

GY

中华人民共和国广播电视和网络视听行业标准

GY/T XXX. 1—XXXX

网络视听收视大数据技术规范 第1部分：总体要求

Technical specification for internet video and audio service big data—
Part 1: General requirements

(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 数据采集要求	3
6.1 数据源要求	3
6.2 采集内容要求	3
6.3 性能要求	3
6.4 数据源与数据采集接口	3
7 大数据平台要求	3
8 数据开放与服务要求	4
8.1 功能要求	4
8.2 开放内容	4
8.3 数据表达	4
8.4 性能要求	4
8.5 交互与数据接口	4
9 数据服务安全要求	4
9.1 概述	5
9.2 数据采集环节	5
9.3 数据传输环节	5
9.4 数据接入平台环节	5
9.5 数据存储环节	5
9.6 数据处理环节	5
9.7 数据开放环节	5
9.8 接口安全	6
9.9 个人信息隐私保护	6
参考文献	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1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、华数传媒网络有限公司、湖南快乐阳光互动娱乐传媒有限公司、浙江传媒学院、未来电视有限公司。

本文件主要起草人：李忠焰、郑冠雯、李庆国、聂明杰、诸葛海标、俞定国、王志豪、曹志、姜蕾、刘琛良、张亦弛、景生军、李鸣、邱文状、遇琪、邱波、刘晓敏、黄瑞婷、葛进平、杨阳。

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、大数据平台、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1部分：总体要求

1 范围

本文件规定了网络视听收视大数据技术规范的总体要求，包括网络视听收视大数据数据采集、大数据平台、数据开放和服务、数据安全等方面的要求。

本文件适用于互联网视听服务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 35273—2020 信息安全技术 个人信息安全规范
- GB/T 37973—2019 信息安全技术 大数据安全管理指南
- GB/T 37721—2019 信息技术 大数据分析系统功能要求
- GB/T 37722—2019 信息技术 大数据存储与处理系统功能要求
- 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
- GY/T XXX. 3—XXXX 网络视听收视大数据技术规范 第3部分：接口
- GD/J 075—2018 电视收视数据交换接口规范

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、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。

[来源：GB/T 35295—2017，2.1.1]

3.2

大数据平台 big data platform

实现大数据参考架构的全部或部分功能的软硬件集合，通常指集成了数据接入、存储、处理、分析、共享，以及各类配套功能组件的软硬件基础设施。

[来源：GB/T 35274—2017，3.6]

3.3

大数据系统 big data system

实现大数据参考体系结构的全部或部分功能的系统，由大数据平台、数据采集终端、网关以及相关辅助功能组件构成。

注：大数据参考体系结构是一种用作工具以便于对大数据内在的要求、设计结构和运行进行开放性探讨的高层概念模型。其定义见GB/T 35295—2017的2.1.3。

3.4

网关 gateway

一种实现网络连接协议和数据格式转换的软件组件或实体设备。

4 缩略语

下列缩略语适用于本文件。

API 应用程序编程接口 (Application Programming Interface)

SQL 结构化查询语言 (Structured Query Language)

5 概述

网络视听收视大数据系统系统由数据源、网关和大数据平台等部分组成。

在互联网视听平台内部，大数据平台通过网关汇聚各业务平台数据和网络舆情、市场信息等其他数据，进行处理、分析后应用于精准营销、精细服务、精确管理等，互联网视听平台内部收视大数据系统模型如图1所示。

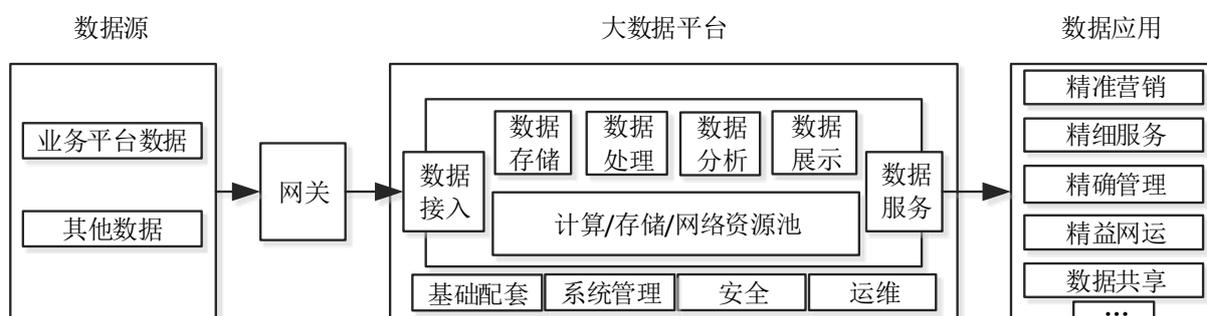


图1 互联网视听平台内部收视大数据系统模型

为支撑行业应用，互联网视听平台应对外共享网络视听收视数据，由网络视听行业收视大系统汇聚各网络平台的媒资数据、用户基本数据和用户行为数据，进行数据集中存储、处理、分析和展示。网络视听行业大数据系统模型架构如图2所示。

