交通运输行业标准
交通运输行业高质量数据集建设导则
（征求意见稿）
编制说明

标准起草组

2026 年 7 月

目　录

一、工作简况 1

二、标准编制原则和确定标准主要内容的依据 4

三、技术经济论证或预期的经济效果 13

四、与国际、国外同类标准水平的对比情况 16

五、与有关法律、法规及相关标准的关系 17

六、重大分歧意见的处理经过和依据 20

七、涉及专利的有关说明 20

八、标准过渡期的建议 20

九、其他应予说明的事项 21

一、工作简况

（一）任务来源

根据《交通运输部关于下达2025年交通运输标准化计划（第三批）的通知》（交科技函〔2025〕798号）安排，制定《交通运输行业高质量数据集建设导则》，计划编号为JT 2025-125。本标准由交通运输信息通信及导航标准化技术委员会（以下简称标委会）提出并归口管理，中国交通信息科技集团有限公司牵头承担编制工作。

（二）编制背景

随着交通运输行业数字化转型深入推进，各类业务系统积累了海量数据资源，但数据质量参差不齐、标准不统一、共享流通不畅等问题日益突出。高质量数据集作为数据要素价值释放的基础载体，其规范化建设已成为行业亟需解决的关键课题。当前，行业主数据库及核心数据资源库持续完善，交通运输数据的汇聚、融合、流动与应用场景大幅增加，但数据供给能力与质量保障水平亟待提升，构建高质量数据集对于夯实行业数据基础、促进数据高效流通具有重要意义。

党的十九届四中全会明确数据为新型生产要素后，国家层面密集出台政策文件，推动数据要素规范化流通与高效开发利用。2025年，国家数据局印发《构建数据基础制度更好发挥数据要素作用2025年工作要点》，并发布九个领域“数据要素×”典型场景指引，推动数据要素与多要素协同发力。同年，国家发展改革委等部门印发《关于完善数据流通安全治理 更好促进数据要素市场化价值化的实施方案》（发改数据〔2025〕18号）、《公共数据资源登记管理暂行办法》（发改数据规〔2025〕26号），为数据流通安全与公共数据资源管理提供制度保障。在高质量数据集建设核心领域，2025年8月国家数据局正式印发《高质量数据集建设指引》，配套发布建设指南、格式要求、分类指南、质量评测规范四项核心技术文件，同步启动标准验证试点、先行先试及典型案例征集工作，完成首轮标准落地验证，形成标准化建设流程与行业应用模板。2025年底，五项技术文件启动国家标准立项工作，2026年陆续形成征求意见稿，标准体系持续迭代完善。交通运输部积极响应，于2025年印发《交通运输行业高质量数据集建设方案》及《关于加快交通运输公共数据资源开发利用的实施

意见》（交科技发〔2025〕120号），明确行业高质量数据集建设的目标、任务与路径。截至2026年6月，依托国家数据集管理服务平台，整套数据集建设标准体系已全面落地，交通、工业等重点行业均完成首批标准化高质量数据集建设，数据常态化评测、规范化入库、场景化赋能的发展格局基本成型。

当前，我国交通运输行业数据规模与增速迅猛增长，构建超大规模行业高质量数据集的条件已具备。然而，行业数据集建设仍面临标准不统一、质量参差不齐、流通共享不畅等问题，数据供方“不愿、不敢、不想”的困局依然存在。数据流通安全是破解这一困局的关键，而高质量数据集的建设则是释放数据要素价值的基础。随着国家信息平台部省联动深化与行业数字化转型加速，制定《交通运输行业高质量数据集建设导则》，明确数据集建设的总体要求、分类分级、技术规范、质量评价与安全管理等内容，对于规范行业高质量数据集建设、促进数据要素合规高效流通、支撑智慧交通高质量发展具有重要意义。

（三）主要工作过程

1.起草阶段

2025 年 12 月 19 日，正式成立标准起草组；召开第一次编制工作会议；起草工作大纲。

2026年1月～3月，中国交通信息科技集团有限公司项目人员开展国内外文件收集和资料整理，分析研究国内外高质量数据集建设工作与政策法规、数据流通安全标准工作情况等相关内容，为研究起草标准内容奠定基础；同步向各级交通运输部门、其他相关部委、行业科研机构和高校、运输企业、研究机构等单位进行摸底调研，组织多次技术研讨会，围绕数据集建设中的关键环节开展技术研讨。了解相关现状、问题等基本信息；编写调研报告；完成标准草案。

2026年4月，召开第二次编制工作会议，起草组集体讨论标准草案；组织召开中期咨询会，与有关技术支持单位进行交流，了解部高质量数据集建设有关政策最新动向。形成了标准征求意见稿初稿。

2026年4月，组织主审专家初审会议，针对标准征求意见稿提出修改建议和讨论修改方案，起草组进一步优化标准内容。

2026年5月，标准征求意见稿初稿提交标委会秘书处，经标委会秘书处审核、校对，对相关问题再次修改完善，形成正式的征求意见稿。

（四）标准编制起草单位、起草人员及其所做的具体工作

本标准起草单位有：中国交通信息科技集团有限公司、交通运输部规划研究院、中国交通通信信息中心、上海库帕思科技有限公司、深城交科技集团股份有限公司、人民交通出版社股份有限公司、中交星宇科技有限公司、支付宝（杭州）数字服务技术有限公司、北京百度网讯科技有限公司。

