

# 团体标准

T/XAZN XXX—2022

## 智能运输系统 大数据 数据质量评估模型

Intelligent transport systems-Big data-Data quality evaluation model

(征求意见稿)

2022 - XX - XX 发布

2022 - XX - XX 实施

雄安新区智能城市创新联合会 发布



## 目 次

前言.....	I
1 范围.....	2
2 规范性引用文件.....	2
3 术语和定义.....	2
4 交通数据质量需求.....	3
5 数据质量评估模型.....	3
附录 A (资料性附录) 交通数据质量评估指标及规则示例.....	8
参考文献.....	10

## 前 言

本文件按照GB/T 1.1—2020给出的规则起草。

本文件由雄安新区智能城市创新联合会（XAZN）提出并归口。

本文件起草单位：

本文件主要起草人：

# 智能运输系统 大数据 数据质量评估模型

## 1 范围

本文件规定了智能运输系统中涉及的交通大数据的质量评估模型和数据质量需求。  
本文件适用于智能运输系统中产生和应用的交通大数据核心数据质量评估。

## 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 36344—2018 信息技术 数据质量评价指标

GB/T 28788 公路地理信息数据采集与质量控制

GB/T 29101 道路交通信息服务 数据服务质量规范

ISO 19157:2013 地理信息 数据质量 (Geographic information—Data quality)

## 3 术语和定义

下列术语和定义适用于本文件。

### 3.1

**交通大数据质量** transport big data standard

在指定条件下使用时，交通大数据的特性满足智能运输系统中数据应用分析及业务流转中明确的和隐含的要求的程度。

### 3.2

**交通大数据标准** transport big data standard

智能运输系统中产生或应用的交通大数据的命名、定义、结构、取值以及有关更新频率的规则，包括已有的行业数据传输要求及标准。

### 3.3

**元数据** metadata

关于智能运输系统中交通大数据或数据元素的数据(可能包括其数据描述)，以及关于交通大数据拥有权、存取路径、访问权和数据易变性的数据。

### 3.4

**主数据 master data**

智能交通运输系统中的交通大数据涉及交通运输部门需要跨系统、跨部门进行共享的核心业务实体数据及核心字段。

**3.5****评估对象 assess target**

评估对象是指需要进行评估的数据集对象，智能运输系统产生的所有数据对象或数据对象的子集都可以作为评估对象。

**4 交通数据质量需求****4.1 交通数据质量需求的流程描述**

交通数据质量需求的流程描述如下：

- a) 定义交通数据质量管理目标，依据交通数据分析应用及业务数据流通的需求，明确交通数据质量需求目标，确定数据评价量级、时空范围；
- b) 选取和定义交通数据质量评价维度，依据交通数据质量管理的目标，参考常用交通数据质量评估指标及规则，制定交通数据质量评估维度，指导数据质量评价工作的开展；
- c) 明确交通数据质量评价应用范围，依据数据分析应用及业务发展的需求以及常见数据问题的分析，明确组织数据质量管理的范围，梳理各类数据的优先级以及质量需求；
- d) 设计数据质量规则，依据交通数据质量管理需求及目标，识别交通数据质量特性，定义各类交通数据的质量评价指标、校验规则与方法，并根据业务发展需求及数据质量检查分析结果对交通数据质量规则进行持续维护与更新。

**4.2 能力需求等级****4.2.1 初始级**

在初始级情况下，交通数据需求的管理主要是在交通数据分析前，数据质量基本要求，即符合数据总量要求、时间段要求、字段标准要求等。

**4.2.2 稳健级**

在稳健级时，交通数据已被当作实现交通数据分析应用及相关业务持续应用过程，数据质量符合动态监管要求，包括数据持续的运维机制、动态更新机制、数据自动纠正要求等。

**4.2.3 优化级**

在优化级的等级中，交通数据被认为是面向智能运输系统的重要支撑基础，数据质量应符合大规模交通发展规律，数据符合实际真实情景效率提升，大规模数据或数据分析结果可以进行共享交换。

**5 数据质量评估模型****5.1 评估流程**

在得到交通数据质量评估模型的对象关系后，交通数据质量的评估可划分为下面几个步骤，评估流程

见图 1:

