ICS 35.240.01

CCS L70

|  |
| --- |
|  |

团 体 标 准

T/CECC XXX—2024

|  |
| --- |
|       |

数据质量评估 指标体系

第1部分：通用指标要求

Data quality evaluation—— Indicator system——
Part 1: Common indicator requirements

|  |
| --- |
| （草案讨论稿） |
|  |

2024-XX-XX发布

2024-XX-XX实施

**中国电子商会** 发 布

目 次

目 次 I

前 言 II

1 范围 3

2 规范性引用文件 3

3 术语和定义 3

4 通用指标要求 4

参 考 文 献 12

前 言

《数据质量评估 指标体系》分为以下3个部分：

‑‑‑‑第1部分：通用指标要求；

‑‑‑‑第2部分：开放共享要求；

‑‑‑‑第3部分：交易要求。

本文件 T/CECC XXX—2024 为第1部分。

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这些专利的责任。

本文件由中国电子商会数据要素发展工作委员会提出。

本文件由中国电子商会归口。

本文件起草单位：。

本文件主要起草人：。

数据质量评估 指标体系 第1部分：通用指标要求

1. 范围

本文件规定了数据质量评估指标体系，给出了通用指标的要求。

本文件适用于数据生存周期各个阶段的数据质量评价。

1. 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 5271.17—2010 信息技术 词汇 第17部分：数据库

GB/T 35295—2017 信息技术 大数据 术语

GB/T 36344—2018 信息技术 数据质量评价指标

1. 术语和定义

GB/T 5271.17—2010、GB/T 35295—2017和GB/T 36344-2018所界定的以及下列术语和定义适用于本文件。

* 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。

[来源：GB/T 5271.17—2000，定义01.01.02]

元数据 metadata

关于数据或数据元素的数据（可能包括其数据描述），以及关于数据拥有权、存取路径、访问权和数据易变性的数据。

[GB/T 5271.17—2020，定义17.06.05]

* 1.

数据质量 data quality

在制定条件下使用时，数据自身、数据生产过程、数据应用过程应满足的相关要求。

数据生存周期 data lifecycle

将原始数据转化为可用于行动知识的过程。

数据集 dataset

具有一定主题，可以标识并可以被计算机化处理的数据集合。

数据模型 data model

对分析的图像和文本表述，该分析识别了组织为完成其使命、功能、目标、目的和战略，以及管理和评价组织所需要的数据。

主题域 subject domain

数据库中的主题域是联系较为紧密的数据主题的集合，用于描述和分析特定的业务领域或主题。

概念域 conceptual domain

数据库的概念域是数据模型的高层次抽象，代表了现实世界中一个特定领域或者业务领域中所有相关概念和实体的集合。

逻辑模型 logic model

数据库设计中的抽象层，=定义了数据的组织方式，包括数据的类型、属性、实体之间的关系以及数据的约束关系。

物理模型 physical model

数据库设计中的实现层，描述了数据在存储介质上的具体存储方式和访问方法，包括数据的存储结构、存储参数、访问路径和硬件布局等。

被评估方 organization subject to the appraisal

申请数据质量评估的组织个体。

1. 通用指标要求
	1. 概述

数据质量评估指标体系如图1所示。



图1 数据质量评估指标体系

* 1. 合规性
		1. 管理

被评估方应建立以下相关管理机制：

1. 被评估方应建立完善的数据安全规划，确保数据在整个生命周期中的安全性、完整性和可用性；
2. 被评估方应建立数据分类分级管理机制，实现对数据资产的科学管理；
3. 被评估方应建立数据访问权限管理制度，确保数据的安全性和合规性，防止未授权访问和数据泄露；
4. 被评估方应建立数据脱敏管理制度，被评估方应建立数据脱敏管理制度，明确数据脱敏的原则、范围、方法和技术要求；
5. 被评估方应建立数据共享和开放安全管理制度，明确数据共享和开放的条件、范围、流程、权限和责任；
6. 被评估方应建立数据安全销毁管理制度，确保不再使用的数据得到安全、彻底的销毁，防止数据泄露或被未授权的第三方恢复；
7. 被评估方应建立供应方安全管理制度，确保供应链中的数据安全和服务连续性；
8. 被评估方应建立安全监督检查制度，定期对数据安全管理的有效性进行审核和评估；
9. 被评估方应建立安全事件管理与应急响应制度，在发生数据安全事件时能够迅速、有效地采取应对措施；
10. 被评估方应设立专职的数据管理岗位，负责组织数据规划、数据治理、数据质量控制等工作，数据管理岗位人员应接受数据安全培训；
11. 敏感数据应进行脱敏与加密处理，以保护个人隐私和商业秘密，防止数据在非授权环境中被滥用；
12. 数据中采集的个人数据应得到相关用户的授权，保证数据采集的合法性；
13. 数据应建立内容安全检查机制，在文本、图片、音频、视频等内容时防止违法不良信息的传播，保障用户隐私和数据安全。
	* 1. 技术

