

# 数据质量管理实践指南

## (1.0版)

CCSA TC601 大数据技术标准推进委员会

2025年12月



## 版权声明

---

本报告权属于CCSA TC601 大数据技术标准推进委员会，并受法律保护。转载、摘编或利用其他方式使用本报告文字或者观点的，应注明“来源：CCSA TC601 大数据技术标准推进委员会”。违反上述声明者，将追究其相关法律责任。

# 编制说明

本指南的编写得到了各数据治理领域多家企业与专家的支持和帮助，主要参与单位与人员如下：

## 参编单位：

大数据技术标准推进委员会、联通数据智能有限公司、科大讯飞股份有限公司、中兴通讯股份有限公司、中国移动通信集团北京有限公司、中国联合网络通信集团有限公司、国家管网集团北京智网数科技术有限公司、国家电网有限公司大数据中心、中信证券股份有限公司、中信建投证券股份有限公司、振华智造（西安）科技有限公司、御数坊（北京）科技有限公司、东部机场集团有限公司、四川省大数据中心、海南数造科技有限公司、郑州埃文科技有限公司、安阳数智科技集团有限公司、中国移动通信集团湖北有限公司、中国移动通信集团河南有限公司、河海大学、中移（苏州）软件技术有限公司、中国南方电网有限责任公司、中国核动力研究设计院、北京神州绿盟科技有限公司、云赛智联股份有限公司、人保信息科技有限公司、浙商证券股份有限公司、南京银行股份有限公司、中国平安财产保险股份有限公司、中国移动通信集团青海有限公司

## 参编人员

姜春宇、王妙琼、周京晶、阚鑫禹、张娇婷、周圣文、田明慧、林木森、刘思达、张一鸣、高海暘、王宇龙、李铁峰、王庆、聂兆博、王刚、孙辉、孟宇、吴德亮、程竞超、王世峰、梁雅丽、刘天元、杨瑞、王衍祺、刘佳林、张志鹏、黄岩、武丽莎、任英杰、王桂强、陈辉华、邢默、李可、孟祥凯、李晓燕、王会、刘晨、李宗锴、刘伟平、齐翌、郭涛、王瀚、鲍立飞、吴炜、赵旭、武可嘉、郑丹辉、肖翔、孟晓莉、刘雪婷、叶紫光、王重任、张国宝、罗辉、李勇、徐键、徐欢、张新阳、马小琴、徐敏、郑芒英、陈开杰、张倬、曾未、刘佳、王新洋、禹芳、查毓灵、杨琦、林晖晖、冯四凤、吴蓉蓉、许璐、王宪

# 前言

在数字经济时代，数据已经成为与土地、资本、劳动力同等重要的新型生产要素。国家在顶层设计层面，陆续出台了《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》《数字中国建设整体布局规划》等政策文件，明确提出要提升数据要素价值，推动数据治理体系建设。

在金融、制造、政务、能源等行业对数据质量也提出了更高要求。**金融行业**，中国人民银行联合相关部门印发《银行业保险业绿色金融高质量发展实施方案》《推动数字金融高质量发展行动方案》等专项文件。**制造业**，工业和信息化部出台《关于做好 2025 年工业和信息化质量工作的通知》《制造业质量管理数字化实施指南（试行）》等指导文件。**政务领域**，国务院及相关部门印发《国务院关于加强数字政府建设的指导意见》、出台《政务数据共享条例》等法规政策。**能源行业**，国家能源局等部门发布《关于推进能源装备高质量发展的指导意见》《关于加强电力可靠性数据治理 深化可靠性数据应用发展的通知》等配套政策。

随着数据规模的扩张、AI与大模型新数据的发展、业务节奏加快和合规要求严格的背景下，数据质量管理不仅是企业合规经营和风险控制的基础，更是业务创新的关键驱动力，成为**企业数智化征途中的“生命线”**。

本指南对数据质量管理的定义、方法论、保障措施和实施要点进行了阐述，总结了数据质量管理的价值与误区，分析了当下数据质量管理的困境，并展望了未来数据质量管理的发展趋势。

# 目 录

## 一、 数据质量管理概述

(一) 什么是数据质量 .....	01
(二) 什么是数据质量管理 .....	03
(三) 数据质量管理的意义 .....	07

## 二、 数据质量管理主要内容

(一) 数据质量管理怎么管 .....	10
(二) 数据质量管理的保障措施 .....	15

## 三、 数据质量管理实施要点

(一) 领导支持 .....	19
(二) 认责机制 .....	19
(三) 评价机制 .....	19
(四) 协作机制 .....	20

## 四、 数据质量管理挑战与展望

(一) 主要挑战 .....	21
(二) 未来展望 .....	22

# 一、数据质量管理概述

## （一）什么是数据质量

### 1.数据质量的定义

在数字经济时代，数据已成为与土地、资本、劳动力同等重要的生产要素。然而，并非所有数据都具备使用价值。**数据能否有效支撑业务和决策，其质量就是核心考量因素之一。**

发布年份	权威来源	定义
2009年 (第一版)	DAMA 国际（数据管理协会） 《DAMA 数据管理知识体系指南(DMBOK2)》	数据符合用户需求的程度。
2015年	ISO/IEC 8000-1:2015 数据质量标准	定义数据质量为“数据满足规定或隐含需求的能力”，并强调“业务场景适配性”是衡量数据质量的核心前提。
2020年 (第一版)	ISO 8000-2:2022 《数据质量 第 2 部分：数据质量术语》	Data quality: degree to which a set of inherent characteristics of data fulfill requirements. 一组固有特性的数据满足需求的程度。
2024年	中国电子行业联合会 《数据管理专业术语》	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。

综合以上内容总结，将数据质量定义为：**数据符合目标需求的程度。**

**数据质量并非一个绝对的、统一的、抽象的标准，而是与数据的具体应用场景紧密关联**——即同一组数据在某一业务场景中可能被判定为高质量，而在另一场景中却可能因不符合需求而被视为低质量。

**数据质量并非绝对概念，而是相对的，取决于使用者对数据的要求。**例如，在金融行业，监管报送数据需要高度准确和一致；而在互联网应用场景下，实时性和完整性可能更加重要，这说明数据质量具有场景依赖性和多维度属性

## 2.数据质量的特点

数据质量具有多维度性、动态性、关联性和相对性等显著特点，这些特点决定了数据质量管理的复杂性和持续性。

**多维度性**是数据质量最突出的特点之一。如前文所述，数据质量并非由单一指标衡量，而是涵盖准确性、完整性、一致性、及时性、唯一性、有效性等多个维度。每个维度从不同角度反映数据的质量状况，且不同维度之间可能存在相互影响——例如，为了追求数据的及时性，可能在一定程度上影响数据的准确性；而确保数据的一致性，往往需要对多个数据源进行协同管理。

