ICS 35.240.01
CCS L 70

团　　体　　标　　准

T/CESA XXXX—202X

信息技术 公共数据质量评价规范

Information technology——Public data quality evaluation specification

征求意见案

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。
已授权的专利证明材料为专利证书复印件或扉页，已公开但尚未授权的专利申请证明材料为专利公开通知书复印件或扉页，未公开的专利申请的证明材料为专利申请号和申请日期。

202X-XX- XX 发布　　202X-XX- XX 实施

中国电子工业标准化技术协会　发布

目　　次

前　　言 III
1 范围 1
2 规范性引用文件 1
3 术语和定义 1
4 评价原则 2
4.1 科学性 2
4.2 客观性 2
4.3 系统性 2
4.4 可操作性 2
4.5 可比性 2
5 公共数据质量评价的一般流程 2
5.1 确定评价目标范围 3
5.2 收集数据与资料 3
5.3 制定评价规则 3
5.4 执行评价 3
5.5 分析评价结果 3
5.6 报告与反馈 3
6 评价指标 3
6.1 指标框架 3
6.2 指标编号及编码规则 4
6.3 规范性 4
6.4 完整性 6
6.5 准确性 6
6.6 一致性 7
6.7 时效性 8
6.8 可访问性 9
6.9 安全性 10
7 评价方法 11
7.1 评价准备 11
7.2 评价执行 11
7.3 评价结果分析 12
7.4 评价报告 12
附录A 14
参 考 文 献 15

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。

本文件由国家信息中心提出。

本文件由中国电子工业标准化技术协会数字经济推进分会归口。

本文件起草单位：。

本文件主要起草人：。

信息技术 公共数据质量评价规范

1 范围

本文件规定了公共数据质量评价原则、评价流程、评价指标、评价方法等内容。本文件所定义的公共数据类型为结构化数据。

本文件适用于公共数据采集、汇聚、治理、开放、共享和授权运营等全生命周期活动中的数据质量评价，为推动公共数据开发利用提供指导。

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 36344—2018 信息技术 数据质量评价指标

GB/T 43697—2024 数据安全技术 数据分类分级规则

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

3 术语和定义

GB/T 36344—2018界定的术语和定义以及下列术语和定义适用于本文件。

3.1

公共数据 public data

各级具有公共管理和公共服务职能的组织，在依法履职或提供公共服务过程中产生的数据。

注1：本文件提及的数据均为公共数据。

3.2

公共数据质量 public data quality

数据的特性满足公共管理和公共服务的数据管理需求的程度。

3.3

数据项 data item

在特定上下文内数据的最小可识别单位，其定义、标识、允许值和其他信息是由一组属性指定。

[来源：GB/T 25000.24—2017,4.9]

3.4

数据元素 data element

数据的基本单位，定义了数据的属性和特征，一个数据元素可由若干个数据项组成。

4 评价原则

在对公共数据进行质量评价时，应遵循以下原则：

4.1 科学性

从公共数据采集处理及应用全流程实际出发，基于对公共数据质量评价的理论分析，设置评价指标框架，选取具有代表性、完整性和系统性的指标，明确评价方法，开展综合考核评价，为得出科学合理、真实客观的综合评价结果提供保障。

4.2 客观性

公共数据质量评价中，应按照指标体系框架，选取符合评价标的特征的指标，确定数据计算方式，遵循客观可信、全程可监督的评价方法，保证评价结果的准确性和可靠性，避免主观评价。

4.3 系统性

公共数据质量评价涉及领域点多面广，指标的确定力求做到系统全面、突出重点，需覆盖对公共数据质量评价的关键维度和领域，尽可能从不同的层次、不同的角度来描述被评价对象在各个方面的主要特征和状况，避免片面评价，并避免指标间冲突。

4.4 可操作性

指标应可测度、可评价，需选取代表性高的综合指标和专业指标，准确反映评价内容。指标的数据计算公式科学合理，计算支持机器处理，以利于指标数据的搜集、整理、汇总与历史数据分析。评价方法应简洁明确，便于掌握和操作，能够有效地运用于实际评价分析。

