

KSSJ/ZL13-2023

智能化矿山数据融合共享 数据质量管理规范

Intelligent mine data fusion and sharing

Specifications for data quality management

国家矿山安全监察局
2023年6月

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
3.1. 数据 data	1
3.2. 数据质量 data quality	1
3.3. 数据质量度量指标 data quality measurement indicator	2
3.4. 数据质量规则 data quality rule	2
3.5. 异常数据 abnormal data	2
3.6. 设计质量 design quality	2
3.7. 执行质量 executive quality	2
3.8. 指标度量值 measurement value	2
4. 总体要求	2
4.1. 数据质量管理框架	2
4.2. 数据质量的基本要求	3
4.3. 数据质量管理开展原则	4
4.4. 数据生命周期质量管理原则	5
5. 数据质量管理活动	5
5.1. 数据质量稽核控制	5
5.2. 数据质量评估度量	6
5.3. 数据质量改进	8
6. 数据质量管理组织与职责	9
6.1. 数据管理组织	9
6.2. 数据管理责任人职责	9
6.3. 数据工程技术人员职责	10
附 录 A（资料性） 数据质量模板	11
A.1 数据质量控制规则类型表	11
A.2 设计质量打分模板	13
A.3 数据质量度量指标字典模板	14
A.4 执行数据质量五分衡量标准	15

前 言

本文件参照 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这些专利的责任。

本文件起草单位：陕西煤业化工集团有限责任公司、国家能源投资集团有限责任公司、中国中煤能源集团有限公司、华为技术有限公司、应急管理部信息研究院、中国华电集团有限公司、山东能源集团有限公司、晋能控股集团有限公司、陕煤集团神木张家峁矿业有限公司、陕西陕煤陕北矿业有限公司、国能数智科技开发（北京）有限公司、中煤信息技术（北京）有限公司、陕西陕煤陕北矿业有限公司信息技术运维分公司、西安工程大学、西安科技大学、中国煤炭科学研究总院矿山大数据研究院、新华三技术有限公司、国能神东煤炭集团有限责任公司、矿冶科技集团、北京北矿智能科技有限公司、西安重工装备制造集团有限公司、华电煤业集团数智技术有限公司、精英数智科技股份有限公司、西安重装智慧矿山工程技术有限公司、云鼎科技股份有限公司、陕西云智数维科技有限公司、浪潮数字企业技术有限公司、陕西陕煤黄陵矿业有限公司、西安合智宇信息科技有限公司、中安智讯（北京）信息科技有限公司、美林数据技术股份有限公司、腾讯云计算（北京）有限责任公司。

本文件技术指导：王世斌、屈永利、张忠温、张建安、丁震、王鹏、杨林、邓文革、王前、胡而已、宋文兵、张冬阳、郭军、付强、白明亮、谢旭阳、张碧英、路艳萍、李文俊、金卫多、逯宪彬、任跃武、赵晓勇、曹怀轩、崔小为、王金双。

本文件主要起草人：赵宇波、韩培强、张世磊、徐金陵、呼少平、赵文豪、高子元、牛虎明、郑耀涛、王陈书略、薛忠新、李艳、田丰、龚星宇、王艳艳、朱锐、米波、王曹睿、杨永生、李佩、彭涛、卢欣奇、张元生、赵亦辉、王海、黄金、陈帅领、朱晓宁、于谌彬、刘波、潘涛、李旭、董博、王锦、王川、李俊。

智能化矿山数据融合共享 数据质量管理规范

1. 范围

数据质量管理规范部分内容主要规定智能化矿山数据融合共享的过程中的数据质量规则和质量要求，为智能化矿山数据真实有效提供支撑和保障。

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；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 5271.18-2008 信息技术 词汇 第18部分：分布式数据处理

GB/T 7027-2002 信息分类和编码的基本原则与方法

GB/T 10113—2003 分类与编码通用术语

GB/T 15259—2008 煤矿安全术语

GB/T 18725-2008 制造业信息化 技术术语

GB/T 20529.1-2006 企业信息分类编码导则 第1部分：原则与方法

GB/T 20529.2-2010 企业信息分类编码导则 第2部分：分类编码体系

GB/T 32400-2015 信息技术 云计算 概览与词汇

GB/T 35295-2017 信息技术 词汇 第2部分：大数据及其应用领域术语

GB/T 36073-2018 数据管理能力成熟度评估模型

GB/T 37700-2019 信息技术 工业云 参考模型

3. 术语和定义

3.1. 数据 data

信息的可再解释的形式化表示，以适用于通信、解释或处理。

注：可通过人工或自动手段处理数据。

[GB/T 5271.1-2000，定义01.01.02]