**动态性**体现为数据质量会随着时间和业务变化而发生改变。一方面，数据本身处于不断更新的状态，新数据的接入、旧数据的变更都可能导致数据质量波动；另一方面，业务需求的演变会使数据质量的衡量标准发生调整——过去符合要求的数据，可能因业务规则的升级而不再满足新的使用需求。例如，随着客户信息采集维度的增加，早期仅包含基本联系方式的客户数据，在需要进行精准营销的场景下，就会因完整性不足而降低质量等级。

**关联性**意味着数据质量并非孤立存在，不同数据集之间的关联关系会影响整体质量。在一个复杂的信息系统中，数据往往通过各种逻辑关系相互连接，某一环节的数据质量问题可能会传导至其他环节，引发连锁反应。例如，在供应链管理系统中，供应商信息的错误可能导致采购订单的生成出现偏差，进而影响库存管理和生产计划，最终对整个供应链的运转效率产生负面影响。

**相对性**则强调数据质量的评价标准是相对的，而非绝对的。如定义中所提及的，数据质量的高低取决于其是否满足特定场景的使用需求。对于一些非核心业务场景，可能允许数据存在一定程度的不完美；而对于关键业务决策或高风险场景，则需要极高的数据质量标准。这种相对性要求在数据质量管理过程中，需根据实际业务需求和优先级，制定差异化的质量目标和控制策略，以实现资源的优化配置。

## （二）什么是数据质量管理

### 1.数据质量管理的定义

权威来源	定义
DAMA 国际（数据管理协会） 《DAMA 数据管理知识体系指南(DMBOK2)》	为确保满足消费者的需求，应用数据管理技术进行规划、实施与控制等管理活动
ISO/IEC 8000-1:2015 (数据质量标准)	数据质量管理是一种关于数据质量评估、监控和控制的系统化方法和实践。该标准定义了数据质量的关键要素，包括数据准确性、完整性、时效性、可靠性、有用性和可维护性，还提供了一种数据质量评估框架，以帮助组织评估和改进其数据质量。
ISO 8000-2:2022 《数据质量第 2 部分：数据质量术语》	Coordinated activities to direct and control an organization with regard to data quality. 围绕数据质量对组织进行指导与管控的协调活动。
ISO 8000-2:2022 《数据质量第 2 部分：数据质量术语》	数据质量管理（DQM）是一系列用于提高和维护组织数据质量的实践。有效的数据质量管理有助于企业避免低质量数据，这些低质量数据可能导致业务流程中代价高昂的错误和效率低下。通过数据质量管理，企业可以获得可信、可靠的数据，从而解锁有价值的见解，实现更好的决策，并将人工智能（AI）集成到业务运营中。数据质量管理包括数据剖析、数据清理、数据验证、数据质量监控和元数据管理等实践
中国电子行业联合会 《数据管理专业术语》	组织对数据生命周期的每个阶段里可能出现的各类数据质量问题，进行识别、监控、分析、改进等一系列管理活动，并通过改善和提高管理水平使得数据质量获得进一步提高。

综合以上所述，**数据质量管理**是在数据全生命周期（采集、存储、处理、应用、销毁）中，通过制定标准、部署技术工具、建立组织流程，持续监控、评估、改善数据质量，确保数据符合业务需求与合规要求，最终支撑决策有效性与数据价值挖掘的系统化管理活动。

更关键的是，鉴于数据质量管理的体系性强、数据质量特点多，如果原始的数据资源不能转化为相对稳定且具体的数据产品或数据应用，无法形成某个特定状态，则无法评价数据质量水平。

## 2.数据质量管理的方法论

在企业数据管理实践的推动下，数据质量管理已经形成了多种不同的框架理论，为企业数据质量的规划、建设和管理工作提供了理论和实践指导。下文中列举了国内外典型的数据质量理论框架及方法论，这些内容都是在企业实践中广为传播和使用的，是当下业界开展数据质量相关工作的理论基础。

### 1) 通用质量管理方法论

#### · PDCA循环

PDCA循环是美国质量管理专家Walter A. Shewhart首先提出的，由戴明20世纪30年代提出并推广、宣传，获得普及，所以又称戴明环。PDCA循环是一种系统性管理方法，旨在通过持续改进流程和产品实现目标优化。核心思想是通过计划（Plan）、执行（Do）、检查（Check）、行动（Act）四阶段闭环循环，以数据为依据，全员参与，实现流程或产品的持续优化。

#### · 六西格玛（Six Sigma）

六西格玛（Six Sigma）方法论起源于20世纪80年代美国摩托罗拉公司的质量管控方法，后经通用电气（GE）等企业推广而广泛应用。六西格玛方法论核心是通过“定义—测量—分析—改进—控制（DMAIC）”流程，旨在通过数据驱动的闭环流程，减少过程变异、降低缺陷，将数据质量缺陷率降至极低水平。六西格玛的得名来自“六倍标准偏差（ $6\sigma$ ）”，就是每百万故障率（parts per million，简称PPM）少于3.4个。

## 2) 数据质量管理的方法论

### · 全面数据质量管理计划TDQM

全面数据质量管理计划TDQM（Total Data Quality Management Program）的概念由美国麻省理工学院Richard Y. Wang和Diane M. Strong等学者在20世纪90年代提出。TDQM将数据视为一种特殊产品，把管理产品质量的方法应用于数据质量管理，从根本上改变了“数据是系统的附属品”的传统概念，强调数据质量管理的关键是提高人的素质，充分调动员工积极性，基本思想是将以事后检查把关为主转换为以预防和改进为主，把用户对数据使用的满意度作为评价质量高低的标准。