- a) 首先针对要进行质量评价的交通数据对照国家或行业标准进行整体对照，标定符合标准和不符合标准的数据数量作为评价交通数据基本指标项之一。
- b) 确定评价数据内容，根据业务需求，选取数据全部内容或典型数据字段集作为数据评价对象。
- c) 选取质量评价指标，对于评价维度选择，可以是全面选择也可以根据业务需求选择其中个别维度。
- d) 数据质量评价方法选择，进行质量综合评价的方法有层次分析法、模糊综合评价法、云模型评估法、缺陷扣分法等，要根据数据评价内容进行选择。
- e) 汇总并记录各数据集评价结果，依据各数据集权重和得分计算数据待评价集的得分，确定数据产品质量等级，完成数据质量评价报告。



图 1 评估流程

## 5.2 评估维度

基于智能运输系统中产生或应用中的交通数据的生命周期以及数据生命周期的每个环节，结合交通数据的自身特点，提出从数据规范性、完整性、准确性、有效性、一致性、及时性、可访问性7个维度评估系统的交通数据质量，其具体的评价维度如下：

- a) 数据规范性：智能运输系统中的数据应按照要求的规则存储，包括关系数据库、列式数据库和图数据库等，也要符合数据标准、数据模型和业务规则的要求。
- b) 数据完整性：智能运输系统中必须采集数据不能含有空值，数据缺失包括实体缺失、属性缺失、记录缺失和字段值缺失 4 个方面。
- c) 数据准确性：智能运输系统输出的数据结果值、计算值与网络中真实值之间的差异大小。
- d) 数据有效性：主要包括数据格式、类型、值域和业务规则的有效性。对于采集到的数据，需要考虑这个数据的绝对门限和波动门限。超出门限范围的数据为无效数据。
- e) 数据一致性：数据的属主系统、数据源系统以及消费系统之间的数据差异和相互矛盾的一致性。
- f) 数据及时性：智能交通大数据中涉及得提取、传送、处理、装载、展现的数据的时间属性，体现为及时性和准确性。
- g) 数据可访问性：智能运输系统中的数据应具备可被访问性。

## 5.3 评估指标

数据质量的评估模型通常包括 6 个方面：评估对象、评估指标、评估规则、权值、期望值、评估结果。结合智能运输系统数据的评估模型中的对象关系如图 2 所示。

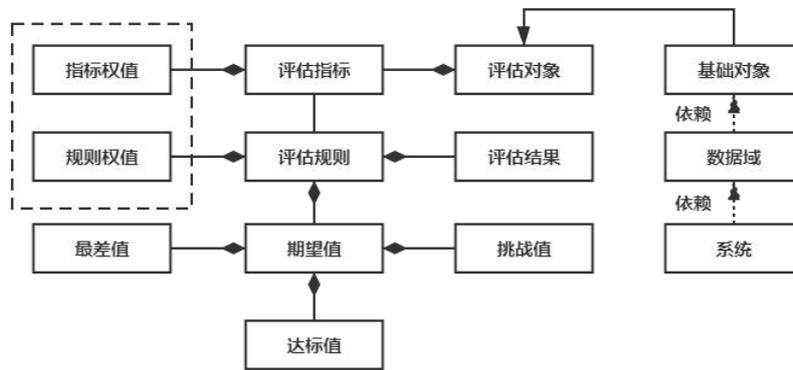


图2 数据质量评估模型

从5个维度评价数据质量，对评估对象自上面下划分为如下几个层次：系统、数据域、对象指标。每层对象的数据质量评估值都由若干个下层评估值计算得到，评估对象的层次如图3所示。

