

ICS 点击此处添加 ICS 号
CCS 点击此处添加 CCS 号

团 体 标 准

T/CIA XXXX—XXXX

数据质量鉴证评价办法

Data Quality Attestation and Evaluation Methodology

(征求意见稿)

XXXX - XX - XX 发布

XXXX - XX - XX 实施

中国信息协会 发布

目 次

前 言	II
引 言	III
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 元数据 metadata	1
3.3 数据质量 data quality	1
3.4 数据质量鉴证 data quality assurance	1
3.5 数据标准 data standard	1
3.6 数据安全 data security	2
3.7 数据司法证明力 judicial probative force of data	2
4 鉴证评价指标构建原则	2
4.1 全面性原则	2
4.2 系统性原则	2
4.3 动态性原则	2
4.4 可比性原则	2
4.5 可操作性原则	2
4.6 权威性原则	2
5 鉴证评价指标框架	2
6 鉴证评价指标说明	3
6.1 规范性	3
6.2 完整性	4
6.3 准确性	4
6.4 一致性	4
6.5 时效性	4
6.6 安全性	4
6.7 权属性	5
7 鉴证评价流程	5
7.1 指标要求和测试方法	5
7.2 根据场景需求的加权计算	10
附 录 A (规范性) xxxxx 要求	11
附 录 B (资料性) XXXX 参考	12
参 考 文 献	13

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。

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专利的责任。

本文件由中国信息协会提出并归口。

本文件起草单位：国家信息中心、中福华会计师事务所有限责任公司、北京天融信网络安全技术有限公司、北京网御星云信息技术有限公司、北京市国信公证处（北京市知识产权公证服务中心）、国家授时中心、北京国际大数据交易所有限责任公司、贵阳大数据交易所有限责任公司、中央财经大学、南京乾恩软件有限公司

本文件主要起草人：魏连、王佳慧、毛立明、安根生、李萌、毛胤奇、王奇飞、沈余峰、韩明畅、陈嘉、关雨呈、李德新、郑嵩、华宇、寇无玷、王敏、温天宁、周源、胡钊、林存银、王伟、刘云松、薛伟成、宋欣明

引 言

大数据蕴藏着重要价值，而数据质量是价值实现的基础保障。低质量数据不仅无法产生有效洞见，更可能导致决策偏差。数据质量鉴证是确保数据合规合法、提升数据决策准确性、增加利益相关者信任以及保障数据司法效力的关键环节。本标准基于全面性、系统性、动态性、可比性和可操作性五大基本原则，构建包含规范性、完整性、准确性、一致性、时效性、安全性以及权属性七个维度的数据质量鉴证评价的指标，旨在实现对数据质量的科学、客观、全面鉴证。本数据质量鉴证评价办法建议应用于数据资产入表、数据价值评估、数据融资增信、数据合规审计、数据产品定价、数据流通交易以及数据司法证明等场景。

数据质量鉴证评价办法

1 范围

本标准作为 GB/T 36344-2018《信息技术 数据质量评价指标》的细化补充，规定了数据质量鉴证评价指标体系的构建原则、指标体系以及评价流程。

本标准适用于数据资产入表、数据价值评估、数据融资增信、数据合规审计、数据产品定价、数据流通交易以及数据司法证明等场景，为数据质量鉴证活动提供评价依据。

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 5271.1 信息技术 词汇 第1部分：基本术语
GB/T 37550—2019 电子商务数据资产评价指标体系
GB/T 36344—2018 信息技术 数据质量评价指标

3 术语和定义

下列术语和定义适用于本文件。

3.1 数据 data

信息的可再解释的信息化表示，以适用于通信、解释和处理。

注：可以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 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。注：可以通过人工或自动手段处理数据。

[GB/T 5271.17—2010,定义 17.06.05]

3.3 数据质量 data quality

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。

[GB/T 36344—2018]