4.5 可比性

评价指标体系各项指标应统一可量化，指标值采用相对数，支持用于不同时期、不同地区、不同细分领域间的对比。同时要考虑与国内现有评价指标、现行及未来一段时期内公共数据工作实践接轨的要求，以实现评价指标体系跨时期、跨区域、跨领域的可用性。

5 公共数据质量评价的一般流程

公共数据质量评价的工作流程通常包括确定评价目标范围、收集数据与资料、制定评价标准、执行评价、分析评价结果、报告与反馈6个步骤。公共数据质量评价是一个持续的过程，需要不断地评价、改进和监控。公共数据质量评价流程可以根据具体业务需求和组织环境进行调整和优化，见图1。

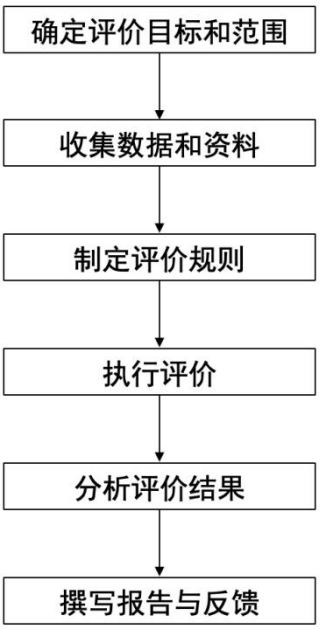

图1 公共数据质量评价的工作流程

5.1 确定评价目标范围

根据工作任务需要，明确公共数据质量评价的目的和范围。

5.2 收集数据与资料

收集被评价的数据及相关资料，对数据进行预处理。

5.3 制定评价规则

根据业务需求和数据特性，选择确定评价指标，制定相应的数据质量评价规则。

5.4 执行评价

根据数据质量评价规则，对数据进行质量评价，可以采用常用的评价方法有数据核对、数据统计分析、数据校准、演绎推算、内部验证、与原始资料（或更高精度的独立原始资料）对比、独立抽样检查、多边形叠加检查、有效值检查等。