被评估方应实行以下技术要求：

1. 被评估方宜将数据存储到本地、大型服务商私有云或中型服务商私有云，在条件允许的情况下宜避免将数据存储在小型服务商私有云或软件开发商服务器等位置，保证数据存储安全性；
2. 被评估方应在数据机房设置备用电源，保障关键信息系统和基础设施在主电源发生故障或中断时能够持续运行，避免数据丢失、系统损坏或业务中断；
3. 数据应支持测试环境搭建，确保系统开发过程中功能验证与性能优化；
4. 数据存储主机应保证独立的网络环境，独立于互联网，以防止外部网络攻击和未授权访问；
5. 被评估方应使用自动化或手动的方式实现数据备份，并实现备份数据的完整性检查，确保在数据丢失、损坏或遭受攻击时能够迅速恢复；
6. 被评估方应建立数据销毁记录，保证数据销毁可查询；
7. 被评估方应定期开展数据安全审计，评估数据管理活动的合规性、有效性以及潜在的安全风险；
8. 被评估方应建立数据分类保护清单，明确不同类别数据的安全保护要求和措施；
9. 被评估方应建立数据访问记录档案，详细记录所有数据访问活动，包括访问者身份、访问时间、访问目的、访问的数据范围和操作类型等关键信息；
10. 被评估方应对数据访问密码进行约束，强化账户安全，防止未授权访问和数据泄露。
	1. 规范性
		1. 数据库结构

位于数据库的评估对象，其结构应符合以下要求：

1. 数据库的命名应简洁、明确，字数宜不大于10个文字，应反映其内容及用途，应避免使用模糊词语或特殊字符、符号；
2. 数据库的描述应提供数据库的总体功能说明，包括数据类型和结构、数据更新频率、使用目的和服务的用户群体；
3. 索引的命名应简洁、明确、易于识别；
4. 索引命名应包含关联表名和字段信息，且应在数据库范围内唯一，避免与表名或其他数据库对象名称冲突；
5. 视图的命名应反映其数据范围或查询目的，结合业务需求和数据内容，采用清晰、易懂的名词或短语；
6. 数据表的表名应反映表中数据的性质与用途，采用名词或短语，同一数据库中的数据表命名应具有统一规则。
	* 1. 数据概念

评估对象的相关数据概念应符合以下要求：

1. 文件名的命名应反映文件内容或类型，遵循一致的格式规则，避免不相关或过于简略的命名；
2. 数据集的命名应反映其所包含的信息内容、来源及时间范围，便于用户快速识别；
3. 数据集的描述应包括数据集的来源、时间跨度、采集方法、质量评估，以及数据及的适用范围、使用条件和限制；
4. 数据元素的命名简洁、具有描述性，能够反映字段所存储数据的类型、单位或度量标准；
5. 数据元素的描述应包括数据元素的定义、数据类型（如字符串、整型、浮点型等）、数据格式（如日期格式）、适用的约束条件（如是否必填）、业务规则与计算逻辑信息；
6. 参考数据的编码方式应符合业务规则，编码应唯一、可读、可扩展；
7. 参考数据的命名应明确、一致，反应数据含义与用途，避免使用模糊或容易引起混淆的词语；
8. 参考数据应覆盖业务所需的关键信息，包括但不限于分类、属性、关系和约束条件；
9. 数据内容应与预期的数据格式和业务逻辑一致，确保实际存储和处理的数据符合数据模型的要求，应避免数据格式错误、空值或逻辑不一致。
	* 1. 数据组织

位于数据仓库的评估对象，其组织应满足以下要求：

1. 主题域分类应对应特定业务领域，分类应保持相对稳定，不宜频繁变化；
2. 主题域分类在口径上应具有一致性和可比性，覆盖数据所涉及的主要业务领域，不应出现较大遗漏或交叉重复；
3. 概念域应覆盖业务领域关键内容和特征；
4. 概念域的实体、属性、关系以及其他元素的命名应明确、简洁、一致，反映数据业务含义，不应使用可能引起混淆的文字。
	* 1. 数据模型
			1. 逻辑模型

数据库的逻辑模型应符合以下要求：

1. 逻辑模型应保持内部的逻辑一致，确保数据结构和业务规则在整个模型中的统一和协调；
2. 逻辑模型应包括所有必要实体、属性、关系以及数据完整性约束，准确表示业务需求和数据结构，避免关键数据元素的遗漏；
3. 逻辑模型应易于理解和解释，采用清晰、直观的命名规则和分组分层设计方法；
4. 逻辑模型应准确反映现实世界的数据特征和业务规则，模型中的实体、属性和关系应真实代表其所模拟的现实世界对象，确保数据模型与业务需求之间的一致与对应。
	* + 1. 物理模型