3.2. 数据质量 data quality

数据质量指数据满足应用的可信程度，主要包括：完整性、及时性、准确性、

一致性、唯一性、有效性。

3.3. 数据质量度量指标 data quality measurement indicator

用于计算数据质量综合得分的指标, 含有数据质量度量指标定义、计算公式、用途等以及该指标对应的数据用户的满意度感知标准。

3.4. 数据质量规则 data quality rule

不符合业务实质、不满足数据应用要求的数据判断逻辑。

3.5. 异常数据 abnormal data

不满足数据标准, 不符合业务实质的数据。

3.6. 设计质量 design quality

数据架构落地情况, 含有数据资产目录、数据标准、数据模型、数据分布 4 个方面。

3.7. 执行质量 executive quality

数据执行结果质量, 根据数据质量度量指标的结果评估其质量情况。

3.8. 指标度量值 measurement value

根据数据质量度量指标取数规则及公式, 计算的结果。

4. 总体要求

4.1. 数据质量管理框架

4.1.1. 数据质量框架

数据质量管理框架图如图 1 所示:

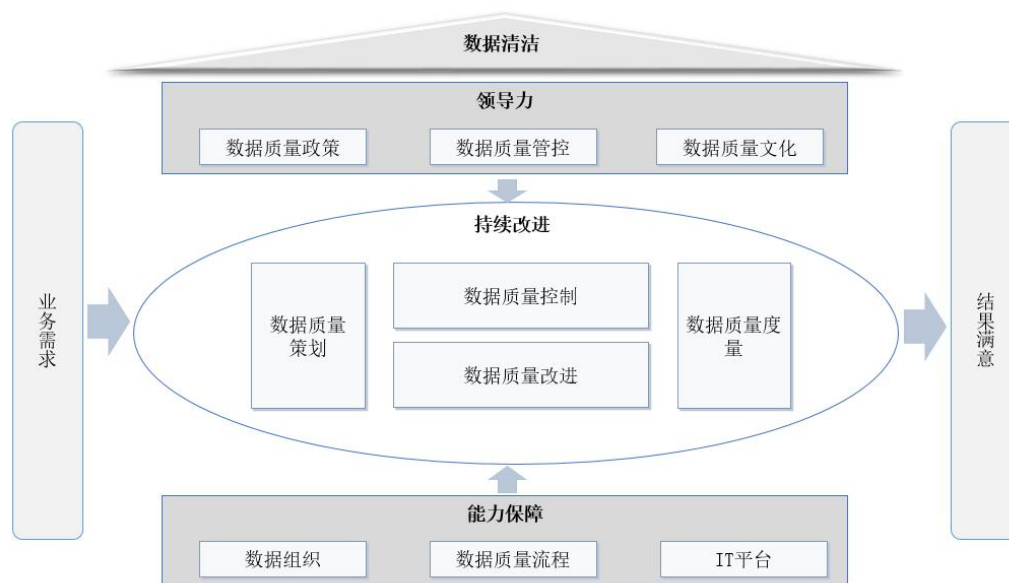


图1 数据质量管理框架

4.1.2. 数据质量领导力

数据质量政策应有不同的层次，数据质量的管控应兼顾宏观方面的指导原则以及微观层面的具体操作要求，引导正确的业务行为，提升企业成员的数据质量意识。

4.1.3. 数据质量持续改进

数据质量管理应是动态、持续的循环过程以满足不断变化的业务需求。

4.1.4. 数据质量能力保障

通过技术工具将数据质量自动融入日常业务，不断提升数据质量管理组织的管理水平、改善数据质量工具平台，使企业数据质量持续提高。

4.2. 数据质量的基本要求

智能化矿山数据质量包含以下六个方面的基本要求：

- a) **完整性：**数据在创建、传递过程中应无缺失和遗漏，包括实体完整、属性完整、记录完整和字段值完整四个方面。

示例：掘进进度统计中，全部统计有效进尺数据。

- b) **及时性：**应及时记录和传递相关数据，满足业务对信息获取的时间要求，数据交付要及时，采集要及时，应用要及时。

示例：例如某矿的产量、进尺数据及时上报。

- c) **准确性：**应真实、准确地记录原始数据，无虚假数据及信息，数据要准确反映其所建模的“真实世界”实体，数据记录信息与业务实际相符。

示例：掘进进度统计中，生产进尺数据与生产实际一致。

- d) **一致性：**应遵循统一的数据标准记录和传递数据信息，主要体现在数据记录应规范，数据应符合逻辑。

示例 1：班组上报的产量与皮带秤采集的产量数据一致。

示例 2：财务统计的年度采购总金额和各单位每月上报的采购金额汇总值一致。

- e) **唯一性：**同一数据应有唯一的标识符。各数据源系统在给其他系统提供数据时，应保证主键和联合主键具有唯一性，并且根据业务逻辑应具有唯一性的字段，在发布前应进行全量校验，确保唯一性的要求。

