

数据交易 第3部分：数据产品质量评估规范

Data exchange Part 3: The quality evaluation specification for data product

(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

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
4 评估原则.....	2
5 评估框架.....	2
5.1 质量模型.....	2
5.2 指标项组成.....	3
6 指标体系.....	3
6.1 基础质量指标.....	3
6.2 技术使用指标.....	6
6.3 市场应用指标.....	7
7 评估方法.....	8
7.1 评估方法类型.....	8
7.2 定量评估.....	8
7.3 定性评估.....	8
7.4 结果评分.....	9
8 评估流程.....	9
8.1 流程模型.....	9
8.2 明确评估主体.....	10
8.3 设计指标方案.....	10
8.4 实施质量评估.....	11
8.5 提升产品质量.....	11
8.6 数据产品运营.....	11
8.7 质量需求变更.....	11
8.8 质量评估完成.....	11
附 录 A （资料性） 数据产品质量评估报告编制要求.....	12
A.1 总体要求.....	12
A.2 首页.....	12
A.3 正文.....	12
A.4 附录.....	12
参考文献.....	13

前 言

《数据交易》分为下列6个部分：

- 第1部分：数据流通交易合规指南；
- 第2部分：数据产品权益认定指南；
- 第3部分：数据产品质量评估规范；
- 第4部分：数据资产评估规范；
- 第5部分：数据产品定价方法；
- 第6部分：数据产品可信交付技术要求与评估规范。

本文件为DB31/T XXXX的第3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专利的责任。

本文件由上海市经济和信息化委员会提出并组织实施。

本文件由上海市信息标准化技术委员会归口。

本文件起草单位：

上海数据交易所有限公司、中远海运科技股份有限公司、上海质量管理科学研究院有限公司、中国建筑第八工程局有限公司、上海计算机软件技术开发中心、上海富数科技有限公司、上海浦东软件平台有限公司、蚂蚁科技集团股份有限公司、上海芯化和云数据科技有限公司、北京市中闻（上海）律师事务所、复旦大学、上海商学院、上海华东电信研究院、上海市软件评测中心有限公司、欧冶云商股份有限公司、国网上海市电力公司、中国工商银行软件开发中心、上海浦东发展银行股份有限公司、上海生腾数据科技有限公司、上海久事（集团）有限公司、上海智慧城市发展研究院、上海信投数字科技有限公司、上海数字产业发展有限公司

本文件主要起草人：

卓训方、韩懿、张瑞韬、司萌萌、杨琳、孟凯、王超毅、李远刚、张绍华、昌文婷、杨天雅、李川、侯冕、关淘、彭莉、张诚、程夏莹、林力、苏运、苏亚武、丁瑾、陆燕、金晶、唐飞、朱喜华、李周平、彭晋、卞阳、常永波、李浩林、李凡、沈蓓瑾、徐宏杰、谈云骏、姜善定、杨晓倩、徐香君、王冠群、胡琼方、胡力旗、罗凌、杨冠军

引 言

本文件根据《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》《“十四五”数字经济发展规划》《中共中央 国务院关于加快建设全国统一大市场的意见》和《上海市数据条例》等要求制定，旨在促进数据要素流通、培育数据要素市场、推动数字经济高质量发展和加快数据质量标准化体系建设。

本文件的落地实施，有利于保障数据的可靠性和实用性，规范数据产品质量评估领域工作，提高数据产品质量水平，促进本市数据产品交易，推进数据要素市场建设，完善数据要素流通交易市场制度体系建设，加快数字经济发展。

数据交易 第3部分：数据产品质量评估规范

1 范围

本文件提供了可交易数据产品的质量评估框架、指标体系、评估方法和评估流程。

本文件适用于各类数据交易场景下的数据产品质量评估，为数据交易中相关方开展质量评估与自评估提供参考。

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 数据质量评估指标

DB31/T XXXX-XXXX 数据交易 第1部分：数据流通交易合规指南

3 术语和定义

下列术语和定义适用于本文件。

3.1

属性 attribute

目标数据的固有性质或特性，能用人工或自动手段定量或定性地区分。

3.2

元数据 metadata

定义和描述其他数据的数据，主要用来指示数据类型、内容概要、存储途径、数据访问权、资源查找、信息记录等。

[来源：GB/T 36344，2.2，有修改]

