价能力弱。	存在竞争优势，文化数据有场景化差异化，但核心技艺内容仍有部分重叠，用户可根据场景选择，而非完全依赖本数据。	强竞争优势，文化数据核心价值维度不可替代。在“完整性+专业性”上形成独家壁垒，用户无法通过同类数据满足核心需求。	绝对竞争优势，几乎无同类文化数据，如独家考古发现影像，市场无内容、价值可匹配的替代数据，用户获取渠道唯一。
----------	------------------------------	----------------------------------	---	--	---

5.2.7 技术可行性

技术可行性用于衡量文化数据的质量、技术适配性及存储传输的可行性，反映其技术层面的可利用性。具体指标包括：

- a) 数据质量
- b) 数据兼容性
- c) 数据处理难度
- d) 存储与传输成本

在对文化数据的技术可行性评价过程中，其各类指标的具体分值可采用表 8 至表 18 的建议的方式进行确定。

表 8 数据质量评分指南

评分指标	指标释义	分值			
		0~10分	10~40分	40~70分	70~100分
7.1 数据质量	衡量文化数据针对视频、音频、图片、三维、文本等不同数据格式，在内容准确性、格式适配性、完整性、传输存储可靠性及各格式专属质量要求方面，满足文化领域应用需求的程度。	质量极差，多数数据格式的核心质量指标不达标，完全无法满足文化领域基础应用需求。	质量较差，部分数据格式的核心质量指标达标，仅能勉强支撑文化领域简单基础应用，需针对短板格式进行大量调整优化。	质量良好，多数数据格式的核心质量指标达标，能满足文化领域常规应用需求，仅需针对个别不足格式进行少量优化。	质量优秀，主要数据格式的核心质量指标均高标准达标，可直接满足文化领域复杂高端应用需求，无需额外调整。

在对文化数据的数据质量评价过程中，根据文化数据涉及的数据格式，其数据质量具体分值可采用表 9 至表 15 的建议方式进行确定。

表 9 视频数据质量评分指南

评分 指标	分值			
	0~10 分	11~40 分	41~70 分	71~100 分
分辨率	720p	1080p	2K	≥4K
帧率	≥24FPS	≥24FPS	≥24FPS	≥30FPS
文件格式	mp4、mov、mkv	mp4、mov、mkv	mp4、mov、mkv	mp4、mov、mkv
编码格式	H.264/AVC	H.264/AVC	H.264/AVC	H.265/HEVC
数据质 量要求	稳定性	避免剧烈抖动。		
	连贯性	视频片段在时序上应连贯，无频繁跳切。		
	多模态对齐	音画应同步。		
数据标 签要求	时序定位	标注不仅包括类别，还包括动作的起止时间戳，并达到秒级精度。		
	密集标注	视频分割、姿态估计等任务需要逐帧标注。		
	复杂性	一个视频片段可能包含多个同时发生的动作和多个物体，标注体系须全部标注。		
数据量 要求	动作识别	大型数据集通常≥10 万个 短视频片段，每个 10 秒左右，涵盖数百个动作类别。		
	视频生成/预 训练	通用多模态大模型训练视频数据量≥100 万小时；特定细分领域模型数据量≥10 万小时。		

表 10 音频数据质量评分指南

评分 指标	分值			
	0~10 分	11~40 分	41~70 分	71~100 分
采样率	44.1kHz/48kHz	44.1kHz/48kHz	44.1kHz/48kHz	96kHz
信噪比	>75%	>85%	>90%	>95%
比特率	≤128kpbs	128~320kpbs (单/双通道)	1411kpbs (双通道)	≥1411kpbs (双通道)
格式	FLAC、ALAC、 WAV、MP3	FLAC、ALAC、 WAV、MP3	FLAC、ALAC、 WAV、MP3	FLAC、ALAC、 WAV、MP3
频率响应	20Hz~20kHz	20Hz~20kHz	20Hz~20kHz	20Hz~40kHz+
数据质 量要求	纯净度	无回声、爆音、电流声等。		
	说话人	对于 TTS 和声纹识别，需要说话人身份一致且音质稳定。		
数据标 签要求	自动语 音识别	转录文本应与音频精确对应，包括标点符号。 去除口语化现象，如重复、犹豫、打断。需制定统一规则(保留或删除)。		
	文本到 语音	需要 <音频, 文本>。 文本需归一化。例：将“100 元”转为“一百元”，将“Dr.”根据上下文转为“Doctor”或“Drive”。		
	声纹/ 情感识 别	每个音频样本需标注说话人 ID 或情感标签。		
数据量 要求	自动语 音识别	单语种语音识别模型≥3000 小时纯净语音数据。大型多语言 ASR 模型≥20000 小时。		
	文本到 语音	对单个说话人，需要 10-20 小时的高质量、高一一致性的音频数据才能合成出自然的声音。		