本标准主要起草人及任务分工见表1。

表 1 本标准主要起草人及任务分工表

序号	姓名	工作单位	任务分工
1	姚俊峰	中国交通信息科技集团有限公司	全面统筹标准编制进度，审核最终成果。负责编写第 6 章。
2	宋鹏飞	中国交通信息科技集团有限公司	配合统筹标准编制进度，协调内外部资源，主持关键会议，审核过程成果。负责编写第 4 章。
3	刘　玲	中交星宇科技有限公司	负责技术框架设计与整体内容衔接，协助组织技术讨论，解决编制过程中的技术争议。负责编写第 7 章，参与编写第 6.1 条和附录 E。
4	杨　树	中交星宇科技有限公司	设计调研方案，汇总分析调研结果。负责第 1～4 章，参与编写第 6.1 条和附录 F。
5	孙　硕	交通运输部规划研究院	设计调研方案，组织问卷发放、实地走访与座谈，汇总分析调研结果。参与编写第 6.2 条。
6	张晓春	深城交科技集团股份有限公司	参与总体思路制定、总体技术框架设计，参与整体技术把关、阶段成果研讨与审核。参与编写第 6.1 条。
7	舒小婷	中交星宇科技有限公司	安排编制标准沟通会议，跟踪任务进度，整理会议纪要并督促落实，负责标准编制。参与编写第 1 章和附录 E。
8	刘雅芝	中交星宇科技有限公司	负责“总体要求”等章节的起草、修改与内容完善。参与编写第 2 章、第 5 章和附录 A。
9	李　军	中国交通通信信息中心	辅助审核最终成果。参与编写第 3 章和第 4 章。
10	沈帝文	支付宝（杭州）数字服务技术有限公司	负责“数据评测”章节的指标体系设计与内容撰写。参与编写第 6.5 条。
11	常　宇	人民交通出版社股份有限公司	负责“数据发布”章节的流通模式、运营维护等内容的编写。参与编写第 6.7 条。

序号	姓名	工作单位	任务分工
12	张　硕	交通运输部规划研究院	设计调研方案，组织问卷发放、实地走访与座谈，汇总分析调研结果。参与编写第6.3条和附录B。
13	张瑞钦	上海库帕思科技有限公司	参与技术细节设计。参与编写第6.3条和附录C。
14	王卿鹏	北京百度网讯科技有限公司	参与技术细节设计。参与编写第6.4条。
15	唐瑞博	中国交通信息科技集团有限公司	组织技术讨论，协助解决编制过程中的技术争议。参与编写第6.6条。
16	王　晶	中国交通信息科技集团有限公司	负责总体技术框架设计。参与编写第6.5条、附录D和附录F。
17	姚　琦	中国交通信息科技集团有限公司	负责编制过程中的技术咨询。参与编写第7章。
18	游　历	中交星宇科技有限公司	组织问卷发放、实地走访与座谈。参与编写第6.4条。
19	周鹏颖	中交星宇科技有限公司	负责“安全要求”章节的合规性及安全管控内容编制。参与编写第5章。
20	曾　玮	中国交通信息科技集团有限公司	负责把关技术细节设计。参与编写第6.1条和附录D。

二、标准编制原则和确定标准主要内容的依据

（一）编制原则

1.规范严谨性原则

严格按照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》编写内容与格式、专业术语的编写要求，与GB/T 41867《信息技术 人工智能 术语》、GB/T XXXXX《高质量数据集 格式要求》等现行国家标准保持术语体系统一，确保内容表达准确、条文措词精准、逻辑清晰严谨，结合交通运输行业公路、水路、铁路、民航、邮政、城市交通等多领域数据集建设实际调研情况，确保标准的规范性、权威性与可执行性。

2.协调兼容性原则

标准严格遵循《中华人民共和国数据安全法》、GB/T 41867、JT/T 1522《交通运输数据安全分级和保护要求》、TC609-5-2025-01《高质量数据集 建设指南》等现行国家法律法规、标准、行业标准，确保术语定义、技术要求、安全规范与上位要求相协调，实现与交通运输行业现有业务信息系统、平台的

无缝衔接，兼容行业现行的数据编码、格式规范、接口标准与安全管控要求，最大化复用行业现有数据资源与基础设施，避免重复建设，保障标准落地的平滑性与经济性。

3.科学先进性原则

标准内容基于人工智能、大数据领域的现代科技成果与交通运输行业数据集建设先进经验，核心技术要求与建设流程经行业多场景试点实践与严格科学验证，充分贴合交通运输行业不同场景高质量数据集建设的实际需求，可有效支撑交通运输行业人工智能模型开发、训练与评测全流程应用。

4.实用导向性原则

标准紧密结合我国交通运输行业数据资源建设现状与人工智能技术落地应用的实际需求，聚焦行业基础设施、交通装备、运输服务、行业治理、绿色低碳、安全保障等重点领域数据集建设痛点，服务于交通运输行业高质量数据集全生命周期建设体系构建与落地应用，全面提升行业数据资源标准化、规模化、高价值化应用水平，赋能行业数字化转型与智能化升级。

5.全面适用性原则

标准内容明确交通运输行业高质量数据集建设的总体原则与全流程技术要求，覆盖数据采集、预处理、标注、评测、发布全生命周期环节，针对行业通识与专识、不同业务领域、不同数据模态、不同模型应用阶段的建设需求，提出通用化建设规范与差异化实施指引，相关技术要求均参照当前行业主流的数据治理、智能标注、质量评测技术，兼具行业普适性与技术先进性。

（二）确定标准主要内容的依据

1.范围

明确了本标准的主要内容及适用范围。依据《国家数据基础设施建设指引》（发改数据〔2024〕1853 号）、《“数据要素 ×” 三年行动计划（2024—2026 年）》（国数政策〔2023〕11 号）和《国家数据标准体系建设指南》（发改数据〔2024〕1426 号），高质量数据集建设的要求，结合交通运输行高质量数据集应用场景，本标准规定了交通运输行业高质量数据集建设的总体要求、

