

GY

中华人民共和国广播电视和网络视听行业标准

GY/T XXX. 2—XXXX

网络视听收视大数据技术规范 第2部分：数据元素集

Technical specification for internet video and audio service big data—
Part 2: Data element set

(报批稿)

XXXX—XX—XX 发布

XXXX—XX—XX 实施

国家广播电视总局

发布

目 次

前言	III
引言	IV
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 缩略语	1
5 概述	2
6 用户基本数据	2
6.1 用户信息	2
6.2 设备信息	3
7 媒资数据	3
7.1 节目 ID ContentID	3
7.2 节目一级类别 ContentType1	3
7.3 节目二级类别 ContentType2	3
7.4 标题 Title	3
7.5 副标题 Subtitle	3
7.6 导演 Director	3
7.7 编剧 Scriptwriter	3
7.8 主要演员 Actor	3
7.9 摘要 Abstract	4
7.10 国家或地区 Area	4
7.11 语言 Language	4
7.12 年份 Year	4
7.13 技术格式 TechnicalSpecification	4
7.14 节目来源 Source	4
7.15 更新状态 UpdateStatus	4
7.16 上线日期 AddDate	4
7.17 更新日期 UpdateDate	4
7.18 节目时长 Duration	4
7.19 剧集 ID SeriesID	4
7.20 子集 ID EpisodeID	4
7.21 剧集数 EpisodeCount	4
7.22 当前剧集所处序号 EpisodeSerialNumber	4
7.23 发布者 ID UploadID	5
7.24 节目付费类型 PayType	5

7.25	节目播放 URL	PlayURL	5
7.26	节目关联 URL	RelatedURL	5
7.27	溯源频道名称	ChannelName	5
7.28	播出日期	PlayDate	5
7.29	播出时间	PlayTime	5
8	用户行为数据		5
8.1	节目播放		5
8.2	播放行为		6
8.3	页面访问		7

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。

本文件为GY/T XXX《网络视听收视大数据技术规范》的第2部分。GY/T XXX已经发布了以下部分：

- 第1部分：总体要求；
- 第2部分：数据元素集；
- 第3部分：接口。

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。

本文件由全国广播电影电视标准化技术委员会（SAC/TC 239）归口。

本文件起草单位：国家广播电视总局广播电视规划院、浙江传媒学院、华数传媒网络有限公司、华数数字电视传媒集团有限公司、湖南快乐阳光互动娱乐传媒有限公司、未来电视有限公司。

本文件主要起草人：李忠炤、郑冠雯、聂明杰、李庆国、王志豪、张玮、李鸣、唐志燕、遇琪、俞定国、诸葛海标、杨阳、殷楚冬、黄瑞婷、景生军、邱波、胡晔宸、刘晓敏、邱文状、葛进平、王祎。

引 言

考虑到规范内容、应用对象和场景不同，《网络视听收视大数据技术规范》分为总体要求、数据元素集和接口3个部分。

《网络视听收视大数据技术规范 第1部分：总体要求》规定了网络视听收视大数据数据采集、数据处理平台、数据开放和服务、数据安全等方面的要求，适用于网络视听收视大数据系统和业务的规划、设计、实施、升级改造和运行维护。

《网络视听收视大数据技术规范 第2部分：数据元素集》根据网络视听节目收视综合评价需要，规定了网络视听收视调查中的数据元素。网络视听平台采集网络视听数据，应至少包含该文件规定的元素。

《网络视听收视大数据技术规范 第3部分：接口》针对数据提供方数据采集手段、收视数据源的差异，规定了具有共性的收视数据元素的交换接口，规范数据接收方和数据提供方以统一的数据格式、消息格式和传输协议交换网络视听收视大数据。

网络视听收视大数据技术规范 第2部分：数据元素集

1 范围

本文件规定了网络视听收视大数据中的数据元素。

本文件适用于互联网视听服务大数据系统和业务的规划、设计、实施、升级改造和运行维护。

2 规范性引用文件

本文件没有规范性引用文件。

3 术语和定义

3.1

用户 user

在互联网视听服务平台收听收视视音频内容的自然人或机构。

3.2

节目 program

在互联网视听平台上供用户收听收视、具有独立主题或意义的视音频内容，包括新闻、电视剧、电影、综艺、网络剧、网络大电影及其片段等长视频、短视频，以及直播内容。

3.3

直播 direct broadcast

声音和（或）图像节目不经预先录制而直接播出的过程。

3.4

操作 operation

用户在互联网视听平台上收听收视视音频内容时，开始播放、倍速、暂停、退出、情节选择和下载等行为，仅针对在线观看时的实时行为，不包括下载后的观看行为。

3.5

用户基本数据 user data

用于描述和识别用户的一组信息。

3.6

媒资数据 media asset data

用于描述和识别媒体内容的一组信息。

3.7

用户行为数据 User operating data

用户操作和评论的连续数据，以用户为单位，精确到秒。

4 缩略语

下列缩略语适用于本文件。

APP 应用程序 (Application)

