

黑 龙 江 省 地 方 标 准

DB XX/T XXXX—XXXX

工业互联网综合平台数据质量管理规范

（征求意见稿）

（联系人：马超 电话：15804514973 邮箱：machao8396@163.com）

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX – XX – XX 发布

XXXX – XX – XX 实施

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 1

5 数据分类分级 2

 5.1 数据分类 2

 5.2 数据分级 2

6 数据质量管理框架 3

7 数据质量管理要求 3

 7.1 数据质量管理模型 3

 7.2 数据质量定义 3

 7.3 数据质量判断 3

 7.4 数据质量评估 4

 7.4.1 数据质量评估准则 4

 7.4.2 数据质量核查 4

 7.4.3 数据质量分析 5

 7.4.4 数据质量监督 5

 7.4.5 数据质量控制 5

8 数据质量管理方法 6

 8.1 基本要求 6

 8.2 数据质量合规准则 6

 8.3 数据质量合规审查任务 6

9 数据质量监测与分析 7

 9.1 质量监控策略 7

 9.2 数据质量警报 7

 9.3 数据质量修复 7

 9.4 数据质量综合分析报告 7

10 数据质量改进 8

 10.1 改进措施 8

 10.2 数据质量合规跟踪 8

参 考 文 献 10

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。

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。

本文件由黑龙江省工业和信息化厅提出并归口。

本文件起草单位：哈尔滨理工大学、黑龙江省标准化研究院、黑龙江省网络空间研究中心、中数(深圳)时代科技有限公司、哈尔滨工业大学、国家电网有限公司、国网黑龙江省电力有限公司电力科学研究院。

本文件主要起草人：马超、刘志伟、吴宏伟、张罗刚、周廷楠、张颜遵、赵天诚、李双翼、叶子、关志博、黄海、方舟、白瑞、于海宁、吴英东、王孝余、乔羽、史心月、尚方、刘生。

工业互联网综合平台数据质量管理规范

1 范围

本文件界定了工业互联网综合平台数据质量管理的术语和定义，规定了数据质量管理的缩略语、分类分级、数据质量管理框架、数据质量管理要求、数据质量管理方法等内容。

本文件适用于工业互联网综合平台数据质量的监督、分析和质量改进。

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 25000.12—2017 系统与软件工程 系统与软件质量要求和评价（SQuaRE） 第12部分：数据质量模型

GB/T 36344 信息技术 数据质量评价指标

GB/T 39400—2020 工业数据质量 通用技术规范

GB/T 42381.8 数据质量 第8部分：信息和数据质量：概念和测量

3 术语和定义

GB/T 39400—2020 界定的以及下列术语和定义适用于本文件。

3.1

工业互联网综合平台

以工业互联网平台为基础，支持汇聚数据、服务、用户等各类资源，具备数据集成分析、应用支撑能力和基础应用能力，支持省域内特色型、专业型、企业级等平台的集成和接入，支撑制造资源泛在连接、弹性供给、高效配置和协同创新的载体。

3.2

数据质量

在指定条件下使用时，数据的特性满足明确的和隐含的要求的程度。

[来源：GB/T 25000.12—2017,4.3]

3.3

数据质量管理

指导和控制某机构数据质量的协调活动。

[来源：GB/T 39400—2020,3.2]

3.4

数据质量管理工具

运用信息化技术和管理方法来规划、实施与控制数据质量的识别、评估、提升等一系列活动。

4 缩略语

下列缩略语适用于本文件。

App: 应用程序 (Application)

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

LLM: 大语言模型 (Large Language Model)

5 数据分类分级

5.1 数据分类

工业互联网综合平台数据包括但不限于电力、制造、物流运输、煤炭、燃气、建筑、水务、现代服务、化工、冶金等行业领域产品和服务全生命周期产生和应用的数据，结合行业要求、业务规模、数据复杂程度等实际情况，可将数据分为以下类型：