建设流程以及安全与质量要求。本标准适用于交通运输行业管理部门、企事业单位、高等院校、科研机构、第三方服务机构及社会组织开展重点领域高质量数据集的建设。

2.规范性引用文件

规范性引用文件与本标准中的规范性技术要素具有同等效力。本标准规范性引用了多项国家和行业标准，主要有：

GB/T 4754 国民经济行业分类

GB/T 35274 数据安全技术 大数据服务安全能力要求

GB/T 35295 信息技术 大数据 术语

GB/T 37988 信息安全技术 数据安全能力成熟度模型

GB/T 41867 信息技术 人工智能 术语

JT/T 1522 交通运输数据安全分级和保护要求

3.术语和定义

为便于本标准的阅读和理解，在引用GB/T 35274、GB/T 35295、GB/T 41867界定的术语和定义之外，本标准还定义了交通运输行业高质量数据集、元数据及数据标注3个术语。

“交通运输行业高质量数据集”的术语定义主要是参考《中共中央国务院关于构建数据基础制度更好发挥数据要素作用的意见》以及国家数据局发布的TC609-5-2025-01，将其定义为：结合交通运输行业人工智能模型开发、训练等需求，经过系统化采集、标准化加工、专业化标注等 数据治理过程形成的，具备完整性、准确性、一致性、多样性等质量特征，可直接用于模型训练与应用的数据集合。

“元数据”的术语定义主要是参考 GB/T 18391.1《信息技术 元数据注册系统（MDR）第 1 部分：框架》以及国家数据局发布的 GB/T XXXXX《高质量数据集 格式要求》，将其定义为：描述数据信息资源特征的数据，以及关于数据所有权、存取路径、访问权限和数据易变性的数据。

“数据标注”的术语定义主要是参考 GB/T 42755《人工智能 面向机器学习的数据标注规程》，将其定义为：给数据样本指定目标变量和赋值的过程。

4.缩略语

从规范本标准缩略语的要求角度，采用国际通用缩写及专用缩写，给出了“AIS”“BIM”和“CNN”等缩略语，确保文档表述的简洁性和一致性。

5.总体要求

参考《国务院关于深入实施“人工智能+”行动的意见》以及《交通运输部 国家发展改革委 工业和信息化部 国家数据局 国家铁路局 中国民用航空局 国家邮政局关于“人工智能+交通运输”的实施意见》，并结合交通运输的业务特点，规定了总体要求的内容。

6.建设流程

6.1 流程框架

参考TC609-5-2025-01等标准，围绕数据采集、数据预处理、数据标注、数据质量评测、数据集发布等方面，规定了交通运输行业高质量数据集总体建设流程。

6.2 数据集分类与构成

6.2.1条为数据集分类与编码，参考《交通运输行业高质量数据集建设方案》及TC609-5-2025《高质量数据集》系列标准，规定了数据集依照知识领域覆盖范围、业务领域、数据模态、模型阶段的原则进行分类和编码的原则要求。

6.2.1.1条为知识覆盖范围分类，参考GB/T XXXXX《高质量数据集 分类指南》，规定了围绕知识领域覆盖范围的数据集分类原则。

6.2.1.2条为行业通识数据集分类，参考《交通运输行业高质量数据集建设方案》及JT/T 1058《交通运输信息系统 数据库字段命名及属性定义》和GB/T 4754《国民经济行业分类》，规定了围绕行业通识的数据集分类方法与编码规则。

6.2.1.3条为行业专识数据集分类，参考《交通运输行业高质量数据集建设方案》及JT/T 1058《交通运输信息系统 数据库字段命名及属性定义》，规定了围绕行业专识的数据集分类方法与编码规则。

6.2.1.4条为业务主题分类，参考《交通运输行业高质量数据集建设方案》及JT/T 1058—2026《交通运输信息系统 数据库字段命名及属性定义》和GB/T 4754-2017《国民经济行业分类》，并结合交通运输行业特征业务数据情况，规定了围绕业务领域的数据集分类方法与编码规则。

6.2.1.5条为数据模态分类，参考GB/T XXXXX《高质量数据集 分类指南》，规定了围绕数据模态类型的数据集分类原则。

6.2.1.6条为模型阶段分类，参考TC609-5-2025-01、GB/T XXXXX《高质量数据集 分类指南》，规定了围绕模型训练阶段的数据集分类原则。

6.2.2 条为数据集命名规则，参考 JT/T 1058—2026《交通运输信息系统 数据库字段命名及属性定义》、JT/T 747.5—2020《交通运输信息资源目录体系 第 5 部分：标识符编码规则》、GB/T XXXXX《高质量数据集 格式要求》、GB/T XXXXX《高质量数据集 分类指南》，规定了覆盖知识领域覆盖范围、行业、模态、数据集标识符、训练阶段及数据集简称信息的数据及编码和命名规则。

6.2.3条为数据集构成，参考GB/T 45288《人工智能 大模型》系列标准、JT/T 747《交通运输信息资源目录体系》系列标准及TC609-5-2025系列技术文件，规定了包含元数据、内容、说明文档的数据集构成。

6.2.3.1条为数据集元数据，参考JT/T 747. 3—2020《交通运输信息资源目录体系 第3部分：核心元数据》、GB/T XXXXX《高质量数据集 格式要求》，规定了5个数据集元数据的实体内容。

6.2.3.2条为数据集内容，参考GB/T 45288.1《人工智能 大模型 第1部分：通用要求》、GB/T 45288.2《人工智能 大模型 第2部分：评测指标与方法》、GB/T XXXXX《高质量数据集 格式要求》、GB/T XXXXX《高质量数据集 质量评测规范》，规定了以训练集、 验证集和测试集为划分的数据集内容。

