

附件 3:

## 证券期货业资讯数据质量评估标准研究报告

**【摘要】** 本课题依据证券期货业数据要素、标准化等相关政策要求，基于国内外数据质量评估体系发展现状，并结合证券期货业资讯数据特性，建立证券期货业资讯数据质量评估标准体系，产出资讯数据质量评估方法标准草案，并以实际使用场景为例，实践资讯数据质量评估过程，验证资讯数据质量评估方法的可行性。本课题研究旨在推动国家和行业相关要求的落实，提升行业资讯数据的质量水平，为行业的高质量发展贡献力量。

**关键词：** 证券期货业；数据质量评估；资讯数据。

# 目录

一、引言	4
(一) 研究背景	4
1. 国家宏观层面对数据质量和价值评估提出要求	4
2. 行业资讯生态稳步发展，资讯数据质量问题日渐突出	4
(二) 研究目标	5
1. 明确资讯数据定义	6
2. 设计资讯数据质量评估维度	6
3. 构建资讯数据质量评估指标体系	6
4. 提出资讯数据质量评估计算方法	6
5. 形成资讯数据质量评估等级划分	6
6. 产出资讯数据质量评估方法标准草案	7
(三) 研究意义与价值	7
1. 制定资讯数据质量评估行业标准的重要性	7
2. 助力证券期货业资讯数据质量提升	7
二、研究方法	9
(一) 研究方法概述	9
1. 调查法	9
2. 案例研究法	9
3. 文献研究法	9
4. 经验总结法	10
(二) 行业资讯数据定义梳理	10
1. 标准规划方面	10
2. 行业标准方面	10
3. 行业实践方面	11
(三) 行业资讯数据质量问题分析	11
(四) 数据质量评估方法研究	13
1. 国际数据质量管理框架	13
2. 国内现行数据质量评估标准	16
3. 数据质量管理实践	19
4. 数据质量评估方法优劣分析	21
(五) 数据质量评估维度调研	22
1. 数据质量评估维度现状	22
2. 评估维度繁多带来的问题	24
三、研究结果	25
(一) 资讯数据定义	25
(二) 资讯数据质量评估维度	25
(三) 资讯数据质量评估指标体系	27
(四) 资讯数据质量评估计算方法	35
(五) 资讯数据质量评估等级划分	36

(六) 《证券期货业资讯数据质量评估方法》(标准草案) .....	37
四、研究实践 .....	39
(一) 定义评估的目标和范围 .....	39
(二) 确定评估数据 .....	39
(三) 选取数据质量评估指标 .....	40
(四) 定义数据质量规则 .....	41
(五) 设定质量权重 .....	42
(六) 实施质量评估 .....	43
(七) 数据质量评分及定级 .....	44
(八) 数据质量问题分析 .....	44
(九) 质量结果反馈 .....	45
五、研究结论与建议 .....	47
(一) 通过构建行之有效的资讯数据质量评估体系,为资讯数据质量的提升提供支撑 .....	47
(二) 通过落地评估体系实践应用,充分验证评估体系可行性和实用性 .....	47
(三) 通过评估体系标准化转化,促进行业资讯数据质量水平进一步提升 .....	47
【参考文献】 .....	49

# 正文

## 一、引言

### （一）研究背景

#### 1.国家宏观层面对数据质量和价值评估提出要求

2022年12月2日，中共中央、国务院印发的《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》指出，要加快推进数据管理能力成熟度国家标准及数据要素管理规范贯彻执行工作，推动各部门各行业完善元数据管理、数据脱敏、数据质量、价值评估等标准体系。

2023年11月2日，中国证监会、国家标准委印发的《关于加强证券期货业标准化工作的指导意见》<sup>[2]</sup>对资本市场标准化工作取得的进步，在规范业务发展、提高市场效率、降低行业成本等方面发挥的重要作用做出肯定，同时对进一步推动证券期货业标准化工作高质量发展，提出新的意见。

#### 2.行业资讯生态稳步发展，资讯数据质量问题日渐突出

随着数字金融的发展，资讯数据在证券期货行业中的影响力愈发显著，资讯生产方、资讯服务方、资讯消费方共同组成了资讯数据的应用生态。行业内三方的参与单位和承担职责如表1所示：

表1 资讯数据参与方情况

参与方角色	参与方单位	参与方职责
资讯生产方	交易所、中国结算、监管机构、 人民银行、外管局等	发布资讯数据

资讯服务方	资讯供应商等公司、行业平台，部分资讯生产方直接提供服务	通过采集技术获取资讯数据，并各自整合成数据（库）产品
资讯消费方	证券期货业各机构	依托资讯数据开展业务

随着数字化的发展，资讯数据在行业发挥重要作用的同时，也逐渐突显出了一些问题。在资讯数据的应用生态体系中，参与者众多且角色多元，单个参与者往往扮演着多重角色。这一复杂多变的服务架构，为整个生态系统的管理带来了显著挑战。同时，由于数据源头众多、机构间数据交换链路多样、数据加工处理复杂等原因，导致行业内资讯数据质量存在较大差异。数据质量的参差不齐会对整个行业的高质量发展产生影响，低质量的数据无法为行业及用户提供有力的支持和指导。

## （二）研究目标

随着数字金融的飞速发展，资讯数据的质量对行业的高质量发展有重大意义。由于行业目前缺失对资讯数据质量的评估体系，导致资讯数据质量的提高缺乏方向指引，构建资讯数据质量评估体系是解决质量问题的关键一步。