3.3

数据文件 data document

一个单元化的相关数据记录。

3.4

数据格式 data format

对于数据的储存或记录的管理编排。

3.5

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，对应于数据源中一行信息的一组完整内容，与字段同义词。

3.6

数据记录 data record

对应于数据源中一行信息的一组完整的内容。

3.7

数据值 range

通过进行测量对目标实体的属性所赋予的数值或者类别，是数据项的内容，也称为值域。

3.8

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，保障数据使用和交换规范性约束。

[来源： GB/T 36344， 2.8， 有修改]

3.9

数据质量 data quality

作为数据产品的内容时，数据的特性满足明确的和隐含的要求的程度。

[来源： GB/T 36344， 2.3， 有修改]

4 评估原则

数据产品质量评估应遵循下列原则：

- 科学性。评估应能正确、客观地反映数据产品质量状况，主要体现在正确的质量指标选择，以及采用科学合理的评测方法等方面。评估应有一定的理论基础，但又不能脱离实际。指标既要相互联系，相互制约，又要相互独立，不互相包容，避免指标间冲突；
- 严谨性。质量评估过程中应根据业务需求制定清晰、准确的指标计算方式并进行评估实施，避免产生歧义；
- 可操作性。应考虑评估的可行性、操作的便捷性，选取可量化、易获取、可靠的指标进行评估。

5 评估框架

5.1 质量模型

数据产品质量模型包括基础质量、技术使用及市场应用指标三方面。基础质量要求包括完整性、规范性、一致性、准确性和时效性；技术使用指标包括可用性、可靠性、高效性和可理解性。市场应用质量指标包括转化率、复购率、满意度、改进效率和持续高质量。每个质量特性包含若干个质量测度，见图1。

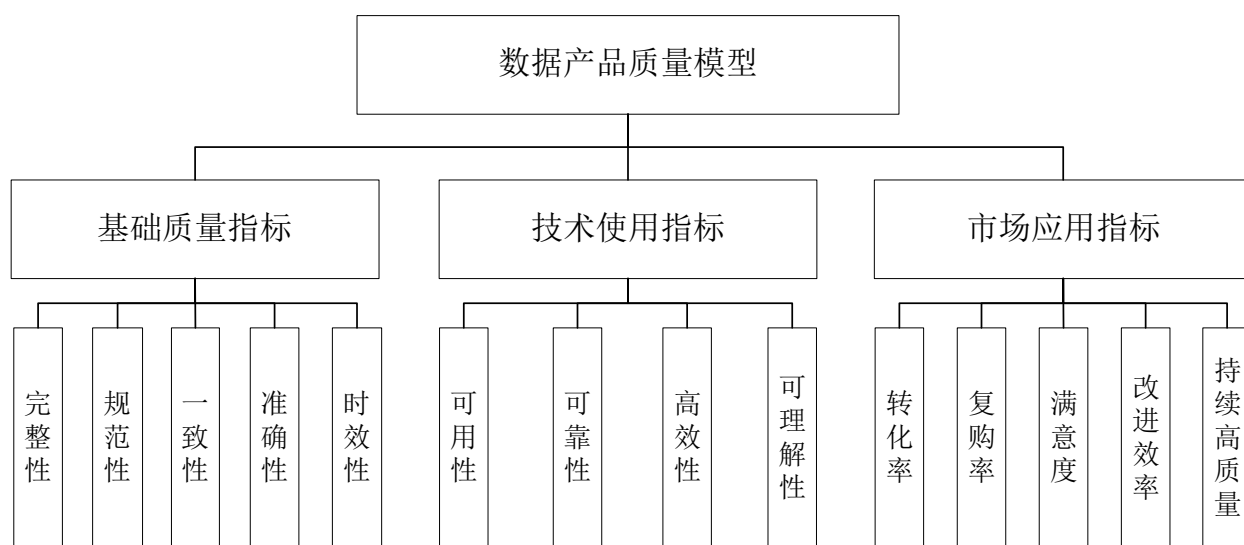


图1 数据产品质量模型

5.2 指标项组成

本文件给出评价表的表头信息如下：

- 指标编号：采用 4 位阿拉伯数字，第 1 位代表所属一级指标，为 1 表示是通用质量质量指标，为 2 表示是使用质量指标，为 3 表示是价值质量指标。第 2 位代表该二级指标在所属一级指标中的排序，3、4 位表示该三级指标在所属二级指标里的排序；
- 指标名称：三级指标的名称；
- 指标描述：指标的定义、解释与示例；
- 计算/实施方法：指标的计算公式。