6.2.3.3条为说明文档，参考GB/T XXXXX《高质量数据集 质量评测规范》，规定了数据集说明文档的相关要求。

6.3 数据采集

6.3.1 条为数据规划，参考 TC609-5-2025-01、GB/T XXXXX《高质量数据集 分类指南》、JT/T 697《交通信息基础数据元》系列标准以及交通运输行业高质量数据集建设需求，规定了数据集目标明确、数据来源梳理及数据架构设计等数据规划要求。

6.3.1.1 条为数据集目标确定，依据 TC609-5-2025-01 中数据规划阶段要求，并结合交通运输行业模型训练需求，明确应根据模型用途、训练方法及应用场

景，界定任务类型、数据类型及组织方式，确保数据集建设目标与模型能力需求一致。

6.3.1.2 条为数据来源规划，参考 TC609-5-2025-01、JT/T 1415《交通运输数据资源交换与共享》系列标准以及交通运输行业多源数据特点，规定应梳理政府与行业权威数据、交通运营企业数据、物联网与感知设施数据、互联网与商业数据、仿真与合成数据及政策标准与图书论文等知识类数据来源，并构建数据来源清单与获取路径。

6.3.1.3 条为数据架构设计，依据 TC609-5-2025-01 及 JT/T 697 系列标准，规定应围绕数据需求与来源建立统一数据资源目录与数据规范体系，确保数据组织结构统一、可管理与可扩展。

6.3.2 条为采集工作量评估，参考 TC609-5-2025-01 及交通运输行业数据工程实施要求，规定了交通运输行业高质量数据采集工作量评估方法，明确应从已有数据利用、外部采购、数据转化及接口建设等方面评估采集规模、实施周期及成本，并结合视频流、实时信号及多模态数据等不同类型数据的存储与传输需求制定资源分配与优先级策略。

6.3.3 条为数据采集实施，参考 TC609-5-2025-01、JT/T 1415 系列标准、JT/T 1534《国家综合交通运输信息平台视频资源接入技术要求》及交通运输行业数据采集相关规范，并结合交通运输行业多源异构数据特点，规定了数据采集方式及不同来源数据的采集要求。

6.3.3.1 条为通用采集要求，参考 TC609-5-2025-01、JT/T 1522 及 GB/T 43697《数据安全技术 数据分类分级规则》，规定数据采集可通过自有数据加工、外网汇聚、行业共建共享及依法采购等方式开展，并从安全性、时效性、数据质量、可扩展性及成本控制等方面提出数据采集技术要求。

6.3.3.2 条为不同来源数据采集要求，参考 JT/T 1415 系列标准、JT/T 1534 及 GB/T 35273《个人信息安全规范》，规定应针对文本数据、图像数据、音频数据、视频数据、雷达及点云数据、结构化数据、文本与图像多模态数据、文本与音频多模态数据、文本与视频多模态数据、融合类多模态数据、仿真与合成生成数据等多种来源，提出差异化数据采集方式及技术要求，明确多源异构数据统一接入、接口对接、流式采集与批量同步等实施方法。

6.4 数据预处理

6.4.1 条为数据转换，参考 TC609-5-2025-01，并结合交通运输行业高质量数据集建设的实际需求，规定了文本数据、图像数据、语音数据、视频数据、坐标数据、雷达数据等数据转换宜采取的方法。

6.4.2 条为数据验证，参考 TC609-5-2025-01，并结合交通运输行业高质量数据集建设的实际需求，规定了文本数据、图像数据、语音数据、视频数据、坐标数据、雷达数据等数据验证宜采取的方法。

6.4.3 条为数据清洗，参考 TC609-5-2025-01，并结合交通运输行业高质量数据集建设的实际需求，规定了修正错误与缺失数据等数据清洗宜采取的方法。

6.4.4 条为数据聚合，参考 TC609-5-2025-01，并结合交通运输行业高质量数据集建设的实际需求，规定了两个或多个数据集汇总聚合宜采取的方法。

6.4.5 条为数据抽样，参考 TC609-5-2025-01，并结合交通运输行业高质量数据集建设的实际需求，规定了文本、图像、语音、视频、坐标、雷达、信号控制等数据抽样宜采取的方法。

6.4.6 条为特征创建，参考 TC609-5-2025-01，并结合交通运输行业高质量数据集建设的实际需求，规定了文本、图像、语音、视频、坐标、雷达、信号控制等数据特征创建宜采取的方法。

6.4.7 条为数据增强，参考 TC609-5-2025-01，并结合交通运输行业高质量数据集建设的实际需求，规定了文本、图像、语音、视频、坐标、雷达等数据增强宜采取的方法。

6.4.8 条为数据融合，参考 TC609-5-2025-01，并结合交通运输行业高质量数据集建设的实际需求，规定了多源异构数据进行数据融合宜采取的方法。

6.5 数据标注

6.5.1条为明确标注需求，依据 GB/T 42755《人工智能 面向机器学习的数据标注规程》 数据标注前期准备要求，参考 JT/T 1522制定，规定了标注用途与对象的确定应回溯至模型设计目标，基于任务模式、工具与成本预估数据量；以及对重要数据、敏感信息开展安全审核，对个人信息去标识化和匿名化处理，并对接 JT/T 1522 及《中华人民共和国个人信息保护法》合规要求。

6.5.2条为制定标注规则，依据 GB/T 42755 任务执行框架，从数据模态与任务类型两个维度定义标注规则，任务类型划分为检测、分类、分割、跟踪、预测、信息抽取、生成共7类，覆盖行业人工智能应用主要需求。

