

ICS

CCS 点击此处添加 CCS 号

DB 11

北京市地方标准

DB XX/T XXXX—XXXX

数据资产 质量评估要求

Data asset — Requirements for quality assessment

（征求意见稿）

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。

XXXX – XX – XX 发布

XXXX – XX – XX 实施

北京市市场监督管理局 发布

目 次

前言 II

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 总体要求 2

5 评估指标及方法 2

6 评估流程 3

7 数据资产质量评估管理 4

8 评估结果应用效果反馈 6

附录 A（资料性） 数据资产质量评估二级指标说明及计算方式 7

附录 B（资料性） 数据抽样方法操作方式及适用条件说明 9

附录 C（资料性） 数据资产质量评估报告框架 10

参考文献 11

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。

本文件由北京市政务服务和数据管理局、北京市经济和信息化局提出。

本文件由北京市政务服务和数据管理局、北京市经济和信息化局归口管理并组织实施。

本文件起草单位：北京国际大数据交易所有限责任公司。

本文件主要起草人：

数据资产 质量评估要求

1 范围

本文件规定了数据资产质量评估的范围、原则和流程，评估指标及方法、评估管理、评估结果应用等要求。

本文件适用于各行业结构化数据资产质量评估工作。

2 规范性引用文件

本文件没有规范性引用文件。

3 术语和定义

下列术语和定义适用于本文件。

3.1

结构化数据资产 Structured data asset

特定主体合法拥有或者控制的，以结构化数据为载体和表现形式的，能进行货币计量的，且能带来经济利益或社会效益的数据资源。

[来源：国家数据局 数据领域常用名词解释（第二批），有修改]

3.2 数据资产质量 data quality

在指定条件下使用时，数据资产的特性满足明确的和隐含的要求的程度。

[来源：GB/T 25000.12-2017，4.3，有修改]

3.3

质量规则 quality rule

对数据质量进行检查、分析、计算的方法和说明。包括但不限于检测规则、问题统计规则、权重规则、指标计算（计算指标、统计指标、维度指标、总得分、汇总等）。

3.4

数据模型 data model

一种抽象的、逻辑性的表示，用于描述数据的结构、关系、约束和操作，数据模型为数据的存储、查询和操作提供了一种组织方式，同时也为数据分析和应用开发提供了基础。

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。

[来源：GB/T 36344-2018，2.8]

3.6

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。

[来源：GB/T 35295-2017，2.2.7]

3.7

数据资产质量评估机构 data asset quality assessment organization

提供数据资产质量评估服务的机构，其主要职责是对数据资产进行全面、客观、准确的质量评估。

4 总体要求

- 4.1 数据资产质量评估的范围为结构化数据资产。
- 4.2 数据资产质量评估应基于事实和数据，避免主观偏见，全面覆盖数据资产的各维度数据质量指标。
- 4.3 数据资产质量评估是一个持续的过程，应随数据资产的变化而进行相应的调整和更新。
- 4.4 评估机构应具有数据资产评估的专业能力和经验；独立于数据交易双方，以保证评估结果的客观性。

5 评估指标及方法

5.1 指标框架

数据资产质量评估指标框架宜参照GB/T 36344建立，指标框架图见图1，包括规范性、准确性、完整性、唯一性、一致性、时效性、可访问性7个一级指标以及17个二级指标。其中，规范性是指数据资产符合数据标准、数据模型、业务规则、元数据或权威参考数据的程度；准确性是指数据资产准确表示其所描述的真实实体（实际对象）真实值的程度；完整性是指按照数据资产规则要求，数据元素被赋予数值的程度；唯一性是指数据资产唯一标识程度；一致性是指数据资产与其他特定上下文中使用的数据无矛盾的程度；时效性是指数据资产在时间变化中的正确程度；可访问性是指数据资产能被访问的程度。数据资产质量评估二级指标说明及计算方式见附录A。

除上述通用指标外，评估机构及被评估方可根据具体场景增加可选指标，如在数据资产内部使用的应用场景下可补充管理信息完整性、应用场景完整性等指标，在数据资产对外应用的场景下可采用调用成功率、数据来源权威性、采集方式、流通交易情况、数据资产登记情况、投资收益率稳定性等指标。

