

团 体 标 准

T/CAMIR 003—2022

媒体大数据分类分级指南

Guidelines for media big data classification and grading

2022 - 11 - 16 发布

2022 - 11 - 16 实施

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 媒体大数据分类分级原则	2
4.1 多维分类	2
4.2 多层分级	2
4.3 就高从严	2
4.4 自主分类定级	2
5 媒体大数据分类	2
5.1 分类方法	2
5.2 分类过程	2
5.3 分类维度	2
6 媒体大数据分级	4
6.1 分级方法	4
6.2 分级要素	4
6.3 分级规则	5
6.4 级别变更	5
附录 A（资料性） 媒体大数据分类列表	7
参考文献	9

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。

本文件由中国市场信息调查业协会提出并归口。

本文件起草单位：中国传媒大学、清华大学、广东南方新媒体股份有限公司、杭州凡闻科技有限公司、中科网联数据科技有限公司、北京闻信智远信息科技有限公司、广州数说故事信息科技有限公司、闻客信息（深圳）有限公司、北京易观数智科技股份有限公司、网智天元科技集团股份有限公司、华媒百象传媒科技（北京）有限公司、北京格陵兰瑞信息科技有限公司、北京蓝象标准咨询服务公司。

本文件主要起草人：沈浩、钟茂华、王兵、张燕、李传珍、朱立谷、刘畅、陈法涌、张余、海玲、李旭日、张业军、杨帆、张传文、徐彦超、韩凌、段小莉、马建红、张德保、乔华阳。

媒体大数据分类分级指南

1 范围

本文件提供了媒体大数据分类分级原则、媒体大数据分类和媒体大数据分级的建议。
本文件适用于指导媒体大数据活动中的分类分级。

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 7027—2002 信息分类和编码的基本原则与方法
GB/T 38667—2020 信息技术 大数据 数据分类指南

3 术语和定义

下列术语和定义适用于本文件。

3.1

大数据 big data

具有体量巨大、来源多样、生成极快、且多变等特征，并且难以用传统数据体系结构有效处理的包含大量数据集的数据。

注：国际上，大数据的4个特征普遍不加修饰地直接用volume、variety、velocity和variability予以表述，并分别赋予了它们在大数据语境下的定义：

- a) 体量 volume：构成大数据的数据集的规模；
- b) 多样性 variety：数据可能来自多个数据仓库、数据领域或多种数据类型；
- c) 速度 velocity：单位时间的数据流量；
- d) 多变性 variability：大数据其他特征，即体量、速度和多样性等特征都处于多变状态。

[来源：GB/T 35295—2017，2.1.1]

3.2

媒体大数据 media big data

以互联网、广播电视网和移动网络等为载体，具有大众传播属性的媒体机构在提供信息服务过程中产生的数据。

3.3

媒体大数据活动 media big data activity

组织针对媒体大数据开展的一组特定任务的集合。

注：媒体大数据活动主要包括采集、存储、处理、分发和删除等活动。

3.4

大数据分类 big data classification

根据大数据的属性或特征，将其按一定的原则和方法进行区分和归类，并建立起一定的分类体系和排列顺序的过程。

[来源：GB/T 38667—2020，3.3]

3.5

大数据分级 big data grading

根据大数据的敏感程度和影响程度，将其按遭篡改、破坏、泄露或非法利用后可能造成的危害程度进行级别划分的过程。

注：破坏是指由恶意攻击、存储介质损坏、磁干扰等导致的数据毁坏或丢失。

3.6

分类主体 classification subject

大数据收集、存储、使用、分发、删除等过程中对大数据进行梳理归类的组织或个人。

[来源：GB/T 38667—2020, 3.4]

3.7

分类维度 classification dimension

用于实现分类的数据所具有的某个或某些共同特征。

注：常见数据分类维度包括产生来源、结构化特征、业务归属、处理时效性要求等。

[来源：GB/T 38667—2020, 3.6]

4 媒体大数据分类分级原则

4.1 多维分类

媒体大数据分类宜考虑国家、行业、组织和个人等多个视角，从便于媒体大数据流通、使用和管理

4.2 多层分级

媒体大数据分级宜明确分级要素和分级规则，便于根据不同分级要素的不同组合划定不同级别。

4.3 就高从严

媒体大数据分级宜采用就高不就低的原则，以确保媒体大数据能够得到安全有效的管理和使用。

4.4 自主分类定级

媒体大数据活动相关方宜依据本文件，结合工作实际，自主完成数据分类分级工作。

5 媒体大数据分类

5.1 分类方法

媒体大数据分类建议遵循GB/T 7027—2002中6.3规定的面分类法，并按GB/T 38667—2020中第6章

的要求，结合媒体大数据行业特点和实际需求划分出媒体大数据的分类维度。不同分类主体可以根据实际业务需求，对媒体大数据分类维度进行扩展，也可进行下一级别的细分。媒体大数据分类列表见附录A。