6.5.3条为分配标注任务，依据 GB/T 42755 标注人员管理要求制定，规定要求标注人员具备交通运输行业基础知识、建立技能矩阵并动态调整、明确过程监测指标、并实行交叉验证并对标注平台实施安全管理，对接 JT/T 1522 要求。

6.5.4条为实施标注任务，依据 GB/T 42755 任务执行要求，明确"机器预标注+人工审核"与"人工主标注+机器辅助"两种模式，并对预标注模型提出性能验证与退化监测要求；明确了结构化数据标注应覆盖时空特征、交通事件、设施状态等行业核心要素；明确了单模态数据按任务类型标注；规定了多模态数据实施联合标注、构建跨模态关联关系。

6.5.5条为质量控制，依据 GB/T 42755 质量控制要求，参考 GB/T 36344《信息技术 数据质量评价指标》数据质量评价指标框架制定。本条要求根据质量、成本与进度制定组合式策略，可通过多人验证、埋题验证、以及自动化验证的方式进行。

6.5.6条为质量检查，依据GB/T 36344 数据质量评价指标框架并参考 GB/T 2828.1 《计数抽样检验程序 第1部分：按接收质量限（AQL）检索的逐批检验抽样计划》抽样检验标准的通行做法制定。本条将质量检查划分为逐条检查、按比例抽查、抽样检验三类，并明确适用场景。同时规定当检查结果低于测标指标预期时应重新标注，确保质量问题可兜底处理。

6.5.7条为反馈修正，依据 GB/T 42755标注迭代优化要求制定，规定对问题进行分类统计；对接 GB/T XXXXX《高质量数据集 质量评测规范》 要求逐条记录原始标注人员、纠正人员、原始与纠正内容及时间等信息，确保全过程可追溯；并建立双重校验机制，未通过复核的启动二次修正。

6.6 数据质量评测

6.6.1条为文档完整性，结合高质量数据集在人工智能模型训练中对可追溯性、可复现性的实际需求，规定了数据集应具备基础信息完整性、内容特征完

整性、建设过程完整性、应用说明完整性，确保数据集配套文档的规范性和完整性。

6.6.2条为质量符合性，参考GB/T 37988、JT/T 1522等标准，结合交通运输行业多源异构数据的特点，从数据完整性、现时性、安全性、内容一致性四个维度提出质量合规性要求，确保数据集在采集、存储、标注等环节满足行业标准和业务需求。

6.6.3条为场景适用性，参考GB/T 41867中关于模型训练数据质量的相关概念，结合交通运输典型应用场景（如路面病害检测、交通事件识别、客流预测等），从内容多样性、数据均衡性、标注准确性、规模完整性、模型适配性五个方面提出场景适用性要求，确保数据集能够有效支撑人工智能模型的训练与评估。

6.7 数据集发布

6.7.1条为数据集封装，参考TC609系列技术文件，结合交通运输行业数据流通的实际需求，规定了数据集封装的基本要求。6.7.1.1条为数据集配置，根据预训练、后训练、评测等不同类型数据集在模型生命周期中的不同用途，明确了各类数据子集的配置策略；6.7.1.2条为数据标准化与安全预处理，规定了数据格式统一、加密存储及版本标识等要求；6.7.1.3条为产品化封装，提出了实时服务接口、离线数据包、在线数据平台等封装形态；及与知识图谱结合的相关要求；6.7.1.4条为版本管理，采用“主版本号.次版本号.修订号”规则，确保数据产品与文档版本对应、变更可追溯。

6.7.2条为数据集流通，依据《中共中央 国务院关于构建数据基础制度更好发挥数据要素作用的意见》《“数据要素×”三年行动计划（2024—2026年）》等政策文件，参考JT/T 1415系列标准，结合交通运输行业数据分级分类管理要求，规定了数据开放共享、数据交易、定向共享与协同、可信数据空间数据流通等流通方式，明确了各方式下的安全要求和适用范围。

6.7.3条为数据集运营，依据《交通运输政务数据共享管理办法》（交科技发〔2021〕33号）等文件，参考GB/T 36073—2018《数据管理能力成熟度评估模型》，结合高质量数据集持续更新、迭代优化的实际需求，规定了组织管理、更新与维护等运营要求，确保数据集在发布后的长期可用性和价值延续。

7. 安全与质量要求

7.1条为数据安全与合规处理，依据《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》《数据出境安全评估办法》等法律法规，参考JT/T 1522、GB/T 37988等标准，结合交通运输行业数据集建设全生命周期安全管理需求，规定了数据来源合法性、敏感数据差异化加密存储与脱敏处理、数据投毒防范、基于角色的访问控制、版权数据合规处理、数据出镜安全评估及合规审计等要求。特别增加了对图书、论文等受版权保护数据的合规处理要求，明确应依法取得授权并进行版权标识与存证；同时强调应充分利用隐私计算、区块链等技术实现数据“可用不可见”，为数据共享和开放提供分类分级、数据水印、数据脱敏等安全防护能力，确保高质量数据集在建设、流通、使用全过程中的安全可控、可追溯。

7.2条为建设质量控制与持续运维，依据《中华人民共和国数据安全法》等法律法规，参考JT/T 1522的规定，规范了数据集全生命周期质量管控与运维管理工作。建设阶段明确过程自查、问题闭环、档案留存三项要求：在数据采集、预处理、标注、评测、发布等关键环节设置阶段性质量检查节点并形成检查记录；对自查、评测或使用反馈中发现的问题建立整改台账，实行问题发现、整改、验证、归档的闭环管理，对系统性或重复性问题开展根因分析并制定预防措施；全过程留存完整档案记录，保存期限不宜少于5年，涉及重要数据或核心业务的档案永久保存。运维阶段确立迭代更新、用户反馈与服务、退役管理机制：根据业务需求变化、技术发展及使用反馈定期评估与迭代更新，记录更新原因、范围及影响评估；建立用户反馈渠道并及时响应处理；数据集退役时按规定进行数据归档、迁移或安全销毁，涉及重要数据或敏感信息的退役遵循JT/T 1522相关要求。通过全流程标准化管控与常态化运维，切实保障数据质量、可用性与安全性。