6 指标体系

6.1 基础质量指标

6.1.1 完整性

完整性是指数据产品中数据文件记录的数据项被赋予数值程度，评价指标定义见表1。

表1 完整性评价指标

指标编号	指标名称	指标描述	计算方法
1101	记录完整性	数据产品记录的数据字段被赋值数量完整程度，即行的完整性。	$X = A/B$ 式中： A = 数据产品通过检查被赋值的字段个数； B = 被评价的数据产品总字段数的个数。
1102	属性完整性	数据产品记录的条数被赋值数量完整程度，即列的完整性。	$X = A/B$ 式中： A = 数据产品通过检查被赋值的记录条数； B = 被评价的数据产品总条数。
1103	数据值完整性	数据产品记录的数据值被赋值数量的完整程度。	$X = A/B$ 式中： A = 数据产品通过检查的数据值的数量； B = 被评价的数据产品数据元素总数量。

6.1.2 规范性

规范性是指数据产品符合数据标准、业务规则、元数据及格式等要求的规范程度，评价指标定义见表2。

表2 规范性评价指标

指标编号	指标名称	指标描述	计算方法
1201	值域规范性	数据产品是否符合值域标准规范。	$X = A/B$ 式中： A = 数据产品通过值域标准规范检查的数量； B = 被评价的数据产品数据元素总数量。
1202	元数据规范性	数据产品是否符合元数据标准规范。	$X = A/B$ 式中： A = 数据产品通过元数据标准规范检查的数量； B = 被评价的数据产品数据元素总数量。
1203	格式规范性	数据产品是否符合格式标准规范。	$X = A/B$ 式中： A = 数据产品通过格式标准规范检查的数量； B = 被评价的数据产品数据元素总数量。
1204	安全规范性	数据产品是否符合法律法规和行业安全标准规范。	$X = A/B$ 式中： A = 数据产品通过法律法规和行业安全标准规范检查的数量； B = 被评价的数据产品数据元素总数量。

6.1.3 一致性

一致性是不同数据产品描述同一件事物的无矛盾程度，评价指标定义见表3。

表3 一致性评价指标

指标编号	指标名称	指标描述	计算方法
1301	数据赋值一致性	数据产品具有相同的意义的数据在同一时点、存储在不同位置的赋值一致程度。	$X = A/B$ 式中： A = 数据产品通过相同含义数据赋值一致检查的数； B = 被评价的数据产品数据元素总数量。
1302	语义一致性	数据产品的数据项符合语义规则一致程度。	$X = A/B$ 式中： A = 数据产品通过数据项语义规则检查的数量； B = 被评价的数据产品数据元素总数量。

6.1.4 准确性

准确性是数据产品准确描述事件的真实程度，评价指标定义见表4。

表4 准确性评价指标

指标编号	指标名称	指标描述	计算方法
1401	内容准确性	数据产品说明书揭露事项与数据产品中所列的数据项内容一致情况	$X = A/B$ 式中： A = 数据产品通过内容表述正确检查的数量； B = 被评价的数据产品数据元素总数量。
1402	数据重复率	数据产品之特定字段、记录或数据文件重复记录情况	$X = A/B$ 式中： A = 数据产品重复记录的条数； B = 被评价的数据产品记录总条数。
1403	脏数据出现率	数据产品存在非法字符和业务含义错误等无效数据情况	$X = A/B$ 式中： A = 数据产品无效数据数量； B = 被评价的数据产品数据元素总数量。

6.1.5 时效性

时效性是数据产品真实反应事件的及时程度，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
1501	更新频率及时性	数据产品符合业务周期更新频率要求情况	$X = A/B$ 式中： A = 数据产品通过业务周期频率要求的数量； B = 被评价的数据产品数据元素总数量。
1502	更新即时性	数据产品符合延迟时间要求更新情况	$X = A/B$ 式中： A = 数据产品通过延迟时间更新要求的数量； B = 被评价的数据产品数据元素总数量。

