

DB61

陕西省地方标准

DB 61/T 1896—2024

高速公路数据集成规范

Specification for expressway data integration

2024 - 09 - 24 发布

2024 - 10 - 24 实施

陕西省市场监督管理局 发布

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 缩略语..... 2

5 总体要求..... 2

6 数据集成流程..... 2

7 数据源和集成内容选择..... 3

8 集成数据获取..... 3

9 集成数据预处理..... 4

10 集成数据质量评估..... 4

11 集成数据存储..... 4

附录 A （资料性） 集成数据内容..... 5

附录 B （资料性） 集成数据质量评估指标..... 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。

本文件由陕西省交通运输厅提出并归口。

本文件起草单位：陕西交通电子工程科技有限公司、西北工业大学、陕西交控工程技术有限公司、帆软软件有限公司。

本文件主要起草人：樊吉飞、孙楠、王鑫、陈思行、田龙、张高峰、刘烨、孙杰、王一博、吴天文、刘浩、沈钧戈、罗叶军、屈战辉、任伟伟。

本文件由陕西交通电子工程科技有限公司负责解释。

本文件首次发布。

联系信息如下：

单位：陕西交通电子工程科技有限公司

电话：029-88450858

地址：陕西省西安市高新区唐延路6号2楼10号

邮编：710065

高速公路数据集成规范

1 范围

本文件规定了高速公路数据集成的总体要求、数据集成流程、数据源和集成内容选择、集成数据获取、集成数据预处理、集成数据质量评估和集成数据存储的要求。

本文件适用于高速公路建设、管理、养护、运营、服务等数据的集成，包括数据获取、预处理、质量评估和存储。

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 35295—2017 信息技术 大数据 术语
- GB/T 36344 信息技术 数据质量评价指标
- DA/T 82—2019 基于文档型非关系型数据库的档案数据存储规范
- JT/T 697.2—2014 交通信息基础数据元 第2部分：公路信息基础数据元
- JT/T 697.7—2014 交通信息基础数据元 第7部分：道路运输信息基础数据元
- JT/T 697.9—2016 交通信息基础数据元 第9部分：建设项目信息基础数据元
- JT/T 697.10—2016 交通信息基础数据元 第10部分：交通统计信息基础数据元
- JT/T 697.13—2016 交通信息基础数据元 第13部分：收费公路信息基础数据元

3 术语和定义

GB/T 35295—2017，DA/T 82—2019中界定的以及下列术语和定义适用于本文件。

3.1

高速公路数据集成 **expressway data integration**

通过系统化的方法，将来自不同来源及系统的高速公路相关数据进行获取、预处理、质量评估和存储的过程。

3.2

结构化数据 **structured data**

一种数据表示形式，按此种形式，由数据元素汇集而成的每个记录的结构都是一致的并且可以使用关系模型予以有效描述。

[来源：GB/T 35295—2017，2.2.13]

3.3

半结构化数据 semi-structured data

具有结构性，但结构变化大，且难以用结构化数据的处理方法将其放进二维表的数据。

[来源：DA/T 82—2019，2.8]

3.4

非结构化数据 unstructured data

不具有预定义模型或未以预定义方式组织的数据。

[来源：GB/T 35295—2017，2.1.25]

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）

NoSQL：非关系型数据库（Not Only SQL）

5 总体要求

数据集成的总体要求如下：

- a) 数据集成不应对生产系统的正常使用产生影响；
- b) 应明确数据集成接口和方式；
- c) 数据集成应进行数据预处理及质量评估；
- d) 应对原始数据进行存储，建立可伸缩、稳定可靠的数据存储架构。

