



中华人民共和国国家标准

GB/T 36625.3—2021

智慧城市 数据融合 第3部分：数据采集规范

Smart city—Data fusion—Part 3: Data acquisition specifications

2021-04-30 发布

2021-11-01 实施

国家市场监督管理总局 发布
国家标准化管理委员会

目 次

| | |
|--------------------|---|
| 前言 | 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 | 1 |
| 5 数据采集过程 | 2 |
| 6 数据采集内容 | 3 |
| 7 数据采集技术 | 3 |
| 7.1 数据采集技术要求 | 3 |
| 7.2 数据采集方法 | 3 |
| 8 数据采集质量控制 | 4 |
| 8.1 数据质量控制原则 | 4 |
| 8.2 数据质量控制方式 | 4 |
| 8.3 数据质量评价方法 | 4 |
| 9 数据采集安全控制 | 4 |
| 9.1 数据安全要求 | 4 |
| 9.2 安全控制方法 | 5 |

前 言

GB/T 36625《智慧城市 数据融合》分为以下五个部分：

- 第1部分：概念模型；
- 第2部分：数据编码规范；
- 第3部分：数据采集规范；
- 第4部分：开放共享要求；
- 第5部分：市政基础设施数据元素。

本部分为 GB/T 36625 的第3部分。

本部分按照 GB/T 1.1—2009 给出的规则起草。

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这些专利的责任。

本部分由中华人民共和国工业和信息化部提出。

本部分由全国通信标准化技术委员会(SAC/TC 485)归口。

本部分起草单位：中国城市科学研究会、中城智慧(北京)城市规划设计研究院有限公司、中城智慧科技有限公司、软通智慧科技有限公司、北京清华同衡规划设计研究院有限公司、上海竹唛信息技术有限公司、华为技术有限公司、深圳市华傲数据技术有限公司、中国电子科技集团公司信息科学研究院、讯飞智元信息科技有限公司、中兴通讯股份有限公司、中冶京诚工程技术有限公司、广东珠光集团有限公司、吉林吉大通信设计院股份有限公司、中国电子技术标准化研究院、中国信息通信研究院、山东省标准化研究院、建设综合勘察研究设计院有限公司。

本部分主要起草人：万碧玉、吴丽丽、马蓉、王飞飞、王圣波、聂明、崔昊、龚健、李赟、苏平、朱崇亚、姜栋、李玲玲、曹余、李宁丽、黎俊茂、毛汉平、刘雅晶、佟敏、刘朝晖、李海龙、单峰、张红卫、方可、臧磊、杜青峰、李公立、刘棠丽、孙郁頔、陈慧文、王建新、于富东、王树东。

智慧城市 数据融合

第3部分:数据采集规范

1 范围

GB/T 36625 的本部分规定了智慧城市数据融合过程中数据采集过程、数据采集内容、数据采集技术、数据采集质量控制及数据采集安全控制。

本部分适用于智慧城市各系统的规划设计、建设和管理,其他信息化领域的数据采集过程也可参考使用。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件,仅注日期的版本适用于本文件。凡是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 5271.1 信息技术 词汇 第1部分:基本术语

GB/T 22239 信息安全技术 网络安全等级保护基本要求

GB/T 31916.1—2015 信息技术 云数据存储和管理 第1部分:总则

GB/T 36625.1—2018 智慧城市 数据融合 第1部分:概念模型

3 术语和定义

GB/T 5271.1、GB/T 36625.1—2018 界定的以及下列术语和定义适用于本文件。为了便于使用,以下重复列出了 GB/T 36625.1—2018 中的某些术语和定义。

3.1

数据融合 data fusion

集成多个数据源以产生比任何单独的数据源更有价值信息的过程。

[GB/T 36625.1—2018,定义 3.1]

3.2

数据采集 data acquisition

从数据源中得到原始数据,通过标准化处理并转化为满足数据共享与利用需求的过程。

4 缩略语

下列缩略语适用于本文件。

CCKS:组合诚信密钥系统(Combined Credit Key System)

CPK:组合公钥(Combined Public Key)

FTP:文件传输协议(File Transfer Protocol)

HTTP:超文本传输协议(HyperText Transfer Protocol)

HTTPS:超文本传输安全协议(HyperText Transfer Safe Protocol)

JDBC:Java 数据库连接(Java DataBase Connectivity)

- MQTT:消息队列遥测传输(Message Queuing Telemetry Transport)
- ODBC:开放数据库互连(Open Database Connectivity)
- OPC:对象连接和嵌入技术在过程控制规范(Object Linking and Embedding for Process Control)
- PKI:公钥基础设施(Public Key Infrastructure)
- PDA:数据采集器(Personal Digital Assistant)
- 注: 又称掌上电脑。
- REST:表述性状态转移(Representational State Transfer)
- TCP:传输控制协议(Transmission Control Protocol)
- UDP:用户数据报协议(User Datagram Protocol)
- VPN:虚拟专用网络(Virtual Private Network)