为响应国家层面对于数据质量、价值评估等标准体系的完善要求，同时以质量评估的方式助力资讯数据质量持续改进，课题组设立的研究目标如下：

### 1.明确资讯数据定义

界定清晰、划分明确的研究对象对于课题研究至关重要。本课题将调研行业已发布的资讯数据相关标准，结合资讯供应商的实践对资讯数据定义进行研究，为评估工作奠定基础。

### 2.设计资讯数据质量评估维度

从数据质量评估维度现状分析、问题梳理、国内外数据质量评估维度调研、资讯数据质量评估维度设计等方面开展研究，确定评估维度，为评估指标体系的构建提供理论支持。

### 3.构建资讯数据质量评估指标体系

课题组对国内外的一系列数据质量评估标准及数据实践进行研究，并结合资讯数据的特点，确定适合于资讯数据的指标，形成多级指标体系，为数据质量评估的量化提供支持。

### 4.提出资讯数据质量评估计算方法

为了客观地衡量某一维度数据质量的优劣，课题组在数据质量评估的实际操作中应用量化的方式，深入探究指标体系中各项指标的计算方法，从而为评估体系奠定实践基础。

### 5.形成资讯数据质量评估等级划分

通过实施等级划分策略，衡量不同维度的质量水平，从而提供一个清晰直观的视角，以便准确把握数据质量的优劣状况。基于各维度指标的综合考量，明确多级指标体系中各项维度得分与等级划分之间的关联性。

## 6.产出资讯数据质量评估方法标准草案

通过课题研究，产出《证券期货业资讯数据质量评估方法》的标准草案，为行业内的资讯数据质量评估提供可操作的规范与指导。这一标准的制定对于提升整个证券期货业的资讯数据管理水平、推动行业高质量发展具有重要意义。

### （三）研究意义与价值

#### 1.制定资讯数据质量评估行业标准的重要性

鉴于资讯数据的复杂性和多样性，从资讯数据质量评估方法的角度出发，制定一套针对性的行业标准显得尤为重要。通过实施这一科学的资讯数据质量评估流程，行业能够及时发现资讯数据中的潜在问题，并采取有效的改进措施，从而逐步提升整个行业的资讯数据质量水平。此举不仅有助于降低行业在资讯数据使用过程中的风险，还能构建持续健康的数据生态环境，深入贯彻已建立的数据标准，为行业的高质量发展奠定坚实基础。

#### 2.助力证券期货业资讯数据质量提升

本课题研究以国内外数据质量评估的先进维度为基础，深入参考证券期货业数据治理的相关政策，并结合该行业数据的独特性质，充分考虑了我国证券期货业资讯数据的应用生态现状。在研究过程中，课题组严格遵循全面性、简洁性、目的性及可计量性的原则，对证券期货业资讯数据质量评估体系进行了细致化构建，并制定了相应的评估标准。这一体

系的建立旨在规范和指导证券期货业资讯数据的高质量发展，推动《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》<sup>[1]</sup>的深入落实。

通过引入可量化的评估方法，有助于行业识别证券期货业资讯数据中的质量问题，并持续改进，从而显著提升资讯数据质量。这一体系的建立有助于资讯数据作为行业核心资源的有效利用和高质量发展。未来，随着该体系的不断完善和推广，证券期货业将能够更加高效地利用数据资源，推动行业的数字化转型和高质量发展。

## 二、研究方法

### （一）研究方法概述

证券期货行业资讯数据质量评估标准研究以 DAMA 数据管理知识体系<sup>[3]</sup>、《信息技术 数据质量评价指标》<sup>[4]</sup>（GB/T 36344—2018）数据质量评价指标体系等为基础，充分借鉴了国内外理论和实践成果，并结合我国证券期货行业资讯数据管理特点，以确保评估标准的科学性和专业性。通过资讯数据质量评估，对资讯数据质量做出科学评价，有利于资讯数据质量提升和价值的实现。

本课题研究使用了以下研究方法：

#### 1. 调查法

通过问卷、咨询等多种方式，调查了解行业内核心机构、资讯供应商、经营机构等多方的资讯数据质量管理方式，收集并分析潜在的问题，归纳原因并提出建议。

#### 2. 案例研究法

针对关键数据种类、核心问题进行深入分析，做好案例撰写和分析记录，持续为研究主题提供最新的实践经验和理论支持，确保研究的深度和广度。

#### 3. 文献研究法

通过查阅国内外的数据质量相关报告文献，广泛收集与课题相关的素材，掌握最新的研究动态和研究成果并进行整理和分析，结合调研的实际案例，不断为课题的研究提供最

新经验和理论依据，保证课题研究深度。

#### 4.经验总结法

在已有研究和实践经验的基础上，及时总结、反思、提炼、优化评估体系，逐渐从理论上升到实践，探索并总结资讯数据质量评估体系的落地策略。

#### （二）行业资讯数据定义梳理

界定清晰、划分明确的研究对象对于课题研究至关重要。课题组对于资讯信息的分类开展研究，调研到行业内已有的研究成果：

##### 1.标准规划方面

证标委发布了《证券期货业数据标准规划（2022-2025）》<sup>[5]</sup>，该规划的重点任务之一是制定包含资讯信息的数据元标准。该规划将资讯信息分为宏观数据、行业数据、资本市场数据、货币及外汇市场、新闻公告研报五个方面。

