

GY

中华人民共和国广播电视和网络视听行业标准

GY/T XXX.3—XXXX

网络视听收视大数据技术规范

第3部分：接口

Technical specification for internet video and audio service big data—
Part 3: Interface

(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	II
引言	III
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 缩略语	2
5 概述	2
6 接口技术实现方式	3
7 数据交换格式	4
7.1 数据提供方基本信息	4
7.2 用户数据	5
7.3 媒资数据	6
7.4 用户行为数据	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。

本文件为GY/T XXX《网络视听收视大数据技术规范》的第3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、湖南快乐阳光互动娱乐传媒有限公司、华数传媒网络有限公司、浙江传媒学院、广东南方新媒体股份有限公司。

本文件主要起草人：李忠炤、郑冠雯、聂明杰、曹志、李庆国、黄元浩、刘琛良、王志豪、姜蕾、景生军、葛进平、杨阳、诸葛海标、张亦弛、张元迪、王祎、俞定国、邱波、刘晓敏。

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、数据元素据和接口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3部分：接口

1 范围

本文件规定了网络视听收视大数据提供方对外输出收听收视数据时应遵循的数据格式及接口规范。本文件适用于互联网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。

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 2260—2007 中华人民共和国行政区划代码

GB/T 7408—2005 数据元和交换格式 信息交换 日期和时间表示法（ISO 8601:2000，IDT）

IETF RFC 3629 UTF-8，ISO 10646的一种转换格式（UTF-8，a transformation format of ISO 10646）

IETF RFC 8259 JSON数据交换格式（The JavaScript Object Notation (JSON) Data Interchange Format）

国家新闻出版广电总局. 地级以上广播电视播出机构及频道频率名录(截至2017年底) <http://www.sapprft.gov.cn/sapprft/govpublic/9770/357800.shtml>

国家新闻出版广电总局. 县级广播电视播出机构名录(截至2017年底) <http://www.sapprft.gov.cn/sapprft/govpublic/9770/357798.shtml>

国家新闻出版广电总局. 付费频道名录(截至2017年11月) <http://www.sapprft.gov.cn/sapprft/govpublic/9771/353860.shtml>

3 术语和定义

下列术语和定义适用于本文件。

3.1

用户 user

在互联网视听服务平台收听收看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看、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)

HTTP 超文本传输协议 (Hypertext Transfer Protocol)

HTTPS 安全的超文本传输协议 (Hyper Text Transfer Protocol over SecureSocket Layer)

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)

IP 网际互联协议 (Internet Protocol)

JSON JavaScript对象标记 (JavaScript Object Notation)

NAT 网络地址转换 (Network Address Translation)

PC 个人计算机 (Personal Computer)

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

URL 统一资源定位符 (Uniform Resource Locator)

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

5 概述

本文件规定了互联网视听收视大数据提供方对外输出收听收视数据时应遵循的数据格式及接口规范。接口的应用场景示意图见图1。

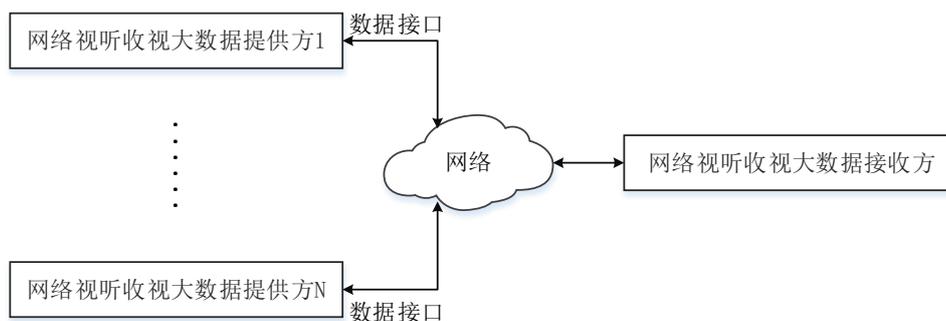


图1 接口应用场景示意图

互联网视听收视大数据提供方通常为单一平台或单个区域的互联网视听收视大数据采集机构，互联网视听收视大数据接收方通常为互联网视听收视大数据汇聚、处理机构。

互联网视听收视大数据提供方可以有多个，可以相互独立地与互联网视听收视大数据接收方进行数据交换，数据提供方之间也可以直接交换数据。

数据提供方通过本文件定义的接口和传输协议交换用户互联网视听收视大数据时，交换方式应支持实时交换模式、准实时交换模式和非实时交换模式：

——实时交换模式下，数据交换双方通过网络采用 HTTPS 或 HTTP 传输协议进行数据实时交互和传输；

——准实时交换模式下，数据提供方与数据接收方进行协商，双方在规定的时间内将一段时间内生成的收视数据通过网络采用 HTTPS 或 HTTP 传输协议进行数据的传输；