### · ISO 8000

ISO 8000是国际标准化组织（ISO）为数据质量的评估、管理和改进提供国际通用框架，确保数据在交换、共享和使用中的准确性、一致性和可靠性。适用于所有需要管理关键数据的组织，规范和管理数据质量活动、数据质量原则、数据质量术语、数据质量特征（标准）和数据质量测试等内容。ISO 8000数据质量标准由通用原则、主数据质量、交易数据质量、产品数据质量四个部分组成。

### · DMBOK

DMBOK是国际数据管理协会（DAMA）对数十年数据管理领域知识和实践的总结，是当前业界开展数据管理工作的重要参考理论，也是数据治理从业人员的学习教材。指南从数据治理、数据架构、数据质量、数据安全、数据管理、数据建模设计、大数据等方面介绍了数据管理的知识体系。DMBOK将数据质量视为数据管理的核心支柱，强调其是实现数据价值的前提条件，同时阐述了如何通过结构化方法确保数据满足业务需求，并提供了从评估到改进的全生命周期管理框架。

### · 《数据管理能力成熟度评估模型》

GB/T-36073《数据管理能力成熟度评估模型》（简称“DCMM”）是我国首个数据管理领域正式发布的国家标准。在DCMM中数据质量是数据治理的重要目标，需要在数据标准、元数据等功能的协同支持下，实现对组织数据资产全生命周期的管理，用以提升数据质量，展现数据治理的价值。DCMM的数据质量能力域包括数据质量需求、数据质量检查、数据质量分析和数据质量提升。随着DCMM贯标评估工作在全国范围内的持续、深入开展，DCMM中对数据质量管理的能力要求得到了企业的广泛重视和采纳，并以DCMM框架作为企业自身开展数据质量管理的重要指导理论。

### 3.我国数据质量专项标准

我国已发布的数据质量专项标准规范涵盖了国家标准、行业标准以及团体标准等多个层面：

类型	标准名称	标准内容
国家标准	GB/T42381《数据质量》系列标准	本系列标准包括GB/T 42381.62-2023《数据质量 第62部分：数据质量管理：组织过程成熟度评估：过程评估相关标准的应用》、GB/T 42381.63-2023《数据质量 第63部分：数据质量管理：过程测量》等5项标准定义了数据质量的相关特征，并提供了管理、测量和提高数据质量的方法
	GB/T 31594-2015《社会保险核心业务数据质量规范》	对社会保险核心业务数据的指标、数据质量两方面做出了规范性要求
	GB/T 36344-2018《信息技术数据质量评价指标》	规定了数据质量评价的相关指标，包括规范性、完整性、准确性、一致性、时效性和可访问性等指标
	GB/T 39400-2020《工业数据质量 通用技术规范》	对工业数据质量持续改进的模型、质量的描述、识别、评价、控制和报告等的标准化，提高工业数据的质量
	GB/T 41149-2021《基础地理信息数据质量要求与评定》	规定了基础地理信息数据的质量要求和评定方法，适用于基础地理信息数据的生产、监理、验收和质量评定等
	GB/T 43709-2024《资产管理信息化 数据质量管理要求》	规定了资产管理信息化数据质量管理的相关要求
行业标准	YD/T 4522-2023《面向机器学习的电信数据规范 数据质量》	规定了对电信数据进行质量管理的通用原则、一般流程和评估方法
团体标准	T/CCSA 502-2024《大数据数据质量管理平台技术要求》	构建了通用性数据质量管理平台技术能力要求，明确数据质量管理平台的主要功能点和关键技术要求
	T/CESA 1418-2025《信息技术公共数据质量评价》	构建了公共数据质量评价指标体系，并规定了评价工作流程和评价方法

注：本表所列仅收录部分专项标准，后续将根据标准更新情况持续补充。

### （三）数据质量管理的意义

高质量的数据是数字经济和智能社会的核心基础，质量管理直接关系到运营效率、决策科学性和风险防控，在数据要素市场中，数据质量是资产化和价值转化的前提。组织在推进数据质量管理过程中，应避免陷入“只看源头、忽视全局”“只看技术、忽视业务”的误区，建立以业务需求为导向、以制度和技术为保障、覆盖全生命周期的质量管理体系。

#### 1.数据质量管理的价值

**数据质量管理的价值贯穿组织运营与发展。**

**提升数据价值与业务效能。**数据质量管理是释放数据价值、驱动业务效能的核心抓手。在数字化浪潮中，数据已成为企业核心战略资产，高质量数据才能有效支撑决策与业务创新。通过系统化的数据治理实践，企业可持续提升数据质量，持续优化数据资产利用效率，降低数据应用风险，从而强化企业在数字化转型中的市场洞察力与核心竞争力，实现数据价值向业务成果的高效转化。

**保障分析结果与决策可靠。**通过建立系统化的保障机制，确保数据质量核心维度满足分析需求，从而为可靠决策奠定基础。它贯穿数据全生命周期，从采集、存储到处理、应用各环节实施严格管控，防止低质量数据导致分析偏差。尤其在大模型应用中，高质量的数据能避免特征提取偏差、模型过拟合等问题，保障预测结果的真实性，最终提升决策的科学性和有效性。

**平衡共享效能与风险管控，保障合规与风险管理。**在数据共享过程中，通过流程整合与可视化呈现质量评估、检核、整改及报告环节，能够显著提升数据流通效率与业务价值。同时，嵌入标准化风险管控机制，依托实时监控识别数据缺陷，防范决策风险与安全隐患，确保共享行为符合合规要求。这种平衡既强化了数据作为核心资产的战略驱动作用，又为数字化转型奠定了坚实的质量根基。

#### 2.数据质量管理的理解误区和过程冲突

##### 1) 数据质量管理的误区

**仅关注源头数据的技术指标，忽视业务价值导向。**企业常将数据质量管理简化为对准确性、有效性、一致性等指标的源头数据校验，如清洗重复值、修正格式错误等。然而，这种技术导向的局限在于，不仅割裂业务场景，数据质量标准脱离业务实际需求（如特定字段允许

空值的业务逻辑）导致无效质量告警，同时未将数据质量与业务目标深度关联，忽视业务部门在数据质量管理中的核心作用，仅依赖技术部门通过工具处理数据，导致数据处理结果与业务需求脱节，无法解决数据在业务场景中的实际应用问题即使技术达标的的数据也难以有效支撑决策，最终削弱数据资产的实际价值。