##### 2.行业标准方面

证标委发布的行业标准《证券期货业机构内部接口 资讯数据》（JR/T 0275-2023）<sup>[6]</sup>，该标准将资讯数据分为三大模块：基础数据、金融产品数据、宏观及行业数据，其中基础数据包括证券基本信息、公司基本信息、市场基准及利率、各类业务资格会员、人员信息标准、行业分类列表、交易日志；金融产品数据包括股票数据、基金数据、债券数据、理财数据、金融衍生品；宏观及行业数据包括宏观数据、行业

数据。

### 3.行业实践方面

行业资讯供应商在实践中，从资讯数据库组织角度出发，将资讯分为股票数据库、固收数据库、基金数据库、理财数据库、衍生品数据库、宏观数据库、衍生指标库、资讯/舆情数据库、指数数据库、财务数据库、机构数据库等。

综上所述，《证券期货业数据标准规划（2022-2025）》从顶层设计角度给出了资讯信息划分建议，《证券期货业机构内部接口 资讯数据》（JR/T 0275-2023）自顶向下设计并完善了资讯分类体系，行业资讯供应商在实践中分类建设了资讯数据库。

#### （三）行业资讯数据质量问题分析

摸清当前行业典型资讯数据质量问题，带着问题进行研究事半功倍。课题组开展了针对证券期货业资讯信息标准化的调研，调研上海证券、中信建投证券、中金公司、南方基金、申万宏源证券等5家经营机构和多家资讯供应商。

调研对象反馈了针对资讯信息数据使用中存在的问题，并给出了所采用的事前、事中、事后的应对方案。例如，上海证券提出针对资讯信息数据格式不统一，数据存在缺失的问题，可以通过数据加工及核验机制处理。中信建投证券提出数据变更不及时、遇到过数据错误和数据缺失等问题，一般由客户反馈得知，由数据对接人主动联系资讯供应商沟通

解决。中金公司提出部分数据非结构化，不全面，同时质量参差不齐，经常是客户后置反馈给资讯供应商。南方基金提出同一类数据，偶尔存在不同资讯供应商数据不一致的情况，常规做法是通过资讯比对来解决。申万宏源证券提出数据格式不统一、数据表述不统一、数据缺失、数据重复等，特别是需要跨资讯数据源进行关联使用的时候，数据格式不统一、表述不统一的问题更为突出，遇到以上问题时会以官方（例如基金业协会、国家企业信用信息公示系统）的数据为基准联系资讯数据源厂商进行处理。除上述问题之外，还暴露出不同数据源出现不同时，无法确认权威数据源，数据字典更新不及时等问题。问卷调研结果如表 2 所示：

表 2 调查问卷总结

调查对象	资讯数据使用问题
南方基金	同类数据偶尔存在不同资讯供应商数据不一致的情况
上海证券	资讯信息数据格式不统一、数据存在缺失等
申万宏源	数据格式不统一、数据表述不统一、数据缺失、数据重复等
中金公司	部分数据非结构化、不同数据源出现不同时、无法确认权威数据源、时效性有延迟等
中信建投	数据变更不及时、数据错误和数据缺失、数据结构变更不兼容等
资讯供应商	客户的及时性需求难以统一、数据处理遗漏导致的全面性问题等

从调研结果分析总结得知，资讯数据在使用中，主要出

现数据重复、数据缺失、数据失真、数据时效性差、数据结构差异、数据结构变更不兼容、数据溯源困难等 7 大类问题，具体体现在元数据、数据结构、数据内容、数据时效等方面。

#### （四）数据质量评估方法研究

随着资讯数据应用日渐广泛，资讯数据的质量改进工作迫在眉睫，但目前行业内部现行资讯数据相关标准不足，缺乏针对资讯数据特点的数据质量评估体系来指导资讯数据的质量改进工作，为此，课题组通过研究国内外先进的数据质量理论和实践，并在已有理论和实践的基础上，探索资讯数据质量评估标准的可行性。

##### 1. 国际数据质量管理框架

在数据管理领域，DAMA 国际组织多年来深耕企业信息和数据管理研究，积累了丰富的知识体系。DAMA 国际的数据管理知识体系涵盖了数据治理、数据架构、数据建模和设计、数据存储和操作、数据安全、数据集成和互操作、文档和内容管理、参考数据和主数据管理、数据仓库与商务智能、元数据管理、数据质量管理等多个方面。

其中，数据质量管理基于“计划-执行-检查-处理”的问题解决模型，如下图 1 所示。数据质量的改进过程是通过一组确定的步骤来实现。根据标准测量数据状况，如果数据状况不符合标准，需进一步确定与标准不符的根本原因。无论是技术性、非技术性的根本原因，会出现在处理过程的某一