6 数据集成流程

数据集成流程如图1所示。

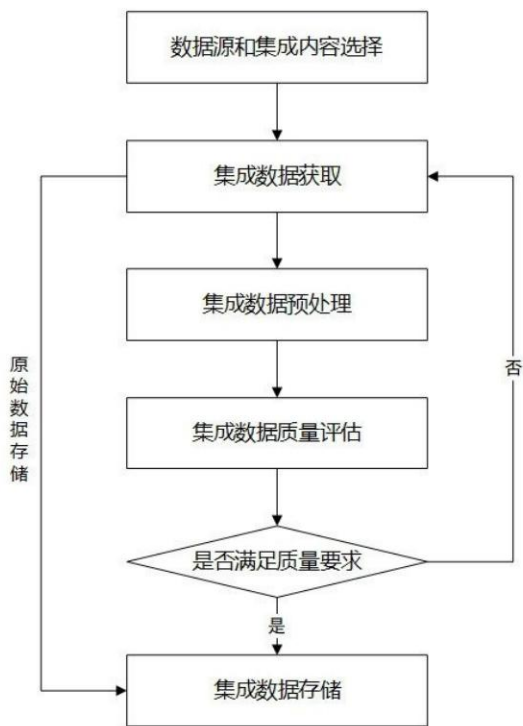


图 1 数据集成流程图

7 数据源和集成内容选择

- 7.1 应结合需求选择数据源，并确定集成内容，包括建设数据、管理数据、养护数据、运营数据、服务数据等。
- 7.2 建设数据应符合 JT/T 697.9—2016 中第 4 章的要求。
- 7.3 管理数据应符合 JT/T 697.7—2014 中 4.1、4.2、4.3、4.5 节，JT/T 697.10—2016 中 4.5.1、4.6、4.7 节的要求。
- 7.4 养护数据应符合 JT/T 697.2—2014 中 4.1、4.2、4.3、4.4、4.5、4.6、4.7、4.8、4.9、4.10 节的要求。
- 7.5 运营数据应符合 JT/T 697.10—2016 中 4.3.1 节，JT/T 697.13—2016 中第 4 章的要求。
- 7.6 服务数据应符合 JT/T 697.2—2014 中 4.11 节的要求。
- 7.7 其他数据可参考附录 A。

8 集成数据获取

8.1 获取数据的类型

获取数据的类型分为结构化数据、半结构化数据和非结构化数据。

8.2 结构化数据获取

结构化数据可使用数据库集成、API集成等方式获取。

- a) 对于通过数据库集成的数据源，需具备数据库类型、数据库结构、字段名称、字段类型、长度和描述信息，并提供数据库的连接信息，包括访问地址、端口、数据库名、用户名和密码；

- b) 对于通过 API 集成的数据源, 需具备 API 说明文档, 包含访问地址、频率、输入参数、输出参数、参数的数据类型、参数的描述信息、访问的用户信息。

8.3 半结构化数据获取

半结构化数据可使用文件集成、NoSQL集成、消息队列集成等方式获取。

- a) 对于通过文件集成的数据源, 需具备文件的访问方式、访问位置、文件格式、文件数据描述;
- b) 对于通过 NoSQL 数据库集成的数据源, 需具备数据库类型、数据库结构、字段名称、字段类型、长度和描述信息, 并提供数据库的连接信息, 包括访问地址、端口、数据库名、用户名和密码;
- c) 对于通过消息队列集成数据源, 需具备消息队列类型、订阅的主题、数据的格式、数据内容的结构信息, 并提供消息队列的连接信息。

8.4 非结构化数据获取

非结构化数据可使用API集成、文件集成等方式获取。

- a) 对于通过 API 集成的数据源, 需具备 API 说明文档, 包含访问地址、频率、输入参数、输出参数、参数的数据类型、参数的描述信息、访问的用户信息;
- b) 对于通过文件集成的数据源, 需具备文件的访问方式、访问位置、文件格式、文件数据描述。

9 集成数据预处理

9.1 对集成数据应从缺失数据的比例、遗漏的重要信息量、数据的错误率、不一致性程度等方面进行分析。

9.2 按照集成数据分析结论为数据提供标记功能, 进行数据建模, 制定预处理规则, 包括缺失值处理、异常值处理、去重处理等。

- a) 缺失值: 检查数据集中某些特征的取值缺失或者未知, 对缺失值的处理方式包括删除缺失值、均值填补、热卡填补法等;
- b) 异常值: 检查数据集中与大多数样本明显不同的数值, 对异常值的处理方式包括删除异常值、替换异常值、分箱处理等;
- c) 重复值: 检查数据集中存在完全相同的记录, 对重复数据进行去重。