**缺乏业务需求视角的全流程管理。**当前数据质量管理存在将责任归属于技术部门的认知误区，缺乏业务需求驱动的全流程管控视角，未构建跨业务、技术、管理的协同治理机制，导致质量规则与业务流程严重脱节，数据问题在跨系统传递过程中持续累积放大，形成管理盲区与责任真空。

**将数据质量管理工作视为“一次性项目”或者“短期修正项目”。**将数据质量管理等同于短期数据清洗、修正项目，未建立持续管控机制，忽视数据随业务发展、系统迭代产生的动态变化。数据质量问题并非只出现在采集环节，而是贯穿于全生命周期。从采集、传输、存储，到加工、共享、应用，任何环节都可能出现错误。如果只在单一环节进行治理，而缺乏全流程的体系设计，就无法真正建立长效机制。

**割裂数据质量与资产化、要素化等战略关联。**数据质量管理易陷入“为质量而质量”的技术闭环，仅关注数据准确性等基础指标，却忽视数据能否满足资产化所需的“可估值、可交易”、要素化所需的“可流通、可赋能”等战略要求，最终导致数据质量提升无法转化为实际价值，使数据难以从“资源”真正升级为“资产”与“要素”。

## 2) 数据质量管理过程中的冲突

**数据质量管控要求与业务应用优先的冲突。**数据质量管理贯穿于数据全生命周期，重点关注在对数据质量问题的预防上，在数据标准的梳理、系统的开发设计阶段需要进行强管控，而在上述阶段中业务相关方尚未感知到数据质量的必要性和重要性，参与通常程度较低；而且为了尽快实现系统的业务应用，也会在系统建设过程中出现“不重视”或者“无意识地忽略”数据质量管理的要求。等到系统建设完成开展应用的过程中又发现数据质量问题会导致系统应用效果不佳，反过来产生了数据质量管理的需求，但是在这一阶段往往数据质量的提升的难度和成本都陡然增加。业务以“结果导向”为核心考核指标，而数据质量的价值多体现为“长期隐性价值”，短期与业务的核心诉求存在优先级差异。例如业务高峰期系统需处理海量数据，若严格执行全量、高频次数据校验，可能导致系统响应延迟，影响业务办理效率。

**数据质量存量的问题和增量问题的治理要求存在冲突。**企业积累的存量数据（尤其是历史遗留数据）普遍存在格式不规范、关键信息缺失、逻辑关联弱等问题，整改需要投入大量人力、时间与技术资源，且整改过程中可能影响现有业务系统稳定，而业务部门常因短期业务连续性受影响、整改成本过高而抵制或拖延。同时，在整改优先级上，专业领域基于“影响范围”排序，业务部门则依据自身业务紧急性要求调整，而增量数据需按新的数据质量标准实时管控，若优先开展存量数据整改，可能分散增量数据管理资源，导致新增数据质量问题。最终整改计划混乱，凸显了存量与增量治理在现实操作中的矛盾。

## 二、数据质量管理主要内容

### （一）数据质量管理怎么管

数据质量管理是企业将数据转化为核心资产、驱动业务价值的关键驱动力。它并非单点任务，而是一个需要持续运转的良性循环：从顶层体系构建出发，经由常态化检查评估、深度分析改进直至端到端流程优化，并需配套健全的保障措施以确保该循环高效运行。本节将深入剖析这一循环的核心模块与支撑其落地的保障体系。

#### 1.数据质量体系框架与机制

##### 1) 企业流程管理体系

构建企业级和业务领域级两级数据质量管理流程架构，企业级流程：制定覆盖数据全生命周期的统一管理主流程，包含“需求定义 - 规则制定 - 稽核执行 - 问题整改 - 评价优化”等核心环节。业务领域级流程：在企业级主流程基础上，结合各业务领域的业务特性，补充专项流程模块。以流程为引擎，推动数据质量管理工作的落地执行，形成常态化数据质量运营模式，发布数据质量管理流程，业务与技术协同，保障数据质量管理的各项举措落地执行，保证数据质量要求和规则的有效执行。

##### 2) 构建常态化监控能力

建立实时与定期相结合的数据质量监控机制，做到“全面覆盖+重点聚焦”。实时监控利用自动化工具对关键数据质量指标进行持续监测，及时发现异常（例如数据进入系统时自动校验字段格式和取值范围，不符合规则则实时警告）。定期监控则通过每日/每周/月度数据质量报告或仪表盘，汇总各项质量指标的趋势和问题汇总，供管理层审阅。ISO 8000强调数据质量监控应贯穿数据供应链的每个环节，并通过证据驱动的数据质量度量来增强信任。例如，可采用数据质量评分卡(Data Quality Scorecard)量化评估不同业务域的数据质量状况，设置红线阈值和预警机制。一旦监测发现质量指标跌破阈值，应触发预案，通知相关责任人调查和处理，确保问题“早发现、早纠正”。

##### 3) 设定数据质量管理的目标

数据质量管理的核心目标是通过系统化管理活动，确保数据在全生命周期内满足业务需求，提升数据在业务决策中的价值，从而为企业创造经济效益。

**短期目标**是解决已出现的数据质量问题（如纠正错误数据、补充缺失信息），通过制度化管控手段，确保数据质量符合业务需求，为数据的有效应用奠定基础。

**中期目标**是建立标准化的质量管控流程与技术体系，实现数据质量问题的主动预防。降低因信息偏差导致的战略决策风险，减少数据问题引发的经济损失，同时通过优化数据流程减少人工干预，提升运营效率与业务可持续性，从而精准赋能业务价值创造。

**长期目标**则是将数据质量融入组织文化，使高质量数据成为业务创新、决策优化的常态化支撑。有效的数据质量管理不仅是一次性的整顿，而是需要建立制度化、流程化、自动化的质量保障机制，确保数据在流转与应用的每个环节中都能保持高质量。数据质量管理作为数据治理战略的核心环节，为数据治理提供可靠的数据资产基础。为释放数据价值，支撑组织决策科学化，助力数字化转型，使数据成为驱动组织战略落地的核心生产要素。