——非实时交换模式下，数据提供方将收视数据按照本文件定义的格式，存储成数据文件并通过 FTP 等非实时网络传输或者其他物理介质存储的方式进行交换。

数据交换时应遵循以下原则：

——真实性原则：数据提供方应确保提供交换数据的客观性和真实性；

——代表性原则：交换的数据应是全部数据或者是经过抽样的对全部数据具有代表性的数据子集；

——用户信息保密原则：应确保交换数据中个人信息得到有效保护。

6 接口技术实现方式

实时交换和准实时交换模式下，接口宜采用 RESTFUL 或 RPC 等方式。

接口中的消息应以 IETF RFC 8259 规定的 JSON 字符串进行描述，字符编码应符合 IETF RFC 3629 定义的 UTF-8。

消息结构如下。

```
{
  "object": {
    "key1": "string",
    "key2": true,
    "array1": ["elem1", "elem2"],
    "array2": [
      {
        "key1": "value1"
      }, {
        "key2": "value2"
      }
    ]
  }
}
```

互联网视听收视数据传输请求由互联网视听收视大数据提供方发起，接口流程如图2所示。

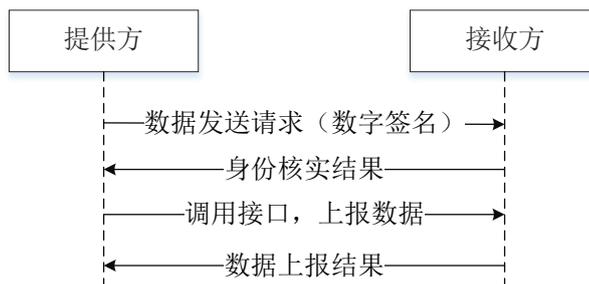


图 2 接口通用流程

互联网视听收视数据传输接口名称如下，URI示例为https://<IP>:<Port>/User_Data:

——数据提供方基本信息上报接口: Provider

——用户基本数据上报接口: User? ProviderID

——媒资数据上报接口: Media_Asset? ProviderID=? &Action=?

——用户行为数据上报接口: Operating? ProviderID

互联网视听收视数据上报响应消息见表1。

表 1 互联网视听收视数据上报响应消息

消息内容项	消息内容项描述
响应状态码	200
响应数据	{ “code” :” 000” ,//返回代码, 000 为数据上报成功 “details” :” GUID” }
code 字段定义	000 数据上报成功 101 数据已经上报过 301 数据有误 401 其他

7 数据交换格式

7.1 数据提供方基本信息

数据进行交换传输时，应先传输数据提供方的基本信息。

数据提供方的基本信息应符合表 2 的规定。

表 2 数据提供方的基本信息

参数名称	字段	类型	参数说明
机构名称	ProviderName	STRING	自定义
机构代码	ProviderID	STRING	18 位，为数据提供方的统一社会信用代码
传输时间	TransTime	STRING	按 GB/T 7408—2005 “基本格式” 中的 “完全表示” 方法，数据格式为：yyyymmddhhmmss

7.2 用户数据

7.2.1 用户信息

用户信息应包含的参数及其取值应符合表 3 的规定。

表 3 用户信息数据应包含的参数及其取值

序号	参数名称	字段	类型	参数说明	备注
1	账号	Account	STRING	用于用户在互联网视听平台中代表自己身份的名称。	必选
2	用户 ID	UserID	STRING	系统自动生成的用于唯一识别用户的一组不重复的编码。	必选
3	账号类型	AccountType	STRING	用于区分不同的用户属性，如个人普通账户、UGC 账户、企业账户、政府账户等。	必选
4	创建时间	UserCreateTime	STRING	在互联网视听平台创建用户账号时的时间戳。 按 GB/T 7408—2005 “基本格式” 中的 “完全表示” 方法，为 yyyyymmdd，如 20180612	必选
5	用户区域标识	UserRegionID	STRING	用户注册区域编码，标识用户所属区域。 使用 6 位十进制数字，应符合 GB/T 2260—2007 的规定。	必选
6	用户年龄	UserAge	UINT	用户注册年龄信息。	必选
7	用户性别	UserSex	UINT	用户注册性别信息。	必选