5 数据采集过程

数据采集过程实现对数据的交换与提取、数据汇聚处理、数据安全加密压缩,并提供质量控制、安全控制等辅助工具,如图 1 所示。

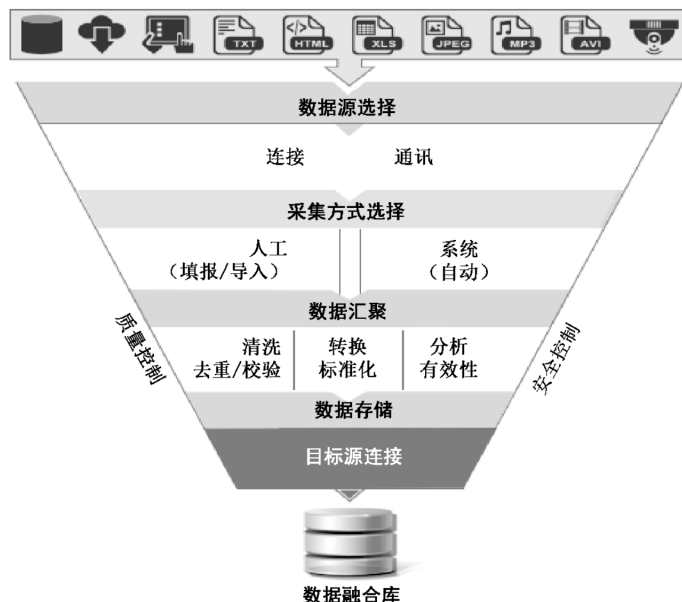


图 1 数据采集过程

数据采集过程包括:

- 数据源选择:根据需要采集数据的数据源类型(如:文件、数据库、传感器等),确定数据源连接通讯的方式,明确采集标准范围及属性。可支持结构化数据和非结构化数据类型;可支持 JDBC、ODBC、OPC、MQTT、Modbus 等多种连接方式;可支持 TCP、UDP、FTP、HTTP 等通讯协议。
- 数据采集方式选择:数据采集分为人工采集和系统采集两种,通过分析相关数据源类型,根据可操作性、成本导向等原则选定数据采集方式。
- 数据汇聚:对采集的原始数据进行清洗、转换、分析等处理,确保数据的完整性、准确性和时效性。
- 数据存储:可按照 GB/T 31916.1—2015 的规定进行,处理后的数据存储应满足海量、安全、高性能、高可靠、易管理。

- 数据质量:数据采集周期内,应保证数据的完整性、准确性、一致性、时效性、可访问性、可追溯性。
- 数据安全:数据采集周期内,应按数据安全及安全控制要求,实现授权访问、可定位溯源、数据加密、安全审计及监测等。

6 数据采集内容

从智慧城市建设与管理需求及我国智慧城市实践经验角度,智慧城市数据采集内容包括但不限于基础数据、专题数据、业务专属数据和其他数据四大类:

- 基础数据,如人口、法人单位、自然资源、地理空间、宏观经济、电子证照等数据;
- 专题数据,如房屋、城市部件、网格等与智慧城市相关的公共共享数据;
- 业务专属数据,如涉及公安、公共卫生和医疗、教育、民政、交通、水利、人力资源和社会保障、市场监管等众多领域的业务数据;
- 其他数据,如互联网、工业、商业等数据。

7 数据采集技术

7.1 数据采集技术要求

数据采集技术应具备复杂网络环境下、不同异构数据源之间高速、稳定、弹性伸缩的数据移动及同步能力。采集技术包括但不限于:

- 针对结构单一、数据量相对较小的结构化数据,可通过数据库表、文件、网络服务(Web Service)、REST、HTTP/HTTPS、消息订阅/发布等技术进行数据采集;
- 针对传感器、智能手机、PDA设备、网络等渠道产生的类型丰富、数据量较大的数据,可通过分布式系统接口、分布式流数据收集、网络爬虫等技术进行数据采集;
- 针对由麦克风、摄像头等设备产生的海量音视频数据,可通过语音图像识别、编解码等技术转化后进行数据采集;
- 针对问卷调查、实地调研、资料分析等产生的数据,可通过在线填报、离线导入等人工转化方式进行数据采集。

7.2 数据采集方法

7.2.1 人工采集

通过人工填报、人工导入等方式获得数据。

7.2.2 系统采集

通过系统自动录入等方式获得数据,并应满足下列要求:

- 支持全量、历史数据采集:应提供数据传输服务、高并发的离线数据上传下载服务,支持TB/PB级别的数据导入(全量数据或历史数据的批量导入)及导出;
- 支持实时或定时增量数据采集:宜提供实时同步、定时采集、数据订阅、日志采集等服务;
- 支持条件过滤:按照指定条件进行指定过滤采集,例如字段内容;
- 支持采集作业管理和调度:采集作业支持条件触发、并发调度、周期循环调度等模式;支持对作业启动、停止、暂停、恢复等操作;
- 支持数据标签:依据数据清洗要求为数据标记数据标签;
- 支持数据建模:提供基于不同业务需求进行数据建模功能。