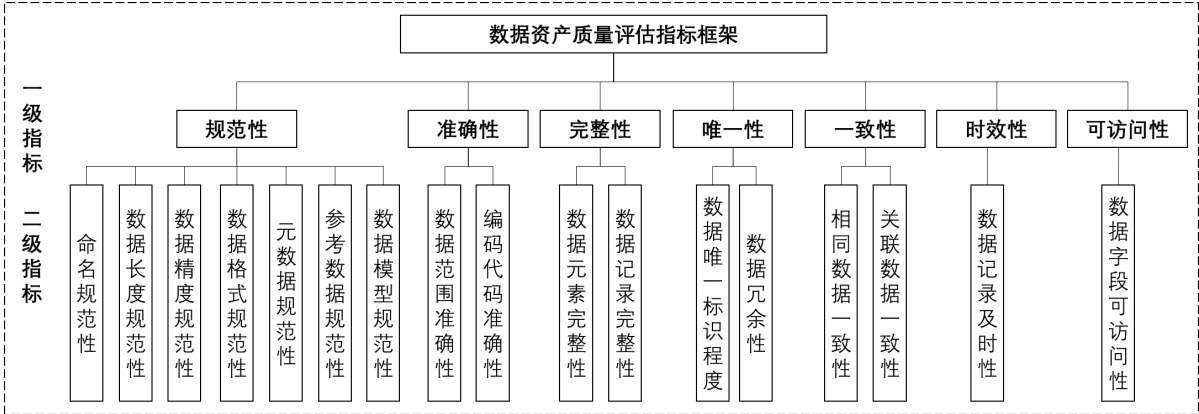


图1 数据资产质量评估指标框架

5.2 评估方法

评估方法如下：

- a) 应按照选定的评估指标及质量规则，结合系统工具评估与人工评估的方法开展数据资产质量评估。其中系统工具评估通过在质量平台等系统工具上部署质量规则进行自动化评估；

- b) 人工检测通过材料审核、人员访谈、自定义检测语句等方式，对相应的数据资产质量指标和评估规则进行比对并客观记录，结合专业知识及行业数据流通的特点进行综合评估；
- c) 根据委评数据资产的数据规模采取全量检测或抽样检测的方式，全量检测对委评数据资产进行全量的数据资产质量评估，抽样检测采用随机抽样、系统抽样、分层抽样等抽样方法对委评数据资产进行抽样，对抽样数据资产进行质量评估，各类抽样方法的操作方式及适用条件详见附录 B；
- d) 为保证数据资产质量检测工作不影响数据生产和应用环境，并保证检测工作的顺利开展，一般需要对检测的数据进行备库。

5.3 分值计算

数据资产质量总体评分通过各数据资产质量指标分值逐级加权平均计算得出。首先根据预先选取的指标和规则对评估对象的数据资产质量进行检测，计算各二级指标的数据资产质量分数 P_{ij} ，其次根据评估指标框架，按照公式（1），通过计算二级指标加权平均得出一级指标的数据资产质量分数 P_i ，最后按照公式（2），将一级指标分值通过加权平均，得出数据资产质量总体评分 P 。公式（3）（4）所示各一级指标及二级指标数据资产质量分数权重之和为1，数据资产质量各级指标及总体评分满分均为100分，权重的设置可以根据评估目的以及评估对象特点进行调整。

公式：

$$P_i = \sum_{j=1}^m P_{ij} X_j, P_{ij} \in [0,100] \quad (1)$$

$$P = \sum_{i=1}^n P_i X_i, P_i \in [0,100] \quad (2)$$

$$\sum_{j=1}^m X_j = 1 \quad (3)$$

$$\sum_{i=1}^n X_i = 1 \quad (4)$$

式中：

P ——评估对象的数据资产质量总分；

P_i ——第 i 个一级指标的数据资产质量分数；

P_{ij} ——第 i 个一级指标下的第 j 个二级指标的数据资产质量分数；

m ——第 i 个一级指标下的二级指标总数；

n ——一级指标总数；

j —— j 个二级指标；

i ——第 i 个一级指标；

X_j ——第 j 个二级指标数据资产质量分数权重；

X_i ——第 i 个一级指标数据资产质量分数权重。

6 评估流程

6.1 评估流程图

数据资产质量评估流程如图2所示，包括评估准备、评估实施、报告撰写、结果评审4个主要阶段。

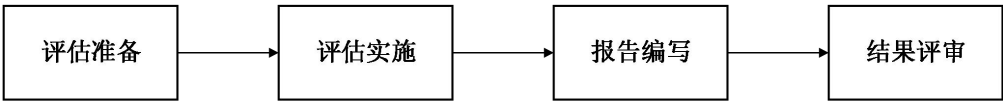


图 2 数据资产质量评估流程图

6.2 评估准备

在评估准备阶段，需要被评估方和评估机构展开调研，确定委评数据，了解评估目的、评估范围（时间阶段、数据项等），制定数据资产质量评估方案，选取合适的评估方式、评估指标及质量规则，配置基于数据安全保护和便于任务作业的相应工具部署环境。