3.4 数据质量鉴证 data quality assurance

数据质量鉴证是指通过系统方法和程序，对数据的安全性、真实性、合法性、关联性等进行检验和验证，以确保其满足司法证明力要求。

3.5 数据标准 data standard

数据的命名、定义、结构和取值规范方面的规则和基准。

3.6 数据安全 data security

为数据处理系统建立和采用的技术和管理的安全保护，保护计算机硬件、软件和数据不因偶然和恶意的原因遭到破坏、更改和泄露。

3.7 数据司法证明力 judicial probative force of data

电子数据作为电子证据在司法程序中进行举证时对于案件中待证事实的证明效果和作用。

4 鉴证评价指标构建原则

4.1 全面性原则

指标体系的构建应当覆盖数据质量的所有关键维度，确保能够对企业数据质量实施全方位的鉴证与评价。

4.2 系统性原则

各评价指标之间应当形成系统的逻辑关联，在结构上形成上层指标由下层具体指标构成的层级架构，在关系上保持各个指标相对独立性与内在关联性的统一，从而系统反映数据质量。

4.3 动态性原则

指标体系的构建应当具备动态适应性，可根据评价目的、计算方法和应用场景的变化，对指标体系进行相应调整，以满足鉴证需求。

4.4 可比性原则

评价指标的选取应当充分考虑不同企业或不同数据集之间的差异，确保评价结果能够实现主体维度的横向可比和时间维度的纵向可比。

4.5 可操作性原则

评价指标应当以定量指标为主，定性定量结合，确保具有明确的数据采集路径和计算方法，同时要求指标定义准确清晰以避免歧义，保证评价结果的可验证性。

4.6 权威性原则

评价指标的选取紧密结合国家颁布的数据相关法律法规以及行业标准，确保所选指标不仅能够全面、准确地反映数据质量的真实状况，而且具备法律上的合规性和权益性，保障指标体系的严谨性和公信力。

5 鉴证评价指标框架

数据质量鉴证评价指标框架见图1。

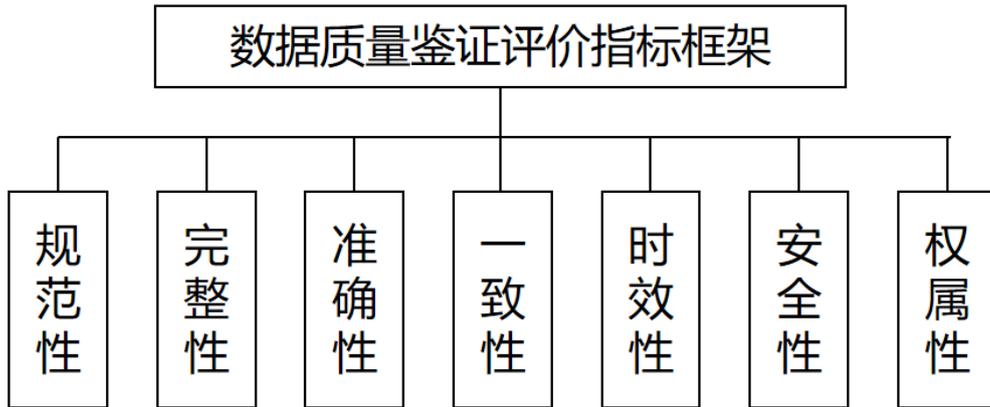


图1 数据质量鉴证评价指标框架

说明：

规范性——数据符合数据标准、元数据或业务规则的程度。

完整性——按照数据规则要求，数据元素被赋予数值的程度。

准确性——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——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——数据在时间变化中的正确程度。