三、技术经济论证或预期的经济效果

（一）经济效果

本标准是交通运输行业首个针对高质量数据集建设全流程的专项指导性标准，其制定与实施将为行业人工智能模型训练、智慧交通场景落地、数据要素

价值释放提供统一的技术规范与建设指引，具备显著的经济效益，核心体现在以下方面：

一是降低行业数据集建设的制度性成本与重复建设投入。本标准明确了交通运输行业高质量数据集从采集、预处理、标注、评测到发布、安全管控的全流程建设规范，统一了数据集分类编码、命名规则、元数据要求、质量评测指标等核心技术标准，解决了当前行业内数据集建设标准不统一、格式不兼容、技术要求不一致导致的“数据孤岛”问题。通过标准化建设，可有效减少行业内各单位在数据集建设过程中的重复研发、重复测试、格式转换适配等无效投入，大幅降低单个企业、科研机构的数据集建设成本，同时提升跨主体、跨区域、跨场景的数据共享与流通效率，减少数据适配环节的资源消耗。

二是提升交通人工智能模型研发与落地的经济效益。本标准围绕交通运输行业基础设施、运输服务、安全保障、绿色低碳等核心场景，明确了高质量数据集的建设要求与质量基准，为行业大模型、垂直领域智能算法的训练、优化与评测提供了高质量数据底座。通过规范数据集建设流程与质量标准，可有效解决因训练数据质量缺陷导致的模型迭代周期长、泛化能力弱、落地效果差等行业痛点，大幅缩短智能算法从研发到规模化应用的周期，降低模型训练与迭代的试错成本。同时，标准化的高质量数据集可推动路面病害检测、交通流量预测、船舶航行安全监测、枢纽旅客智能服务等场景的智能算法规模化商用，提升交通基础设施管养、运输服务调度、安全应急处置的效率，减少因病害发现不及时、应急处置滞后、运输调度不合理带来的直接经济损失，为智慧交通产业规模化发展提供核心数据支撑。

三是赋能交通运输数据要素市场化流通，释放数据资产价值。本标准明确了数据集产品化封装、版本管理、流通共享、合规运营的全流程要求，为交通运输行业数据要素的合规流通、市场化交易提供了标准化依据。通过规范数据集建设的合规性、安全性与质量标准，可有效破解数据供方“不愿共享、不敢流通”的核心困局，推动行业数据资源通过开放共享、市场化交易、定向协同等方式实现高效流通，将交通领域海量数据资源转化为可量化、可交易、可增值的数据资产，培育交通数据要素市场新业态，降低数据要素流通的交易成本与合规风险，为交通运输行业数字化转型注入新的经济增长动力。

（二）社会效果

本标准的制定与实施，具备显著的社会价值与行业示范效应，核心体现在以下方面：

一是填补交通运输行业高质量数据集建设标准空白，完善行业数字化标准体系。当前我国数据要素领域相关标准多为通用型规范，针对交通运输行业场景特性、业务需求的高质量数据集建设专项标准尚属空白。本标准结合交通运输行业公路、水路、铁路、民航、城市交通等多领域业务场景，构建了全流程、全要素的数据集建设规范体系，完善了交通运输信息通信及导航领域的标准体系，为行业内各级管理部门、企事业单位、科研机构开展数据集建设工作提供了统一的遵循依据，推动行业数据资源建设的规范化、标准化发展。

二是保障交通运输行业数据安全与合规，守护公众信息安全与行业数据安全。本标准专章明确了数据集建设全生命周期的安全要求，对数据来源合法性、个人信息脱敏匿名化、敏感数据加密保护、数据投毒防范、数据出境合规、版权保护等提出了明确要求，与《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》及交通运输行业数据安全相关标准形成有效衔接。通过标准实施，可规范行业内数据集建设过程中的数据安全管理行为，防范数据泄露、篡改、滥用等安全风险，既保障了公众个人信息安全，也维护了交通运输行业关键数据与国家重要数据安全，为行业数字化发展筑牢安全底线。

三是推动交通运输行业科技创新与产学研用协同发展。本标准明确了交通运输行业高质量数据集的建设规范与质量要求，为高校、科研机构、科技企业、交通运营单位开展协同创新提供了统一的技术底座。标准化的高质量数据集可有效支撑交通领域人工智能、数字孪生、大数据分析等前沿技术的研发与验证，推动科技创新成果与交通业务场景的深度融合，加快智慧交通技术的成果转化与规模化应用。同时，标准的实施可推动行业内数据资源的共建共享，促进产学研用各方协同合作，培育行业科技创新生态，提升我国交通运输行业数字化、智能化发展的整体水平。

四是提升交通运输行业现代化治理能力，服务交通强国建设。本标准围绕交通强国建设总体要求，聚焦基础设施、运输服务、行业治理、安全保障、绿色低碳等交通强国建设核心领域，规范了相关场景高质量数据集的建设要求，

为行业治理数字化、智能化升级提供了数据支撑。通过标准化的数据集建设，可推动交通运输行业管理从“经验驱动”向“数据驱动”转型，提升行业运行监测、市场监管、应急处置、公共服务的数字化、精准化水平，助力构建现代化综合交通运输体系，为交通强国建设提供坚实的数据要素保障。