5.5 分析评价结果

分析评价结果，确定数据质量的高低，找出数据质量问题，提出相应的数据质量改进措施或建议。

5.6 报告与反馈

编写数据质量评价报告，详细分析被评价数据的质量，包括各项指标的评价结果，给出问题具体描述、改进措施及建议，向利益相关者提供反馈。

6 评价指标

6.1 指标框架

公共数据质量评价指标框架见图2。

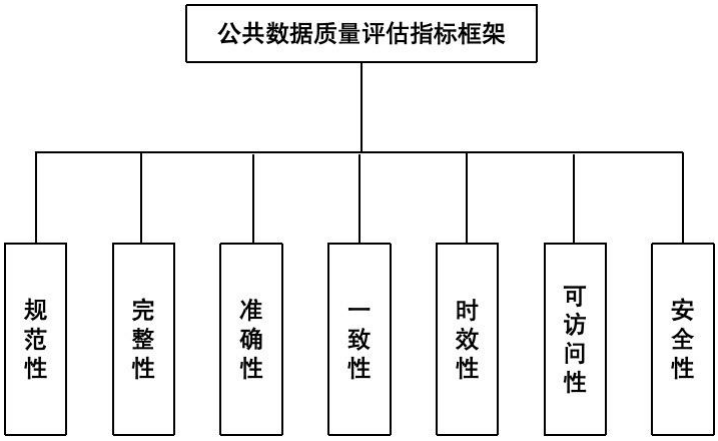

图2 公共数据质量评价指标框架

公共数据质量评价从七方面指标进行评价，包括：

规范性——公共数据符合数据标准、数据模型、业务规则、元数据或权威参考数据的程度。

完整性——按照数据规则要求,公共数据元素被赋予数值的程度。

准确性——公共数据准确表示其所描述的真实实体（实际对象）真实值的程度。

一致性——公共数据与其他特定上下文中使用的数据无矛盾的程度。

时效性——公共数据在时间变化中的正确程度及更新的及时程度。

可访问性——公共数据能被访问和被使用的程度。

安全性——公共数据存储、开放、访问、流转、使用等过程的安全程度。

在七个一级指标框架下,除必选指标外,可根据各部门业务数据实际使用场景选择或增加二级指标,开展数据质量评价。

6.2 指标编号及编码规则

指标编号是公共数据质量评价指标的唯一性编号，由一级指标和二级指标共4位数字组成。编码规则见图3。

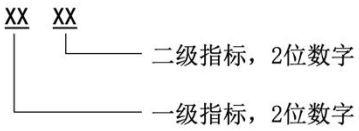

图3 编码规则

一级指标由2位数字组成，01代表规范性指标、02代表完整性指标、03代表准确性指标、04代表一致性指标、05代表时效性指标、06代表可访问性指标、07代表安全性指标。二级指标是由2位数字组成的顺序码，范围为01～99。

6.3 规范性

规范性指标主要用于评价公共数据集在整体结构和内容上符合标准的程度。具体来说包含数据标准符合度、数据模型符合度、元数据符合度、业务规则符合度和权威参考数据符合度等方面，具体评价指标见表1。

表1 规范性评价指标

指标编号	指标名称	指标描述	计算方法	必选/可选指标
0101	数据标准符合度	衡量公共数据集在命名、创建、定义、更新、归档和销毁过程中符合现有国际标准、国家标准、行业标准及地方标准或相关规定的程度。	$X = A/B \times 100$ 式中： A=满足数据标准要求的公共数据集中元素的个数； B=被评价的公共数据集中元素的个数。	必选
0102	数据模型符合度	衡量公共数据集是否符合数据模型要求，包括是否存在清晰可理解的数据模型定义以及这些数据的组织形式。	$X = A/B \times 100$ 式中： A=满足数据模型要求的公共数据集中元素的个数； B=被评价的公共数据集中元素的个数。	必选
0103	元数据符合度	评价公共数据是否符合元数据定义的度量，是否为所有公共数据提供了完整和清晰的元数据文档，包括字段名称、描述和类型值域等内容。	$X = A/B \times 100$ 式中： A=满足元数据定义的公共数据集中元素的个数； B=被评价的公共数据集中元素的个数。	必选
0104	业务规则符合度	评价公共数据是否符合业务规则的度量。数据行为和操作的业务逻辑是否与已有的业务规则相符合。	$X = A/B \times 100$ 式中： A=满足业务规则的公共数据集中元素的个数； B=被评价的公共数据集中元素的个数。	必选
0105	权威参考数据符合度	评价公共数据集中的元素是否使用了权威的参考数据。 注 2：参考数据是系统、应用软件、数据库、流程、报告及交易记录和主记录用来参考的数值集合或分类表。评价数据质量时需要收集参考数据列表。	$X = A/B \times 100$ 式中： A=满足参考数据规则的公共数据集中元素的个数； B=被评价的公共数据集中元素的个数。	必选

6.4 完整性

完整性用于评价公共数据的记录和信息是否完整，是否存在数据缺失情况。该指标包括数据元素完整性、数据记录完整性、元数据标识子集完整性、元数据内容子集完整性，具体评价指标见表2。

表2 完整性评价指标

指标编号	指标名称	指标描述	计算方法	必选/可选指标
0201	数据元素完整性	衡量按照业务规则要求，公共数据集中应被赋值的数据元素的赋值程度。	$X = A/B \times 100$ 式中： A=被赋值的公共数据集中元素的个数； B=预期被赋值的公共数据集中元素的个数。	必选
0202	数据记录完整性	衡量按照业务规则要求，公共数据集中应被赋值的数据记录的赋值程度。	$X = A/B \times 100$ 式中： A=被赋值的公共数据集中记录的个数； B=预期被赋值的公共数据集中记录的个数。	必选
0203	元数据标识子集完整性	评价元数据标识子集的完整程度。	$X = A/B \times 100$ 式中： A=填写完整的公共数据集元数据标识子集个数； B=公共数据集元数据标识子集个数。	可选
0204	元数据内容子集完整性	评价元数据内容子集的完整程度。	$X = A/B \times 100$ 式中： A=填写完整的公共数据集元数据内容子集个数； B=公共数据集元数据内容子集个数。	可选