ID 标识号 (Identification)

IDFA 广告标识符 (Identifier For Advertising)

IMEI 国际移动设备识别码 (International Mobile Equipment Identity)

iOS iPhone操作系统 (iPhone Operating System)

NAT 网络地址转换 (Network Address Translation)

OS 操作系统 (Operating System)

PC 个人计算机 (Personal Computer)

TVOS 电视操作系统 (Television Operating System)

UDID 唯一设备识别符 (Unique Device Identifier)

UUID 通用唯一识别码 (Universally Unique Identifier)

5 概述

数据元素集为互联网视听平台中各类元数据的集合。根据网络视听节目收视综合评价需要,网络视听收视大数据数据元素集应包括用户基本数据、媒资数据和用户行为数据相关数据元素。

网络视听平台采集网络视听数据,应至少包含该文件规定的元素。

6 用户基本数据

6.1 用户信息

6.1.1 账号 Account

用于用户在互联网视听平台中代表自己身份的名称。

6.1.2 用户 ID User ID

系统自动生成的用于唯一识别用户的一组不重复的编码。

6.1.3 账号类型 AccountType

用于区分不同的用户属性,如个人普通账户、企业账户、政府账户等。

6.1.4 创建时间 UserCreateTime

在互联网视听平台创建用户账号时的时间戳。

6.1.5 用户区域标识 UserRegionID

用户注册区域编码,标识用户所属区域。

6.1.6 用户年龄 UserAge

用户注册年龄信息。

6.1.7 用户性别 UserSex

用户注册性别信息。

6.2 设备信息

6.2.1 终端 ID DeviceID

能代表用户设备唯一性的可采集标识码。

6.2.2 终端类型 DeviceType

用户收视收听节目时所使用的终端类型，如智能电视、计算机、平板电脑、手机等。

6.2.3 终端 OS DeviceOS

用户所使用终端设备的操作系统，如 Windows、Android、iOS、TVOS 等。

6.2.4 终端 IP DeviceIP

终端设备实际获得并在网络中使用的外网 IP 地址。

6.2.5 版本号 AppVersion

互联网视听平台的用户在观看节目网络资料的时候，使用的播放软件版本号，如 APP 版本号，PC 客户端版本号等。

7 媒资数据

7.1 节目 ID ContentID

用于唯一标识网络视听节目的一组编码。

7.2 节目一级类别 ContentType1

互联网视听平台的节目所属于的一级分类，如电影、电视剧、综艺等。

7.3 节目二级类别 ContentType2

互联网视听平台的节目所属于的二级分类，如动作、爱情等。

7.4 标题 Title

网络视听节目的名称。

7.5 副标题 Subtitle

在一个数据提供方内部，对于发布在互联网视听平台上面的每一个节目，用副标题对其标题进行说明。

7.6 导演 Director

节目导演的姓名。

7.7 编剧 Scriptwriter

节目编剧的姓名。

7.8 主要演员 Actor

节目主演的姓名。

7.9 摘要 Abstract

网络视听节目的简要描述。

7.10 国家或地区 Area

网络视听节目制作方所属的地区。

7.11 语言 Language

网络视听节目的语言，如中文、英语、日语等。

7.12 年份 Year

网络视听节目发行或出品的年份。

7.13 技术格式 TechnicalSpecification

网络视听节目的技术参数。

7.14 节目来源 Source

网络视听节目的来源，包括用户上传、平台上传、第三方接入等方式。

7.15 更新状态 UpdateStatus

网络视听节目更新进度。

7.16 上线日期 AddDate

网络视听节目资料上线的日期。

7.17 更新日期 UpdateDate

网络视听节目被上传至平台之后，最近一次更新的日期。

7.18 节目时长 Duration

网络视听节目的时间长度。

7.19 剧集 ID SeriesID

互联网视听平台的节目如果是连续剧、纪录片等连载形式，则为可唯一标识该剧的一条编码。

7.20 子集 ID EpisodeID