四、与国际、国外同类标准水平的对比情况

当前，全球范围内针对人工智能高质量数据集建设的标准规范仍处于快速发展阶段，国际标准化组织及主要发达国家多聚焦于通用型数据质量、人工智能数据规范、数据安全相关标准的制定，针对交通运输垂直行业的高质量数据集建设全流程专项标准尚未形成成熟、统一的体系。

国际层面，ISO/IEC JTC 1/SC 42（人工智能分技术委员会）牵头制定了ISO/IEC 5259 系列《人工智能—分析和机器学习的数据质量》、ISO/IEC 20547《信息技术 大数据 参考架构》等通用型标准，明确了人工智能数据集的基础质量要求与数据标注通用规范，为全球数据集建设提供了基础框架；IEEE 针对智能交通场景发布了 IEEE 2846—2022《自动驾驶系统安全相关模型中的假设标准》等数据相关标准，聚焦特定交通场景的技术要求，但未覆盖交通运输全领域、全场景的数据集建设全流程。欧美等发达国家在智慧交通发展过程中，多由企业或行业联盟形成细分场景的数据集建设规范，缺乏政府层面主导的、覆盖交通运输全业务领域的行业级数据集建设统一标准。

国内层面，全国数据标准化技术委员会发布了 TC609-5-2025 系列技术文件，为各行业高质量数据集建设提供了通用框架与基础要求；交通运输行业已发布 JT/T 1522、JT/T 1480《交通运输数据脱敏指南》、JT/T 1415 系列标准等数据相关标准，聚焦数据安全、数据交换共享等单一环节，尚未形成覆盖数据集建设全生命周期、全业务领域的专项行业标准。

本标准在编制过程中，充分对标了 ISO/IEC、IEEE 等国际通用标准的先进理念与核心要求，全面衔接了国内 TC609 系列通用标准与交通运输行业现有数据相关标准，同时深度结合我国交通运输行业的业务特性、场景需求与数字化发展现状，形成了差异化、创新性的行业专项规范，整体技术水平与国际先进标准保持同步，在行业适用性、场景覆盖度、流程完整性上具备显著优势，填

补了交通运输行业高质量数据集建设专项标准的空白。

一是实现了交通运输全业务领域的全覆盖，行业适配性显著优于国际通用标准。本标准突破了国际标准多聚焦自动驾驶单一交通场景的局限，全面覆盖公路、水路、铁路、民航、邮政、城市交通等交通运输全方式，划分了基础设施、交通装备、运输服务、行业治理、绿色低碳、安全保障、科技创新 7 大业务领域，针对不同业务场景的数据集建设提出了差异化要求，更贴合我国综合交通运输体系发展的实际需求。

二是构建了数据集建设全生命周期的闭环规范体系，流程完整性远超现有相关标准。本标准从总体要求、分类及构成、数据采集、预处理、标注、评测、发布到安全要求，构建了“采集-治理-标注-评测-发布-共享-应用”全生命周期的完整规范体系，解决了现有国际、国内标准多聚焦单一环节、缺乏全流程统筹规范的问题，为行业提供了可落地、可执行的全流程建设指引。

三是形成了标准化的分类编码与命名规则，具备更强的实操性与通用性。本标准针对交通运输行业特性，设计了知识领域、业务领域、数据模态、模型阶段的四维分类编码体系，明确了统一的数据集命名规则与元数据要求，解决了行业内数据集分类不统一、命名不规范、跨主体不兼容的痛点，相比国际通用标准，更具备行业实操性，可有效支撑行业数据的跨主体共享与流通。

四是配套了丰富的行业场景建设案例，可落地性显著提升。本标准通过附录形式，补充了公路路面病害检测、道路运维、海上风电海底电缆安全监测、综合交通枢纽旅客智能服务等典型场景的高质量数据集建设案例，将通用技术要求与行业实际场景深度结合，为行业各单位开展数据集建设提供了可直接参考的实操范本，这是现有国际、国内同类标准均未覆盖的内容，大幅提升了标准的落地性与指导性。

五、与有关法律、法规及相关标准的关系

本标准在编制过程中，充分研究了国家现行相关法律、法规、规章，以及国家、行业现行有效标准，严格遵循相关法律法规要求，与现行国家标准、行业标准保持充分协调、有效衔接，无内容冲突与合规性矛盾。

（一）与现行相关法律法规规章的协调性

本标准的编制严格遵循国家现行法律、行政法规与部门规章的核心要求，相关条款设计全面落实上位法规定，具备充分的合规依据，核心对应关系如下：

一是严格遵循数据领域核心上位法要求。本标准全面落实《中华人民共和国数据安全法》《中华人民共和国网络安全法》《中华人民共和国个人信息保护法》《关键信息基础设施安全保护条例》中关于数据全生命周期安全管理、个人信息保护、数据分类分级管理、数据安全合规审计等核心要求，在标准中专门设置安全要求章节，对数据集建设全流程的安全管控、个人信息脱敏匿名化、敏感数据加密保护、数据出境合规、版权保护等作出明确规定，与上位法要求形成全面衔接。

二是全面对接国家数据要素市场化相关政策要求。本标准紧扣《中共中央国务院关于构建数据基础制度更好发挥数据要素作用的意见》《“数据要素×”三年行动计划（2024-2026 年）》《关于构建更加完善的要素市场化配置体制机制的意见》等国家政策文件中关于培育数据要素市场、规范数据要素流通、提升数据要素质量的核心部署，明确了交通运输行业高质量数据集的建设规范、产品化封装要求与流通运营规则，为交通运输行业数据要素合规高效流通提供了标准支撑，与国家数据要素发展战略保持高度一致。