- a) 研发数据：包括研发设计数据、开发测试数据等；
- b) 生产数据：包括控制信息、工况状态、工艺参数、系统日志、生产质量数据、生产实绩数据等；
- c) 供应链数据：包括供需计划数据、仓储物流数据等；
- d) 营销数据：包括投标次数、订单数量、交易金额、客户异议数据等；
- e) 运维数据：包括产品运行状况数据、产品售后服务数据等；
- f) 管理数据：包括客户基本信息、业务合作数据、人事财务数据、系统设备资产信息、产品基本信息、项目进度数据、业务统计数据（如资源量数据、能耗监测数据）等；
- g) 金融服务数据：包括信贷服务数据、融资租赁服务数据、征信服务数据等；
- h) 平台运营数据：接入的设备数据、工业模型库数据、工业 App 数据、平台运行数据、研发数据等；
- i) 外部数据：与其他主体共享的数据。

5.2 数据分级

不同类型的工业互联网综合平台数据在遭受泄露或非法利用后，根据对效益、客户、生产等方面带来的负面影响，可将工业互联网综合平台数据分为以下三个级别：

- a) 一级数据：
 - 1) 对工业互联网系统以及其他设备正常运行影响较小；
 - 2) 受影响的客户和企业数量较少、影响时间较短、产生的影响辐射的生产区域较小；
 - 3) 恢复受损数据或消除负面影响代价较小；
 - 4) 直接经济损失较小，或者给企业造成负面影响较小。
- b) 二级数据：
 - 1) 易引发较大的生产安全事故或突发环境事件；
 - 2) 受影响的客户和企业数量较多、或可导致客户以及供应商信息资源被非法获取；
 - 3) 引发的级联效明显，影响到多个行业、区域；
 - 4) 恢复受损数据或消除负面影响代价较大；
 - 5) 直接经济损失较大，或者给企业造成负面影响较大。
- c) 三级数据：
 - 1) 易引发特别重大的生产安全事故或突发环境事件；
 - 2) 严重影响自身企业运行；
 - 3) 对国民经济、行业发展、公众利益、社会秩序乃至国家安全造成严重影响；
 - 4) 造成直接经济损失巨大。