8 数据采集质量控制

8.1 数据质量控制原则

对数据的质量控制应贯穿整个采集过程,遵循但不限于以下原则:

- a) 完整性:应包含数据规则要求的数据的必要元素;
- b) 准确性:应真实反映数据所描述的实体;
- c) 一致性:应保证数据与其他特定上下文中使用的数据无矛盾;
- d) 时效性:应保证数据发生变化后及时被更新;
- e) 可访问性:应保证数据在需要时能被安全访问;
- f) 可追溯性:应保证数据能够被跟踪和管理。

8.2 数据质量控制方式

8.2.1 数据清洗

数据清洗过程管理应包括但不限于:

- a) 数据分析:应对数据源进行分析,及时发现数据源存在的质量问题;
- b) 定义清洗规则:包括空值的检查和处理、非法值的检测和处理、不一致数据的检测和处理、相似重复记录的检测和处理等;
- c) 执行数据清洗规则:依据定义的清洗规则,补足残缺/空值、纠正不一致、完成数据拆分、数据合并或去重、数据脱敏、数据除噪等;
- d) 清洗结果验证:数据清洗方应对定义的清洗方法的正确性和效率进行验证与评估,对不满足清洗要求的清洗方法进行调整和改进。数据清洗过程宜多次迭代并进行分析、设计和验证。

8.2.2 数据转换

应对数据的标准代码、格式、类型等进行转换。必要时,可建立“数据转换规则表”。

8.2.3 数据分析

应通过数据聚合、数据归类、数据关联等方法,分析采集的数据,形成上下文完整有效的数据。

8.3 数据质量评价方法

数据质量评价方法可分为定性评价法和定量评价法:

- a) 定性评价法可根据事先确定的评价指标,对数据的安全性、目的、用途、日志以及用户自定义项目进行评价;
- b) 定量评价法可采用数据质量检测软件检查数据质量,也可通过辅助工具结合人工识别分析方法进行人工检查。一般可分为全数检查和抽样检查:
 - 1) 针对国家强制要求、特殊要求、其他可能导致严重影响的数据质量项目进行全数检查;
 - 2) 针对质量比较稳定、数据量较大、检查费用与时间有限的情况进行抽样检查。

9 数据采集安全控制

9.1 数据安全要求

数据采集安全贯穿于数据融合整个过程中,应符合数据所属或主管部门的安全要求,包括但不

限于：

- a) 应符合 GB/T 22239 对数据应用安全的相关要求；
- b) 数据在整个采集、转化、传输过程中应依据授权使用，不被非法冒充、窃取、篡改、抵赖；
- c) 应对数据采集环境、设施和技术采取必要的安全管控措施；
- d) 应明确数据采集过程中个人信息和重要数据的知悉范围和安全管控措施，并采取必要的技术手段和管理措施保证数据不被泄露；
- e) 应能够对采集的数据进行定位溯源；
- f) 应能够对数据采集过程进行安全审计及监测；
- g) 应采用通过检测认证的密码产品，来保障采集过程中的安全性。

9.2 安全控制方法

数据采集过程中应全方位防御，避免病毒、攻击、非授权的访问与内部泄密，同时应保障访问记录的审查和监督。应包括但不限于：

- a) 对不同数据进行分类并标识，采用安全技术进行安全维护；
 - b) 监控数据使用情况，防止数据在采集过程中被非法访问、破坏、篡改、丢失、阻止；
 - c) 设立访问和使用权限控制机制；
 - d) 制定应急响应预案及相应处理措施，并定期进行应急演练，及时发现安全问题并处理；
 - e) 定期对数据采集的安全性进行风险评估，并据此制定相应的风险处理计划，及时排查安全漏洞，加固安全技术；
 - f) 采用安全技术维护数据安全，包括但不限于对称与非对称密码技术及其硬化技术、VPN 技术、身份认证与鉴别技术、CPK 技术、CCKS 技术、PKI 技术、完整性验证技术、数字签名技术、秘密共享技术等；
 - g) 制定数据采集操作规程，规范数据采集的数据格式、数据质量、流程和方法等；
 - h) 制定数据采集原则，明确采集数据的目的和用途，确保数据采集的合法性和正当性；
 - i) 建立安全管理规范，避免人为因素导致数据泄露、损坏等安全事故。
-

中 华 人 民 共 和 国
国 家 标 准
智慧城市 数据融合
第 3 部分：数据采集规范
GB/T 36625.3—2021

*

中国标准出版社出版发行
北京市朝阳区和平里西街甲 2 号(100029)
北京市西城区三里河北街 16 号(100045)

网址: www.spc.org.cn

服务热线: 400-168-0010

2021 年 4 月第一版

*

书号: 155066 · 1-66580

版权专有 侵权必究



GB/T 36625.3-2021