9.3 依据预处理规则对数据进行筛查, 对不符合规则的数据进行分类处理。

10 集成数据质量评估

10.1 集成数据质量评估应遵循 GB/T 36344 相关要求。评估指标还应包括唯一性、安全性, 具体评估方法参见附录 B。

10.2 预处理结束后, 应根据数据质量评估方法对定义的预处理结果进行验证及评估。

10.3 对不满足要求的结果应反馈给数据源, 修改后重新获取。

11 集成数据存储

11.1 集成数据存储包括原始数据存储和预处理数据存储。

11.2 数据应集中存储, 提供统一访问端口, 包括关系型数据库、键值存储数据库、列存储数据库、文档数据库、图形数据库及文件系统等。

附 录 A
(资料性)
集成数据内容

A. 1 集成数据内容

集成数据内容见表A.1。

表 A. 1 集成数据内容

集成数据内容分类	集成数据子类	集成数据内容
建设数据	项目管理类数据	组织实施数据
		过程监管数据
		考核评价数据
	勘测类数据	公路勘测数据
		工程地质勘测数据
		工程水文勘测设计数据
	设计类数据	公路新建设计数据
		改扩建公路工程设计数据
	试验类数据	土木实验数据
		岩石实验数据
		结合料实验数据
		合成材料实验数据
	检测类数据	路基路面现场测试数据
		桥梁现场检测数据
		隧道现场检测数据
		机电系统现场检测数据
	监理类数据	监督管理数据
		咨询服务数据
	造价类数据	造价文件编制导则数据
		投资估算数据
		概算预算数据
		工程量清单数据
	其他类数据	其他数据
管理数据	财务类数据	收支数据
		资产负债数据
		预算与决算数据
		审计与监督数据
	办公类数据	办公文件数据
		管理流程数据

表A. 1 (续)

集成数据内容分类	集成数据子类	集成数据内容
管理数据	办公类数据	人员培训数据
	合同类数据	收款合同数据
		付款合同数据
		付款合同数据
	经营决策类数据	市场与销售数据
		供应链与采购数据
		合规与风险管理数据
		战略规划与目标设定数据
	其他类数据	其他数据
养护数据	检测评价类数据	现场检测及监测数据
		技术状况评定数据
		设施性能评价数据
	养护设计类数据	养护工程设计文件数据
		养护项目管理数据
		实验室检测数据
	养护造价类数据	公路养护预算导则数据
		预算数据
		工程量清单数据
	其他类数据	其他数据
运营数据	监测类数据	视频监测数据
		收费监测数据
		结构物监测数据
		机电监测数据
		车路协同数据
	分析类数据	车流量分析数据
		收费金额分析数据
	其他类数据	其他数据
服务数据	出行信息服务类数据	信息发布数据
		信息查询咨询数据
	救援服务类数据	救援处置数据
	服务区服务类数据	服务内容数据
		服务状态数据
		服务统计数据
		服务评价数据
	其他类数据	其他数据

附 录 B
(资料性)
集成数据质量评估指标

B.1 唯一性指标

唯一性主要用于度量与评估数据资产内容或相关属性的重复情况,包括主键唯一性和数据唯一性,见表B.1。

表 B.1 唯一性指标

指标名称	指标描述	计算方法
主键唯一性	数据的主键属性值应该确保唯一，不允许重复	$X = \frac{A}{B}$ 式中： X——主键唯一性 A——满足主键唯一性要求的数据个数 B——满足主键唯一性要求的数据总个数
数据唯一性	数据的全部或部分属性值应该确保唯一，不允许重复	$X = \frac{A}{B}$ 式中： X——数据唯一性 A——满足数据唯一性要求的数据个数 B——满足数据唯一性要求的数据总个数

B.2 安全性指标

安全性主要用于描述数据在传输、存储和访问过程中的安全特性,以确保数据不被未经授权的访问、篡改或泄露。包括传输安全性和存储安全性,见表B.2。

表 B.2 数据安全性指标

指标名称	指标描述	计算方法
传输安全性	数据在传输过程中应防止被截获或篡改	$X = \frac{A}{B}$ 式中： X——传输安全性 A——安全传输的数据量 B——传输的总数据量
存储安全性	数据在存储过程中应防止未经授权的访问和修改	$X = \frac{A}{B}$

表B. 2 (续)

指标名称	指标描述	计算方法
存储安全性	数据在存储过程中应防止未经授权的访问和修改	式中： X——存储安全性 A——安全存储的数据量 B——存储的总数据量

参 考 文 献

- [1] JT/T 498—2019 收费公路车辆通行费车型分类
 - [2] T/ITS 0058—2017 合作式智能运输系统 车用通信系统 应用层及应用数据交互标准
 - [3] T/ITS 0115—2019 城市交通运行状况采集系统数据融合平台技术规范
 - [4] T/ITS 0125—2020 智慧高速公路信息化建设 总体框架
 - [5] T/ITS 0174—2022 智慧高速公路 交通大数据平台 接入数据格式
 - [6] 陕西省公路数据质量评估规范
 - [7] 陕西省公路行业数据采集更新管理办法20211112
 - [8] 计算机科学技术名词2018
-