步骤。通过手段纠正后，持续监控数据以确保其满足要求。

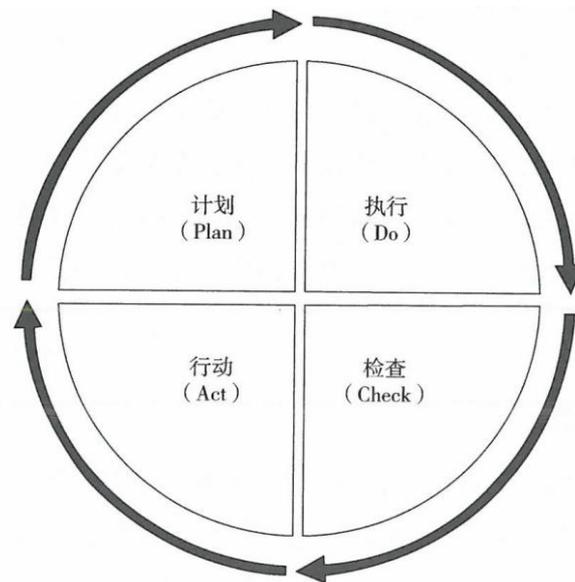


图 1 质量问题解决模型

实现质量问题解决模型有效运转的过程中，一个重要组成部分是开发度量指标。DAMA 数据质量管理框架中给出了设计度量标准时应参考的基本原则：

①可度量性。数据质量指标必须是可度量的。例如，数据相关性是不可度量的。数据完整性这一指标需要得到客观的定义才能测量。预期的结果应在离散范围内可量化。

②业务相关性。数据质量指标需要与数据消费者相关，与数据对关键业务期望的影响相关联。

③可接受性。数据质量指标构成了数据质量的业务需求，根据已确定的指标进行量化提供了数据质量级别的有力证据。根据指定的可接受性阈值确定数据是否满足业务期望。

⑤管理制度。关键利益相关方应理解和审核指标。当度

量的测量结果显示质量不符合预期时，会通知关键利益相关方。业务数据所有者对此负责，并由数据管理专员采取适当的纠正措施。

⑥可控制性。指标应反映业务的可控方面。当度量超出阈值范围，通过触发控制行动来改进数据。

⑦趋势分析。指标使组织能够在一段时间内测量数据质量改进的情况。跟踪有助于数据质量团队成员监控数据质量，并证明数据质量改进活动的有效性。

在数据质量评估方面，国际货币基金组织（IMF）公布了国际通用的数据质量评估框架（DQAF）<sup>[7]</sup>整合了联合国官方统计标准、数据公布特别标准和数据公布通用系统的概念和定义。该框架整体结构呈级联式展开，在第一阶层首先提出了质量的先决条件以及衡量数据质量的五个维度，然后将第一阶层的每个维度分别在第二阶层的评估要素和第三阶层的评估指标中具体化，评估指标后面对统计数据质量评判的标准有更详尽的解释。

DQAF 在结构方面将数据质量的评估标准按照“从共性到特性”的递进顺序组织，仅在第一阶层定义了若干质量维度，子级评估元素和评估指标由用户在具体应用中扩展，证明了其在实践中具有灵活性、可操作性和广泛的适用性，各组织均可根据自身的需要在框架的基础上扩展、丰富或定制该框架。

## 2.国内现行数据质量评估标准

随着数据被确认为生产要素，数据质量的高低、数据价值的挖掘日渐被重视。数据质量评价相关的国家标准、行业标准对各行各业数据质量评估的实施和管理方向的指引具有重要意义。

### ①《信息技术 数据质量评价指标》（GB/T 36344-2018）

该标准旨在规范信息技术领域的的数据质量评估。该标准规定了数据质量评估的指标体系、评估方法和流程。

指标体系建设方面，主要涉及数据准确度、数据完整性、数据一致性和数据可读性等评估指标。其中，数据准确度是指数据是否准确地反映了实际现象；数据完整性则关注数据是否存在缺失或异常值；数据一致性则涉及到不同数据源之间的数据是否匹配；而数据可读性则关注数据的可理解性和易读性。

评估实施方面，定义了组织构建、规范建立、指标设计、评价实施、数据质量提升、交付使用等关键环节。具体数据质量评价过程如下图 2 所示。

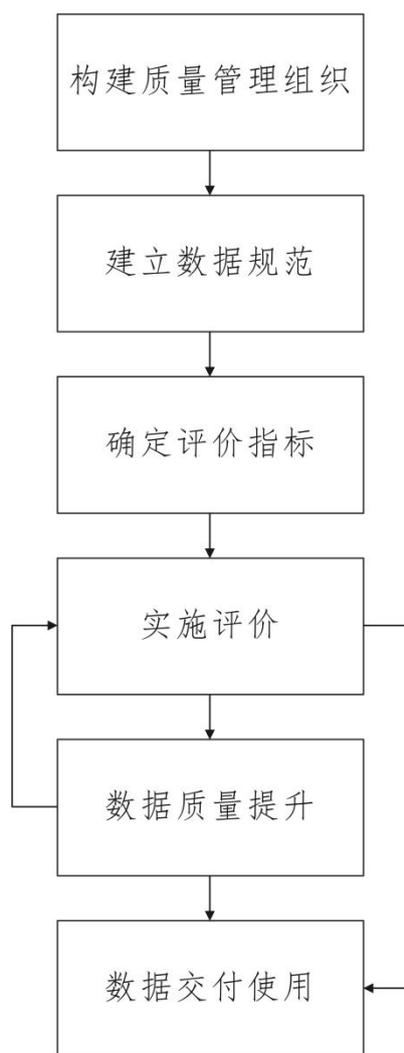


图 2 数据质量评价过程

在实际应用中，组织可以根据该标准的要求，制定适合自身业务需求的数据质量评估方案。该标准还提供了详细的评估方法和流程，帮助组织进行有效的数据质量评估。

该标准虽然为组织提供了数据质量评估的规范，但并不能保证组织的所有数据质量问题都能得到解决。组织还需要结合自身的实际情况，制定更为具体和细致的数据质量管理策略和措施。随着信息技术的发展和数据量的增长，数据质量评估的难度也在逐渐增加。因此，组织需要不断更新和完善自身的数据质量管理体系，以适应不断变化的市场环境和