6.5 准确性

本指标用于评价公共数据所描述的真实实体（实际对象）真实值的程度。该指标包括数据内容正确性、数据格式合规性、数据唯一性、元数据正确性、脏数据出现率，具体评价指标见表3。

表3 准确性评价指标

指标编号	指标名称	指标描述	计算方法	必选/可选指标
0301	数据内容正确性	评价公共数据内容是否是预期数据。	$X = A/B \times 100$ 式中： A=满足数据正确性要	必选

指标编号	指标名称	指标描述	计算方法	必选/可选指标
			求的公共数据集中元素的个数； B=被评价的公共数据集中元素的个数。	
0302	数据格式合规性	评价公共数据格式（包括数据类型、数值范围、数据长度、精度等）是否满足预期要求。	$X = A/B \times 100$ 式中： A=满足格式要求的公共数据集中元素的个数； B=被评价的公共数据集中元素的个数。	必选
0303	数据唯一性	衡量公共数据特定字段、记录、文件或数据集唯一性的程度。	$X = A/B \times 100$ 式中： A=满足唯一性要求的公共数据集中元素的个数； B=被评价的公共数据集中元素的个数。	必选
0304	元数据正确性	评价元数据是否按照所需的正确性对数据进行描述。	$X = A/B \times 100$ 式中： A=提供了合适的需求信息的元数据的个数； B=在数据的需求规则说明中定义的元数据的个数。	可选
0305	脏数据出现率	衡量正确字段、记录、文件或数据集之外无效数据的情况。	$X = A/B \times 100$ 式中： A=有脏数据出现的公共数据集中元素的个数； B=被评价的公共数据集中元素的个数。	必选

6.6 一致性

一致性用于评价公共数据集中数据记录、数据描述和数据逻辑关联等方面无矛盾的程度，包括数据记录一致性、计量单位一致性、数据格式一致性和关联数据一致性，具体评价指标见表4。

表4 一致性评价指标

指标编号	指标名称	指标描述	计算方法	必选/可选指标
0401	数据记录	衡量同一数据项在不同位置	$X = A/B \times 100$	必选

指标编号	指标名称	指标描述	计算方法	必选/可选指标
	一致性	存储或被不同应用/用户使用时，数据记录的一致性水平；数据项在同一时间周期或固定更新频率内发生变化时，存储在不同位置的同一数据被同步修改的程度。	式中： A=达到数据记录一致性要求的元素个数； B=被评价的公共数据集中元素的个数。	
0402	计量单位一致性	衡量数据记录计量单位与元数据计量单位一致的程度。	$X = A/B \times 100$ 式中： A=数据记录计量单位与元数据计量单位一致的数据记录条数； B=数据记录总数。	可选
0403	数据格式一致性	衡量相同数据项数据格式的一致性程度。	$X = A/B \times 100$ 式中： A=相同数据项在不同数据文件中数据格式完全一致的数据项个数； B=可以定义格式一致性的数据项个数。	可选
0404	关联数据一致性	衡量具有逻辑关联的数据项之间语义逻辑符合约束规则的程度。	$X = A/B \times 100$ 式中： A=满足关联数据一致性要求的数据项个数； B=被评价的公共数据集中具有逻辑关联的数据项个数。	必选

6.7 时效性

时效性用于评价公共数据集中的数据更新及时程度和数据内容随时间变化的正确程度，包括基于时间段的正确性、基于时间点的及时性、时序性和更新的及时性，具体评价指标见表5。

表5 规范性评价指标

指标编号	指标名称	指标描述	计算方法	必选/可选指标
0501	基于时间段的正确性	衡量数据记录对应的时间范围或频率分布符合数据描述指定或业务需求的程度。	$X = A/B \times 100$ 式中： A=被评价的公共数据集中满足时间段正确性要求的数据记录数； B=被评价的公共数据	必选