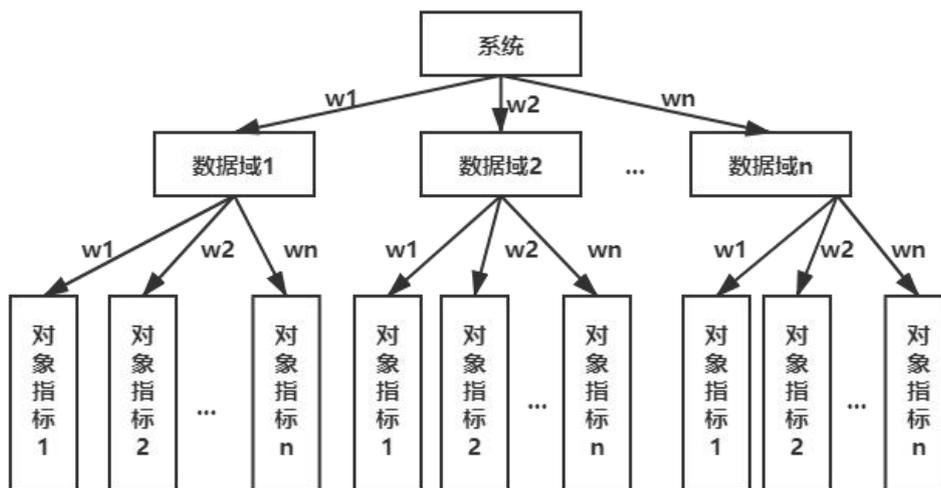


图3 评估对象层次

说明：

对象指标——对象指标是对基础对象数据质量的综合评价，具体体现为某个评估值。

数据域——按照数据类别，属主系统、数据源对数据进行分类，按照智能运输系统交通大数据常用数据资源可分为空间数据、表格数据、制图数据，其中表格数据包括静态基础数据及可以转化为表格形式的动态监测数据。

系统——智能运输系统中某个系统整体数据质量评估值。一个系统包含多个数据域。

评估指标是针对基础对象定义的，数据质量基础的评估指标，一般包含数据规范性、完整性、准确性、有效性、一致性、及时性、可访问性等。

常用交通数据质量评估指标包含如下：

- 数据缺失率：存在缺失数据字段的观测值占总观测值的比例；
- 数据错误率：存在错误数据字段的观测值占总观测值的比例；
- 数据重复率：存在重复数据字段的观测值占总观测值的比例；

- d) 数据异常率: 存在回传间隔异常偏大的观测值占总观测值的比例;
- e) 位置数据漂移率: 存在瞬时速度突变或距离漂移的观测值占剔除缺失、错误及重复数据的观测值比例。
- f) 数据可用率: 数据集每个字段可用率加权平均计算整个数据集的可用率。

## 5.4 评价内容

数据质量评价内容见表1。

表 1 数据质量评价内容

评价维度	构成元素	评价内容
完整性	多余	空间数据: 数据文件、图层、实体、要素、属性
		表格数据: 数据表、表关系、数据项、数据记录
		制图数据: 数据文件、图层、实体、要素、属性、符号、注记、图名、图例、比例尺、坐标网、图廓
	遗漏	空间数据: 数据文件、图层、实体、要素、属性, 及要求数据集完成的其他工作量
		表格数据: 数据表、表关系、数据项、数据记录, 及要求数据集完成的其他工作量
		制图数据: 数据文件、图层、实体、要素、属性、符号、注记、图名、图例、比例尺、坐标网、图廓, 及要求数据集完成的其他工作量
一致性	概念一致性	空间数据: 数据文件、图层、实体、要素、属性, 及其各自间、相互间关系
		表格数据: 数据表、数据项、数据记录, 及其各自间、相互间关系
		制图数据: 数据文件、图层、实体、要素、属性、符号、注记、图名、图例、图廓, 及其各自间、相互间关系
	值域一致性	空间数据: 属性值
		表格数据: 数据项值、数据记录值
		制图数据: 属性值、注记值、比例尺、图例、图廓
	格式一致性	空间数据: 数据文件、图层、实体、要素、属性
		表格数据: 数据表、数据项、数据记录
		制图数据: 数据文件、图层、实体、要素、属性、符号、注记、图名、图例、图廓
	拓扑一致性	空间数据: 图层、实体、要素、属性, 及其各自间、相互间关系
		制图数据: 图层、实体、要素、属性, 及其各自间、相互间关系
	时间一致性	空间数据: 数据集、数据系列的时间逻辑关系
		表格数据: 数据集、数据系列的时间逻辑关系
		制图数据: 数据集、数据系列的时间逻辑关系
	存储一致性	空间数据: 数据文件、图层、实体、要素、属性等数据的存储规则与更新频率
表格数据: 数据表、数据项、数据记录等数据的存储规则与更新频率		
制图数据: 数据文件、图层、实体、要素、属性、符号、注记、图名、图例、图廓等数据的存储规则与更新频率		
位置准确度	绝对位置准确度	空间数据: 坐标系、投影参数、图廓点、实体空间平面精度、高程精度
		制图数据: 坐标系、投影参数、图廓点、实体空间位置
	相对位置准确度	空间数据: 实体空间位置、控制点位置
		制图数据: 实体空间位置、控制点位置、注记位置