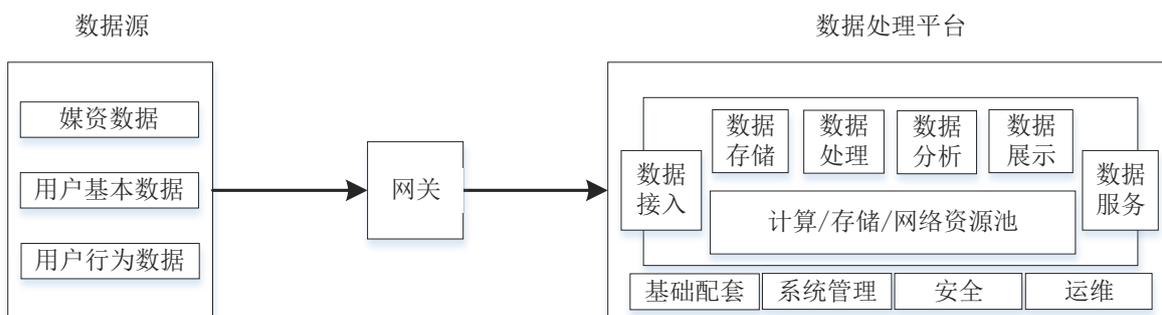


图2 网络视听行业收视大数据系统模型

6 数据采集要求

6.1 数据源要求

网络视听大数据包括网络视听媒资数据、用户基本数据和用户行为数据等。数据源要求如下：

- a) 应具备数据采集的功能；
- b) 应保证数据的完整性、准确性；
- c) 宜具备图形化的数据采集配置或管理界面。

6.2 采集内容要求

采集内容要求如下：

- a) 应具备采集 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》规定的元素集各类数据的功能；
- b) 应具备采集结构化、非结构化、半结构化数据的功能；
- c) 应具备采集实时数据的功能；
- d) 应具备采集定时数据的功能。

6.3 性能要求

性能要求如下：

- a) 采集用户行为数据时，应具备实时和定时传输、接入用户行为数据的能力；
- b) 采集媒资数据时，应具备 1h 内完成当日增量数据的传输、接入的能力；
- c) 采集用户基本数据时，应具备 1 天内完成当日增量数据的传输、接入的能力。

6.4 数据源与数据采集接口

对于互联网视听平台内部收视大数据系统，数据源与数据采集接口可以按 GY/T XXX. 3—XXXX《网络视听收视大数据技术规范 第3部分：接口》或企业标准表达数据。

对于网络视听行业收视大数据系统，数据源与数据采集接口要求如下：

- a) 应按 GY/T XXX. 3—XXXX《网络视听收视大数据技术规范 第3部分：接口》规定的格式表达数据；
- b) 数据源应按 GY/T XXX—XXXX《广播电视和网络视听收视综合评价数据脱敏规则》的要求完成数据脱敏处理；
- c) 数据源与大数据平台间应通过网关进行。网关应按 GY/T XXX. 3—XXXX《网络视听收视大数据技术规范 第3部分：接口》规定的格式完成数据源格式转换。

7 大数据平台要求

大数据平台要求如下：

- a) 应具备接入、处理符合 GY/T XXX. 3—XXXX《网络视听收视大数据技术规范 第3部分：接口》要求的数据的能力，支持接入实时、定时数据，同时可管理数据采集行为；
- b) 应具备数据预处理功能，支持对存放在文件系统和数据库中的数据进行抽取、转换、清洗等操作后加载到数据分析等处理模块；

- c) 应具备数据存储功能，支持为其他处理模块提供上传、下载、查看、删除、权限管理等数据操作，并通过冗余备份等机制提供安全的数据存储；
- d) 应具备数据处理功能，支持数据分析功能，提供 SQL 分析、跨数据源关联分析等数据分析功能，并提供对外开放分析服务的 API；
- e) 大数据平台数据分析功能应符合 GB/T 37721—2019 的要求，数据存储和处理功能应符合 GB/T 37722—2019 的要求；
- f) 应支持计算资源的配置、调度和回收等，支持平台资源的全局管理，支持计算框架的水平扩展，支持任务调度和任务优先级管理等。