5.2 分类过程

媒体大数据的分类过程见GB/T 38667—2020的第5章。

5.3 分类维度

5.3.1 概述

媒体大数据分类维度宜包括但不限于：

- a) 数据内容来源；
- b) 数据传播主体类型；
- c) 数据内容属性；
- d) 数据结构属性；
- e) 数据加工程度；
- f) 数据安全属性；
- g) 数据来源地域；
- h) 应用场景属性。

5.3.2 按数据来源分类

按数据来源进行分类，宜将媒体大数据分为：

- a) 网站数据：包括新闻网站数据、电商网站数据和视频网站数据等；
- b) 社交媒体数据：包括用户关系类社交媒体数据、信息类社交媒体数据和短视频社交媒体数据等；
- c) 社区数据：包括问答类社区数据、知识类社区数据和电商类社区数据等；
- d) 其他数据。

5.3.3 按数据传播主体类型分类

按数据传播主体类型进行分类，宜将媒体大数据分为：

- a) 官方来源数据：包括政府机构数据、新闻单位数据、企事业单位数据、其他机构官方平台数据和团体官方平台数据等；
- b) 非官方来源数据：包括机构非官方平台数据、团体非官方平台数据和个人数据等。

5.3.4 按数据内容属性分类

按数据内容属性进行分类，宜将媒体大数据分为：

- a) 文本类：包括短文本类、长文本类、命名实体类等；
- b) 图形图像类：包括图形类、图像类等；
- c) 音视频类：包括音频类、视频类等；
- d) 数据类：包括转发量、点赞量和阅读量等；
- e) 其他类。

5.3.5 按数据结构属性分类

按数据结构属性进行分类，宜将媒体大数据分为：

- a) 结构化数据：包括数据、表格等；
- b) 半结构化数据：包括日志、电子邮件等；
- c) 非结构化数据：包括文本、图像、音视频等。

5.3.6 按数据加工程度分类

按数据加工程度进行分类，宜将媒体大数据分为：

- a) 未加工类：包括原创数据、未加工转载数据等；
- b) 简单加工类：包括列表、引用描述等；
- c) 深度加工类：包括分析报告、论文等。

5.3.7 按数据安全属性分类

按数据安全属性进行分类，宜将媒体大数据分为：

- a) 开源信息数据；
- b) 个人隐私数据；
- c) 商业情报数据；
- d) 公共利益数据；
- e) 国家安全数据；
- f) 其他数据。

5.3.8 按数据来源地域分类

按媒体大数据来源地域进行分类，宜将媒体大数据分为：

- a) 境内数据：媒体机构、团体或个人在中华人民共和国境内收集或产生的数据；
- b) 境外数据：媒体机构、团体或个人在中华人民共和国境外收集或产生的数据。

5.3.9 按应用场景属性分类

按媒体大数据应用场景属性进行分类，宜将媒体大数据分为：

- a) 消费大数据；
- b) 社交大数据；
- c) 文娱大数据；
- d) 地理位置大数据；
- e) 传感器大数据；
- f) 健康医疗大数据；
- g) 社会治理大数据；
- h) 其他大数据。

6 媒体大数据分级

6.1 分级方法

数据遭到破坏后可能造成的影响（如可能造成的危害、损失或潜在的风险等）是确定数据级别的重要判断依据。媒体大数据分级宜主要考虑影响对象和影响程度两个分级要素。不同分类主体可根据媒体大数据活动中的实际业务需求，对媒体大数据分级要素进行扩展与细化。

6.2 分级要素

6.2.1 影响对象

影响对象是指媒体大数据遭到破坏后受到影响的对象，包括但不限于国家安全、公共利益、个人合法权益和组织合法权益。

影响对象的确定建议主要考虑以下内容：

- a) 国家安全：一般指在媒体大数据活动中，数据一旦遭到破坏，可能对国家利益、国家意识安全、社会秩序等造成影响；
- b) 公共利益：一般指在媒体大数据活动中，数据一旦遭到破坏，可能对公众的政治权利、经济效益等造成影响；
- c) 个人合法权益：一般指在媒体大数据活动中，数据一旦遭到破坏，可能对个人隐私、个人信息、私人合法活动、私有财产、私有领域和个人健康等造成影响；
- d) 组织合法权益：一般指在媒体大数据活动中，数据一旦遭到破坏，可能对组织的合法运营、声誉形象、公信力等造成影响。