表 11 图片数据质量评分指南

评分 指标	分值			
	0~10分	11~40分	41~70分	71~100分
分辨率	(512X512) dpi (384X384) dpi	(512X512) dpi (384X384) dpi	(512X512) dpi (384X384) dpi	(512X512) dpi (384X384) dpi
色彩/深度	16位/65536色	24位/1677万色	24位/1677万色	32位/42.9亿色
图片格式	JPG、PNG、BMP	JPG、PNG、BMP	JPG、PNG、BMP	JPG、PNG、BMP
压缩类型	有损压缩	有损压缩	无损压缩	无损压缩
数据质量要求	清晰度	避免过度模糊、抖动、失焦的图像。		
	相关性	图像内容应与标签高度相关。		
	多样性	在光照条件、天气、角度、背景、遮挡、物体尺度等方面应尽可能丰富，以增强模型鲁棒性。		
	无损压缩	避免因过度压缩产生的伪影。		
数据标签要求	准确性	标注框、多边形、关键点应精准贴合目标边缘。		
	一致性	所有标注员需遵循同一标准。		
	完整性	图像中所有需要标注的目标都应被标注，避免遗漏。		
数据量要求	通用图像分类 (ImageNet 级别)	≥100 百万张的图像，分类≥1000 个类别。		
	目标检测/分割	≥10000 张		
	工业特定场景	≥300 张		
	LoRA 风格化模型	≥30 张		

表 12 点云数据质量评分指南

评分 指标	分值			
	0~10分	11~40分	41~70分	71~100分
尺寸/精度误差	≤0.5mm	≤0.1mm	≤0.05mm	≤0.02mm
平均点间距	≤0.5mm	≤0.1mm	≤0.05mm	≤0.02mm
最大点间距	≤0.3mm	≤0.05mm	≤0.02mm	≤0.02mm
完整度	≥70%	≥85%	95~98%	≥98%
数据质量要求	完整性	点云无明显的结构缺失或者失真，能完整的反应物理实体的形态和结构。		
	密度	分布均匀，无明显稀疏和稠密分布。		
	点数量	根据实体的复杂程度，点云文件的点数量应在 10, 000 及 100, 000+之间。		
	比例和朝向	以原点为中心，上方向为 Y 轴或 Z 轴，模型正面应朝向同一轴方向。		
标签文件质量	高描述粒度	对物体进行详细且精准的刻画，准确反映物体的具体形态与本质特质，避免笼统或简略的概括性描述。		
	多维度描述	覆盖类别、材质、风格、结构、场景等多维度信息。		
	一致性	同一个物体的所有多视角图片应共享同一段文本描述。		

表 13 网格数据质量评分指南

评分 指标	分值			
	0~10分	11~40分	41~70分	71~100分

最大三角面数		≤20 万面	≤60 万面	≤80 万面	≤100 万面
顶点和法线精度		单精度浮点 32 位， 小数点后保留 2 位。	单精度浮点 32 位， 小数点后保留 4 位。	单精度浮点 32 位， 小数点后保留 6 位。	双精度浮点 64 位， 保留全部小数位。
数据质量要求	水密性	网格应完全封闭，没有空洞。			
	拓扑	面片分布均匀，无自相交、无退化面。			
	比例和朝向	以原点为中心，上方向为 Y 轴或 Z 轴，比例适中；单位统一为米或厘米。			
标签文件质量	高描述粒度	对物体进行详细且精准的刻画，准确反映物体的具体形态与本质特质，避免笼统或简略的概括性描述。			
	多维度描述	覆盖类别、材质、风格、结构、场景等多维度信息。			
	一致性	同一个物体的所有多视角图片应共享同一段文本描述。			