指标编号	指标名称	指标描述	计算方法	必选/可选指标
			集中具有时间段正确性要的数据记录数。	
0502	基于时间点的及时性	衡量数据集中基于时间戳的数据元素记录数、频率分布或延迟时间符合业务需求的程度。	$X = A/B \times 100$ 式中： A=满足时间点及时性要求的公共数据集中数据元素的个数； B=被评价的数据集中具有时间点及时性要点的数据元素个数。	必选
0503	时序性	衡量数据集中同一实体的数据元素之间的相对时序关系符合逻辑关系的程度。	$X = A/B \times 100$ 式中： A=数据集中满足时序性要求的数据元素个数； B=被评价的数据集中元素的个数。	必选
0504	更新的及时性	评价数据项更新符合业务需求的情况，按照更新周期计算及时性符合程度，主要包括实时、天、周、月度、季度、年度更新。	$X = A/B \times 100$ 式中： A=数据集中及时更新的数据项个数； B=被评价的数据集中需要按要求更新的数据项的个数。	可选

6.8 可访问性

可访问性用于评价公共数据在特定的使用周期中能被访问的程度。该指标包括数据元素可使用率、数据元素可访问率、数据格式可访问性，具体评价指标见表6。

表6 可访问性评价指标

指标编号	指标名称	指标描述	计算方法	必选/可选指标
0601	数据元素可使用率	评价公共数据在设定有效周期内的可使用性。用于衡量公共数据在设定有效周期内是否能够按预期满足业务需求，反映数据质量和有效性的综合水平。	$X = A/B \times 100$ 式中： A=满足可使用性要求的公共数据集中元素的个数； B=被评价的公共数据集中元素的个数。	必选
0602	数据元素可访问率	评价公共数据元素在需要时可获取性。用于评	$X = A/B \times 100$ 式中：	必选

指标编号	指标名称	指标描述	计算方法	必选/可选指标
		估公共数据是否能够及时、方便地被检索和获取，反映数据的可用性和系统的支持能力。	A=满足可访问性要求的公共数据集中元素的个数； B=被评价的公共数据集中元素的个数。	
0603	数据格式可访问性	衡量排除数据或信息因特定的格式而不能被预期用户访问的情况后，公共数据格式可访问的程度。	$X = (1 - \frac{A}{B}) \times 100$ 式中： A=因格式问题不能被访问的公共数据项的个数； B=能定义格式可访问性的公共数据项的个数。	可选

6.9 安全性

安全性用于评价公共数据的存储、开放、访问、流转、使用等过程中遵守相关安全规定，并实现相关风险隐患管控的程度。该指标包括敏感数据项脱敏、公共数据访问权限管控情况、公共数据防泄露管控情况和非脆弱性，具体评价指标见表7。

表7 安全性评价指标

指标编号	指标名称	指标描述	计算方法	必选/可选指标
0701	敏感数据项脱敏	已公开的公共数据应对敏感数据项进行脱敏，不会对公共利益或个人隐私造成严重影响，此项为扣分项。	基于词组对比分析，已对全部敏感数据项脱敏得100，未对敏感数据项进行脱敏或未完全脱敏得0。	可选
0702	公共数据访问权限管控情况	评价根据公共数据不同类别和级别，对数据管理、审计类账号开通、分配、使用、变更、注销等安全管理要求和规则制订的情况。	已设置分级分类公共数据访问权限管理制度或规则，并施行相应管控措施100，未实现得0。	必选

指标编号	指标名称	指标描述	计算方法	必选/可选指标
0703	公共数据防泄露管控情况	评价制订相应的应急预警管理机制，对公共数据流转、泄露和滥用情况进行监控，并及时对异常数据操作进行预警、定位和阻断的情况。	已制订并实施相应的应急预警管理机制，得50，未制订实施相应的应急预警管理机制，得0。能实现数据防泄漏监控和响应的，得50，未实现得0。	必选
0704	非脆弱性	衡量设置访问权限的公共数据项只能由授权用户访问的程度。	$X = (1 - \frac{A}{B}) \times 100$ A=在特定时间段内非授权用户为了获得目标数据项而尝试正式入侵期间成功完成访问的次数； B=在特定时间段内非授权用户对目标数据项尝试访问的次数。	可选