6.2.2 影响程度

影响程度是指媒体大数据遭到破坏后所产生影响的大小，按其程度从高到低宜划分为：严重危害、中等危害、轻微危害和无危害。

影响程度的确定建议主要考虑以下内容。

- a) 严重危害包括：
 - 1) 可能导致危及国家利益或国家安全的重大事件，造成重大损失；
 - 2) 可能严重危害社会秩序和公共利益，引发重大舆情，或导致媒体行业秩序遭到严重破坏等；
 - 3) 可能导致重大个人信息安全风险、侵犯个人合法权益等；
 - 4) 可能导致组织合法权益造成严重损害，或者影响组织重要/关键业务的开展，遭受严重的经济损失。
- b) 中等危害包括：
 - 1) 可能导致危害社会秩序和公共利益，引发区域或一定规模舆情，或导致媒体行业秩序一定程度失序等；
 - 2) 可能导致一定规模的个人信息泄露、滥用等安全风险，对个人政治或经济权益造成部分或一定的影响；

- 3) 可能导致组织合法权益造成一定程度损害,或者导致组织部分业务临时中断,遭受一定的经济损失。
- c) 轻微危害包括可能对个人合法权益、组织合法权益造成轻微或潜在损害,或对媒体行业秩序有一定潜在影响,但不危及或损害国家安全、社会秩序或公共利益。
- d) 无危害包括对国家利益、社会秩序和公共利益、个人合法权益、组织合法权益均无影响。

6.3 分级规则

媒体大数据级别从高到底宜划分为五个级别,即5级、4级、3级、2级和1级。媒体大数据级别划分宜遵循以下规则:

- a) 一旦遭受破坏后将直接影响到国家安全、社会秩序、公共利益与媒体行业稳定的媒体大数据,其级别建议不低于5级;
- b) 建议数据责任主体高度重视包含个人信息相关的数据,并在数据分级过程中宜从高考虑;
- c) 媒体大数据分级宜根据实际产生的影响范围、规模及可控程度等情况及时确定数据级别。按“就高从严”原则,表1给出了媒体大数据分级规则。

表1 媒体大数据分级规则

级别	分级要素		说明
	影响对象	影响程度	
5	国家安全	严重	国家相关保密规范中规定的保密数据,对国家安全、国家经济发展具有重要战略意义,或者影响国家国际竞争力的数据和资料。一般针对特定人员公开,且仅为必须知悉的对象访问或使用。 数据遭到破坏后,影响国家安全,或对公共利益造成严重的影响。
	公共利益	严重	
4	公共利益	中等	未公开发表的,仅行业内极少量机构拥有的媒体大数据流通、使用和管理过程中的非常重要的数据与资料。 数据遭到破坏后,对公共利益造成中等影响,或对相关个人、组织合法权益造成严重的影响,但不影响国家安全。
	个人合法权益	严重	
	组织合法权益	严重	
3	公共利益	轻微	未公开发表的,行业内机构拥有的媒体大数据流通、使用和管理过程中的重要数据与资料。 数据遭到破坏后,对公共利益造成轻微影响,对个人合法权益或组织合法权益造成中等影响,但不影响国家安全。
	个人合法权益	中等	
	组织合法权益	中等	
2	个人合法权益	轻微	未公开发表的、能够通过多种渠道申请获取的媒体大数据流通、使用和管理过程中的一般性数据与资料。 数据遭到破坏后,对个人合法权益或组织合法权益造成轻微影响,但不影响国家安全和公共利益。
	组织合法权益	轻微	
1	国家安全	无危害	参加公开出版发行的媒体大数据流通、使用和管理过程中的数据与资料。 数据遭到破坏后,可能对组织合法权益造成潜在的影响,但不影响国家安全、公共利益及个人合法权益。
	公共利益		
	个人合法权益		
	组织合法权益		