- a) 确定委评数据：形成委评字段范围、数据集时间跨度、数据规模总量（条数、项数、记录数）的基本信息描述；
- b) 了解评估目的：识别被评估方的评估目的，包括但不限于登记确权、估值入表等；
- c) 校正评估范围：对委评数据进行校正，明确本次评估的数据系统、数据表、数据项等信息；
- d) 工具部署方案：基于委评单位的数据保密等级、行业及地方、国家相关数据安全保护要求，从硬件、软件环境、人员要求、工具要求等方面，形成数据资产质量检测与评估的工具部署方案。

6.3 评估实施

在评估实施阶段，被评估方应根据评估机构确定的数据资产质量评估方案，通过在质量平台等系统工具上部署质量规则或在数据库中自定义检测语句进行机器检测，并结合材料审核、人员访谈等人工检测方式，由评估机构对评估对象开展数据资产质量评估工作。

6.4 报告撰写

在报告撰写阶段，评估机构根据评估结果计算数据资产质量分数并编制数据资产质量评估报告，报告内容宜包括但不限于评估对象、评估范围、评估方法、评估指标及计算方法、评估实施过程、评估结果、质量问题分析等。

6.5 结果评审

在结果评审阶段，数据资产质量评估报告经被评估方确认后，由评估机构进行评审、备案及发布。

7 数据资产质量评估管理

7.1 数据资产质量评估机构管理

7.1.1 制度和规程要求

评估机构制度和规程要求如下：

- a) 应制定数据资产质量评估管理策略，说明数据资产总体目标、范围、原则和评估框架等；
- b) 应建立数据资产质量评估管理制度，包括但不限于数据资产质量评估安全管理制度，数据资产质量评估公证管理制度，数据资产质量评估流程规范，数据资产质量分级制度等；
- c) 应建立严格的数据资产质量评估流程，配备评估工具、设定评估方法、完善评估结果、进行结果复核；
- d) 应为数据资产质量评估人员或操作人员执行的管理或业务操作建立操作规程；

- e) 应定期对数据资产质量评估管理策略、制度和规程进行评审，并及时进行更新；
- f) 未经授权不得擅自留存及使用数据资产拥有方的数据或数据衍生品；
- g) 除质量评估结果规定输出外任何其他数据拥有方信息不得对外披露。

7.1.2 人员管理要求

评估机构人员管理要求如下：

- a) 应建立数据资产质量评估领导小组，由机构最高管理者或授权代表担任组长；
- a) 应建立数据资产质量评估委员会，设立委员会主任，对评估结构进行最后审核；
- b) 应设立数据资产质量评估专员、数据资产质量调研员等岗位，定义各个岗位职责，并配备一定数量的岗位人员；
- c) 应对数据资产质量评估重要岗位人员进行审查和技术考核，确保无违法违规记录；
- d) 应对数据资产质量各类岗位人员制定和实施培训计划，培训内容包括安全意识、专项技能等，具备与岗位要求相适应的质量评估管理知识和专业技术水平。

7.2 数据资产质量评估工作管理

7.2.1 管理要求

数据资产质量评估工作管理要求如下：

- a) 评估机构及其评估专业人员开展数据资产质量评估工作，应独立进行分析和评估并形成专业意见，不得直接以预先设定的质量结果作为评估结论；
- b) 应关注数据资产的安全性和合法性，并遵守保密原则；
- c) 应合理使用评估假设和限制条件；
- d) 应采用适当方法考虑企业数据资产质量管理能力对数据资产质量未来发展趋势的影响，合理评估数据资产质量。