8 数据开放与服务要求

8.1 功能要求

网络视听大数据平台数据开放与服务功能要求如下：

- a) 应具备对外提供数据服务的功能，如提供数据发布等功能；
- b) 应具备对数据服务用户的注册、权限进行管理，监看和审计数据服务用户的行为；
- c) 应提供开放的数据服务接口，如提供用户注册 API、服务应用 API；
- d) 宜提供二次开发接口，允许用户自定义业务；
- e) 应具备开放 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》规定的元素集的功能。

8.2 开放内容

开放内容要求如下：

- a) 应具备开放和共享平台接入的各类数据的功能；
- b) 应具备开放和共享平台处理或分析数据结果的功能。

8.3 数据表达

开放的数据集、收视数据应按 GD/J 075—2018 规定的格式表达，其他数据格式应符合 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》的规定。

8.4 性能要求

大数据平台对外提供数据开放服务时，性能要求如下：

- a) 应用系统向大数据平台发送的注册或认证消息，大数据平台的响应时间应小于 1s；
- b) 应用系统向大数据平台请求数据交换的响应时间、交换的数据包大小、共享的数据容量、多数数据源访问响应时间、并发用户数等性能应满足 GY/T XXX. 3—XXXX《网络视听收视大数据技术规范 第 3 部分：接口》的要求。

8.5 交互与数据接口

平台与应用系统之间的交互和数据通信接口满足 GY/T XXX. 3—XXXX《网络视听收视大数据技术规范 第 3 部分：接口》的要求。

9 数据服务安全要求

9.1 概述

网络视听大数据平台在开展数据服务的过程中，通过创建数据服务的安全策略和规程，明确组织和人员及岗位，建立数据采集、传输、存储、处理、交换共享、审计过程中的安全制度，数据安全应符合GB/T 37973—2019的有关要求。

9.2 数据采集环节

数据采集环节要求如下：

- a) 应明确数据采集的目的和用途，确保数据采集的合法性和正当性；
- b) 应按照元素集规定的范围采集数据，按照数据传输接口规范的要求规范数据格式，明确采集频率；
- c) 应通过对组件、终端、设备等采集器进行必要的技术控制，如在部署前对采集器进行采集能力认证、计量等，确保数据的完整性、一致性和真实性；
- d) 应明确数据收集和获取过程中个人信息和重要数据的知悉范围和安全管控措施，确保采集数据的合法性、完整性和真实性；
- e) 应采取脱敏、加密等技术或管理措施确保采集过程中涉及的个人信息和重要数据不被泄露；
- f) 应对数据采集行为进行权限管理。

9.3 数据传输环节

数据传输环节宜对数据进行加密传输。

9.4 数据接入平台环节

数据接入平台环节要求如下：

- a) 应建立数据采集终端/网关接入平台的身份识别与鉴别策略、权限分配策略和相关操作规程，建立访问控制时效管理机制；
- b) 应具备数据接入的安全审计功能。

9.5 数据存储环节

数据存储环节要求如下：

- a) 应采取冗余存储或多副本存储措施，确保数据的可用性；
- b) 应采取校验等检验和容错技术措施，确保多副本数据存储的一致性；
- c) 应确保存储技术架构具备加密个人信息、重要数据的能力。

9.6 数据处理环节

数据处理发生在大数据平台内部，要求如下：

- a) 应确保处理过程中发生数据迁移、变换时的真实性、完整性和一致性；
- b) 应确保数据处理过程操作行为可审计，数据质量可溯源，重要数据受保护。

9.7 数据开放环节

数据开放环节要求如下：

- a) 应建立应用系统开放平台的身份识别与鉴别策略、权限分配策略和相关操作规程，建立访问控制时效管理机制，具备数据访问记录功能；
- b) 应建立应用系统访问平台数据、服务的控制策略；

- c) 应具备数据交换、共享、发布的安全审计功能。

9.8 接口安全

接口安全要求如下：

- a) 应制定数据服务接口安全控制策略，明确规定使用服务接口的安全措施，如身份鉴别、授权策略、访问控制机制、签名、时间戳、安全协议等；
- b) 应制定数据服务接口安全规范，包括接口名称、接口参数、对接安全要求等，具备限制或过滤不安全参数的能力；
- c) 应具备数据服务接口访问审计功能；
- d) 宜在数据采集终端与平台、平台与应用系统间采用安全通道或加密通道进行接口调用。

9.9 个人信息隐私保护

在采集、传输、存储、展示、共享、管理网络视听收视大数据中涉及个人的信息数据时，应遵循GB/T 35273—2020的要求。

参 考 文 献

- [1] GB/T 34982—2017 云计算数据中心基本要求
- [2] GB/T 35274—2017 信息安全技术 大数据服务安全能力要求
- [3] GB/T 35295—2017 信息技术 大数据 术语