6.2 技术使用指标

6.2.1 可用性

可用性是数据产品能被正常访问的程度，评价指标定义见表6。

表6 可用性评价指标

指标编号	指标名称	指标描述	计算方法
2101	可访问性	数据需方成功请求访问的情况	$X = A/B$ 式中： A = 数据产品被数据需方通过请求访问成功的次数； B = 被评价的数据产品请求访问总次数。
2102	可获取性	数据需方成功使用该数据产品的情况	$X = A/B$ 式中： A = 数据需方成功获取数据产品的次数； B = 总使用次数。
2103	可用性	数据产品满足数据需方业务使用占比情况	$X = A/B$ 式中： A = 满足业务需求数据项数量； B = 总数据量。

6.2.2 可靠性

可靠性是指数据产品在数据需方规定条件下和规定时间内保持正常工作的能力，评价指标定义见表7。

表7 可靠性评价指标

指标编号	指标名称	指标描述	计算方法
2201	故障率	数据需方在使用数据产品过程中，单位时间内发生各类故障的概率	$X = A/B$ 式中： A = 故障次数； B = 总使用时间。
2202	平均无故障时间	数据产品在每两次相邻故障之间平均正常运行的时间	$X = A/B$ 式中： A = 无故障使用时间总和； B = 故障次数。

6.2.3 高效性

高效性是指数据需方在使用数据产品过程中完成业务目标的效率，评价指标定义见表8。

表8 高效性评价指标

指标编号	指标名称	指标描述	计算方法
2301	业务效率提升	数据需方使用数据产品后业务效率的变化情况	$X = (A-B) / A$ 式中： A = 使用数据产品后业务效率； B = 使用数据产品前业务效率。
2302	精度准确性	数据产品相关数据项是否满足数据需方业务精度的要求	$X = A/B$ 式中： A = 数据产品通过精度标准规范检查的数据项数量； B = 被评价的数据项总数量。
2303	平均修复时间	数据产品发生故障后，数据供方修复故障所需平均时间	$X = A/B$ 式中： A = 总修复时间； B = 故障次数。

6.2.4 可理解性

可理解性指标描述的是数据产品在用户理解和使用方面的难易程度。

表9 可理解性评价指标

指标编号	指标名称	指标描述	计算方法
2401	可理解性	数据产品相关数据项是否符合数据需方的理解	$X = (1-A) / B$ 式中： A = 不易于理解的数据项； B = 被评价的数据项总数量。

6.3 市场应用指标

适用于数据产品质量全生命周期管理，衡量数据产品的综合价值，包括满足数据需方的能力、在关键业务场景中的作用、适应不同场景的实用性以及持续改进的潜力。可用于评估数据产品的整体性能、适用性和未来价值提升潜力，为决策和改进提供参考。指标定义见表10。

表10 市场应用指标

指标编号	指标名称	指标描述	计算方法
3101	转化率	在一定时间内达成交易的比例。	$X = A/B$ 式中： A = 交易次数； B = 总访问量
3102	复购率	在数据需方完成使用数据产品后一定时间内达成复购的比例。	$X = A/B$ 式中： A = 在完成使用数据产品后完成复购的次数； B = 在相同时间段内总交易次数。
3103	满意度	数据需方对数据产品用于不同场景下的评价和反馈。	$X = A/B$ 式中： A = 正面评价的数量； B = 总评价的数量。
3104	改进率	针对过往检查中发现问题的指标采取定期检查、审计等方式，对数据产品的质量改进效率进行评估。	$X = A/B$ 式中： A = 已整改指标数量； B = 被检查指标数量。
3105	持续高质量	若该数据产品在一段时间内对通用质量指标和使用质量指标进行过多次评估，加权平均后得到持续高质量性的数值。数值越高，表明数据产品拥有持续高质量的能力。	$X = A/B$ 式中： A = 基础质量指标的加权平均值； B = 技术使用质量指标的加权平均值。

7 评估方法

7.1 评估方法类型

数据产品质量评估方法分为：

- a) 定量评估方法：通过数学或其他科学手段而做出的判断和评估数据产品质量，能够较好地保证评估结果的科学性和客观性；
- b) 定性评估方法：使用与数据相关的外部知识、专业领域知识和专家经验来推断或评估数据产品质量。

7.2 定量评估

7.2.1 定量评估通过定量指标评估数据产品质量，定量指标用来描述数据产品的定量质量信息，用来表达符合数据规范的程度。

7.2.2 定量评估可分为：完全检查方法与抽样检查方法。

- a) 完全检查：完全检查适用于评估对象规模相对较小时，应对评估对象的所有元素进行检查；
- b) 抽样检查：抽样检查适用于评估对象规模相对较大时，应先对评估对象进行采样，然后评估数据产品质量，以样本数据质量代表整体数据产品质量。