安全性——数据免受未经授权的访问、泄露、篡改和破坏的安全防护程度。

权属性——数据具备持有权、加工使用权、产品经营权以及权属可追溯的程度

6 鉴证评价指标说明

从规范性、完整性、准确性、一致性、时效性、安全性以及权属性角度，分别罗列具体评价标准。

6.1 规范性

数据标准规范性、元数据规范性、业务规则规范性、组织架构规范性、分类分级规范性、安全合规规范性、司法证明材料规范性；

- 1) 数据标准规范性，参照[GB/T 36344—2018, 5.2 指标编号 0101]；
- 2) 元数据规范性，参照[GB/T 36344—2018, 5.2 指标编号 0103]；
- 3) 业务规则规范性，参照[GB/T 36344—2018, 5.2 指标编号 0104]；
- 4) 组织架构规范性：组织内部决策层、管理层、执行层、监督层是否设立专职数据管理部门；
- 5) 分类分级规范性：组织是否建立符合国家标准的数据分类分级管理制度，包括制定明确的分类标准、分级规则及配套操作流程，并确保在实际业务中有效执行；
- 6) 安全合规规范性：组织安全职能部门的设置有效性、数据安全总纲的全面性、全流程管控制度的执行度以及应急响应机制的实际保障能力。
- 7) 司法证明材料规范性：是否符合电子数据司法鉴定的检材要求，是否符合公证机构办理数据证明业务的材料要求。

6.2 完整性

数据元素完整性、数据记录完整性、数据完整性证明力；

- 1) 数据元素完整性，参照[GB/T 36344—2018, 5.3 指标编号 0201]；
- 2) 数据记录完整性，参照[GB/T 36344—2018, 5.3 指标编号 0202]；
- 3) 数据完整性证明力，需通过第三方区块链存证或公证鉴定机构哈希比对完整性校验，使数据具备完整性的司法证明力。

6.3 准确性

数据内容正确性、数据格式合规性、数据唯一性、脏数据出现率、数据计算科学性；

- 1) 数据内容正确性：参照[GB/T 36344—2018, 5.4 指标编号 0301]；
- 2) 数据格式合规性，参照[GB/T 36344—2018, 5.4 指标编号 0302]；
- 3) 数据唯一性，参照[GB/T 36344—2018, 5.4 指标编号 0304]；
- 4) 脏数据出现率，参照[GB/T 36344—2018, 5.4 指标编号 0305]；
- 5) 数据计算科学性，数据算法是否准确可靠。

6.4 一致性

相同数据一致性、关联数据一致性、数据记录一致性；

- 1) 相同数据一致性，参照[GB/T 36344—2018, 5.5 指标编号 0401]；
- 2) 关联数据一致性，参照[GB/T 36344—2018, 5.5 指标编号 0402]；
- 3) 数据记录一致性，主要考察单条数据记录内部各字段之间的逻辑正确性、格式合规性以及业务规则符合性。

6.5 时效性

基于时间段的正确性、基于时间点的及时性、时序性、更新频率以及可验证性；

- 1) 基于时间段的正确性，参照[GB/T 36344—2018, 5.6 指标编号 0501]；
- 2) 基于时间点的及时性，参照[GB/T 36344—2018, 5.6 指标编号 0502]；
- 3) 时序性，参照[GB/T 36344—2018, 5.6 指标编号 0503]；
- 4) 更新频率，数据更新频率是否合理，是否符合应用场景需要和成本收益；
- 5) 可验证性。数据与中国标准时间的关联性和一致性可以被溯源验证，使数据具备时效性的司法证明力。

6.6 安全性

数据采集安全性、数据传输安全性、数据存储安全性、数据处理安全性、数据交换安全性、数据销毁安全性；

- 1) 数据采集安全性，在收集数据和产生新数据的过程中，是否采取数据源鉴别、数据源安全认证、数据分类分级标记等安全措施，确保所采集数据的准确、完整、合法、合规；

- 2) 数据传输安全性，数据从一个实体位置到另一个位置的过程中，是否采取安全的传输协议和认证机制等安全措施，确保敏感数据的机密性、完整性和可用性得到保护；
- 3) 数据存储安全性，数据在组织内部以任何数字格式被保存和存储的过程中，是否采取数据加密、数据防泄漏、数据访问控制等安全措施，保护敏感数据不被未经授权访问、篡改、泄露或丢失；
- 4) 数据处理安全性，数据在组织内部进行计算、分析、可视化等操作的过程中，是否采取身份鉴别、访问控制和安全计算环境等安全措施，确保敏感数据的机密性、完整性和可用性得到保护；
- 5) 数据交换安全性，数据在组织与组织或个人进行交换的过程中，是否采取数据监测、数据审计、数据管控等安全措施，确保敏感数据不被恶意获取、不被盗用；
- 6) 数据销毁安全性，数据在删除的过程中，是否采取不可逆的处置措施，确保数据彻底删除且无法通过任何手段恢复。