业务需求。

②《信息资源 数据规范质量技术基础》（GB/T 41795-2022）<sup>[8]</sup>

该标准规定了质量技术基础信息资源的分类和编码、数据元目录、数据采集、数据清洗、数据质量评价、数据溯源、集成服务平台技术实现和数据交换，适用于质量技术基础集成服务平台建设中信息资源的发布、交换、存储和管理，也适用于质量技术基础集成服务平台的设计、开发和验收。

对数据质量管理过程中评价指标体系的顶层设计、评价模型、评价方法、评价流程等方面提出了较为明确的要求，并根据综合得分划分质量技术基础集成服务数据质量等级，建议将质量技术基础集成服务数据质量综合评价结果划分为四级：优秀、良好、合格和不合格。

③《风电信息管理数据质量评估及治理技术规范》（NB/T 11083-2023）<sup>[9]</sup>等行业标准

国家标准搭建了数据质量评估的顶层框架，课题组进一步调研了《风电信息管理数据质量评估及治理技术规范》（NB/T 11083-2023）、《地震观测数据质量评价规范全球导航卫星系统观测》（DB/T 97-2024）<sup>[10]</sup>、《ISLI 数据质量检测》（CY/T 239-2021）<sup>[11]</sup>等行业数据质量评估相关标准。

各行业标准立足于其所在行业数字化转型过程中，在维护数据价值的基本理念、管理理论和技术框架的真空，旨在

提升行业共识、维护数据价值、降低校验成本和架构评估体系。如《风电信息管理数据质量评估及治理技术规范》(NB/T 11083-2023)对已建风电场监控系统所采集的实时运行数据进行质量评估、评级,并对可修复数据进行合理修订和治理的行业技术规范,有利于进一步完善风电场经济效益评估和推广科学数据管理体系。

### 3.数据质量管理实践

随着“数据质量”被更多人提及,众多企业给予充分重视,探索如何体系性地认识和解决数据质量问题。

①《华为数据之道》<sup>[12]</sup>详细阐述了华为公司在数据管理方面的实践经验,包括数据治理框架、数据治理组织架构和数据治理度量评估体系,以及数据治理在华为的实践案例和效果。

数据质量整体框架方面。华为通过一个整体性的框架,来确保企业能体系性地应对和解决这个问题。下方的数据质量整体框架,通过领导力、持续改进、能力保障等方面的有机结合,给出了整体性解决思路。如图3所示。