7.3 定性评估

定性评估通过非定量指标评估数据产品质量，非定量指标用来描述数据产品的非定量质量信息。

7.4 结果评分

7.4.1 当所有已选择的数据产品质量评估经过基础质量指标、技术使用指标和市场应用指标（可选）计算完成之后，数据产品在每个评估维度上都形成一个百分制的评分，可采用加权平均法的方式，将所有的评分聚合成最终质量评分。

7.4.2 根据最终质量得分，数据产品可以划分为不同质量等级。具体的数据产品质量等级见表 11。

表 11 数据产品质量等级

质量得分	质量等级	具体含义
90 分(含)以上	A 级	应全面满足数据需求方的需求，确保高质量和高满意度，具备出色的数据产品性能。在多种场景中发挥关键性作用，为组织创造长期且显著的回报，展现出卓越的数据产品价值。
80 分(含)-90 分	B 级	应满足大部分需求方的要求，提供符合其需求的数据，具备相对优秀的性能表现。在特定场景中表现出色，能够有效地支持业务决策或解决特定问题，展示出一定的数据产品价值。
60 分(含)-80 分	C 级	应满足部分需求方的要求，能够理解并满足部分数据需求方的需求，但仍存在一些指标不合格。在有限场景中有所表现，但适用范围相对有限，有待提升数据产品质量和挖掘其价值。
60 分以下	D 级	难以满足需求方的要求，多数指标不合格，存在较大的质量缺陷。应用场景单一，难以在多种业务场景中发挥作用，具有较低资产价值，难以在组织内部产生明显的效益。

数据产品质量等级可作为数据产品定价依据，直接影响质量系数的测算。

8 评估流程

8.1 流程模型

数据产品质量评估具体流程见图2。

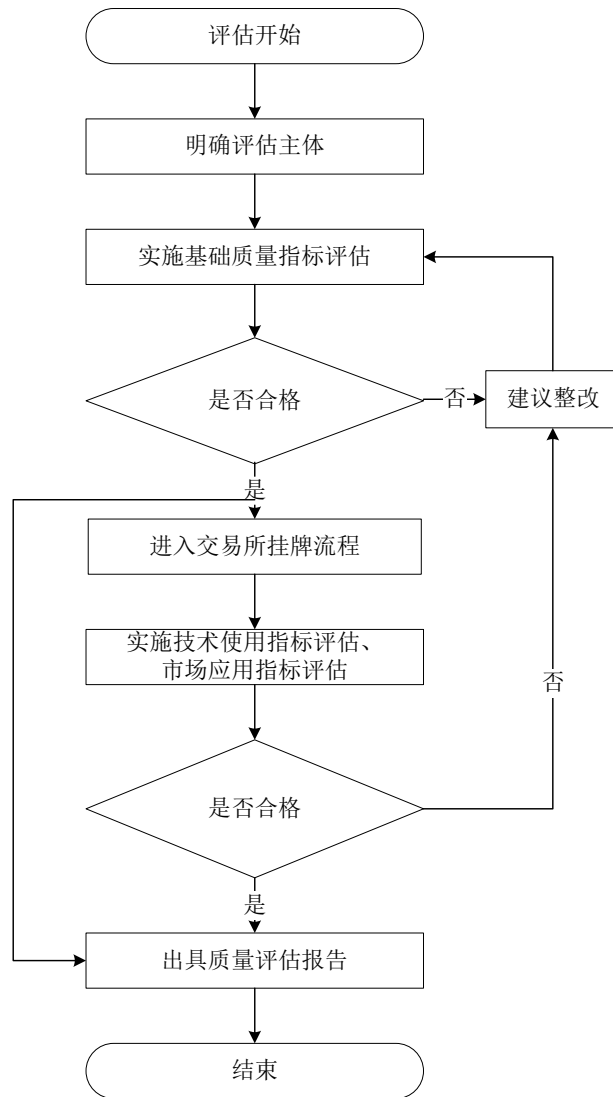


图2 数据产品质量流程模型

8.2 明确评估主体

开展数据产品质量评估，应明确数据产品质量评估主体。数据供方可委托经过数据交易所认证的数据产品质量评估专业服务机构进行评估，并把完成的评估报告提交至给数据交易所。数据产品质量评估主体专门负责数据产品质量评估的计划、实施与跟踪，其职责主要包括：