6 数据质量管理框架

数据质量管理框架主要由数据质量管理要求和数据质量管理方法构成,见图1。

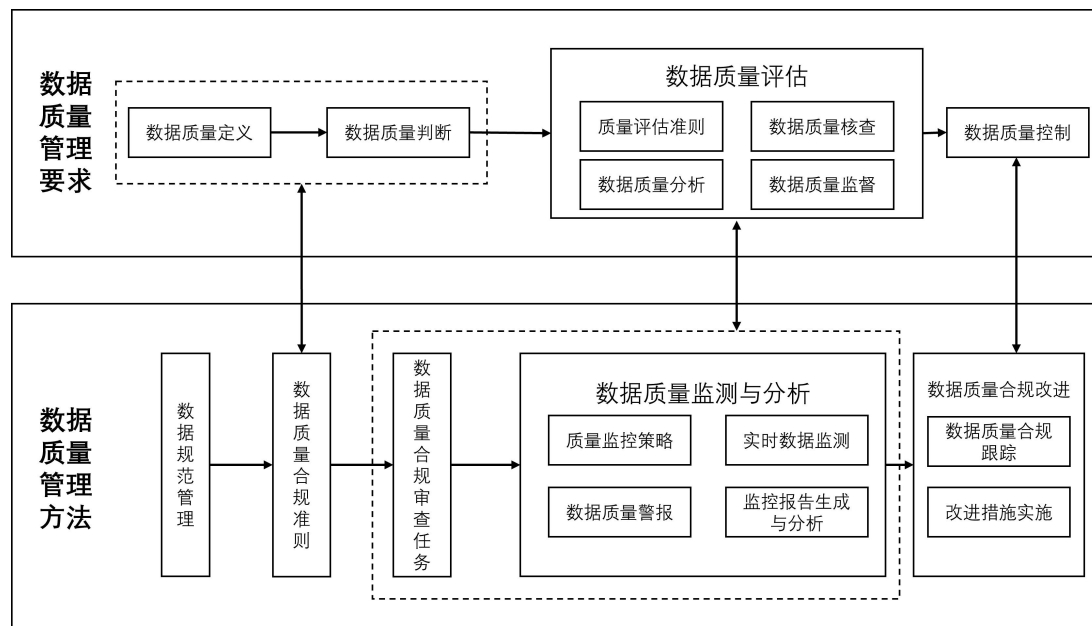


图 1 数据质量管理框架

7 数据质量管理要求

7.1 数据质量管理模型

数据质量管理模型采用“定义-测量-分析-改进-控制”循环（DMAIC），其中包括定义、测量、分析、改进和控制5个阶段：

- a) 定义 (Define)：明确项目范围和目标，识别并定义需要改进的问题或机会；
- b) 测量 (Measure)：收集当前过程的数据和度量，以了解过程的性能和识别关键质量特性；
- c) 分析 (Analyze)：理解数据，识别问题的根本原因，以及确定影响过程输出的关键因素；
- d) 改进 (Improve)：开发和实施解决方案，消除存在的问题；
- e) 控制 (Control)：确保改进的成果得以维持，并防止过程回归到原来的状态。

7.2 数据质量定义

按GB/T 42381.8数据质量定义要求，包括但不限于：

- a) 应依据平台企业的管理需求, 对企业运营过程中的数据需求进行详尽的分析, 并参考外部监管机构的相关规范, 以确保数据质量管理目标与企业战略相匹配;
- b) 应依据平台企业的数据质量管理目标, 结合数据标准规范和数据质量特性, 全面评估数据的来源、类型、使用情境及用户需求, 以界定数据质量评估的关键维度。

7.3 数据质量判断

数据质量判断要求包括但不限于:

- a) 应根据平台企业的业务发展需求，对数据质量管理的范围进行界定。在此基础上，针对常见的数据问题，进行评估以确定各类数据的优先级，并界定不同数据的质量要求；
- b) 应针对平台企业的数据质量管理目标及其具体需求，对不同类别的数据制定相应的质量评价指标；
- c) 应随着业务发展需求的演变以及数据质量检查分析结果的反馈，持续地对数据质量规则进行优化和更新。

7.4 数据质量评估

7.4.1 数据质量评估准则

应根据数据质量规则，按GB/T 36344制定相关数据质量评估包括但不限于：

- a) 真实性：描述数据必须真实准确反应客观实体或真实业务；
- b) 一致性：描述数据是否遵循了统一的规范，数据集合是否保持了统一的格式；
- c) 规范性：描述数据应遵循预定的语法规则，数据类型、格式、取值范围、存储格式等应符合定义；
- d) 完整性：描述数据是否存在缺失记录或缺失字段；
- e) 及时性：描述数据的产生和供应是否及时；
- f) 准确性：描述数据是否与其对应的客观实体的特征相一致
- g) 有效性：描述数据是否满足用户定义的条件或在一定的域值范围内；
- h) 关联性：描述数据是否存在函数关系、相关系数、主外键关系、索引关系等数据关联性问题；
- i) 唯一性：描述数据是否存在重复记录；
- j) 稳定性：描述数据的波动是否是稳定的，是否在其有效范围内；
- k) 可访问性：描述数据是否易于检索和使用，以及数据的可访问性是否符合授权和安全要求；
- l) 安全性：描述数据保护措施是否充分，包括数据的保密性、防止未授权访问和数据泄露的机制；
- m) 可靠性：描述数据在存储、传输和处理过程中的可靠性，以及数据备份和恢复的能力；
- n) 可维护性：描述数据的维护和更新过程是否高效，以及数据结构是否便于未来的维护和扩展；
- o) 互操作性：描述数据是否能够与其他系统或数据源兼容，以及数据交换的流畅性和准确性；
- p) 合规性：描述数据的收集、处理和存储是否符合相关法律法规和行业标准；
- q) 持续性：描述数据支持业务连续性的能力，以及在面对系统故障或灾难时数据的恢复和持续供应能力；
- r) 可追溯性：描述数据的来源、处理历史和分发路径是否可以被追踪，以及数据的变更记录是否完整。

7.4.2 数据质量核查

应根据数据质量规则中的有关技术指标和业务指标、校验规则与方法，对平台企业的数据质量情况进行核查，若发现数据质量问题应及时向数据管理人员进行反馈并生成问题报告。数据质量核查要求包括但不限于：