7 评价方法

7.1 评价准备

根据评价工作任务要求，编制数据质量评价方案，准备测试环境。数据质量评价方案包括但不限于如下内容：

a）数据评价的范围，包括数据表及数据质量维度等。多个应用领域的可拆分为对应的数据集；

b）采用的测试方法，包括自动化测试、人工测试；

c）采用的测试方式，包括全量测试、增量测试、抽样测试；

c）测试设计，包括数据规则定义、测试工具准备、质量维度权重等；

d）测试环境要求；

e）测试步骤；

f）测试记录要求；

g）测试结束或中止规则。

注3：自动化测试是使用数据质量测试评价工具实现自动化评价，人工测试是根据评价指标，结合专家专业判断进行数据质量检核。

注4：全量测试是对涉及的所有数据逐一进行数据质量检核，增量测试是对涉及的数据在特定的范围内和时间段内新增的数据逐一进行数据质量检核，抽样测试是按照抽样方案对抽取的数据逐一进行数据质量检核。

7.2 评价执行

应按照评价方案进行数据质量评价，准确记录测试过程及结果；
一旦开始测试，被测的数据、测试环境、数据质量维度不宜更改。

7.3 评价结果分析

7.3.1 计算数据表的数据质量得分

a）根据数据实际使用场景设定各二级指标权重，对不适用的指标权重可配置为0，加权求和得到各一级指标得分。

$$s_i = \sum_{j=1}^{n_i} d_j \times w_j$$

s_i是第i个一级指标的数据质量得分；
d_j是第i个一级指标下第j个二级指标的数据质量得分，n_i是第i个一级指标下二级指标的总数；
w_j是第j个二级指标的权重，满足$\sum_{j=1}^{n_i} w_j = 1$。

b）根据数据实际使用场景设置一级指标权重，加权求和确定数据表的数据质量得分。

$$S_k = \sum_{i=1}^{7} s_i \times W_i$$

S_k是第k个数据表的数据质量得分；
s_i是第i个一级指标的数据质量得分；
W_i是第i个一级指标的权重，满足$\sum_{i=1}^{7} W_i = 1$。

7.3.2 计算数据集的数据质量得分

根据数据实际使用场景确定各数据表重要程度及对应的调节系数，确定评价的数据集数据质量得分。
注5：对数据表重要程度进行评定时，调节系数可参考表8，不做强制要求，根据公共数据实际使用场景自行设定调节系数。

$$SC = \frac{\sum_{k=1}^{m} S_k \times a_k}{\sum_{k=1}^{m} a_k}$$

TC是数据集的数据质量得分；
m是该数据集中包含的数据表个数；
S_k是第k个数据表的数据质量得分；
a_k是第k个数据表的重要程度权重。

表 8 数据表重要程度和调节系数参考

重要程度	调节系数
核心	3
重要	2
一般	1

7.4 评价报告

编制数据质量评价报告，内容包括但不限于：
a）记录所有错误现象、错误描述、错误数量；
b）每个数据表数据质量得分；

c）各数据集数据质量得分；
d）测量结果分析、对数据质量的评价和建议。
评价报告的参考格式见附录A。

附 录 A

（资料性）

数据质量评价报告参考格式

表A1 数据质量评价报告样例

报告编号：　　评价日期：

数据来源单位			
数据评价单位			
被评价的数据集描述（数据集包含的数据表数量、类型等信息）			
数据质量总得分			
各数据表质量得分（可另附页）			
数据表编号	数据集名称		数据集得分
问题清单（可另附页）			
错误编号	错误现象	错误描述	所属数据表编号
评价结果分析			
改进措施建议			
评价人员签字			
备注			

参 考 文 献

[1] GB/T 36344—2018 信息技术 数据质量评价指标

[2] GB/T25000.24—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

[3] TR-REC-064 数据质量评测方法与指标体系

[4] DB52/T 1540.4—2021 政务数据 第4部分：数据质量评估规范

[5] 《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》