示例：某类物料在主数据系统已经有编码，物料使用单位不用申请新编码，直接使用主数据系统存在的编码，确保一物一码。

- f) **有效性：**数据的值、格式和展现形式应符合数据定义和业务定义的要求。各数据源单位在提供数据时，应在发布前对其字段进行分析，保证字段值具有有效格式，并且其值在有效范围之内和符合逻辑要求。

示例：冲击地压计量单位为 MPa，某煤矿测量数据按照 MPa 单位计算。

4.3. 数据质量管理开展原则

智能化矿山数据质量管理工作应遵循以下原则开展：

- a) **过程受控：**数据质量的管理工作应对数据的全生命周期过程进行管控，包括数据的产生、存储、传递、使用和失效销毁等，应对每个阶段进行全面的過程质量管理；
- b) **标准管理：**数据质量管理应在矿山企业统一的数据质量管理框架内进行，统一制定和发布数据质量标准、数据质量管理策略及管理规范；
- c) **多方参与：**数据质量管理是系统性的工作，需要多方的共同参与。各级业务部门和信息化部门应参与到数据质量管理的工作中，提升数据质量；
- d) **逐步推进、持续改善：**矿山企业应对已发现的数据质量问题逐步改进，应将数据质量管理列为长期工作持续改善，建立常态化的工作支撑。

4.4. 数据生命周期质量管理原则

数据生成、维护、应用是矿山企业数据生命周期管理的关键活动，应遵循以下要求：

- a) 信息系统建设应考虑数据质量要求，将数据质量检查加入系统的日常运维环节；
- b) 应基于数据使用要求来制定数据质量标准，并且征得关键使用部门的同意；
- c) 数据生成应确保录入正确，关键数据应进行复核或审批。录入、复核和审批人员应掌握数据质量要求；
- d) 对影响关键经营指标真实性事件应采取零容忍态度；
- e) 数据生成部门和单位应保证数据的真实、完整并及时传递到数据使用单位。数据使用单位为核实数据质量问题，可调阅所需的数据生成单位的数据；
- f) 数据质量应持续度量。矿山企业应解决长期影响安全生产和经营管理的数据问题。

5. 数据质量管理活动

智能化矿山数据质量建设工作数据质量稽核控制、数据质量评估度量及数据质量持续改进三个部分组成。

5.1. 数据质量稽核控制

5.1.1. 识别控制对象

基于重要性和成本效应原则识别出需要进行质量控制的数据，分阶段有序的将数据质量控制规则落地。

- a) 基于重要性原则，数据质量稽核控制对象主要有：
 - 1) 关键主数据和基础数据：如产品、客户、供应商、组织、人员、物料等；
 - 2) 矿山企业管理重点关注的数据和管理人员分析决策的数据；
 - 3) 重点业务高频次、广泛应用的数据，如安全生产、生产辅助、安全管理、生产技术等涉及的数据。

- b) 基于成本效应原则数据质量稽核控制对象主要有：
 - 1) 运作成熟且质量较高的数据，或度量成本很高但预期的改进很少的数据，可不优先考虑；
 - 2) 应通过收集业务需求、数据质量问题等其他途径从中筛选当前需监控的数据。