6.4 级别变更

6.4.1 变更原则

媒体大数据级别变更宜遵循以下原则：

- a) 从原始媒体大数据汇编、复制出来的新数据级别建议不高于原有最高数据级别；
注：若原始媒体大数据包含多个不同等级，则将这些原始媒体大数据进行汇编、复制之后的新数据级别不宜高于多个等级中的最高级别。
- b) 从多个数据来源进行加工、处理和分析的新数据建议不低于原有数据级别；
- c) 以媒体大数据作为内容生产的核心要素提供信息服务时，宜根据新数据的关键要素重新判定级别；
- d) 数据内容因环境发生变化时，宜明确变更依据并及时变更级别；
- e) 媒体大数据信息服务对象发生改变时，宜重新判定级别；
- f) 媒体大数据级别变更时，宜建立相应的审批机制。

6.4.2 变更类型

6.4.2.1 等级提升

媒体大数据发生以下情况时，宜考虑提升数据级别：

- a) 聚合多个数据来源的媒体大数据；
- b) 媒体大数据活动涉及舆情行业和国家利益；
- c) 媒体大数据活动涉及境外、港澳台地区的信息服务；
- d) 经过时间积累，可能增加危害程度的媒体大数据；
- e) 发生特定事件导致数据影响对象范围扩大。

6.4.2.2 等级降低

媒体大数据发生以下情况时，宜考虑降低数据级别：

- a) 数据已被媒体公开报道或披露；
- b) 数据进行不同等级的脱敏处理后，可考虑相应降低数据级别；
- c) 数据经过较长时间沉淀，已无明确数据含义和时间点；
- d) 发生特定事件导致数据影响对象范围缩小。

附 录 A
(资料性)
媒体大数据分类列表

媒体大数据分类列表见表A.1。

表A.1 媒体大数据分类列表

大类	中类	小类
数据内容来源	网站数据	新闻网站数据
		电商网站数据
		视频网站数据
		其他网站数据
	社交媒体数据	用户关系类社交媒体数据（如微信）
		信息类社交媒体数据（如微博、播客）
		短视频类社交媒体数据（如抖音、快手）
		其他社交媒体数据
	社区数据	问答类社区数据（如百度贴吧、知乎）
		知识类社区数据（如豆瓣）
		电商类社区数据（如小红书）
		其他类社区数据
	其他数据	——
数据传播主体类型	官方来源数据	政府机构数据
		新闻单位数据
		企事业单位数据
		其他机构官方平台数据
		团体官方平台数据
		其他官方来源数据
	非官方来源数据	机构非官方平台数据
		团体非官方平台数据
		个人数据
		其他非官方来源数据
数据内容属性	文本类	短文本类（如事件话题、关键词、搜索语句、问答文本、评论文本等）
		长文本类（如文章、新闻）
		命名实体类（如姓名、机构名称）
		其他文本类
	图形图像类	图形类（如动画、计算机生成图形）
		图像类（如照片、合成图像）
		其他图形图像类
	音视频类	音频类（如语音、音乐、背景音）
		视频类（如短视频、长视频、合成视频）
		其他音视频类
	数据类	转发量
		点赞量
		阅读量

		其他数据
	其他数据	---
数据结构属性	结构化数据	数据
		表格
		其他结构化数据
	半结构化数据	日志
		电子邮件
		其他半结构化数据
	非结构化数据	文本
		图像
		长视频
		短视频
语音		
	其他非结构化数据	
数据加工程度	未加工类	原创数据
		未加工转载数据
		其他未加工数据
	简单加工类	列表
		引用描述
		其他简单加工数据
	深度加工类	分析报告
		论文
		其他深度加工数据
数据安全属性	开源信息数据	----
	个人隐私数据	
	商业情报数据	
	公共利益数据	
	国家安全数据	
	其他数据	
数据来源地域	境内数据	媒体机构、团体或个人在中华人民共和国境内收集或产生的数据
	境外数据	媒体机构、团体或个人在中华人民共和国境外收集或产生的数据
应用场景属性	消费大数据	----
	社交大数据	
	文娱大数据	
	地理位置大数据	
	传感器大数据	
	健康医疗大数据	
	社会治理大数据	
	其他大数据	

参 考 文 献

- [1] GB/T 20527.1—2006 多媒体用户界面的软件人类工效学 第1部分：设计原则和框架
 - [2] GB/T 29804—2013 信息技术 学习、教育和培训 平台与媒体分类代码
 - [3] GB/T 35295—2017 信息技术 大数据 术语
 - [4] GB/T 37973—2019 信息安全技术 大数据安全管理指南
 - [5] 中华人民共和国网络安全法
 - [6] 中华人民共和国数据安全法
 - [7] 中华人民共和国个人信息保护法
 - [8] 网络安全标准实践指南——网络数据分类分级指引
-