7.2.2 数据资产质量检查

数据资产质量检查要求如下：

- a) 根据数据资产质量评估体系中的有关指标、评估方法与流程，规范数据资产质量评估机构及其评估专业人员在数据资产质量评估工作中的实务操作；
- b) 对评估对象进行针对性地现场调查，资料收集，并进行核查验证、分析整理和记录；
- c) 对数据资产质量评估情况进行实时监控，实现数据资产评估监督管理。

7.2.3 数据资产质量分析

数据资产质量分析要求如下：

- a) 应分析数据资产的规范性、准确性、完整性、唯一性、一致性、时效性和可访问性，并找出影响数据资产质量的原因；
- b) 宜执行初始数据资产质量评估，通过评估结果确认问题及优先级。

7.2.4 数据资产质量评估报告

数据资产质量评估报告要求如下：

- a) 应支持展示每类数据资产的详细评估内容；

- b) 应支持通过多维度的评估报告展示质量检测总体结果；
- c) 应支持评估工作日志追踪，在必要时进行有效调整；
- d) 评估报告内容应包含但不限于评估对象、评估目的、评估方法、评估基准日、评估范围、评估结果、应用场景和资产质量评估报告使用人，报告框架见附件 C。

8 评估结果应用效果反馈

数据资产质量评估结果应用效果反馈包括以下方面：

- a) 登记信息完善：数据资产质量评估结果应作为登记信息的补充依据，用于更新和完善数据资产登记信息，将评估发现的质量问题与登记信息中的相关数据项建立关联，客观反映数据资产质量的整体状况；
- b) 价值评估支撑：评估结果应对数据资产价值评估形成有效支撑，明确质量要素作为价值评估参考指标或调节参数的实际应用效果；
- c) 质量提升成效：应体现数据资产质量提升的实际效果，包括但不限于业务决策可信度的提升幅度、业务流程效率的提升比例以及风险成本降低的量化指标；
- d) 管理成本优化：应反映数据资产管理成本的降低情况，包括自动化质量稽核任务实施后的效率提升数据和管理成本节约数据；
- e) 资源配置改进：应说明基于评估结果对数据管理资源配置的优化情况，包括资源重新分配后产生的实际效益提升数据；
- f) 资产化推进效果：应展示数据资产化进程的推动成效，提供数据赋能业务的具体应用案例及量化的实施效果。

附录 A
(资料性)

数据资产质量评估二级指标说明及计算方式

数据资产质量评估指标说明及计算方式详见表A. 1

表 A. 1 数据资产质量评估指标说明及计算方式

一级指标	二级指标	指标描述	计算方式
规范性	命名规范性	表或者数据项命名的规范性	$X=A/B$ A: 满足数据命名规范要求的数据项数量 B: 被评价的所有数据项个数
	数据长度规范性	数据长度要求: 即对数据长度的约束;	$X=A/B$ A: 满足数据长度要求的数据元素数量 B: 被评价的所有数据元素个数
	数据精度规范性	数据精度要求: 即对数据精度的约束;	$X=A/B$ A: 满足数据精度要求的数据元素数量 B: 被评价的所有数据元素个数
	数据格式规范性	数据格式要求: 即对数据中各位取值的约束, 如日期的格式	$X=A/B$ A: 满足数据格式要求的数据元素数量 B: 被评价的所有数据元素个数
	元数据规范性	表或者数据项相关元数据的规范性约束	$X=A/B$ A: 满足元数据规范要求的表/数据元素数量 B: 被评价的所有表/数据元素个数
	参考数据规范性	数据项参考数据的规范性约束 其中参考数据是指该数据项的权威参考来源, 包括国际标准、国家标准、行业标准等	$X=A/B$ A: 满足参考数据规范要求的的数据项数量 B: 被评价的所有数据项个数
	数据模型规范性	表或者数据元素符合数据模型的度量	$X=A/B$ A: 满足数据模型规范要求的表/数据元素数量 B: 被评价的所有表/数据元素个数
准确性	数据范围准确性	数据的值域约束, 即数据的取值应在其值域范围内	$X=A/B$ A: 满足数据范围要求数据元素数量