5.1.2. 确定控制规则

依据数据在数据库落地时的质量特性及数据质量规则类型，应遵从如下四类数据质量控制规则：

- a) 单列数据质量规则，关注数据属性值的有无以及是否符合自身规范的逻辑判断；
- b) 跨列数据质量规则，关注数据属性间关联关系的逻辑判断；
- c) 跨行数据质量规则，关注数据记录之间关联关系的逻辑判断；
- d) 跨表数据质量规则，关注数据集关联关系的逻辑判断。

在四类数据质量分类框架下，围绕数据质量六大特性共设计了 15 类数据质量规则，参考《附录 A.1 数据质量控制规则类型表》。

5.1.3. 促进规则落实

矿山企业应促进数据质量控制规则，提升数据质量：

- a) 部署数据质量规则：通过数据质量管理平台将规则进行部署和有效落实，实现日常质量监控；
- b) 异常数据获取和分析：应定期收集异常数据，采用质量方法和可视化工具对异常数据结构进行分析，包括柏拉图、控制图、鱼骨图、散布图及直方图等；
- c) 处理异常数据：应通过工具和平台监测出来的异常数据，进行异常数据纠正；
- d) 拟制数据质量控制报告：应全面立体的反映当前数据质量情况，做好数据质量监控。

5.2. 数据质量评估度量

5.2.1. 数据质量度量的角度

数据质量评估度量旨在规范矿山各级单位的过程，确保度量结果可信。数据

质量度量应从“设计质量”和“执行质量”两个方面开展，评估数据质量综合水平，揭示数据质量问题。

5.2.1.1 设计质量度量

为确保设计质量标准稳定，从数据架构的四个角度进行综合评估。

a) 数据资产目录

- 1) 业务对象应有明确唯一的数据管理责任人，并对该业务对象全流程端到端质量负责；
- 2) 业务对象的元数据质量，如数据分类是否完整、业务定义是否准确等；
- 3) 资产目录完整性。

b) 数据标准

- 1) 数据标准元数据质量，如数据标准是否唯一、业务用途及定义是否准确等；
- 2) 所有业务对象应准确关联数据标准；
- 3) 数据标准在IT系统及其对应的业务流程中应得到应用和遵从。

c) 数据模型

- 1) 开发概念模型和逻辑模型，并通过矿山企业数据管理组织评审；
- 2) 物理数据模型设计应遵从逻辑数据模型设计，数据库中物理表的落地应遵循物理模型。

d) 数据分布

- 1) 已认证数据源，并通过矿山企业数据管理组织评审；
- 2) 交易侧应具备完整的信息链和数据流，并通过矿山企业数据管理组织评审；
- 3) 交易侧业务资产、数据湖、主题联接、数据服务、自助分析之间应具备完整准确的血缘关系。

e) 设计质量打分模型

质量打分模型可参考《附录 A.2 质量打分模型》，结合矿山企业实际情况进行设计。

5.2.1.2 设计质量度量

执行质量度量宜考虑完整性、及时性、准确性、一致性、唯一性、有效性 6

个维度，从客户关注重要性、法律财务风险性、业务流程战略性三要素评估数据内容的清洁度，业务领域可根据阶段性的管理重点和诉求调整评估的要素。

5.2.2. 数据质量度量过程管理

a) 制定数据质量目标

矿山企业依据当前存在的数据质量问题制定数据质量目标。

b) 确定数据质量度量范围

矿山企业审视数据标准，根据数据质量目标确定数据质量度量范围。

c) 设计数据质量度量指标

矿山企业根据数据质量控制异常数据情况、业务运作及数据使用中存在的问题，从 IT、流程多方面视角分析，设计数据质量度量指标。

数据质量度量指标模板可参考《附录 A.3 数据质量度量指标字典模板》，结合矿山企业实际情况进行设计。

d) 确定执行数据质量衡量标准

矿山企业根据统一的 5 分制，对新建数据质量度量指标确定“差、中、良、优、满分”的达标标准，确定相应的档次区间水平。

数据质量五分衡量标准详见附录 A.4。

e) 实施数据质量度量

矿山企业应针对指标度量值，根据 5 分制对应关系进行各指标度量打分。

f) 发布数据质量度量报告

矿山企业根据质量度量结果，拟制年度数据质量度量报告。

5.3. 数据质量改进

5.3.1. 数据质量改进专项工作

矿山企业数据质量改进专项工作包括：

- a) 信息系统改进：确认有信息系统改进的需求，应实施信息系统改进；
- b) 专项工作落实：应确保数据质量问题解决方案得到落实。

5.3.2. 数据质量改进日常工作

矿山企业数据质量改进日常工作包括：

- a) 制定改进措施：矿山企业应根据数据质量报告，组织数据工程技术人员、业务负责人、信息中心人员等共同制定数据质量改进措施，改进措施应明确到具体改进责任人、改进目标及完成时间；
- b) 跟踪改进情况：矿山企业应就数据质量改进情况，按照年度发布数据质量问题改进和闭环情况；
- c) 刷新目标基线：根据本年度度量情况，结合下一年度业务管理改进诉求，制定新的年度度量目标，刷新目标基线。