6.7 权属性

数据权属证明程度、数据权属可追溯性：

- 1) 数据权属证明程度，是否取得政府批准的公共数据登记机构、数据交易所颁发的数据登记证书，公证处、电子数据司法鉴定中心颁发的数据司法存证函，知识产权管理部门颁发的数据知识产权登记证书，或其他具备司法证明力的数据权属证明；
- 2) 数据权属可追溯性，数据生成及应用链路的可追溯性和可验证性，数据与生成数据系统的关联性和对应性。

7 鉴证评价流程

数据质量鉴证评价应当按照以下流程：依据既定评价指标体系和测试方法，采用 0-100 分制对各指标逐项评分；采用“基准权重约束、场景动态调整”的原则确定各指标权重，权重总和为 100%；最终通过加权计算获得综合评分（总分 100 分制）。

7.1 指标要求和测试方法

6.1.1 规范性评价指标

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
------	------	------	------

0101	数据标准规范性	数据符合数据标准的度量。 注 1:评价数据质量时需要收集数据在命名、创建、定义、更新和归档时遵循的标准,包括国际标准、国家标准、行业标准、地方标准或相关规定等。 注 2:和数据归档一样甚至更重要,在一个完整的数据规则中旧数据的销毁一般也有一个比较详细且具有可执行性的规定。	$X=A/B$ 式中: A=满足数据标准要求的数据集中元素的个数; B=被评价的数据集中元素的个数
0102	元数据规范性	数据符合元数据定义的度量。 注:元数据标注、描述或刻画其他数据、以使检索、或使用信息更容易。评价数据质量时需要检查是否提供可解读的元数据文档。 示例:包含各字段名称、描述、类型、值域等内容的数据字典为一种元数据文档。	$X=A/B$ 式中: A=满足元数据定义的数据集中元素的个数; B=被评价的数据集中元素的个数
0103	业务规则规范性	数据符合业务规则的度量。 注 1:业务规则是一种权威性原则或指导方针,用来描述业务交互,并建立行动和数据行为结果及完整性的规则。 注 2:评价数据质量时需要检查是否存在良好归档的业务规则。	$X=A/B$ 式中: A=满足业务规则的数据集中元素的个数; B=被评价的数据集中元素的个数
0103	组织架构规范性	组织内部决策层、管理层、执行层、监督层是否设立专职数据管理部门。	$X=A/B$ 式中: A=满足组织架构规范的数据集中元素的个数; B=被评价的数据集中元素的个数
0104	分类分级规范性	组织内部建设数据分类分级规范。	$X=A/B$ 式中: A=满足分类分级规范的数据集中元素的个数; B=被评价的数据集中元素的个数
0105	安全合规规范性	组织是否具备安全职能部门设置制度、数据安全总纲制度、数据全流程管控制度及数据应急响应制度。	$X=A/B$ 式中: A=满足安全合规规范的数据集中元素的个数; B=被评价的数据集中元素的个数
0106	司法证明材料规范性	是否符合电子数据司法鉴定的检材要求,是否符合公证机构办理数据证明业务的材料要求。	$X=A/B$ 式中: A=满足司法证明要求的数据集中元素的个数; B=被评价的数据集中元素的个数