评价维度	构成元素	评价内容
		格网数据位置准确度 空间数据, 坐标系、投影参数、控制点坐标
专题准确度	分类正确性	空间数据: 实体属性
		表格数据: 数据项
		制图数据: 符号、注记
	定性属性准确度	空间数据: 稀疏、密集
		表格数据: 高低、大小、分级
		制图数据: 相邻、包含、围绕
	定量属性准确度	空间数据: 实体属性值
		表格数据: 数据项
		制图数据: 符号、注记、坐标网、图廓
时间有效性	时间准确性	空间数据: 数据源时间、数据采集时间、数据加工时间
		表格数据: 数据采集时间、数据加工时间
		制图数据: 数据源时间、数据采集时间、数据加工时间

## 附录 A

## (资料性附录)

## 交通数据质量评估指标及规则示例

## A.1 常用交通数据质量评估指标及规则示例

## A.1.1 IC 卡数据质量评估指标及规则

常用交通数据中 IC 卡数据质量评估指标及规则示例见表 A.1。

表 A.1 IC 卡数据质量评估指标及规则表

序号	数据质量指标	规则描述
1	数据总量	数据总量不得少于要求量
2	数值缺失率	关键字段 (IC 卡号、卡类型、交易时间、上下车站标、线路编号、车辆编号、时间、运行方向等) 数值缺失
3	数据错误率	//下车刷卡时间错误, 与上车时间存在很大差距的现象。//站点多站或少站, 站名错误//前一个节点的下车时间和后一个节点的上车时间的差小于等于 0, //
4	数据重复率	存在同一卡编号同一收费机号同一时间有两个及以上的观测值即为重复上传, 第二次及以后的上传记录为重复记录

## A.1.2 位置数据质量评估指标及规则表

常用交通数据中位置数据质量评估指标及规则示例见表 A.2。

表 A.2 位置数据质量评估指标及规则表

序号	数据质量指标	规则描述
1	数据总量	数据总量不得少于要求量
2	数值缺失率	关键字段 (线路编号、车辆编号、时间、经纬度、方向、方向角) 数值缺失
3	数据错误率	包括时间不在采样时间内及经纬度的值与实际范围不符
4	数据重复率	同一时间上传两条及以上的记录 (即公司名称、车辆号、线路号、方向、时间字段均相同)
5	回传间隔异常率	四分位数法则

表 A.2 位置数据质量评估指标及规则表 (续)

序号	数据质量指标	规则描述
6	GPS 漂移率	<p>计算说明：设定超速阈值，将阈值 <math>vover</math> 设为 70km/h (可设定)，阈值由经验确定。将任意时刻 (不包括起讫时刻) 回传的相邻三个点的数据时间分别记为 <math>t_1</math>、<math>t_2</math>、<math>t_3</math>，将由 <math>t_1</math>、<math>t_2</math> 的经纬度数据算得的距离 <math>l1</math> 和由 <math>t_2</math>、<math>t_3</math> 的经纬度数据算得的距离 <math>l2</math>，如果 <math>l1 &gt; vover * (t_2 - t_1)</math> 且 <math>l2 &gt; vover * (t_3 - t_2)</math>，则判断此点为漂移点，否则为正常。</p> <p>(注：北京 GPS 数据阈值设为 60)</p>

## 参 考 文 献

[1]网管数据质量评估模型研究 方津

---

T/XAZN XXXXX—XXXX

T/XAZN XXXXXX-XXXX

雄安新区智能城市创新联合会  
标准  
智能运输系统 大数据 数据质量评估模型  
T/XAZN XXX-2022

河北省雄安新区容城县奥威路 100 号 (071700)  
雄安新区智能城市创新联合会印刷  
网址: [www.xaicif.org.cn](http://www.xaicif.org.cn)

2022 年 XX 月第一版 2022 年 XX 月第一次印刷