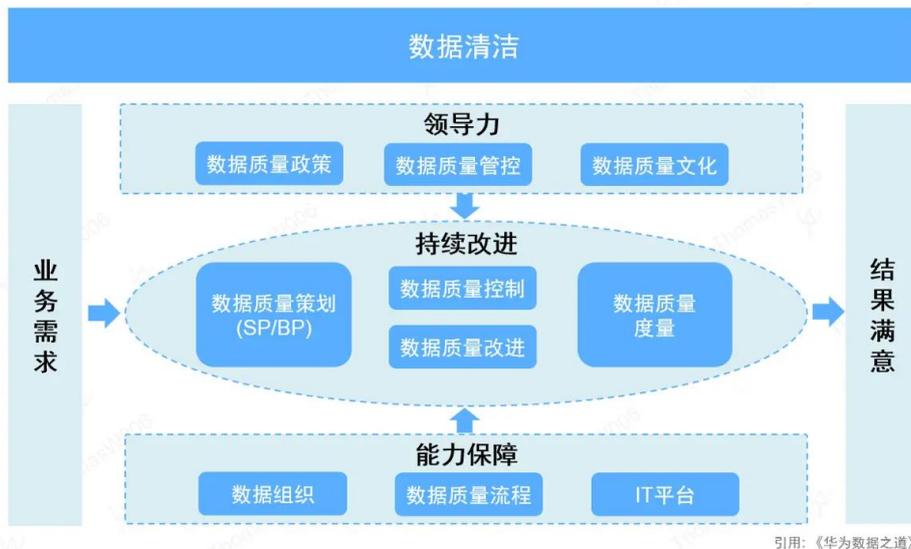


图 3 华为数据质量整体框架

数据质量规则方面，华为在数据六性的基础上，进一步梳理出 15 个具体的质量规则类型和质量特性的关系，明晰了具体的操作思路。如图 4 所示。

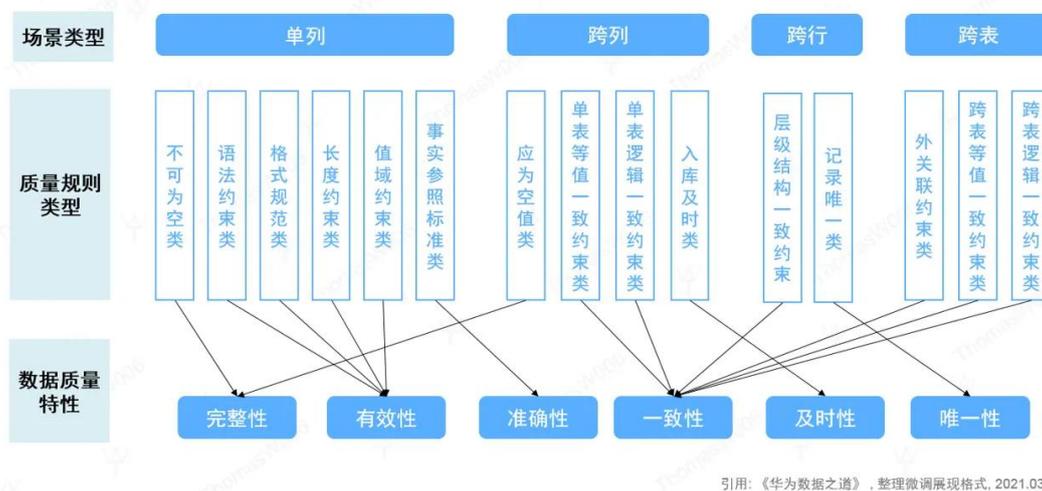


图 4 华为质量规则体系

度量质量方面，华为给出了“设计质量 40% + 执行质量 60%”的思路。尽可能在“源头”确保数据质量，加强对“设计质量”环节的重视和管控。如图 5 所示。

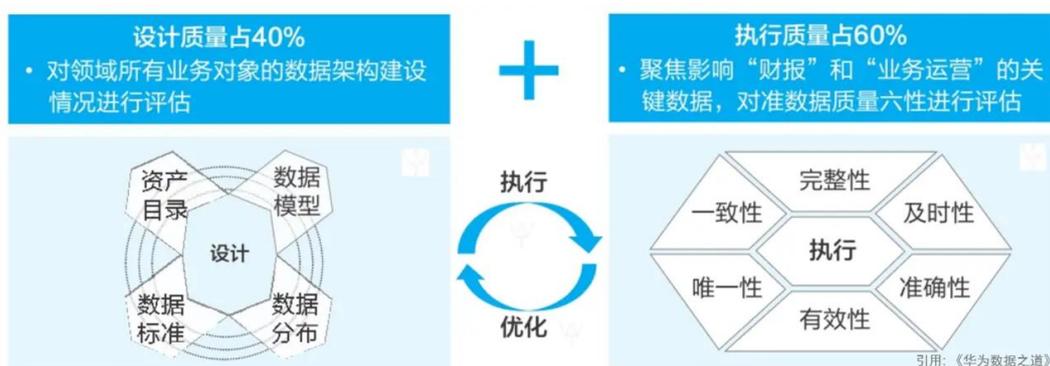


图 5 华为设计质量+执行质量框架

## ② 资讯供应商数据质量管理实践

课题组对行业资讯供应商调研过程中发现，资讯供应商内部多采用技术和行为标准集合对资讯数据质量进行约束。技术标准包括诸如针对单属性字段的格式、符号、取值上下限等规范，针对多属性间的关联验证规范；行为标准包括根据业务场景进行人为的资讯数据质量判断的标准及数据生成后资讯数据质量的日常监测标准，如由专业的人维护并判断其专业领域数据的质量。

实施过程中，往往结合资讯数据应用特点，设置资讯数据质量等级，分类分级对资讯数据实施质量检测、数据规则控制、双路校审、质量追踪等质量管理措施。各家资讯供应商的资讯数据质量管理过程差异较小，基本符合上述国内外研究成果管理框架。

### 4. 数据质量评估方法优劣分析

通过分析上述研究结果，基于国内外的知识体系所建立的数据质量管理方法论已经趋于成熟。其中，数据质量评估

作为关键环节，在各行各业实践中，往往结合所属行业或企业特点，构建特色的数据质量评估体系。

资讯数据这一行业数据，广泛应用于证券期货业各类机构。机构在使用资讯数据时，既存在同其他基础数据类似的共性问题，也在溯源、时效、多源异构等方面存在个性化问题。

因此，现有的数据治理经验并不能完全适用于资讯数据，特别是资讯数据的非标准化特点，使其存在的问题难以通过国内外已有的标准、实践方案得到有效发现和解决。

#### （五）数据质量评估维度调研

数据质量维度的划分是构建数据质量评估体系的基础。课题组从数据质量评估维度现状分析、问题梳理到资讯数据质量评估维度设计等方面开展研究。

##### 1.数据质量评估维度现状

经调研，不同单位、组织或个人结合自身实际需求提出了不同的评估维度。如表 3 所示。

DAMA 提出的数据质量评估维度主要包括：准确性、完整性、一致性、时效性、精确度、隐私（访问控制和使用监控）、合理性、参照完整性、及时性、有效性等，定义了具有普遍一致性的数据质量维度。

国际货币基金组织（IMF）在其通用性数据质量评估框架（DQAF）中提出通过相关性、准确性、可靠性、适用性、

可访问性等五个维度来评估数据质量，侧重对统计活动进行全面质量管理。

《信息技术 数据质量评价指标》（GB/T 36344-2018）规定的数据库质量评价指标框架包括：规范性、完整性、准确性、一致性、时效性、可访问性等，适用于数据生存周期各个阶段的质量评价。

《华为数据之道》数据质量指“数据满足应用的可信程度”，从数据使用者的角度定义，满足业务、用户需要的数据即为“好”数据，从完整性、及时性、准确性、一致性、唯一性、有效性等六个维度对数据质量进行描述。

资讯供应商将积累的资讯数据质量规则划分到全面性、准确性、及时性等三个维度，在问题发现、规则定义到监控预警的质量管理过程中提高了资讯数据质量，其定义的资讯质量维度更具落地实践意义。

表 3 数据质量评估维度对照表

质量维度	DAMA	DQAF	《信息技术 数据质量评价指标》（GB/T 36344-2018）	《华为数据之道》	资讯供应商实践
准确性	√	√	√	√	√
完整性	√		√	√	√
一致性	√		√	√	
时效性	√		√		
精确度	√				
隐私	√				
合理性	√				

及时性	√			√	√
有效性	√			√	
唯一性				√	
规范性			√		
关联性		√			
可靠性		√			
适用性		√			
可访问性		√	√		