6.1.2 完整性评价指标

完整性评价指标定义见表 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
0201	数据元素完整性	按照业务规则要求，数据集中应被赋值的数据元素的赋值程度。	$X=A/B$ 式中： A=被赋值的数据集中元素的个数； B=预期被赋值的数据集中元素的个数
0202	数据记录完整性	按照业务规则要求，数据集中应被赋值的数据记录的赋值程度。	$X=A/B$ 式中： A=被赋值的数据集中元素的个数； B=预期被赋值的数据集中元素的个数
0203	数据完整性证明力	通过区块链存证或第三方公证机构验证哈希值一致性的程度，使数据具备完整性的司法证明力。	$X=A/B$ 式中： A=被赋值的数据集中元素的个数； B=预期被赋值的数据集中元素的个数

6.1.3 准确性评价指标

准确性评价指标定义见表 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
0301	数据内容正确性	数据内容是否是预期数据。	$X=A/B$ 式中： A=满足数据正确性要求的数据集中元素的个数； B=被评价的数据集中元素的个数
0302	数据格式合规性	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。 示例：性别一栏不能出现男/女以外的内容；身份证号不能出现标点符号；以及对字符编码的一些限制，都需要通过规定内容的格式来实现。	$X=A/B$ 式中： A=满足格式要求的数据集中元素的个数； B=被评价的数据集中元素的个数
0303	数据唯一性	正确字段、记录、文件或数据集之外无效数据的度量。	$X=A/B$ 式中： A=满足唯一性要求的数据集中元素的个数； B=被评价的数据集中元素的个数
0304	脏数据出现率	正确字段、记录、文件或数据集之外无效数据的度量。	$X=A/B$ 式中： A=有脏数据出现的数据集中元素的个数； B=被评价的数据集中元素的个数

0305	数据计算科学性	数据算法是否准确可靠。	$X=A/B$ 式中： A=满足数据计算科学性要求的数据集中元素的个数； B=被评价的数据集中元素的个数
------	---------	-------------	---------------------------------------------------------------

6.1.4 一致性评价指标

一致性评价指标定义见表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
0401	相同数据一致性	同一数据在不同位置存储或被不同应用或用户使用时，数据的一致性；数据发生变化时，存储在不同位置的同一数据被同步修改。	$X=A/B$ 式中： A=满足一致性要求的数据集中元素的个数； B=被评价的数据集中元素的个数
0402	关联数据一致性	根据一致性约束规则检查关联数据的一致性。	$X=A/B$ 式中： A=满足一致性要求的数据集中元素的个数； B=被评价的数据集中元素的个数
0403	数据记录一致性	单条数据记录内部各字段之间的逻辑正确性、格式合规性以及业务规则符合性。	$X=A/B$ 式中： A=满足一致性要求的数据集中元素的个数； B=被评价的数据集中元素的个数

6.1.5 时效性评价指标

时效性评价指标定义见表5。

表5 时效性评价指标

指标编号	指标名称	指标描述	计算方法
0501	基于时间段的正确性	基于日期范围的记录数或频率分布符合业务需求的程度。	$X=A/B$ 式中： A=满足有效性要求的数据集中元素的个数； B=被评价的数据集中元素的个数
0502	基于时间点的及时性	基于时间戳的记录数、频率分布或延迟时间符合业务需求的程度。	$X=A/B$ 式中： A=满足及时性要求的数据集中元素的个数； B=被评价的数据集中元素的个数
0503	时序性	数据集中同一实体的数据元素之间的相对时序关系。	$X=A/B$ 式中： A=满足时序性要求的数据集中元素的个数； B=被评价的数据集中元素的个数

0504	更新频率	数据更新频率是否合理，是否符合应用场景需要和成本收益。	$X=A/B$ 式中： A=满足更新频率要求的数据集中元素的个数； B=被评价的数据集中元素的个数
0505	可验证性	数据与中国标准时间的关联性和一致性可以被溯源验证，使数据具备时效性的司法证明力。	$X=A/B$ 式中： A=满足可验证性要求的数据集中元素的个数； B=被评价的数据集中元素的个数