6. 数据质量管理组织与职责

6.1. 数据管理组织

智能化矿山数据质量的责任主体应由矿山企业信息化部门承担，任命各级单位的数据管理责任人和数据工程技术人员。

- a) 矿山企业信息化管理部门：应贯彻落实矿山企业制定的数据质量管理规章制度，制定信息系统数据质量管理要求，并适时配合数据工程技术人员开展检查监督、数据清洗、数据质量考核等，保障数据质量管理符合规范性。根据信息系统管理部门和信息系统使用部门的数据质量管理需求，负责数据质量管理在信息化建设方面的支撑和保障；
- b) 矿山企业信息系统使用部门：应严格执行数据质量管理规章制度，确保数据采集质量，加强对采集、录入人员的业务培训和管理，提高数据录入的准确率。按照规定进行数据质量考核、督查等工作；
- c) 矿山企业信息系统数据录入部门：应根据业务规则和数据标准填写、上报数据。按照原始记录，准确、完整、及时地将数据录入系统。确保符合企业数据质量要求。针对数据质量问题，分析问题根本原因，制定改进举措并实施。

6.2. 数据管理责任人职责

矿山企业数据管理责任人应负责下达数据质量目标、签发数据质量度量报告，对企业数据质量结果奖励或问责。

矿山企业数据管理责任人应依据要求设定的企业数据质量目标，并对企业数据质量负责。明确数据质量问题改进责任人，推动问题闭环。覆盖数据从生成到

销毁的全生命周期，实现数据质量持续提升。

6.3. 数据工程技术人员职责

数据质量日常工作由矿山企业数据工程技术人员承担，应对业务所需的数据，进行全面质量管理。

具体职责如下：

- a) 分解落实数据质量目标，制定数据质量标准；
- b) 建立数据质量控制，负责分工及职责划分，指导数据质量工作开展；
- c) 设计质量指标，执行数据质量度量，编制数据质量年度度量报告；
- d) 推动跨单位数据质量问题解决，验收数据质量问题闭环。

附录 A

(资料性)

数据质量模板

A.1 数据质量控制规则类型表

表A.1 数据质量控制规则类型表

序号	规则类型	业务对象	质量特性	类型描述	示例
1	不可为空类	单列	完整性	属性不允许或在满足某种条件下不允许出现空值	如工作面名称不可为空
2	语法约束类		有效性	属性值满足数据语法规范取值约束	如身份证号满足国家标准
3	格式规范类		有效性	属性值必须满足展现格式约束	如日期有多重格式,对于同一属性指定同一类格式
4	长度约束类		有效性	属性值须满足约定的长度范围	如密码的长度至少要8位以上,不超过16位
5	值域约束类		有效性	属性值必须存在于已定义的枚举值列的约束	煤质类别必须在煤质分类列表中存在
6	事实参照标准类		准确性	存在事实数据或者事实参考标准数据,与该事实或事实参照标准对比一致的约束	材料申购所使用的物料编码需和主数据平台物料编码相对应
7	应为空值类	跨列	完整性	属性满足某种条件下不能维护值	如矿山企业要求销售数据在某些系统不展示
8	单表等值一致约束类		一致性	某一属性值与本实体其他属性计算值相等的约束	如集团企业已汇总产量与各矿业公司上报详细产量的再汇总保持一致
9	单表逻辑一致约束类		一致性	某一属性值与本实体其他属性满足逻辑关系约束(大于或小于)	如项目的关闭日期不能早于立项日期
10	入库及时类		及时性	数据进入系统的及时性约束,通常要包括数据原材料获取时间和入库时间才能进行规则设计	如通过内部大市场系统库存信息应及时根据出入库信息进行更新
11	外关联约束类	跨表	一致性	引用其他业务对象属性时,所维护的属性值必	如合同的签约客户必须为客户主数据