用户信息数据的JSON消息示例如下：

```
{
  "Account": "12d41345039a11eaac64525400f3186f",
  "UserID": "60215739463920000483790",
  "AccountType": "12",
  "UserCreateTime": "20180117142334",
  "VIPType": "注册会员",
  "VIPActiveTime": "\N",
  "UserRegionID": "玉林市",
  "UserAge": "\26",
  "UserSex": "\male"
}
```

7.2.2 设备信息

设备信息应包含的参数及其取值应符合表 4 的规定。

表 4 设备信息数据应包含的参数及其取值

序号	参数名称	字段	类型	参数说明	备注
1	终端 ID	DeviceID	STRING	能代表用户设备唯一性的可采集标识码，如设备的 IMEI、UDID、UUID、IDFA，或者浏览器的 Cookie。 设备终端 ID 优先采用 UDID，其次采用 UUID，如 UDID、UUID 都无法提供，可采用 IMEI 或 IDFA。 终端 ID 取值为“标识码类型/标识码”，如“UDID/bb4d786633053bcdda4gh7”。	必选
2	终端类型	DeviceType	STRING	用户收听收视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。	必选
3	终端 OS	DeviceOS	STRING	用户所使用终端设备的操作系统，如 Windows、Android、iOS 等。	必选
4	终端 IP	DeviceIP	STRING	终端设备实际获得并在网络中使用的外网 IP 地址。	必选
5	版本号	AppVersion	STRING	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。	必选

设备信息数据的JSON消息示例如下：

```
{
  "DeviceID": "03078Q980001A9AE2FF7F",
  "DeviceType": "03078Q980",
  "DeviceOS": "Android ",
  "DeviceIP": "47.98.74.54",
  "AppVersion": "A02B15"
}
```

7.3 媒资数据

节目资源数据参数及其取值应符合表 5 的规定。

表 5 节目资源数据应包含的参数及其取值

序号	参数名称	字段	类型	参数说明	备注
1	节目 ID	ContentID	STRING	在一个数据提供方内部，可以唯一标识互联网视听平台上供用户观看的节目的一条编码，由数字组成。	必选

表5（续）

序号	参数名称	字段	类型	参数说明	备注
2	节目一级类别	ContentType1	STRING	互联网视听平台所提供的可供用户观看的节目所属于的一级分类，如电影、电视剧、综艺等。	必选
3	节目二级类别	ContentType2	STRING	互联网视听平台所提供的可供用户观看的节目所属于的二级分类，如动作、爱情等。	必选
4	标题	Title	STRING	在一个数据提供方内部，可以唯一标识互联网视听平台上供用户观看的节目的一条编码。	必选
5	副标题	Subtitle	STRING	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。	可选
6	导演	Director	STRING	节目导演的姓名。	必选
7	编剧	Scriptwriter	STRING	节目编剧的姓名。	必选
8	主要演员	Actor	STRING	节目主演的姓名。	必选
9	摘要	VodAbstract	text	网络视听节目的简要描述。	必选
10	国家或地区	Area	STRING	网络视听节目制作方所属的地区。	必选
11	语言	Language	STRING	网络视听节目的语言，如中文、英语、日语等。	必选
12	年份	Year	UINT	网络视听节目发行或出品的年份。	必选
13	技术格式	TechnicalSpecification	STRING	网络视听节目的技术参数。	必选
14	节目来源	Source	STRING	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。	必选
15	更新状态	UpdateStatus	UINT	网络视听节目更新进度。取值如下，默认为“1”： 1：是； 2：否。	必选
16	上线日期	AddDate	STRING	网络视听节目资料上线的日期。 按 GB/T 7408—2005 “基本格式” 中的“完全表示”方法，为 yyyyymmdd，如 20180612	必选
17	更新时间	UpdateDate	STRING	网络视听节目被上传至平台之后，最近一次更新的日期。 按 GB/T 7408—2005 “基本格式” 中的“完全表示”方法，为 yyyyymmdd，如 20180612	必选