## 2. 评估维度繁多带来的问题

从表 3 可以看出，各组织/单位从不同的出发点考虑，对数据质量评估的角度是不一样的，且差异较大。虽然数据质量评估的维度越多越能更全面地评估数据质量的好坏，但是太多的评估维度，给实际操作会带来极大的不便。

主要问题如下：

① 太多相似的度量名称，容易混淆，不利于数据质量评估工作的开展，如及时性与时效性。

② 部分度量维度在实际操作中具有相似的计算方法，如准确性、合规性、规范性、精确度等几个维度，都是基于质量规则库，校验数据是否符合预先配置的规则，其目的都是为了判断数据是否准确，因此这类度量维度可以考虑合并。

③ 部分维度属于主观判断，难以客观分析，缺乏可操作性，如：合理性、适用性等。

### 三、研究结果

#### （一）资讯数据定义

基于上述关于行业资讯数据定义的梳理结果，课题组发现《证券期货业机构内部接口 资讯数据》（JR/T 0275-2023）中自顶向下设计并完善了资讯分类体系，明确了各类资讯的定义，且资讯分类体系与行业机构的实践过程较为契合。因此，课题组根据课题研究目标，明确以该标准的资讯定义及分类为基准，将资讯信息按照基础数据、金融产品数据、宏观及行业数据进行划分。具体分类如图 6 所示。

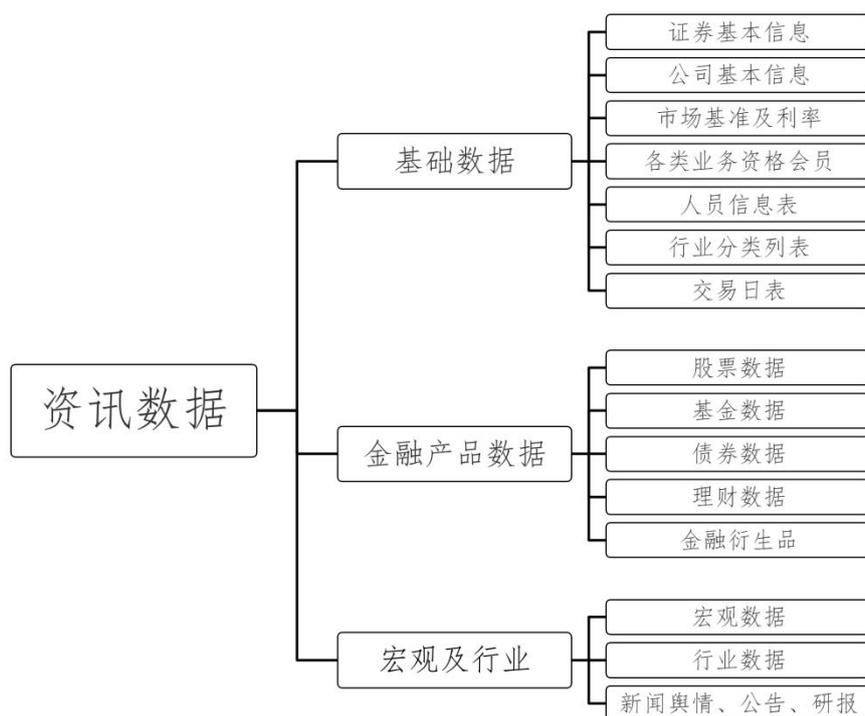


图 6 JR/T 0275-2023 资讯数据分类体系

#### （二）资讯数据质量评估维度

资讯数据的质量评估维度应按照客观反映真实情况、避

免含义混淆、满足资讯特有场景的原则进行选取。因此，结合资讯数据特点，课题组基于上述各类评估维度，将准确、唯一、有效等针对数据单行或跨行内容的要求整合为准确性维度。完整性维度在数据集完整度量的基础上，补充了度量数据丢失或与外部信息关联缺失、未建立关联等。一致性维度支撑资讯数据上下游频繁交换时易产生的数据结构、数据内容不一致问题的评估。可审计性维度针对资讯数据解释性差、数据结构变更频繁、溯源难等特点。最终选取了准确性、完整性、一致性、时效性、可审计性等 5 个维度来评估资讯数据质量。资讯质量评估维度如表 4 所示。

表 4 资讯评估维度表

评估维度	说明
准确性	1.准确：度量已有数据内容的准确性，检查人工输入、跨链路复制等错误。 2.唯一：度量数据重复或者数据的属性重复。 3.有效：度量数据的定义、值、格式和展现形式符合定义要求。如范围有效性、日期有效性等
完整性	度量数据丢失或与外部信息关联缺失、未建立关联等
一致性	度量跨源、库、表数据信息含义上不一致
时效性	度量单条数据延迟时点、多条数据延迟时段等
可审计性	数据可理解（定义、加工逻辑等）、度量数据结构和内容的变更、溯源、安全等

### （三）资讯数据质量评估指标体系

课题组对国内外的一系列数据质量评估标准及数据实践进行研究，并结合资讯数据的特点，设计适合于资讯数据的指标，形成多级指标体系，为资讯数据质量评估的量化提供支持。形成的资讯数据质量评估指标体系如图 7 所示。