表 14 纹理贴图数据质量评分指南

评分指标	分值				
	0~10 分	11~40 分	41~70 分	71~100 分	
尺寸/精度误差	≤0.5mm	≤0.1mm	≤0.05mm	≤0.02mm	
纹理贴图分辨率	≤1024×2048	≥1024×2048	≥4096×4096	≥8192×8192	
格式	JPEG、PNG	JPEG、PNG	EXR、TIFF	TGA、PNG	
纹理贴图色差	≤5	≤5	≤3	≤1.5	
完整度	≥70%	≥85%	95~98%	≥98%	
数据质量要求	高分辨率	≥512x512			
	一致的照明	所有视角使用完全相同的灯光设置。			
	纯净背景	使用纯色背景，最好是透明背景。			
	准确的相机参数	精确记录每张渲染图的相机外参和内参。			
标签文件质量	高描述粒度	对物体进行详细且精准的刻画，准确反映物体的具体形态与本质特质，避免笼统或简略的概括性描述。			
	多维度描述	覆盖类别、材质、风格、结构、场景等多维度信息。			
	一致性	同一个物体的所有多视角图片应共享同一段文本描述。			

表 15 文本数据质量评分指南

评分指标	分值				
	0~10 分	11~40 分	41~70 分	71~100 分	
数据量	≥100 万 tokens	≥100 万 tokens	≥100 万 tokens	≥100 万 tokens	
错误率	<5‰	<3‰	<1‰	<0.5‰	
质量要求	清洁度	去除无关字符、乱码。			
	完整性	语义完整，避免文本截断。			
	有害内容	已经过滤掉暴力、仇恨、色情、隐私泄露、涉政涉恐内容。			
语法要求	语言、语义、语法要求表达准确，符合规则。				
标签要求	标签体系应互斥且完备；采用标准的标注体系并明确实体类型； 关于生成任务的标签：答案应该具备事实准确性、流畅性、相关性； 对于同一输入，提供多个高质量输出。				

文本数据格式	.txt,.pdf	.txt,.pdf	.txt,.pdf	.txt,.pdf
	文档编码格式: UTF-8			
标签数据格式	.txt,.csv,.json,.jsonl	.txt,.csv,.json,.jsonl	.txt,.csv,.json,.jsonl	.txt,.csv,.json,.jsonl

在对文化数据的数据兼容性评价过程中,具体分值可采用表 16 的建议的方式进行确定。

表 16 数据兼容性评分指南

评分指标	指标释义	分值			
		0~10 分	11~40 分	41~70 分	71~100 分
7.2 数据兼容性	评估文化数据在不同操作系统、软件平台、硬件设备间的读取、解析、调用能力,反映数据在跨场景、跨工具应用中的适配性与易用性。	与主流软件和平台兼容性低于 60%,转换需依赖定制化技术方案或专业技术团队支持,技术难度极大,转换后易出现严重数据损耗,甚至部分格式无法实现有效转换以满足常规应用需求。	与 60%-80%主流软件和平台兼容,需借助专业转换工具并具备基础技术能力完成转换,转换后可能存在轻微数据损耗,且转换效率中等、操作门槛较高。	与 90%以上主流软件和平台兼容,仅在少数小众工具或特殊硬件上需通过常规转换工具进行简单转换,转换过程耗时短、操作便捷,且无明显数据质量损耗。	与各类主流操作系统、硬件设备完美兼容,无需任何格式转换即可直接读取、解析与调用,无数数据损耗或功能限制。