- a) 数据合法性：验证数据字段的取值是否限定在预设的合法范围内，并确保字段类型与实际数据内容相符；
- b) 编码规范性：审查数据字段的编码是否遵循既定的规则，确保编码的一致性；
- c) 数据完整性：评估数据记录是否包含所有必要的关键属性，无遗漏；
- d) 数据准确性：对比数据字段的实际数值精度与定义的数据类型精度，确保精度一致；

- e) 数据及时性：确保数据按照既定的频率及时更新，满足时效性要求；
- f) 主键唯一性：确保数据表中的物理主键和逻辑主键的唯一性，避免重复记录；
- g) 外键关联性：核查数据的外键是否符合引用完整性的要求，保证数据之间的正确关联；
- h) 数据可读性：确保数据的表达和呈现方式易于用户理解和解读，无论是通过视觉格式、语言清晰度还是数据的组织结构；
- i) 数据可操作性：检查数据是否能够被系统或用户有效地操作和处理，包括数据的导入、导出、转换和分析能力；
- j) 数据可扩展性：评估数据结构和数据库设计是否支持未来的增长和扩展，确保数据能够适应业务发展和变化的需求。

7.4.3 数据质量分析

数据质量分析要求包括但不限于：

- a) 应对数据质量检查过程中发现的问题进行深入分析，识别并探究影响数据质量的根本原因；
- b) 应对识别出的数据质量问题依据严重性和影响范围进行优先级排序，确保关键问题得到及时解决；
- c) 应对数据质量趋势进行定期监测和分析，以便预测潜在的数据风险，并制定预防措施；
- d) 应对数据质量改进措施的效果进行评估，确保所采取的行动能够实质性地提升数据质量，并根据评估结果调整改进策略。

7.4.4 数据质量监督

应根据数据质量规则中的有关技术指标和业务指标、校验规则与方法，对平台企业的数据质量情况进行实时监控、检查异常并完善规则。数据质量监督要求包括但不限于：

- a) 应对数据核查中发现的异常内容及时触发结果告警机制，确保问题能够迅速被识别和响应；
- b) 应对数据处理的每个环节进行详尽的跟踪，确保数据流动的透明性和可追溯性；
- c) 应对数据质量报告提供灵活的订阅选项，允许用户根据需求定制化接收相关信息；
- d) 应对数据质量检查结果进行系统化管理，包括记录、存储和易于访问的查询；
- e) 应根据平台企业的数据质量规则进行定期检查，发现数据异常时及时进行规则的调整和完善；
- f) 应根据质量评估结果调整或改进数据质量控制措施；
- g) 应对数据质量监督过程中使用的技术指标和业务指标进行持续的评估和优化，以适应不断变化的业务需求；
- h) 应对数据质量监督工具和系统进行定期的审查和升级，确保其有效性并采用最新的监控技术。

7.4.5 数据质量控制

数据质量控制要求包括但不限于：

- a) 应根据数据质量分析的结果，制定详尽的改进方案，并严格执行，包括对错误数据的及时更正、业务流程的持续优化以及应用系统中发现的问题的修复；
- b) 应建立综合性的数据质量问题预防体系，通过主动监测和风险评估，降低数据问题发生的风险，保障数据质量的稳定性和可靠性；
- c) 应定期对数据质量管理流程进行全面评审和评估，确保流程与业务需求和行业最佳实践同步，必要时进行策略的调整和更新；
- d) 应对数据质量控制措施的效果进行持续监控，通过定量分析和反馈循环，不断优化控制措施，提高数据质量管理的效率和效果。

8 数据质量管理方法

8.1 基本要求

包括但不限于：

- a) 应实施数据标准的统筹规划与统一管理，确保能够依据国家标准、行业标准及黑龙江省地方标准等不同来源制定相应的数据标准，涵盖标准从编制到审核再到发布的各个环节；
- b) 应确保数据标准的执行力，提供数据质量管理所需的标准要求，以及根据已录入的数据标准形成质量检测规则，确保标准得到实际应用；
- c) 应持续维护数据标准，包括对标准内容进行适时的变更、迭代完善和严格的版本控制，以适应不断变化的业务和技术需求；
- d) 应定期对数据标准的执行情况和实施效果进行监控，以便及时发现并解决执行过程中的问题，确保标准的有效性；
- e) 应根据监控和评估的反馈，对不再适用或存在缺陷的数据标准进行修订或重新制定，以保证数据标准的先进性和适用性；
- f) 应建立数据标准与业务流程的对应关系，确保数据标准能够与业务实践紧密结合，提高数据的业务相关性和实用性；
- g) 应加强数据标准的宣传教育和培训工作，提高全员对数据标准重要性的认识，确保数据标准的理解和执行达到预期效果。