## 2.搭建数据质量问题分析框架

数据质量分析是分析问题原因、确认影响范围并指导数据质量提升的过程，一般包括以下几个内容：一是建立组织层面统一的数据质量分析方法，明确统一的数据质量分析要求；二是汇总数据质量检查结果，开展统计分析，找出问题源头，分析问题根因，确定问题影响范围并形成数据质量分析报告，为后续数据质量提升改进提供方案。

常见的数据质量分析方法包括根本原因分析（RCA）、5Why分析法、鱼骨图（石川图）、故障树分析（FTA）、帕累托分析（80/20法则）。

数据质量分析方法	优势	适用场景
根本原因分析（RCA）	<ol style="list-style-type: none"> <li>1. 根源解决问题，避免“头痛医头、脚痛医脚”，降低问题复发率；</li> <li>2. 提升团队的问题解决能力，培养系统性思维；</li> <li>3. 减少因重复问题导致的资源浪费（如返工、投诉处理成本）。</li> </ol>	<p><b>产品质量缺陷</b>：分析不合格品的底层成因；</p> <p><b>安全事故</b>：追溯工伤、设备故障的根本原因；</p> <p><b>客户投诉</b>：挖掘导致客户不满的核心问题；</p> <p><b>流程效率低下</b>：找到影响流程卡顿的关键瓶颈。</p>

5Why 分析法	<ol style="list-style-type: none"> <li>1. 简单直观，易操作，无需复杂工具；</li> <li>2. 聚焦问题本质，逐步深入核心；</li> <li>3. 适合团队快速共识。</li> </ol>	单一问题、线性因果关系场景（如设备故障、简单流程失误）
鱼骨图 (石川图)	<ol style="list-style-type: none"> <li>1. 结构化分类原因（如人、机、料、法、环），全面性强；</li> <li>2. 适合团队协作 brainstorm，激发多角度思考；</li> <li>3. 可视化呈现，便于梳理因果关联。</li> </ol>	多因素导致的问题（如质量缺陷、客户投诉、流程效率低）
故障树分析 (FTA)	<ol style="list-style-type: none"> <li>1. 逻辑严谨，通过“与 / 或”门量化故障路径；</li> <li>2. 适合复杂系统，可追溯多层次因果链；</li> <li>3. 支持概率计算，便于风险评估。</li> </ol>	复杂系统故障（如航空航天、核电设备、软件系统崩溃）
帕累托分析 (80/20 法则)	<ol style="list-style-type: none"> <li>1. 聚焦“关键少数”原因，优先级明确；</li> <li>2. 简单易用，基于数据统计（如缺陷频率）；</li> <li>3. 快速定位核心改进点。</li> </ol>	多原因并存、需优先解决主要问题的场景（如生产缺陷分类、客户投诉原因排序）
因果矩阵	<ol style="list-style-type: none"> <li>1. 量化评估原因对结果的影响（如通过权重打分）；</li> <li>2. 兼顾多个输出指标，综合分析关联性；</li> <li>3. 减少主观偏差。</li> </ol>	多输入多输出的复杂流程（如制造业生产参数优化、服务流程质量提升）

### 3.开展专项数据质量检查与评估

数据质量的评估检查是发现问题的关键手段，通过系统性、多层次的检查与评估机制，实现对数据质量问题的全面识别、及时预警与持续改进。构建**专项检查、日常监测、过程评估**相结合的评估体系，形成“点面结合、高低频互补”的治理闭环，切实保障数据在生产、流转与使用过程中的可靠性，实现全面覆盖和重点突破。

## 1) 专项检查

围绕核心业务场景或高风险数据定期开展深入的专项质量检查，以发现和解决重大隐患。通常频率为季度或半年度一次，如每季度对客户主数据、每年末对财务决算数据进行专项核查。检查范围聚焦对业务影响重大的关键数据（如客户信用等级、交易流水、财务报表数据），采用技术+业务并行的方法：一方面利用SQL查询、数据比对工具等技术手段批量扫描不一致、缺失等问题；另一方面联合业务人员对数据样本进行人工核验，验证数据与现实是否相符。专项检查结束后输出详细的检查报告，列明发现的问题清单、影响范围，并对问题可能原因做初步分析。这种深度检查相当于“审计”，能够定期摸清数据质量的短板，为后续改进提供依据。

## 2) 日常检查

**日常监测：**将数据质量检查融入日常运营，做到高频次、轻量化的常态检查。可设定每日/每周的巡检任务，例如每日自动检查当日新增订单的完整性，每周抽样检查新录入客户数据的准确性。在业务系统中嵌入校验规则（如手机号必须11位数字，不符合则禁止提交）实现事前防控，同时安排数据专员定期抽样复核关键字段，进行事后监督，弥补系统校验的不足。所有日常发现的问题应记录在案，包括问题数量、类型、处理结果，形成数据质量日报/周报。通过持续的日常监测，及时发现即时性、局部性的问题，做到“治病于初始”，防止小问题演变成大事故。

## 3) 过程评估

**过程评估：**除了对数据本身进行检查，企业可定期评估数据质量管理过程的有效性，即对数据质量治理机制做体检，避免流程流于形式。建议每半年或年度进行一次数据质量管理成熟度或效果评估，从以下维度着手：

**流程合规性：**检查各部门是否严格执行了既定的数据质量规范和流程。例如业务部门是否按标准录入数据，IT部门是否及时修复系统校验规则漏洞等。

**工具适用性：**评估现有数据质量工具和平台是否满足需求，如能否支持实时监控、处理海量数据，是否需要升级改造。

**问题整改率：**统计前期发现的数据问题是否及时有效地整改闭环。如上季度专项检查发现的问题项，有多少达到整改完成。该指标反映了问题管理的闭环程度。

**人员执行力：**评估数据质量相关角色履职情况，如数据专员是否按计划完成日常检查，业务团队对发现的问题是否积极配合整改等。

#### 4.数据质量改进和维护

本小节数据质量的改进与维护主要围绕数据质量改进方法与流程维护。

**数据质量的改进**首先需要分析数据质量问题的根本原因，明确数据质量改进方法，对业务流程优化、系统问题修正、制度和标准完善等方面提出具体的行动建议，明确数据质量问题责任人，制定《数据质量改进方案》。其次，采用数据清洗、数据转换、数据整合、业务流程优化、系统问题修复、制度和标准完善等手段和技术，对不符合质量要求的数据进行处理，纠正处理数据质量问题，并及时反馈执行进度与结果信息。最后，记录数据质量事件的评估、初步诊断和后续行动等信息，跟踪问题解决情况，对实施结果进行评估和检查，验证数据质量提升的有效性，输出数据质量提升工作总结报告。

