ICS 35.240.01

CCS L70

|  |
| --- |
|  |

团 体 标 准

T/CECC XXX—2024

|  |
| --- |
|       |

数据质量评估 实施方法

Data quality evaluation——Implementation methods

|  |
| --- |
| （征求意见稿） |
|  |

2024-XX-XX发布

2024-XX-XX实施

**中国电子商会** 发 布

目 次

前言 [Ⅱ](#_Toc60239894)

目 次 I

1 范围 3

2 规范性引用文件 3

3 术语和定义 3

4 评估流程 3

5 数据质量判定方法 6

附　录　A （资料性） 数据质量评估指标参考 8

参 考 文 献 9

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 实施方法

1. 范围

本文件给出了开展数据质量评估的评估流程及数据质量判定方法。

本文件适用于数据质量评估活动。

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。

T/CECC XXX—2024 数据质量评估 指标体系 第1部分：通用指标要求

1. 术语和定义

T/CECC XXX—2024所界定的以及下列术语和定义适用于本文件。

* 1.

被评估方 organization subject to the appraisal

申请数据质量评估的组织个体。

评估机构 assessment authority

统筹数据质量评估工作的组织个体。

1. 评估流程
	1. 概述

数据质量评估流程包括评估前沟通、评估计划制定、评估实施、质量分析以及质量改进环节，如图1所示。



图1 数据质量评估实施流程

* 1. 评估准备阶段
		1. 受理评估

受理评估申请是评估机构对被评估方提交的数据质量评估申请材料进行审核，确认被评估方符合数据质量评估条件，条件包括：

1. 被评估方应是在中国境内合法注册成立并运营的组织，其业务活动符合国家相关法律法规和政策要求；
2. 被评估方应当具备一定的数据管理能力，包括但不限于数据的存储、处理等；
3. 被评估方应具备一定的技术支持能力，向评估机构必要的数据质量评估支持；
4. 被评估方应保持开放和合作的态度，积极配合评估机构的工作。
	* 1. 组建评估组

评估机构根据被评估方提供的数据情况，选派评估人员组建评估组，评估人员应符合相关要求，包括：

1. 具有经济、统计、计算机相关的工作背景或工作经验；
2. 遵守评估工作要求，掌握评估方法与评估工具的使用；
3. 按照评估计划开展评估活动；
4. 维护数据在评估过程中的保密性与安全性；
5. 能够识别数据质量相关问题。
	* 1. 确认评估数据

评估机构应向被评估方明确数据质量评估目的及用途。

评估机构应根据评估目的及用途，明确被评估数据范围，包括但不限于评估数据时间范围、内容范围等。

评估机构应明确被评估数据的来源及获取方式，确定数据来源的合法性与准确性。

评估机构应明确数据存储位置及数据读取方式，保证数据在评估过程中的可访问性。

* + 1. 预评估

评估机构对被评估数据开展预评估，分析数据质量基本情况，达到数据评估基本要求则继续开展评估，未达到数据评估基本要求的应完成数据治理再开展数据质量评估。

* + 1. 明确业务要求

数据的质量与数据所支持的业务需求密切相关，评估准备阶段评估机构应深入分析被评估方业务需求，明确业务运行过程中对数据质量的各项关键要求，包括但不限于数据精度、数据完整程度、数据时效等。

* + 1. 标准法规分析

数据评估机构应根据被评估数据的内容及用途，明确数据评估工作使用的标准及相关法规。

* 1. 评估实施阶段
		1. 计划制定

评估机构应根据被评估数据内容及数据量，合理分配评估人员及时间，制定合理的评估计划，包括评估目的、评估范围、评估任务、评估时间、评估人员、评估日程安排、评估所需要场地和设备等内容。

* + 1. 评估方法

评估机构应按照被评估数据的内容与形式，选择适当的数据质量评估方法，包括：

1. 系统评估：使用数据质量管理系统等数据质量工具实现自动化评估，如自定义SQL语句检核、数据集成工具等；
2. 人工评估：根据评估指标，结合评估人员数据质量评估经验进行数据评估，如资料判断、经验判断等。
	* 1. 评估方式

评估机构应根据委托主体及被评估数据，确定数据评估方式，评估方式包括：

1. 全量评估：对涉及的所有数据进行逐一评估；
2. 增量评估：对涉及的数据，在特定的范围和时间段内新增的数据进行逐一评估；
3. 抽样评估：按照抽样方案，对抽取的数据进行逐一评估。
	* 1. 评估工具

评估机构应根据数据类型，如文本、图像、音频、视频以及三维模型等，选用相应的数据质量评估软件和工具，相关的数据评估软件和工具可参考以下工具：

1. 文本数据：DiVoMiner X AI、ChineseWebText、CTAP、KH Coder、Wordij；
2. 图像数据：ImageJ、Fiji、Image Pro Plus、iQstest、GIQA、Image-Quality-Assessment；
3. 音频数据：AudioAnalyze、MATLAB with Signal Processing Toolbox、Python with LibROSA、Raven Pro、OpenSMILE；
4. 视频数据：Amazon Rekognition、Google Cloud Video Intelligence、Adobe Premiere Pro、DaVinci Resolve、Video Indexer；
5. 三维模型数据：Autodesk ReCap、ArcGIS、DREAM.3D、CloudCompare、ArcGIS 3D Analyst、SuperMap 3D GIS。
	* 1. 指标确定