8.2 数据质量合规准则

数据质量合规准则包括但不限于：

- a) 应允许企业根据自身业务需求和数据特性，自定义数据质量评估的维度，支持数据质量特性的多维度评估；
- b) 应依据数据标准规范，定义一系列基础数据质量规则，包括空值、值域、格式、长度、精度、唯一性以及主外键等方面的校验；
- c) 应根据既定的基础规则和特定平台的业务场景，支持用户自定义特定的业务规则，以适应多样化的数据质量需求；
- d) 应利用自定义 SQL 语句等技术手段，执行数据质量的检查和审核工作；
- e) 应建立数据质量反馈机制，确保可以收集和整合用户反馈，以持续优化数据质量评估标准；
- f) 应探索将大型语言模型（LLM）集成到数据质量合规准则中，利用其高级文本分析能力，辅助理解复杂的数据质量报告，提供自然语言生成的改进建议，从而提高数据合规性的评估效率和准确性。

8.3 数据质量合规审查任务

数据质量合规审查任务要求包括但不限于：

- a) 应支持创建质量检测任务，使用户能够依据数据质量管理规范和具体业务需求，定制并配置检测任务，确保数据质量评估严格遵循既定标准；
- b) 应支持设定检测任务的调度规则，使用户能够基于业务需求和数据变更周期，调整任务的运行时间、频率及触发条件，实现周期性或事件驱动的检测；
- c) 应支持检测任务的手动和自动执行，提供多样化的执行选项，包括即时触发、定时任务、连续监控、周期性执行及自定义调度，以满足多样化的检测需求；

- d) 应支持检测任务的运维监管，包括审查任务执行记录、调度统计等，以监控任务执行状况，及时发现并解决执行中的问题；
- e) 应提供问题数据的查询和导出工具，使用户能够依据检测结果迅速识别并提取合格率不达标的数据记录，便于深入分析和后续处理；
- f) 应允许用户查询和导出未达标数据，获取所有未满足平台数据质量标准的数据记录，以进行全面的数据质量评审和必要的改进；
- g) 应利用先进的数据分析技术，如大型语言模型（LLM），对检测任务结果进行深入分析，自动识别数据中的模式和异常，提供更精准的数据质量洞察；
- h) 应建立数据质量合规性评估的反馈循环，通过持续收集用户反馈和系统性能数据，不断优化审查任务的准确性和效率；

注1：问题数据特指那些在平台设置的数据库表中，其合格率未达到预设范围的数据记录。

注2：未达标数据指的是在数据库表中，所有未满足平台设定的数据质量特性合格率要求的数据记录。

9 数据质量监测与分析

9.1 质量监控策略

质量监控策略要求包括但不限于：

- a) 应对数据流向进行细致监控，确保能够清晰识别数据的来源和目的地；
- b) 应对数据检查中发现的异常情况及时触发结果告警机制，以便快速响应；
- c) 应对数据处理的每个环节进行严格跟踪，确保数据处理活动的透明性和可追溯性；
- d) 应对规则校验结果进行实时监控和深入分析，确保校验结果能够准确关联到具体的原始数据项；
- e) 应建立多维度的数据质量评估体系，从不同角度监控数据质量，以全面识别潜在问题；
- f) 应实施动态调整机制，根据监控结果和业务发展需求，灵活调整监控策略和校验规则，以适应不断变化的数据环境。

9.2 数据质量警报

数据质量警报要求包括但不限于：

- a) 应对识别出的问题数据或异常情况实施即时结果告警，确保相关人员能够迅速获得通知并采取相应措施；
- b) 应提供定制化的数据质量警报设置，允许用户根据特定需求和偏好定义告警条件、通知方式和紧急程度；
- c) 应确保数据质量警报系统的高可用性和实时性，以便在数据问题发生时，能够不间断地向相关人员发送准确及时的警报信息。