**数据质量流程管理**是确保数据质量管理“常态化、标准化”的关键，需覆盖数据全生命周期，明确每个环节的“操作规范、责任主体、校验节点”。

数据生命周期阶段	核心操作	质量管理要点	责任部门
数据产生阶段	业务人员录入、系统自动生成（如订单数据）	<ol style="list-style-type: none"> <li>1. 制定录入规范（字段格式、必填项）；</li> <li>2. 系统嵌入实时校验（如身份证号合法性校验）；</li> <li>3. 新数据录入后1小时内完成初次复核。</li> </ol>	业务部门（录入）、技术部门（系统支持）如身份证号合法性校验
数据传输阶段	跨系统同步（如CRM→数据仓库）、跨部门共享	<ol style="list-style-type: none"> <li>1. 采用加密传输（如HTTPS、AES加密）；</li> <li>2. 传输前后校验数据完整性（如MD5校验）；</li> <li>3. 实时监控传输成功率，失败时自动重试并告警。</li> </ol>	技术部门（传输监控）、业务部门（需求确认）
数据存储阶段	数据库存储、数据湖归档、备份	<ol style="list-style-type: none"> <li>1. 按数据敏感度分级存储（敏感数据加密）；</li> <li>2. 制定备份策略（每日增量备份、每周全量备份）；</li> <li>3. 数据生命周期管理（过期数据归档/销毁，如3年前的非核心订单数据归档）。</li> </ol>	技术部门（存储管理）、数据管理部门（策略制定）

数据使用阶段	报表生成、决策支持、业务应用（如风控模型）	1. 数据使用前校验质量（如报表生成前检查数据完整性）； 2. 建立数据溯源机制（明确数据来源、加工规则）； 3. 标注数据质量等级（如A级：可直接用于决策，B级：需复核后使用）。	业务部门（使用校验）、数据管理部门（溯源管理）
数据销毁阶段	过期数据删除、废弃系统数据清理	1. 安全销毁（如硬盘物理粉碎、数据库彻底删除，避免泄露）； 2. 销毁前审批（需业务部门与数据管理部门双签）； 3. 留存销毁记录（便于审计）。	技术部门（执行销毁）、数据管理部门（审批监督）

## （二）数据质量管理的保障措施

数据质量管理的落地需要在组织层面和技术层面提供配套保障措施，包括组织架构、制度流程、工具平台和企业文化等方面的建设。本节从实践角度提出实施要点。

### 1. 组织架构

组织架构是阐明角色、责任和决策过程的框架，它描述了人们如何互相协作，确保组织内部的正确职能得到体现，促进了沟通，并提供了解决问题的流程。

数据质量的管理应构建“决策 - 管理 - 执行”三级组织架构，明确各层级职能与跨部门协作机制，避免出现多头管理和职责缺位，实现纵向贯通、横向协同的数据质量架构。

**1) 决策层：**由企业高层（如CEO、CIO）及各核心业务部门负责人组成“数据质量管理委员会”，作为领导机构，主要职责包括：

- 审定数据质量管理战略、政策及年度目标；
- 批准重大资源分配方案，如IT投资预算与业务部门协同支持计划；
- 评审涉及系统改造或流程重构的重大数据质量治理方案；
- 对数据质量整体状况和管理成效进行定期督导与评估。

**2) 管理层：**设立常设“数据管理部门”（如数据治理办公室或数据质量工作组），承担统筹协调与日常管理职能，具体包括：

- 制定、修订并推广数据质量管理制度、标准规范与流程；
- 组织落实数据质量检查、评估、根因分析及改进项目；

- 协调业务部门与技术部门之间的协作，推动跨领域问题解决；
- 建立数据质量报告机制，定期向决策层汇报整体进展与关键结果。

**执行层：**由业务部门“数据联络员”与技术部门“数据工程师”共同组成，作为数据质量工作的前线实施单元：

- 业务侧：负责严格执行数据录入和维护规范，参与数据质量自查与整改，及时反馈业务操作中的质量问题与改进需求；
- 技术侧：承担数据质量管理工具的研发、运维与优化，实施数据清洗、规则配置、系统改造等技术任务，为质量保障提供持续支撑。

## 2.制度体系

数据质量管理的制度体系是一套覆盖数据全生命周期、规范各环节操作标准、明确权责分工的规则集合。



**战略层：**明确数据质量管理的核心目标、原则与愿景，奠定体系建设的方向性基础。

《数据质量管理总纲》：定义数据质量管理的战略定位、核心价值、长期目标及组织保障，统筹规划数据质量的顶层设计与资源分配，确保与组织整体战略目标一致。

**管理层：**构建数据质量管理的管理框架、流程规范与协作机制，解决“如何管”的问题。

《数据质量管理办法》：明确数据质量管理的基本规则、组织职责分工、流程管控要求（如数据采集、处理、存储、应用等全生命周期管理）、问题响应机制及跨部门协作流程，形成系统化的管理规范。

**执行层：**细化管理要求，提供具体可执行的操作指南与技术标准，解决“怎么做”的问题。

《数据质量管理实施细则》：针对数据质量管理的具体场景。如数据清洗、校验、修复、监控等，制定标准化操作步骤、责任分工及工具使用规范，确保各环节操作一致性与可追溯性。

《数据质量管理工具操作说明》：规定数据质量管理工具（如数据质量检测平台、ETL工具、报表系统等）的操作流程、功能应用及安全要求，提升技术执行效率与规范性。

考核层：建立数据质量评估体系与考核机制，通过量化指标驱动持续改进，解决“好不好”的问题。

《数据质量管理绩效考核规范》：制定数据质量的考核目标、考核指标体系、评估方法及奖惩机制，将数据质量管理成效与组织/个人绩效挂钩，形成“执行 - 评估 - 改进”的闭环管理。

### 3.人员管理

人员管理的核心是“建立责任体系、规范协作流程、完善人员培养路径、构建激励机制”。数据质量管理涉及业务、技术、管理多环节，需明确不同角色的核心职责，避免交叉或遗漏。