在对文化数据的数据处理难度评价过程中,具体分值可采用表 17 的建议的方式进行确定。

表 17 技术可行性—数据处理难度评分指南

评分指标	指标释义	分值			
		0~10 分	11~40 分	41~70 分	71~100 分
7.3 数据处理难度	衡量对文化数据进行关键信息提取、内容加工、形态转换等操作的技术难易程度,涵盖图片类数据的元素识别(如纹样、场景)、音视频类数据的语音转写与内容拆分、电子文本类数据的分词与关键词抓取、三维文件类数据的模型拆解与材质分析等操作,反映数据后续开发利用过程中技术	极难处理。图片关键元素无法自动提取,音视频处理准确率<70%,电子文本分词受严重干扰(如乱码、方言俚语密集),三维文件模型拆分需顶级专业工具及资深技术人员,且难以实现完整拆分。	较难处理。图片提取关键元素需人工标注占比 30%以上,音视频处理准确率 70%-89%,电子文本分词受明显干扰,三维文件模型拆分需专业工具及技术人员操作。	较易处理。图片提取关键元素需少量人工校准,音视频处理准确率达 90% 以上,电子文本分词受轻微干扰,三维文件模型拆分需简单工具辅助。	极易处理。图片可自动精准提取关键元素,音视频可实时完成语音识别/字幕提取/内容分割,电子文本无干扰分词与关键词提取,三维文件可一键完成模型拆分/尺寸测量/材质提取,无需人工干预。

	实现的便捷性与成本投入水平。				
--	----------------	--	--	--	--

在对文化数据的存储与传输成本评价过程中，具体分值可采用表 18 的建议的方式进行确定。

表 18 技术可行性—存储与传输成本评分指南

评分指标	指标释义	分值			
		0~10分	11~40分	41~70分	71~100分
7.4 存储与传输成本	评估文化数据在长期存储与跨场景传输过程中的资源消耗与经济成本，包括数据存储所需的硬件空间占用量、传输时对网络带宽的最低要求，以及由此产生的设备采购、网络租赁、维护管理等成本，反映数据在规模化应用、远程协作等场景中落地的经济性与可行性。	较高成本。存储所需空间极大，需部署大型存储阵列或专用云存储集群。传输时对网络带宽要求极高，需专线网络或专业网络加速技术保障。	中等成本。存储所需空间大幅提升，需部署中小型存储服务器或租用基础云存储服务。传输时对网络带宽要求较高，需稳定的企业级网络，大文件传输需借助专业工具。	较低成本。存储所需空间较大，需配置中小型存储设备。传输时对网络带宽有一定要求，高峰时段可能出现轻微卡顿，需简单优化网络。	极低成本。存储所需空间小，传输时对网络带宽要求低，无需额外扩容设备与特殊网络配置。

5.2.8 合作体量

合作体量用于衡量文化数据的规模大小及数据集中符合“高质量数据”标准的内容占比，反映数据规模与质量的匹配度，二者共同构成对合作体量的完整评估。具体指标包括：

- a) 数据体量
- b) 高质量数据占比

在对文化数据的合作体量评价过程中，其各类指标的具体分值可采用表 19 的建议的方式进行确定。

表 19 合作体量评分指南

评分指标	指标释义	分值			
		0~10分	11~40分	41~70分	71~100分
8.1 数据体量	以数据量（GB）为核心指标，衡量文化数据的规模大小，反映数据在合作开发、批量应用及规模化推广中的潜力和可行性。	0-99GB，体量极小，常规情况下合作规模受限，仅能支撑零星开发或单次小额授权，几乎无规模性合作价值。	100-4999GB，体量中等偏下至中等，适用于中小型合作场景，如单一文化项目的数字化开发、小范围商业授权、地方级文化	5000-9999GB，体量较大，可满足中大型合作需求，如区域级文化 IP 开发、行业垂直领域的批量内容授权等，规模优势较	10000GB 及以上，体量庞大，可支撑大规模产业级合作，如覆盖全国范围的文化资源数据库、超大型数字藏品合集等，具备

			素材整理等，具备基础规模潜力，规模化应用需结合数据质量优化。	明显。	显著规模效应。
8.2 高质量数据占比	评估文化数据中符合“高质量数据”标准的内容占比，反映数据规模的“质效”，避免“低质数据堆砌”的规模陷阱。	高质量数据占比<20%，低质数据占比高，合作开发需承担大量筛选成本，甚至可能因优质内容不足导致开发失败。	高质量数据占比20%~49%，存在一定比例低质数据。	高质量数据占比50%~69%，质量与规模均衡，具备较好合作潜力。	高质量数据占比≥70%，数据规模与质量匹配度极高，可直接支撑高价值批量开发。

文化数据的高质量数据占比，按公式（1）计算。

$$P = \left(\sum_{i=1}^n a_i \times b_i \right) \times 100\% \quad \dots\dots\dots (1)$$