数据库物理模型应符合以下要求：

1. 物理模型应能够适应特定的运行环境和性能要求，充分考虑数据库的存储结构、访问方法、索引策略和硬件资源；
2. 物理模型应具有完整性，在物理存储层面保证数据的准确和一致；
3. 物理模型应具有可读性，模型清晰可读、易于理解；
4. 物理模型应具有一致性，确保数据库状态符合预定的规则和约束；
5. 物理模型应准确实现和映射逻辑模型的设计，确保数据在存储、处理和访问中的真实性和正确性。
	* 1. 元数据
			1. 元模型

元模型应符合数据库设计与业务逻辑要求，确保其能够准确、全面描述系统中各类数据和关系。

元模型中的实体、属性、关系等元素的命名应遵循明确的命名规则，保持结构的一致性和完整性，确保模型能清晰、直观地展现业务逻辑和数据依赖关系。

元模型应保持技术与业务逻辑的一致性，确保数据的存储、处理和访问均能满足业务需求。

* + - 1. 业务元数据

业务元数据应准确描述实际业务流程、概念和实体，确保其能够真实反映业务操作。

业务元数据项的含义、适用范围和用途应与实际业务保持一致。

业务元数据应涵盖所有与业务相关的关键信息，各业务模块的元数据均应得到全面描述，支持业务操作的全生命周期。

* + - 1. 技术元数据

技术元数据应准确描述数据库的技术实现细节，包括但不限于表结构、索引、存储方法等，确保数据库系统正确执行设计预期。

技术元数据应涵盖系统中所有技术细节，确保所有与系统性能、存储、访问相关的技术要素均得到清晰描述，包括但不限于硬件资源、存储架构、访问路径和优化策略。

* + - 1. 操作元数据

操作元数据应准确记录数据库系统的运行状况和操作历史，如操作日志、备份记录用户行为等，确保其能够为系统管理与维护提供可靠依据。

操作元数据应涵盖系统操作的所有相关信息，包括用户行动、系统资源利用、故障记录、警报等，确保系统运行中所有操作均记录且可追溯。

* + 1. 业务规则
			1. 业务规则

业务规则的定义和描述应符合一致、简洁、明确的原则，确保业务规则清晰反映业务逻辑，应避免使用模糊或不确定的表述。

业务规则应覆盖所有业务流程的关键逻辑，确保业务系统能够准确执行实际操作中所有必要逻辑。

* + - 1. 业务规则文档

业务规则文档应包括所有业务逻辑和例外处理情况，确保业务流程各环节得到清晰记录，为开发、测试和维护提供详细的操作指南。

业务规则文档的描述应与实际业务流程一致，应避免错误、过时信息，保障业务规则在系统中的执行与文档描述一致。

* + 1. 数据精度

数据精度应满足基于数据所开展业务的使用需求，支持业务决策的准确与高效。

* 1. 完整性

数据内容应具备完整性，确保数据在存储、处理和传输过程中不被非法篡改、破坏或丢失。

数据档案应具备完整性，确保档案中的数据信息是全面、准确、未被篡改。

* 1. 准确性
		1. 数值准确

数据的数值准确性方面应符合以下要求：

1. 数据统计口径应一致，确保不同业务开展过程中所采用的标准、定义、方法和流程保持统一；
2. 数据内容应处于业务认定的合理取值范围，确保数据的准确性和有效性，防止错误数据对业务决策、运营和报告造成负面影响；
3. 数据采集应具有无偏性，确保收集的数据能够客观、公正地反映实际情况；
4. 数据分布应符合正态性，保证统计分析的有效与可靠；
5. 数值型数据超出均值加减三倍标准差范围时，应具有业务的可解释性。
	* 1. 重复率

数据相关内容的重复率应符合以下要求：

1. 数据应控制数据接口重复率，保持重复率处于低水平，确保系统稳定性和数据准确性。
2. 数据应控制文件重复率，优化存储空间使用效率和数据管理成本。
3. 数据应控制数据表重复率，维护数据库的完整性与查询效率。
4. 数据应控制数据库重复率，保持数据的一致性，优化存储资源。
5. 数据应控制数据记录重复率，保障数据的准确性和可靠性，提升数据库运行效率。
	1. 一致性