#### 角色与职责定义

明确数据质量管理中涉及的各类角色及其职责，确保责任到人，确保有人“想管、敢管、会管”。

数据质量管理负责人：制定数据质量战略和目标，审批资源投入（人力、技术预算），推动跨部门协作等。

数据质量管理专员：负责整体数据质量管理策略的制定、实施和监督，组织跨部门协调，报告数据质量状况。

数据质量管理工程师：负责设计和维护支持数据质量管理活动的技术架构、数据模型、ETL流程和数据存储。

业务人员：负责提供质量业务规则，验证数据的准确性和可用性，反馈数据质量问题。

技术人员：负责数据质量管理工具、平台的运维和技术支持。

### 4.人员培训

建立持续的培训机制，加强全员数据质量意识培训，将数据质量理念融入企业文化，提升人员数据素养和专业技能。

入职培训：为新员工提供数据质量管理基础知识、公司政策和流程的培训。

定期培训：组织内部专家或外部聘请的专家开展关于数据质量方法论、前沿技术融合、最佳实践、案例分析等培训。

技能提升：支持员工参加外部认证或专业课程。

知识共享：建立内部知识库、社区或定期分享会，促进经验交流。

#### 绩效考核与激励机制

将数据质量相关指标纳入人员绩效考核，激励员工积极参与。

KPI设定：为数据管理员等角色设定明确的数据质量KPI。

## 5. 数据质量评价体系

数据质量评价指标的设计需以业务为导向，结合企业实际需求选择核心维度，避免指标体系过于复杂或脱离实际。应根据不同业务环节对数据的依赖重点，合理设定动态权重，突出核心数据的质量要求，设计可监控、可改进的质量管理目标。

## 6. 平台工具

数据质量管理（Data Quality Management, DQM）实践中，平台工具是核心支撑，其功能覆盖数据探查、清洗、监控、规则管理、问题溯源等全流程，帮助企业高效发现并解决数据质量问题（如缺失值、重复值、逻辑错误、不一致性等）。根据工具的核心定位和适用场景，可划分以下几类：

### 1) 全流程数据质量管理平台类

此类工具是数据质量管理的“核心中枢”，整合了从数据诊断到持续监控的全环节能力，适合企业级、规模化的数据质量管控，支持多数据源接入（如数据库、数据仓库、湖仓、API 等）。

### 2) 数据探查与诊断工具类

此类工具聚焦“数据摸底”，帮助用户快速了解数据结构、分布特征及潜在质量问题，是数据质量管理的“第一步”（通常在清洗、建模前使用）。

### 3) 数据清洗与标准化工具类

此类工具针对已发现的质量问题（如重复、格式混乱、逻辑错误），提供自动化或半自动化的修复能力，是数据“提质”的核心环节。

### 4) 数据质量监控与告警工具类

此类工具聚焦“持续保障”，通过实时/定时检查数据质量规则，及时发现新增问题并触发告警，避免低质量数据流入下游（如报表、分析模型、业务系统）。

## 三、数据质量管理实施要点

数据质量管理不仅是一套理念或方法，更是需要在组织中落地的实践。很多组织在推动数据治理过程中，常常遇到“目标明确但执行不到位”的问题，往往是因为缺乏清晰的实施要点和落地路径。本章围绕六个方面提出实施要点，以便政企组织能够形成“有顶层设计、有制度保障、有技术抓手、有协作机制”的整体治理格局。

### （一）领导支持

**战略决策与方向引领。**企业高层领导的支持和重视是数据质量管理成功的关键因素。高层领导需将数据质量管理纳入企业整体战略框架，确立数据作为核心资产的战略定位。通过制定与业务发展目标一致的数据管理方针，明确数据质量管理长期目标与实施路径。

**资源保障与区域协同。**数据质量管理的实施离不开坚实的资源保障与高效的区域协同机制。高层领导需统筹调配技术工具、预算资金及专业人才等资源，为数据质量管理提供基础保障，还需协调业务部门与数据研发团队的深度协作，通过资源整合实现数据质量管理的全流程覆盖。

**文化塑造与责任落实。**高层领导需解决“数据是部门私产”的认知壁垒，构建“数据质量人人有责”的组织文化，通过明确数据质量准则，推动全员参与数据质量管控，形成自上而下的责任体系。同时，高层需监督责任落实情况，建立数据质量绩效评估与问责机制，确保数据质量管理要求贯穿业务全链条。

### （二）认责机制

建立清晰的认责机制是数据质量管理体系落地的关键举措，需明确从数据产生、处理、流转应用到全生命周期中各环节的责任主体，例如业务部门对原始数据录入的准确性负责，IT部门对数据传输与存储的完整性负责，分析团队对数据建模的合理性负责。此外，认责机制还应包含责任追溯与协同机制，当出现数据质量问题时，能快速定位责任环节与相关人员，既确保问题能够及时整改，也通过跨部门责任协同推动根源性问题的解决，从而在组织层面构建起数据质量文化氛围。

### （三）评价机制

**制度基础。**常态化评价机制需明确责任主体与制度框架，由专职团队统筹执行，制定定期与不定期相结合的评价计划，依据行业规范或企业内部制度定义指标阈值，形成制度化保障。将评价结果纳入考核以落实责任，作为数据治理优化的输入依据。通过定期培训提升全员数据质量意识，对标行业最佳实践，并根据业务变化动态调整评价规则与指标，保持机制适应性。

## （四）协作机制

数据质量管理并非**某一单独部门的职责**，其成效高度依赖技术与业务部门的深度协作。当前，部分组织普遍面临业务部门与技术部门在数据管理关系不对等、业数融合困难、数据质量执行人员与业务部门脱节等问题，导致数据质量规则落地难、问题反复出现，严重制约数据价值释放。

业务部门与数据部门的协作不畅，本质上是组织管理、目标导向与专业认知差异共同作用的结果。组织管理主要是由于层级间不对等导致的执行乏力，而业务部门与数据部门的目标导向差异，进一步导致了协作堵点加剧，最终因业务与数据部门的专业背景差异造成了较难打破的沟通困境。