三是充分契合交通运输行业发展相关法规与政策要求。本标准严格遵循《中华人民共和国公路法》《中华人民共和国港口法》《中华人民共和国铁路法》《中华人民共和国民用航空法》等交通运输行业专项法律，以及《交通强国建设纲要》《国家综合立体交通网规划纲要》等政策文件中关于交通运输数字化、智能化发展的总体要求，聚焦交通基础设施建管养运、运输服务提质升级、行业治理现代化、安全生产保障等核心业务，规范相关场景高质量数据集建设要求，为交通强国建设提供数据标准支撑，与交通运输行业发展顶层设计全面契合。

四是有效衔接政务数据共享与开放相关管理规定。本标准落实《交通运输政务数据共享管理办法》《政务信息资源共享管理暂行办法》等规章要求，在数据发布章节明确了数据集分级分类开放共享、定向协同共享的相关要求，与交通运输行业政务数据共享管理体系形成有效衔接，为行业政务数据资源的标

准化治理与合规共享提供了规范依据。

（二）与现行标准的协调性

本标准的技术要求与国内信息安全、大数据、人工智能、交通运输数据管理领域现行有效标准保持高度一致，无技术冲突，同时针对交通运输行业高质量数据集建设场景进行了细化与补充，形成了“通用标准为基础、行业专项为补充”的协调体系，核心衔接关系如下：

1.规范性引用文件的全面衔接

本标准规范性引用了 11 项现行有效国家及行业标准、技术文件，包括 GB/T 35274、GB/T 35295、GB/T 41867、JT/T 1522 以及 TC609-5-2025 系列。本标准中关于术语定义、数据安全要求、数据质量评测、数据集分类等核心条款，均与上述引用标准保持一致，相关技术要求在引用标准基础上，结合交通运输行业特性进行了细化与落地，确保标准体系的统一性与协调性。

2.与人工智能、大数据领域通用国家标准的协调性

本标准全面衔接 GB/T 37988、GB/T 43697、GB/T 22239《信息安全技术网络安全等级保护基本要求》、GB/T 35273 等国家标准，在数据分类分级、数据安全管控、个人信息保护、数据全生命周期管理等方面，严格遵循上述国家标准的通用要求，同时针对交通运输行业数据集建设的特定场景，提出了更具针对性的实施要求，是通用国家标准在交通运输垂直领域的细化落地与行业延伸。

3.与交通运输行业现行数据相关标准的协调性

本标准与 JT/T 1415 系列标准、JT/T 697 系列标准、JT/T 1480、JT/T 1534 等行业现行标准保持充分协调。上述标准为交通运输行业数据元、数据交换共享、数据脱敏、视频数据接入等提供了基础规范，本标准在数据集建设的采集、预处理、数据格式规范等环节，全面沿用了上述行业标准的核心要求，确保与交通运输行业现有数据标准体系的兼容性与一致性，避免重复建设与标准冲突。同时，本标准填补了现有行业标准在高质量数据集全流程建设、分类编码、标注规范、质量评测等方面的空白，与现有行业标准形成互补，共同完善了交通运输行业数据领域标准体系。

4.与其他相关领域标准的协调性

本标准在场景化建设要求中，衔接了公路工程、水运工程、海上风电、综合交通枢纽等领域的现行技术标准，如 JTG 5210-2018《公路技术状况评定标准》、NB/T 11299-2023《海上风电场工程光纤复合海底电缆在线监测系统设计规范》等，确保标准中相关场景的数据集建设要求与对应领域的技术规范、评定标准保持一致，保障数据集建设与行业实际业务需求的深度契合，提升标准的落地性与实操性。

六、重大分歧意见的处理经过和依据

截至目前，本标准在编制过程中，未出现重大分歧意见。

七、涉及专利的有关说明

在本标准编制过程中，暂未接到任何涉及相关专利或知识产权争议的信息、文件。

八、标准过渡期的建议

为确保本标准平稳落地、有效实施，结合实际执行需求，提出以下过渡期建议：

（一）过渡期时长设定

考虑到数据集建设涉及数据采集、预处理、标注、评测、发布等多个环节，各单位需按本标准要求完善相关技术流程和管理制度，预计耗时 6 个月；同时需预留 3 个月用于标准宣贯培训，确保相关主体准确理解标准要求；另预留 3 个月用于各单位试点应用和问题反馈。建议本标准发布后设置 12 个月过渡期，期满后正式实施。

（二）过渡期配套措施

标准发布后，建议主管部门牵头组织多形式宣贯培训。通过线上线下结合、分层分类覆盖的方式，面向相关单位、网络平台运营者等开展解读，推动各方全面掌握标准要求。同时强化贯彻落实的引导与监督，充分发挥本标准的规范引领作用。建议选择 3-5 家具有代表性的单位开展标准试点应用，总结实践经

验，形成典型案例，为标准全面推广提供参考。过渡期结束后，建议对标准实施情况进行评估，收集各方反馈意见，为后续标准修订完善提供依据。

九、其他应予说明的事项

（一）废止现行有关标准的建议

本标准为新制定的标准，无废止现行的有关标准。

（二）公平竞争审查结论的有关说明

根据部有关规定及《公平竞争审查条例实施办法》要求进行公平竞争审查说明如下：本标准为行业标准，旨在规范交通运输行业高质量数据集建设的技术要求，不涉及市场推入、资质限制或行政垄断内容。本标准内容不含有排除、限制市场竞争的条款，遵循开放、透明、非歧视原则，避免对特定企业、技术路径或区域形成不合理限制。本标准符合《中华人民共和国标准化法》《中华人民共和国反垄断法》相关规定，内容不涉及行政垄断或市场分割。技术条款与现有国家标准、行业标准兼容，支持主流数据格式和处理技术，未形成技术壁垒。