表 5（续）

序号	参数名称	字段	类型	参数说明	备注
18	节目时长	Duration	STRING	网络视听节目的时间长度。 按 GB/T 7408-2005“基本格式”中的“完全表示”方法，数据格式为“hhmmss”	必选
19	剧集 ID	SeriesID	UINT	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。	必选
20	子集 ID	EpisodeID	STRING	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。	必选
21	剧集数	EpisodeSerialNumber	UINT	网络视听节目剧集的子集数量。	必选
22	当前剧集所处序号	EpisodeSerialNumber	UINT	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。	必选
23	发布者 ID	UploadID	STRING	代表网络视听节目上传者的一组编码。	必选
24	节目付费类型	PayType	UINT	标识网站所提供的节目是否付费及其付费方式，取值如下： 1：免费； 2：VIP 免费； 3：用券； 4：按次付费； 5：其他。	必选
25	节目播放 URL	PlayURL	STRING	网络视听节目播放的 URL 地址。	必选
26	节目关联 URL	RelatedURL	STRING	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。	必选
27	溯源频道名称	ChannelName	STRING	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。 电视频道采用行政主管部门批准的“台名”+“节目设置”作为频道名称，台名和节目设置见《地级以上广播电视播出机构及频道频率名录（截至 2017 年底）》《县级广播电视播出机构名录（截至 2017 年底）》，如“中央电视台综合频道”“上海广播电视台东方卫视”；付费频道采用《付费频道名录（截至 2017 年 11 月）》中的频道名称，如“风云音乐”。	必选

表 5（续）

序号	参数名称	字段	类型	参数说明	备注
28	播出日期	PlayDate	STRING	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。 按 GB/T 7408—2005 “基本格式” 中的“完全表示”方法，为 yyyyymmdd，如 20180612。	必选
29	播出时间	PlayTime	STRING	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。 按 GB/T 7408—2005 “基本格式” 中的“完全表示”方法，为 hhmmss，如 133300。	必选

媒资数据的JSON消息示例如下：

```
{
  "ContentID": "CP23010020171209062900",
  "ContentType1": "电影",
  "ContentType2": "恐怖悬疑",
  "Title": "空中危机",
  "Subtitle": "\N",
  "Director": "罗伯特",
  "Scriptwriter": "彼得",
  "Actor": "朱迪·福斯特彼得·萨斯加德肖恩·宾",
  "Abstract": "带领观众揭开飞机每一个构造的同时，揭开了一个惊天的阴谋……",
  "Area": "美国",
  "Language": "英语",
  "Year": "2005",
  "TechnicalSpecification": "高清",
  "Source": "\N",
  "UpdateStatus": "1",
  "AddDate": "20171216",
  "UpdateDate": "20171216",
  "Duration": "004530",
  "SeriesID": "074563",
  "EpisodeID": "785342",
  "EpisodeCount": "2",
  "EpisodeSerialNumber": "1",
  "UploadID": "张三",
}
```

```

"PayType": "3",
"PlayURL": " XIYMyi1BLLYDA0oAe0QseMVM_wdm_13662659_1574027322582 ",
"RelatedURL": "\N",
"ChannelName": "\N",
"PlayDate": "\N",
"PlayTime": "\N"
}
    
```

7.4 用户行为数据

7.4.1 节目播放数据

节目播放数据应包含的参数及其取值应符合表 6 的规定。

表 6 节目播放数据应包含的参数及其取值

序号	参数名称	字段	类型	参数说明	备注
1	账号	Account	STRING	于用户在互联网视听平台中代表自己身份的名称。	可选
2	终端 ID	DeviceID	STRING	能代表用户设备唯一性的可采集标识码。	必选
3	用户 ID	UserID	STRING	系统自动生成的用于唯一识别用户的一组不重复的编码。	必选
4	版本号	AppVersion	STRING	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。	必选
5	节目 ID	ContentID	STRING	用于唯一标识互联网视听平台节目的一组编码。	必选
6	标题	Title	STRING	互联网视听平台的节目资料名称。	必选
7	终端 IP	DeviceIP	STRING	终端设备实际获得并在网络中使用的外网 IP 地址。	必选
8	节目一级类别	ContentType1	STRING	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。	可选
9	节目二级类别	ContentType2	STRING	互联网视听平台的节目所属于的二级分类,如动作、爱情等。	可选
10	节目时长	Duration	STRING	网络视听节目的时间长度。 按 GB/T 7408—2005 “基本格式”中的“完全表示”方法,为 hhmmss,如 004600,精确到秒。	必选
11	记录时间	RecordTime	STRING	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。 按 GB/T 7408—2005 “基本格式”中的“完全表示”方法,数据格式为: yyyymmddhhmmss。	必选

表 6 (续)