评估机构应参考T/CECC XXX—2024 《数据质量评估 指标体系 第1部分：通用指标要求》建立数据质量评估指标体系，包括合规性、规范性、完整性、准确性、一致性、共享性以及时效性方面的指标。

* + 1. 指标权重

评估机构应根据数据所支持业务对指标属性要求的不同，包括合规性、规范性、完整性、准确性、一致性、共享性以及时效性，利用相关权重确定方法，确定不同类型指标权重，不同类型指标权重之和为1。

评估机构应根据基于数据所开展的业务对指标内容的要求，利用相关权重确定方法，对指标赋予分值，同一类型指标分值相加为100分，部分指标赋值情况可参考附录A。数据内容与评估对象应用场景相关性较大或与评估日期较为接近时，相关数据内容在完整性、准确性与一致性上应赋予较大权重。

相关权重确定方法包括：

1. 专家评分法：通过领域内专家对各个指标的重要性评价计算指标权重；
2. 层次分析法：通过数据层次结构，比较不同指标之间的相对重要性确定指标权重；
3. 熵权法：基于数据信息熵，通过指标的信息熵大小确定指标权重。

评估机构宜采用主客观结合的方法确定权重。

* 1. 评估结束阶段
		1. 质量分析

评估机构应针对数据质量结果开展分析，识别数据中的全部质量问题。

评估机构应基于数据质量评估过程及结果编制详实的数据质量评估报告，报告内容应包括但不限于评估结果、数据质量分析、数据质量存在问题以及数据质量改进建议。

* + 1. 质量改进

评估机构应及时向委托主体反馈数据质量评估结果与数据质量问题。

评估机构应针对数据质量结果制定数据质量改进计划，针对数据质量问题制定具体改进措施，并根据数据质量问题严重程度及对业务影响大小对改进措施进行优先级排序。

1. 数据质量判定方法

数据质量评分采用加权平均的方法计算，具体计算方法如公式（1）所示。

$D=\sum\_{i}^{n}C\_{i}α\_{i}$ ……………………（1）

式中：

D——数据质量评分；

n——指标类别数量；

i——指标类别，包括合规性、规范性、完整性、准确性、一致性、共享性以及时效性；

C——单个指标类别内指标分数之和；

α——单个指标类别权重。

数据质量评分与数据质量等级的对应关系可参考表1。

表1 数据质量评分与数据质量等级对应表

|  |  |  |
| --- | --- | --- |
| 数据质量等级 | 分数 | 描述 |
| 优秀 | 90～100 | 数据精准可靠，全面高效支撑业务决策，显著提升运营效率。 |
| 良好 | 80～90 | 数据质量较高，有效支持业务分析，助力业务稳定发展。 |
| 合格 | 60～80 | 数据基本满足业务需求，对业务有一定支持，但存在改进空间。 |
| 不合格 | 60以下 | 数据问题频发，严重阻碍业务运行，影响决策效果。 |

1. （资料性）
数据质量评估指标参考

部分指标赋值情况如表A.1所示。

表A. 1 数据质量评估指标及赋值

|  |  |  |
| --- | --- | --- |
| 指标类别 | 指标 | 分值 |
| 合规性 | 建立了安全规划 | 5 |
| 建立了分类分级管理机制 | 5 |
| 建立了访问权限管理制度 | 5 |
| 数据进行了脱敏处理 | 5 |
| 数据进行了加密处理 | 5 |
| …… | …… |
| 规范性 | 数据库命名简洁明确 | 3 |
| 主题域分类稳定 | 5 |
| 概念域覆盖业务关键内容与特征 | 5 |
| 逻辑模型内部逻辑一致 | 5 |
| 元模型符合数据库设计与业务逻辑要求 | 5 |
| …… | …… |
| 完整性 | 数据内容完整 | 5 |
| 数据档案完整 | 5 |
| …… | …… |
| 准确性 | 数据内容符合业务认定合理取值范围 | 2 |
| 数据采集具有无偏性 | 2 |
| 数据分布具有正态性 | 3 |
| 数据接口重复率低于业务要求 | 2 |
| …… | …… |
| 一致性 | 数据内容具有一致性 | 5 |
| 数据交叉验证具有一致性 | 5 |
| 数据命名具有一致性 | 3 |
| …… | …… |
| 共享性 | 数据接口多样性符合业务要求 | 3 |
| 数据兼容性与互操作性符合业务要求 | 3 |
| 数据支持并发访问 | 5 |
| 数据支持多平台访问 | 5 |
| …… | …… |
| 时效性 | 数据采用实时方式进行更新维护 | 5 |
| 数据备份时效性符合业务要求 | 5 |
| …… | …… |

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

**━━━━━━━━━━━**