式中：

P——高质量数据占比， $S \leq 100\%$ ；

n——数据格式的总个数；

a_i ——第 i 个数据格式权重， $\sum_{i=1}^n a_i = 1$ ；

b_i ——第 i 个数据格式的数据质量得分，得分标准参考 7.1 数据质量评分指南表 9 至表 15。

例：某中心有三类数据，视频数据占比 50%，数据质量得分 80 分、图片数据占比 20%，数据质量得分 50 分、文本数据占比 30%，数据质量得分 80 分，即： $P = (50\% \times 80 + 20\% \times 50 + 30\% \times 80) \times 100\% = 74\% > 70\%$ ，即该批数据的高质量数据占比 $\geq 70\%$ ，分值在 71~100 分的评分范围内。

6 评价方法

6.1 指标取值过程

文化数据总体价值得分是各一级指标得分及对应的一级指标权重的乘积之和。各一级指标得分是各二级指标得分与其对应二级指标权重的乘积之和。文化数据总体价值得分，按公式（1）计算。

$$S = \sum_{i=1}^n [W_i \times \left(\sum_{j=1}^{m_i} v_{ij} \times w_{ij} \right)] \quad \dots\dots\dots (1)$$

式中：

S——文化数据总体价值得分， $0 \leq S \leq 100$ ；

n——一级指标的总个数；

W_i ——第 i 个一级指标的权重， $\sum_{i=1}^n W_i=1$ ；

m_i ——第 i 个一级指标下二级指标的总个数；

v_{ij} ——第 i 个一级指标下第 j 个二级指标得分；

w_{ij} ——第 i 个一级指标下第 j 个二级指标权重， $\sum_{j=1}^{m_i} w_{ij}=1$ 。

6.2 等级划分方法

文化数据价值等级可分为五级，从高到低依次为 AAAAA 级、AAAA 级、AAA 级、AA 级、A 级，根据文化数据总体价值得分（S）划分等级。按照表 20 的规定执行。

表 20 文化数据价值划分等级

等级	A	AA	AAA	AAAA	AAAAA
文化数据总体价值得分 (S)	<40	40~<55	55~<70	70~<85	≥85

注：文化数据价值等级使用字母 A 标识价值等级。

6.3 评价过程

6.3.1

评价人员可参照 5.2 给出的评分指南，将文化数据的现状与各评价指标对应的说明描述进行对比，在最契合的等级内，根据水平给出实际的得分。

6.3.2

评价人员根据 6.1 给出的计算方法测算文化数据价值，并形成文化数据评价报告。

6.3.3

评价组由多名评价人员组成，宜采取平均分或合议的方式对多个评价结果进行汇总处理，达成一致的评价结果，并依据表 20 划分的等级区间确定文化数据价值等级。

附录 A

(资料性附录)

文化数据价值评价指标体系取值示例

在针对具体的文化数据进行价值评价时，可自行设定一级指标与二级指标分值的权重，如表 A.1 所示。需要注意的是所有一级指标的全重之和为 100%；在每一类一级指标下，相应的所有二级指标的权重之和为 100%。

表 A.1 文化数据价值评价指标体系取值示例表

一级指标	二级指标	一级指标权重	二级指标权重
1 本体价值		10%	
	1.1 独特性		30%
	1.2 历史性		30%
	1.3 艺术性		20%
	1.4 文化符号性		10%
	1.5 文化多样性		10%
2 功能价值		8%	
	2.1 传承性		35%
	2.2 社会性		25%
	2.3 教育性		25%
	2.4 科学性		15%
3 情感价值		6%	
	3.1 故事性		30%
	3.2 情感共鸣性		30%
	3.3 情感主流性		40%
4 传播价值		6%	
	4.1 传播力		45%
	4.2 知名度		35%
	4.3 系列性		20%
5 经济价值		25%	

	5.1 版权与知识产权		40%
	5.2 收入与盈利模式		20%
	5.3 商业应用潜力		40%
6 市场需求		25%	
	6.1 目标受众规模		40%
	6.2 市场热度与趋势		40%
	6.3 竞争态势		20%
7 技术可行性		10%	
	7.1 数据质量		30%
	7.2 数据兼容性		30%
	7.3 数据处理难度		20%
	7.4 存储与传输成本		20%
8 合作体量		10%	
	8.1 数据体量		60%
	8.2 高质量数据占比		40%