互联网视听平台的节目如果是连续剧、纪录片等连载形式，则为可唯一标识每一个分集的一条编码。

7.21 剧集数 EpisodeCount

网络视听节目剧集的子集数量。

7.22 当前剧集所处序号 EpisodeSerialNumber

互联网视听平台提供的多集连载形式的剧集，标识出本集的排序。

7.23 发布者 ID UploadID

代表网络视听节目上传者的一组编码。

7.24 节目付费类型 PayType

标识网络视听节目是否付费及其付费方式。

7.25 节目播放 URL PlayURL

网络视听节目播放的 URL 地址。

7.26 节目关联 URL RelatedURL

用户可以进入节目播放 URL 的关联页面，如首页、分类推荐页面等。

7.27 溯源频道名称 ChannelName

网络视听节目如果来自于电视直播，需要标识该电视节目首播（直播）时的频道名称。

7.28 播出日期 PlayDate

网络视听节目如果来自于电视直播，需要标识该电视节目首播（直播）时的播出日期。

7.29 播出时间 PlayTime

网络视听节目如果来自于电视直播，需要标识该电视节目首播（直播）时的播出时间。

8 用户行为数据

8.1 节目播放

8.1.1 账号 Account

用于用户在互联网视听平台中代表自己身份的名称。

8.1.2 终端 ID DeviceID

能代表用户设备唯一性的可采集标识码。

8.1.3 用户 ID UserID

系统自动生成的用于唯一识别用户的一组不重复的编码。

8.1.4 版本号 AppVersion

互联网视听平台的用户在观看节目网络资料的时候，使用的播放环境版本号，如 APP 版本号，PC 客户端版本号等。

8.1.5 节目 ID ContentID

用于唯一标识互联网视听平台节目的一组编码。

8.1.6 标题 Title

互联网视听平台的节目资料名称。

8.1.7 终端 IP DeviceIP

终端设备实际获得并在网络中使用的外网 IP 地址。

8.1.8 节目一级类别 ContentType1

互联网视听平台的节目所属于的一级分类，如电影、电视剧、综艺等。

8.1.9 节目二级类别 ContentType2

互联网视听平台的节目所属于的二级分类，如动作、爱情等。

8.1.10 节目时长 Duration

网络视听节目的时间长度。

8.1.11 记录时间 RecordTime

互联网视听平台用户在观看节目的时候，会在平台生成观看记录，可以标识其在平台生成记录时的时间戳。

8.1.12 地区 Province

互联网视听平台用户在观看节目的时候，其所处于的地区，以直辖市、省份为单位，如北京市、浙江省等。

8.1.13 播放时长 PlayDuration

互联网视听平台用户观看节目的总时长。

8.2 播放行为

8.2.1 账号 Account

用于用户在互联网视听平台中代表自己身份的名称。

8.2.2 终端 ID DeviceID

能代表用户设备唯一性的可采集标识码。

8.2.3 用户 ID UserID

系统自动生成的用于唯一识别用户的一组不重复的编码。

8.2.4 播放操作类型 PlayOperateType

用户节目播放和页面操作行为，包括拖拽、暂停、倍速播放、情节选择、评论、弹幕、收藏、下载、分享、顶、踩等。

8.2.5 播放操作时间 PlayOperateTime

用户行为产生时，可以唯一标识行为发生时间的系统时间戳。

8.2.6 播放操作描述 PlayOperateDescribe

用户行为发生时，描述行为相关参数，如拖拽行为发生时，用户松开鼠标停止拖拽的时间点；倍速播放时，用户选择的播放速度等。

8.3 页面访问

8.3.1 账号 Account

用于用户在互联网视听平台中代表自己身份的名称。

8.3.2 终端 ID DeviceID

能代表用户设备唯一性的可采集标识码。

8.3.3 用户 ID UserID

系统自动生成的用于唯一识别用户的一组不重复的编码。

8.3.4 版本号 AppVersion

互联网视听平台的用户在浏览节目网络资料所在页面的时候，使用的播放环境版本号，如 APP 版本号，PC 客户端版本号等。

8.3.5 链接地址 PageURL

用户浏览页面的 URL 地址。

8.3.6 页面操作类型 PageOperateType

用户页面浏览操作，包括搜索、点击链接等。

8.3.7 前一链接地址 ReferURL

用户跳转至当前页面前，所浏览页面的 URL 地址。

8.3.8 记录时间 RecordTime

用户页面访问记录生成时间。