序号	参数名称	字段	类型	参数说明	备注
12	地区	Province	STRING	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。	必选
12	播放时长	PlayDuration	STRING	互联网视听平台用户观看节目的总时长。按 GB/T 7408—2005 “基本格式” 中的 “完全表示” 方法,为 hhmmss,如 004600,精确到秒。	必选

节目播放数据的JSON消息示例如下:

```
{
  "Account": "12d41345039alleaac64525400f3186f",
  "DeviceID": "03078Q980001A9AE2FF7F",
  "UserID": "60215739463920000483790",
  "APPVersion": "A02B15",
  "ContentID": "CP23010020171209062900",
  "Title": "空中危机",
  "DeviceIP": "47.98.74.54",
  "ContentType1": "电影",
  "ContentType2": "恐怖悬疑",
  "Duration": "004530",
  "RecordTime": "201911181429005",
  "Province": "北京",
  "PlayDuration": "002230"
}
```

7.4.2 视频播放行为数据

视频播放行为数据应包含的参数及其取值应符合表 7 的规定。

表 7 视频播放行为数据应包含的参数及其取值

序号	参数名称	字段	类型	参数说明	备注
1	账号	Account	STRING	于用户在互联网视听平台中代表自己身份的名称。	可选
2	终端 ID	DeviceID	STRING	能代表用户设备唯一性的可采集标识码。	必选
3	用户 ID	UserID	STRING	系统自动生成的用于唯一识别用户的一组不重复的编码。	必选

表 7（续）

序号	参数名称	字段	类型	参数说明	备注
4	播放操作类型	PlayOperateType	UINT	用户节目播放和页面操作行为，行为编码表示如下： 1：拖拽； 2：暂停； 3：倍速播放； 4：情节选择； 5：评论； 6：弹幕； 7：收藏； 8：下载； 9：分享； 10：顶； 11：踩； 12：退出； 13：其他。	必选
5	播放操作时间	PlayOperateTime	STRING	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。 按 GB/T 7408—2005“基本格式”中的“完全表示”方法，数据格式为： yyyyymmddhhmmss	必选
6	播放操作描述	PlayOperateDescribe	STRING	根据不同操作，描述操作参数。 以下为必选操作项目： 拖拽：记录定位的时间点，格式为 h:mm:ss； 倍速播放：描述播放倍速，取值为 0.5，0.75，1，1.25，1.5，2； 情节选择：描述用户选择的情节； 分享：分享的方式 退出：播放时长。	必选
				以下为可选操作项目： 评论：评论的内容； 弹幕：弹幕的内容。	可选

视频播放行为数据的JSON消息示例如下：

```
{
  "Account": "12d41345039a11eaac64525400f3186f ",
  "DeviceID": "03078Q980001A9AE2FF7F ",
  "UserID": "60215739463920000483790",
  "OperateType": "3",
  "OperateTime": "20191118142905",
```

```

"OperateDescribe": "1.5"
}

```

7.4.3 页面访问数据

页面访问数据应包含的参数及其取值应符合表 8 的规定。

表 8 页面访问数据应包含的参数及其取值

序号	参数名称	字段	类型	参数说明	备注
1	账号	Account	STRING	于用户在互联网视听平台中代表自己身份的名称。	可选
2	终端 ID	DeviceID	STRING	能代表用户设备唯一性的可采集标识码。	必选
3	用户 ID	UserID	STRING	系统自动生成的用于唯一识别用户的一组不重复的编码。	必选
4	版本号	AppVersion	STRING	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。	可选
5	链接地址	PageURL	STRING	用户浏览页面的 URL 地址。	可选
6	操作类型	PageOperateType	STRING	用户页面浏览操作,包括搜索、点击链接等。	可选
7	前一链接地址	ReferURL	STRING	用户跳转至当前页面前,所浏览页面的 URL 地址。	必选
8	记录时间	RecordTime	STRING	用户页面访问记录生成时间。 格式符合 GB/T 7408-2005 “基本格式”中的“完全表示”方法,数据格式为: YYYYMMDDHHMMSS	可选

页面访问数据的JSON消息示例如下:

```

{
  "Account": "12d41345039a11eaac64525400f3186f",
  "DeviceID": "03078Q980001A9AE2FF7F ",
  "UserID": "60215739463920000483790",
  "APPVersion": "A02B15",
  "PageURL": " WzLTm2zKoFEDAFPSWyo+cu4A_wdm_13662912_1573920119803",
  "PageOperateType": "\N",
  "ReferPageURL": " XTefiiy/wUDACb9av0wm6b3_wdm_13662912_1563926410098 ",
  "Record": "20191117130021"
}

```