		(具有业务意义的连续范围)	B: 被评价的所有数据元素个数
	编码/代码准确性	编码取值约束, 编码要符合数据标准规定的编码规则。	$X=A/B$ A: 满足编码/代码格式要求数据元素数量 B: 被评价的所有数据元素个数
完整性	数据元素完整性	数据集中应被赋值的数据元素比例	$X=A/B$ A: 被赋值的数据元素数量 B: 被评价的所有数据元素个数
	数据记录完整性	数据集中应被赋值的数据记录比例	$X=A/B$ A: 被赋值的数据记录数量 B: 被评价的所有数据记录个数
唯一性	数据唯一标识的程度	数据的业务唯一性约束, 即一组业务主键在表内保持唯一。 【示例】: 一个客户代码只能对应一条客户信息记录。	$X=A/B$ A: 具有唯一主键的数据记录数量 B: 被评价的所有数据记录个数
	数据冗余性	度量数据项或数据记录的冗余程度	$X=A/B$ A: 冗余的数据项或者数据记录数量 B: 被评价的所有数据项或者数据记录数量
一致性	相同数据一致性	同一数据在不同存储位置或者存储时间的一致性	$X=A/B$ A: 满足一致性要求的数据元素数量 B: 被评价的所有数据元素个数
	关联数据一致性	不同数据对于约束规则的一致性	$X=A/B$ A: 满足一致性要求的数据元素数量 B: 被评价的所有数据元素个数
时效性	数据记录及时性	数据记录满足及时性要求的程度	$X=A/B$ A: 满足及时性要求的数据记录数量 B: 被评价的所有数据记录个数
可访问性	数据字段可访问性	可访问的数据字段的数量与总数的比例	$X=A/B$ A: 满足可访问性要求的数据字段数量 B: 被评价的所有数据字段个数

附录 B
(资料性)

数据抽样方法操作方式及适用条件说明

数据抽样方法操作方式及适用条件说明详见表B. 1。

表 B. 1 数据抽样方法操作方式及适用条件说明

抽样方法	操作方式	适用条件
随机抽样(单纯随机抽样、简单随机抽样)	利用随机数表、随机数骰子或计算机产生的随机数进行抽样。从总体为 N 个个体的被评估单位中抽取 n 个个体作为样本。	适用于总体中个体数量有限以及个体差异不明显的被评估单位。
系统抽样(机械抽样、等距抽样)	将总体被评估单位分成均衡的几个部分,按照预先定出的规则,从每一部分抽取一个个体,得到所需要的样本。假设要从容量为 N 的总体中抽取容量为 n 的样本,具体步骤如下: (1) 先将被评估单位总体的 N 个个体编号。有时可直接利用数据项个体自身所带的号码,如 ID 等; (2) 确定分段间隔 k,对编号进行分段,当 $\frac{N}{n}$ (n 是样本容量)是整数时,取 $k = \frac{N}{n}$; (3) 在第一段用随机抽样确定第一个个体编号 1 (1 ≤ k); (4) 按照一定的规则抽取样本。通常是将 1 加上间隔 k 得到第 2 个个体编号 (1 + k),再加 k 得到第 3 个个体编号 (1 + 2k),依次进行下去,直到获取整个样本。	适用于总体中的个体数较多的被评估单位。
分层抽样	将所有样本分成多个部分之后,再从每个部分按一定方法抽取部分个体得到所需样本,具体步骤如下: (1) 将样本按照突出的分类特征分成不同的类别,记录下每一类的样本数量,可采用数据分级分类情况作为分类参考。 (2) 确定每个类别的数据量占总体数据量的比例,利用这个比例计算出样本中每类应抽取的数据量。 (3) 采用简单随机抽样或系统抽样方法从每个类别中抽取独立随机样本。	当总体是由差异明显的几部分组成时,选择分层抽样的方法可以降低总的抽样误差。

附 录 C
(资料性)
数据资产质量评估报告框架

- 一、评估对象介绍
 - (一)评估对象基本情况
 - (二)数据资产情况
 - (三)数据资产管理情况
- 二、评估过程
 - (一)评估目标
 - (二)评估依据
 - (三)评估指标
 - (四)评估方法
 - (五)评估过程
 - (六)评估结果
- 三、评估结论和质量提升建议
 - (一)评估结果
 - (二)问题分析
 - (三)质量提升建议
 - (四)评估结论

参 考 文 献

- [1] GB/T 25000.12—2022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
 - [2] GB/T 35295—2017信息技术 大数据 术语
 - [3] GB/T 36344—2018信息技术 数据质量评价指标
 - [4] 《中华人民共和国数据安全法》
 - [5] 国家数据局 数据领域常用名词解释（第一批）
 - [6] 国家数据局 数据领域常用名词解释（第二批）
-