序号	规则类型	业务对象	质量特性	类型描述	示例
				须在其他业务对象中存在的约束	中定义的法人客户
12	跨表等值一致约束类		一致性	某一属性值与其他实体的一个或多个属性值的函数计算结果相等的约束	如某煤矿当日产量与调度部门皮带秤汇总一致
13	跨表逻辑一致约束类		一致性	某一属性值满足其他实体的一个或多个属性值的函数关系的约束（大于或小于）	当日煤矿生产用电量小于变电所计量的用电量
14	记录唯一类	跨行	唯一性	记录不重复，存在可识别的业务主键进行唯一性判断，是对数据集内部是否存在相似或重复记录的约束规则	如某矿业上报产量只能有唯一一行记录
15	层级结构一致约束类		一致性	存在层级结构的属性，同层级属性结构一致	如物资类型分为大类，中类，小类结构，物资管理时需按照此规则多层次进行选择

A.2 设计质量打分模板

表A.2 设计质量打分模板

对象得分	得分要素	打分标准	分值分配	评估对象得分
1	数据资产目录	数据对象没有明确的管理责任人,无法对该数据对象全流程端到端质量负责	1	
2	数据资产目录	数据对象有明确管理责任人,对该数据对象全流程端到端质量负责	0.25	
	数据标准	开发完整的数据字典,并通过矿山企业数据管理组织评审	0.25	
	数据模型	及时上传评审后的概念模型文件至相关 IT 系统	0.25	
	数据分布	数据源认证通过矿山企业数据管理组织评审	0.25	
3	数据标准	数据标准通过矿山企业数据管理组织签发	0.25	
	数据标准	数据标准在 IT 系统及其对应的业务流程中得到应用和遵从	0.25	
	数据模型	及时上传评审后的逻辑模型文件至相关 IT 系统	0.25	
	数据分布	完整的数据血缘关系	0.25	
4	数据模型	物理表/物理模型中包含逻辑实体中的所有属性	0.5	
	数据分布	完成数据血缘信息链治理	0.5	
5	领域	管理面向未来 3-5 年的数据架构蓝图	0.3	
	领域	全面实现满足业务主体数据化、业务场景数据化、业务规则数据化、业务决定算法化、业务流程自动化的要求	0.4	
	领域	数据资产成为矿山企业战略核心竞争力	0.3	

A.3 数据质量度量指标字典模板

表A.3 数据质量度量指标字典模板

指标名称	指标字典说明	举例
设置目的	指标度量值体现矿山企业经营管理、安全生产某方面预期要达到的效果，或是体现业务某方面的情况	
指标定义	指标来源于哪个数据质量标准的表达	
计算公式	计算指标值的公式，如：百分比度量，分子定义、分母定义	
数据来源	取数的可信数据源所在的系统	
计量单位	最能反映质量情况，结果值的单位，常用的有比率“%”、金额“美元”、数量“笔”等	
统计纬度	统计角度，反映不同维度的质量状况	
测量对象	指标管理所属层级	
统计部门	指标统计部门	
使用部门	指标使用部门	
统计周期	多久度量一次、月度、季度、半年等	
备注说明	其他备注。	
指标解释部门	指标解释部门	

A.4 执行数据质量五分衡量标准

表 A.4 执行数据质量五分衡量标准

评估得分	用户感知	得分说明	主数据	事务数据
1分	差	指标未度量，或度量结果远不能达到数据质量要求，存在严重的数据质量问题	小于三西格玛 (93.32%)	事务数据指标按照主业务流拉通原则： 1. 按业务流：客户交易流、产品配置流、交付作业流、集成计划流等 2. 按六性：结合六性情况对指标相对拉通
2分	中	度量结果不能达到数据质量要求，存在较多的数据质量问题、影响较大	三西格玛(93.32%) -四西格玛 (99.3797%)	
3分	良	度量结果基本达到数据质量要求，存在少量的数据质量问题、影响一般	四西格玛 (99.379%) -五西格玛 (99.977%)	
4分	优	度量结果完全达到数据质量要求，且基本不存在数据质量问题，达到业界领先水平	五西格玛 (99.977%) -七西格玛(100%)	
5分	满分	零缺陷，所有属性都没有任何异常，且长期不存在任何数据质量问题	零缺陷	