9.3 数据质量修复

数据质量修复要求包括但不限于：

- a) 应确立最佳数据修复方法，根据最佳数据修复方法对损伤数据进行修复；
- b) 应根据不同业务场景进行数据整改。

9.4 数据质量综合分析报告

数据质量综合分析报告要求包括但不限于：

- a) 应对校验结果进行详细记录，并生成全面的检测报告，报告内容应涵盖：
 - 1) 历史检测记录的数据波动情况，展示数据质量随时间的演变趋势，为趋势分析提供数据支持；
 - 2) 本次质量检测的执行时间、检测结果等基础信息，以及数据质量的动态变化情况，帮助识别改进或退化的领域。
 - 3) 每项业务规则的检测数据量、问题数据量、合格率等详细信息，提供量化的数据质量管理指标。
- b) 应支持在线查看每项业务规则问题数据的详细情况，以及打包下载对应的问题数据，便于进一步分析和处理；
- c) 应提供数据质量检测报告的查看和下载功能，以及数据质量监控报告的访问，确保用户能够方便地获取所需信息；
- d) 应具备对检测报告进行综合性分析的能力，包括但不限于：
 - 1) 通过多维度分析报告展示质量检测的总体结果，包括检测数据来源、时间、综合得分等基本信息，以及各质量维度下的得分和短板问题；
 - 2) 展示每项业务规则的详细检测内容，如检测数据总数、问题数据数目、合格率及问题数据分布情况；
 - 3) 通过评分报告详情展示每项规则的得分情况及比分权重，体现规则的重要程度，并对高权重规则进行重点分析。
- e) 应支持对检测报告进行统计分析，包括但不限于：
 - 1) 通过图表直观展示数据质量的概览信息，包括关键指标的趋势和分布；
 - 2) 按照数据源目录层级，即分层分域原则，进行统计分析，识别数据质量问题的模式和趋势。
- f) 应支持定制化报告功能，允许用户根据特定需求选择报告内容、格式和详细程度，以满足不同场景下的报告需求；
- g) 应利用先进的数据分析技术，如大型语言模型（LLM），对检测报告进行自然语言生成的总结和解释，提高报告的可读性和用户的理解度。

10 数据质量改进

10.1 改进措施

改进措施包括但不限于：

- a) 应根据质量分析结果，制定并提供针对问题数据的质量提升建议，涉及数据的修正、格式的统一标准化、以及数据源的优化改进措施；
- b) 应提供数据标准化、数据清洗、数据转换和数据整合等工具和功能，对不符合质量标准的数据显示进行有效处理，并纠正存在的数据质量问题；
- c) 应建立一个系统化的数据质量整改流程，包括整改任务的分配、执行、审核和反馈，确保整改措施得到有效实施；
- d) 应支持对整改过程进行监控和记录，包括整改效果的评估和记录，以便持续追踪数据质量改进的成效，并为未来的质量控制提供参考。

10.2 数据质量合规跟踪

数据质量跟踪要求包括但不限于：

- a) 应对数据质量进行持续跟踪，详细记录数据质量事件的评估过程、初步诊断结果以及采取的后续行动，确保能够验证数据质量改进措施的有效性；
- b) 应对系统中发现的问题进行及时修正，并对数据标准进行持续完善，以提高数据质量的整体水平；
- c) 应建立数据质量跟踪的反馈机制，收集数据使用过程中的反馈信息，为数据质量的持续改进提供依据。

参 考 文 献

- [1] GB/T 36073—2018 数据管理能力成熟度评估模型（2018年）
 - [2] GB/T 42021-2022 工业互联网 总体网络架构国家标准（2022年）
 - [3] 工业互联网创新发展行动计划（2021-2023年）
 - [4] 工业互联网标准体系（版本 3.0）（2021年）
 - [5] 工业互联网综合标准化体系建设指南（2021版）
 - [6] 工业互联网标识解析体系“贯通”行动计划（2024-2026年）
-