6.1.6 安全性评价指标

安全性评价指标定义见表6。

表6 安全性评价指标

指标编号	指标名称	指标描述	计算方法
0601	数据采集安全性	组织采取对数据源鉴别、数据源安全认证、数据分类分级标记等安全措施，确保所采集数据的准确、完整、合法、合规。	$X=A/B$ 式中： A=满足数据采集安全性要求的数据集中元素的个数； B=被评价的数据集中元素的个数
0602	数据传输安全性	组织采取安全的传输协议和认证机制等安全措施，确保敏感数据的机密性、完整性和可用性得到保护。	$X=A/B$ 式中： A=满足数据传输安全性要求的数据集中元素的个数； B=被评价的数据集中元素的个数
0603	数据存储安全性	组织采取数据加密、数据防泄漏、数据访问控制等安全措施，确保数据存储安全。	$X=A/B$ 式中： A=满足数据存储安全性要求的数据集中元素的个数； B=被评价的数据集中元素的个数
0604	数据处理安全性	组织采取身份鉴别、访问控制和安全计算环境等安全措施，确保敏感数据的机密性、完整性和可用性得到保护。	$X=A/B$ 式中： A=满足数据处理安全性要求的数据集中元素的个数； B=被评价的数据集中元素的个数
0605	数据交换安全性	组织采取数据监测、数据审计、数据管控等安全措施，确保敏感数据不被恶意获取、不被盗用。	$X=A/B$ 式中： A=满足数据交换安全性要求的数据集中元素的个数； B=被评价的数据集中元素的个数
0606	数据销毁安全性	组织采取不可逆的数据销毁技术，确保数据销毁安全。	$X=A/B$ 式中： A=满足数据销毁安全性要求的数据集中元素的个数； B=被评价的数据集中元素的个数

6.1.7 权属性评价指标

权属性评价指标定义见表7。

表 7 权属性评价指标

指标编号	指标名称	指标描述	计算方法
0701	数据权属证明程度	是否取得政府批准的公共数据登记机构、数据交易所颁发的数据登记证书，公证处、电子数据司法鉴定中心颁发的数据司法存证函，知识产权管理部门颁发的数据知识产权登记证书，或其他具备司法证明力的数据权属证明。	$X=A/B$ 式中： A=满足权属证明要求的数据集中元素的个数； B=被评价的数据集中元素的个数
0702	数据权属可追溯性	数据生成及应用链路的可追溯性与可验证性。 示例：追溯技术的先进性，是否采用了先进的数据权属追溯技术手段，如区块链、哈希算法等，追溯数据的采集、存储、处理、分析、管控等过程以及流转关系。	$X=A/B$ 式中： A=满足权属可追溯性要求的数据集中元素的个数； B=被评价的数据集中元素的个数

7.2 根据场景需求的加权计算

本标准在不同场景需求下，采用“基准权重约束、场景动态调整”的原则，通过评议确定各指标权重，权重总和为 100%。七项指标基准权重系数均为 10%，剩余 30%的权重依据场景进行动态分配，最后计算数据质量的加权总分，总分为 100 分。具体见表 8。

表 8 数据质量鉴证评价指标加权计算表

序号	指标名称	权重系数（单位：%）	单行得分	加权总分
01	规范性			
02	完整性			
03	准确性			
04	一致性			
05	时效性			
06	安全性			
07	权属性			

注：同一评价目的应使用统一权重方案，特殊场景调整需专家评审确认，以保证评价结果的可比性。

附录 A
(规范性)
XXXX 要求

附录 B
(资料性)
XXXX 参考

参 考 文 献

- [1] 《电子商务数据资产评估指标体系》（GB/T 37550—2019）
 - [2] 《信息技术 数据质量评价指标》（GB/T 36344—2018）
 - [3] 中国资产评估协会《数据资产评估指导意见》
 - [4] 财政部《企业数据资源相关会计处理暂行规定》（财会〔2023〕11号）
 - [5] 电子数据司法鉴定通用实施规范
 - [6] 《中华人民共和国个人信息保护法》（2021年11月1日施行）
 - [7] 《中华人民共和国网络安全法》（2017年6月1日施行）
 - [8] 《中华人民共和国数据安全法》（2021年9月1日施行）
-