- 确定待评估的可交易数据产品；
- 整理数据产品质量的评估需求，对其合理性、可操作性进行评估；
- 确定参与数据产品质量改进的参与角色，包括但不限于数据专家、业务专家、质量管控组织负责人员；
- 编制数据产品质量评估方案，包括但不限于数据供方、评估对象、评估目标、评估体系、评估方法以及评估计划；
- 编制并发布数据产品质量评估报告；
- 跟踪监督数据产品质量提升。

8.3 设计指标方案

根据评估对象具体情况，设计适宜于待评估数据产品的评价指标方案，选取合适的评估维度与指标。确定评估度量和评分标准，应考虑下列影响因素：

- 内部因素：数据产品的使用目的、内容清单、安全级别等；
- 外部因素：数据需方的业务需求、供方介绍等。

8.4 实施质量评估

8.4.1 根据评估方案，利用人工或工具对数据产品质量每个指标进行测评，根据每个度量下各评估指标的评估结果综合得到待评估数据产品的整体评估结果。

8.4.2 评估完成后，应根据最终评估结果，生成相关数据产品质量评估报告。数据产品进入数据交易所挂牌前应至少进行一次基础质量评估，才可进入后续流程。场内交易应根据不同等级采取相应措施：数据产品质量评估未达合格标准的，质量等级为D级，建议整改提升质量；达C级以上的，允许挂牌。数据产品质量评估报告是完整记录数据产品质量评估过程与结果的文件，其编制要求如附录A所示。

8.5 提升产品质量

如数据产品质量未达到合格要求，数据供方应根据评估报告，分析确定需要改进的产品内容，并组织相关人员对数据产品进行改进提升。数据产品质量改进应考虑下列因素：

- 数据产品质量评估报告；
- 数据需方质量需求；
- 同类数据产品质量情况。

8.6 数据产品运营

数据产品质量达到合格要求时，如同时满足数据交易所其他要求，即可进入挂牌运营阶段。其他要求包括但不限于数据产品的登记、合规评估、权益认定、定价等。

8.7 质量需求变更

运营过程中，可根据数据产品质量需求变更，进行单次或定期评估，评估方案可根据评估频率、评估需求进行调整。数据产品质量需求变更包括但不限于：

- 数据产品内容发生改变；
- 数据需方提出的评估需求变更，如评估主体变更、指标方案变更等；
- 数据产品持续运营时间较长，需提供近期数据产品质量评估报告。

8.8 质量评估完成

8.8.1 质量合格的数据产品，应根据数据交易所及数据需方要求，提交质量评估报告，供数据产品交易场景使用。

8.8.2 数据需方可向数据交易所申请查看数据产品质量评估报告，如对数据产品质量有异议，可由经过数据交易所认证的数据产品质量评估专业服务机构复核并提交质量评估报告。

附录 A

(资料性)

数据产品质量评估报告编制要求

A.1 总体要求

数据产品质量评估主体在完成评估之后，应出具相应评估报告，包含首页、正文和附录部分，各部分应满足下列要求：

- 总体反映数据产品质量评估工作，文字简洁，重点突出，结论明确；
- 宜采用有助于理解的图表和数据，资料引用表示清晰；
- 应客观、准确反映评估过程与结果。

A.2 首页

首页可作为数据产品质量评估报告的封面，应包括完整的数据产品名称、数据供方名称、评估机构名称及其公章。

A.3 正文

正文是数据产品质量评估报告主体，一般包括：

- 数据产品基本情况，如数据产品名称、内容、格式和大小；
- 评估的规模；
- 选取的评估指标体系及其定义；
- 各个指标的评估结果与得分。

A.4 附录

附录是数据产品质量评估中有助于理解的附加信息，一般包括：

- 各个评估指标的错误示例；
- 相关的行业标准或规范；
- 其他支持性文件。

参考文献

- [1] GB/T 5271.1-2000 信息技术 词汇 第1部分:基本术语
 - [2] GB/T 5271.17-2000 信息技术 词汇 第17部分:数据库
 - [3] 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(引用)
 - [4] 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(引用)
 - [5] GB/T 39400-2020 工业数据质量 通用技术规范
 - [6] 上海市数据条例
 - [7] 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
 - [8] 上海市数据交易场所管理实施暂行办法
-