数据属性的一致性应符合以下要求：

1. 数据应具有内容的一致性，确保数据在不同位置保持相同的含义与结构，维护数据的准确性和可信度；
2. 数据应具有交叉验证一致性，保证数据的准确性与可靠性；
3. 数据命名应具备一致性，使用统一的命名规则；
4. 数据计量单位应具备一致性，确保数据交流的准确性和效率；
5. 数据类型应具备一致性，相同含义、属性的数据字段应保持相同类型；
6. 数据格式应具备一致性，便于数据的共享与交换；
7. 数据来源应具备一致性，确保数据质量与可比性；
8. 相同用途数据的处理应具备一致性，确保数据处理结果的可预测性与公正性；
9. 数据的档案应具备一致性，确保数据的管理、检索的效率与可靠。
	1. 共享性
		1. 环境支持

数据在对外部环境的支持方面应符合以下要求：

1. 数据接口多样性应符合业务要求，适应业务对不同数据源、数据格式以及传输协议的需求；
2. 数据接口标准化程度应符合业务要求，确保数据交换的顺畅性；
3. 数据的兼容性与互操作性应符合业务要求，确保数据在不同系统、平台以及应用之间的切换与利用；
4. 被评估方应对数据共享性能开展监控，确保数据共享过程的顺畅性与高效性；
5. 被评估方应对数据共享访问进行控制，确保数据安全与隐私，防止未授权数据的访问与泄露；
6. 数据应具备多平台支持能力，满足业务开展过程中不同系统、设备和用户群体的需求；
7. 数据对并发访问的支持应符合业务要求，保证数据在高流量与多用户环境下的稳定性与高效响应。
	* 1. 数据属性

数据属性在共享性上应符合以下要求：

1. 数据共享策略应符合业务要求，保证数据能够安全、高效地被用户访问和使用；
2. 数据的更新与同步应符合业务要求，保证数据在业务使用中的一致性与实时性；
3. 数据的可访问性应符合业务要求，确保用户对数据使用的及时性与便利性；
4. 数据的可组合性应符合业务要求，确保不同数据源和类型能够灵活地融合与重组，满足复杂业务场景下的分析需求和创新应用；
5. 数据结构化程度应符合业务要求，确保数据能够以有序、逻辑性强的方式组织和存储；
6. 数据的文件格式应可转换，满足不同业务场景下对数据格式与传输的要求；
7. 数据应具备可溯源性，确保在数据生命周期内能够追踪其来源、处理过程和流转路径；
8. 数据的共享应遵从法规要求，满足业务在合法合规的前提下高效利用数据资源的要求；
9. 数据应具备反馈与改进机制，保证数据的修正与持续优化；
10. 数据的共享性报告与透明度应符合业务要求，支持用户了解数据共享的范围、规则、效果以及潜在风险，增强对数据共享活动的信任与支持。
	* 1. 数据易读性

数据的易读性应符合以下要求：

1. 被评估方应对数据用户开展需求调研，提升数据的实用性；
2. 数据应具备相应的说明文档与示例，文档与示例应符合业务要求；
3. 被评估方应开展数据共享的用户培训与支持，增加用户对数据共享价值的理解，正确使用共享数据并遵守政策法规要求。
	1. 时效性
		1. 数据更新

数据更新应符合以下要求：

1. 数据应采用实时或定期的方式进行更新与维护；
2. 数据更新频率应满足业务要求，确保数据的时效性与准确性；
3. 数据更新延时应符合业务要求，保证数据在业务规定时间的得到处理与反馈；
4. 数据的更新应进行备份与恢复设置，确保在数据更新过程中能够有效防止数据丢失或损坏，并在发生问题时能够快速恢复数据；
5. 数据的更新应有记录，以便追踪数据的变更历史，包括更新时间、更新人员、变更内容等信息。
	* 1. 其他

数据的时效性应符合以下要求：

1. 数据采样速度应符合业务要求，满足业务对高精度与动态变化需求的分析；
2. 数据输出速度应符合业务要求，满足业务在紧急或实时控制场景下对数据的分析；
3. 数据的时间戳稳定性应符合业务要求，确保数据的准确性和一致性；
4. 数据时间戳的流畅性应符合业务要求，确保数据记录能够连续、无间断地记录时间信息；
5. 数据备份的时效性应符合业务要求，确保在数据意外丢失或系统故障时能够及时恢复最新数据。

参 考 文 献

* 1. 《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》
	2. ISO 8000—1:2022 Data quality — Part 1: Overview
	3. ISO 8000—2-2022 Data quality — Part 2: Vocabulary
	4. 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
	5. GB/T 25000.24—2017 系统与软件工程 系统与软件质量要求和评价(SQuaRE) 第24部分:数据质量测量
	6. GB/T 34960.1—2018 信息技术服务 治理 第1部分：通用要求
	7. GB/T 36344—2018 信息技术 数据质量评价指标
	8. GB/T 38667—2020 信息技术 大数据 数据分类指南

**━━━━━━━━━━━**