构建业务与数据部门的高效协作机制，需要超越传统数据部门主导的模式，通过嵌入式治理将数据质量要求融入业务管理的全链条，实现组织、流程与工具的深度融合。

在本小节中提出了嵌入式的治理方式，从单向约束到双向赋能的实践路径。

### 1.组织嵌入：建立跨部门协作实体

打破层级壁垒的核心是在组织架构中设立跨部门数据治理实体，赋予其超越单一部门的协调权与决策权。这类实体通常分为三级：一是由高管牵头，以数据管理认责方为主体，业务与数据部门负责人共同参与，负责审批数据战略与重大规则；二是数据工作组，由各数据业务认责方组成，负责制定域内数据标准与解决跨部门问题；三是岗位级数据联络人，以数据操作认责方、数据技术认责方为主体，对所经手的数据质量负责，也是数据问题的“第一响应人”。

### 2.流程嵌入：全生命周期数据管控

流程嵌入的核心是将数据质量管控节点融入业务流程，实现业务开展即数据受控。为保障流程嵌入的落地，组织可在重要业务流程环节中建立数据质量检查点，在业务流程的关键节点设置数据准入条件。例如，某制造企业规定生产订单下达前必须完成物料清单数据完整性校验，未达标的订单无法进入生产环节，让数据质量成为业务流程不可分割的一部分。

### 3.工具嵌入：技术平台支撑协同效率

通过构建和现有组织架构相融合的数据管理技术工具，实现数据规则共享、问题协同与价值可视化，降低协作门槛。通过直观的界面与自动化功能，减少业务部门的操作负担。

## 四、数据质量管理挑战与展望

在数字化转型加速和人工智能迅猛发展的时代背景下，数据作为关键生产要素的地位凸显，数据的质量管理作为保障数据价值的关键环节，将直接影响着数据价值的释放效率与决策可靠性。

### （一）主要挑战

数据质量管理的主要挑战将伴随技术演进、业务需求升级以及数据环境复杂化而持续演变，集中体现以下四大维度。

#### 1.数据量激增加大质量校验成本

庞大的数据规模给数据质量管理带来严峻挑战，治理成本正呈现攀升态势，不仅对企业的硬件、算力支撑提出了更高要求，也使得参与全流程数据治理的人力与时间投入持续增加。从数据流转链路来看，海量数据的采集、清洗、转换、集成以及存储等基础操作耗时大幅增加，并极有可能出现数据丢失、重复或不一致的情况。从数据质量监测来看，数据量激增会加大质量校验难度，规则校验无法全面且及时覆盖，数据质量问题隐蔽性增强，人工排查成本高且易遗漏。

#### 2.复合型数据质量人才短缺

近年来随着数字化转型的不断深入，各行业对数据质量的要求越来越高，对复合型数据质量人才的需求也急剧增加，但当前数据质量人才队伍的建设面临多重挑战。一方面，数据人才专业虽已得到高校等教育机构的重视，但从事数据分析应用等价值显现度高的领域人才偏多，愿意从事数据基础管理领域的人才偏少，供给的增速远难匹配要求提升的步伐；另一方面，由于数据治理本身的跨领域特性，对于人才本身的业数技综合素质要求较高，一是技术能力，熟练掌握数据清洗、建模、校验工具等，能识别数据冗余、缺失、不一致等问题；二是业务认知能力，深入理解行业业务逻辑，可将业务需求转化为数据质量标准；三是跨部门协同能力，能协调技术、业务、风控等部门，推动数据质量问题落地解决。满足此类要求的复合型人才在市场上较为稀缺，且培养周期长、跨领域学习实践门槛高，导致复合型数据质量人才的缺口愈发显著。

#### 3.隐私保护与伦理法律带来的合规约束困境

隐私保护的核心目标是最小化个人可识别信息的暴露风险，数据质量管理的目标是最大化数据的可用性、准确性，这使得数据质量管理必须在“保护隐私”与“保障质量”之间寻找动态

平衡点。企业需要确保数据在收集、存储、处理和共享的全过程中都符合法律法规和行业标准，这进一步增加了数据质量管理的复杂性和难度。

#### 4.多模态融合带来的治理难度升级

随着全社会信息化数字化水平的飞速发展，数据应用要求日益提高，数据产生来源繁多、形式日益多元，业务系统输出规整的结构化数据，网页、日志等产生半结构化数据，图像、文本、音视频等构成非结构化数据，逐渐形成以结构化数据、半结构化数据和非结构化数据共存的多元化数据生态格局。不同类型、不同渠道的数据相互补充，为全面挖掘数据潜力，满足多样化业务需求，提供了海量的信息要素，却给数据质量管理及高质量高可用数据的萃取带来多重挑战。数据类型从结构化数据向非结构化、半结构化数据等多模态扩展，不同模态数据的语义差异大、处理逻辑复杂，传统单一模式的数据清洗与质量校验规则难以直接复用。多模态数据融合需解决格式（结构化/非结构化）、标准（字段定义、编码规则）、采集频率（实时/批量）等基础问题，增加数据清洗、校核与融合等治理难度。

## （二）未来展望

数据质量管理未来发展将呈现技术革新、政策赋能与生态协同的深度融合趋势，同时，数据质量管理范围将延伸到非结构化数据。

### 1.技术革新：智能化与自动化作为辅助

在数据规模飞速扩张、业务节奏加快和合规要求愈加严格的背景下，在技术层面，智能化与自动化将有效辅助数据质量管理的工作，例如实时动态监测、自主分析数据特征和质量问题、构建智能修复模型等，有效减缓数据治理的资源分配问题。

智能化数据质量管理的价值主要体现在：

一是大幅提升效率，自动化的质量检测与修复替代了大量人工操作，显著降低了治理成本；

二是提高准确性，通过多维度的特征分析研判，质量问题的识别准确率相比传统规则有大幅提升；

三是实现前瞻性预防，能够基于历史数据和实时特征预测潜在的质量问题，将质量问题的发现时间从“事后”提前到“事中”甚至“事前”，为业务决策提供了更加可靠的保障。

### 2.政策驱动：夯实企业数据管理根基

国家层面持续推动数据产业高质量发展，通过数据授权运营机制、公共数据开放目录等

## 联系方式:

大数据技术标准推进委员会

地 址:北京市海淀区花园北路52号

邮 编:100191

邮 箱:zhoujingjing@caict.ac.cn

官网:www.tc601.com