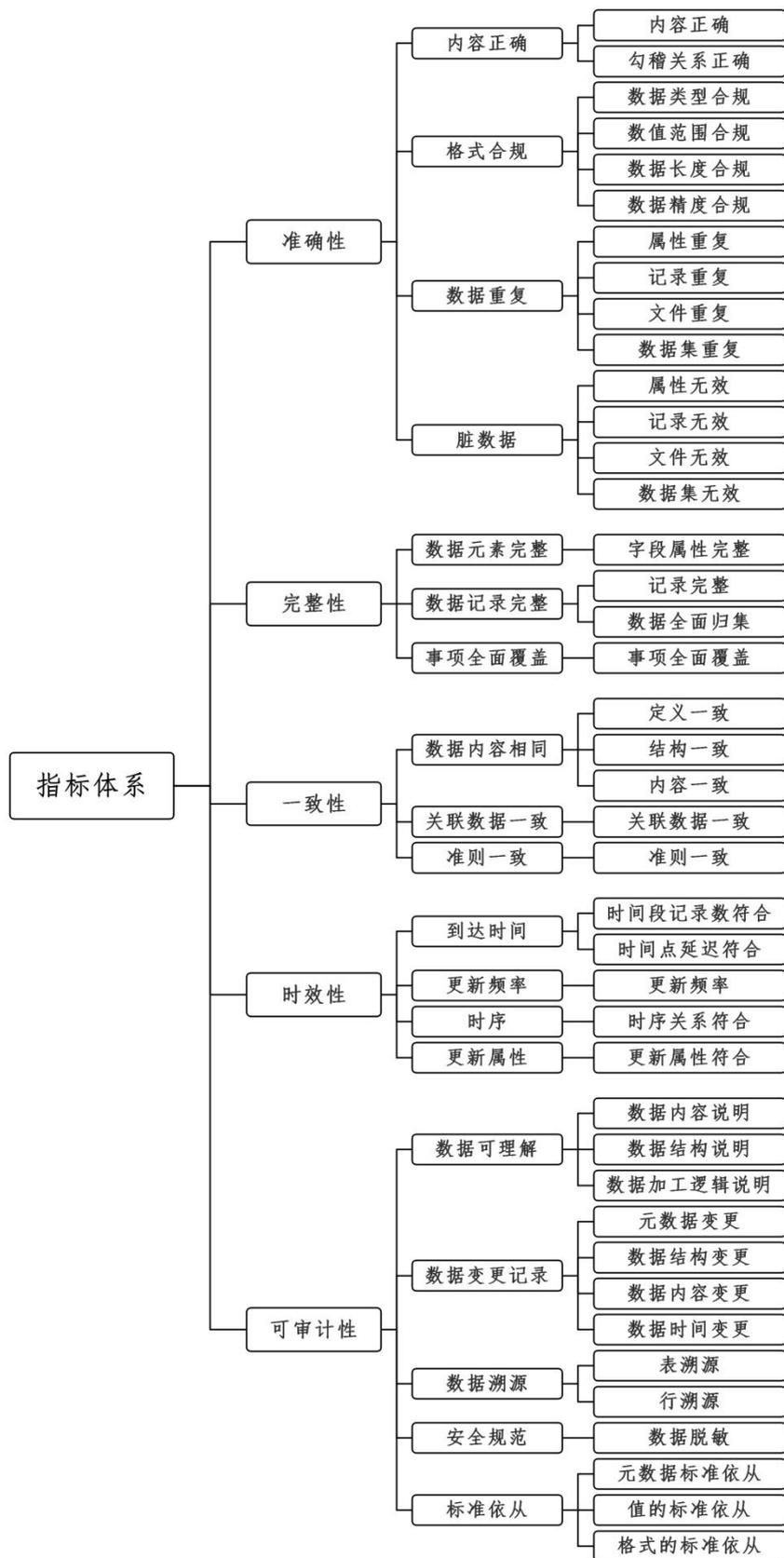


图 7 资讯数据质量评估指标体系

度量已有数据内容的准确性，检查人工输入、跨链路复制等错误。度量数据重复或者数据的属性重复。度量数据的定义、值、格式和展现形式符合定义要求，如范围有效性、日期有效性等。准确性维度下的指标定义如表 5 所示。

表 5 准确性评估指标定义

一级指标	定义	二级指标	计算方法
内容正确	内容是否符合预期	数据内容正确	$X=A/B$ 式中： A=满足数据正确性要求的数据集中数据内容的个数； B=被评价的数据集中数据内容的个数
		勾稽关系正确	$X=A/B$ 式中： A=满足数据正确性要求的数据集中关系的个数； B=被评价的数据集中元素的个数
格式合规	数据类型、数值范围、数据长度、精度等是否符合预期	数据类型合规	$X=A/B$ 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
		数值范围合规	$X=A/B$ 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
		数据长度合规	$X=A/B$ 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
		数据精度合规	$X=A/B$ 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
数据重复	特定字	属性重复	$X=1-A/B$

	段、记录、文件或数据集意外重复的度量		式中： A=重复的数据集中属性的个数； B=被评价的数据集中属性的个数
		记录重复	$X=1-A/B$ 式中： A=重复的数据集中记录的个数； B=被评价的数据集中记录的个数
		文件重复	$X=1-A/B$ 式中： A=重复的数据集中文件的个数； B=被评价的数据集中文件的个数
		数据集重复	$X=1-A/B$ 式中： A=重复的数据集个数； B=被评价的数据集的个数
脏数据	正确字段、记录、文件或数据集之外无效数据的度量	属性无效	$X=1-A/B$ 式中： A=有脏数据出现的数据集中属性的个数； B=被评价的数据集中属性的个数
		记录无效	$X=1-A/B$ 式中： A=有脏数据出现的数据集中记录的个数； B=被评价的数据集中记录的个数
		文件无效	$X=1-A/B$ 式中： A=有脏数据出现的数据集中文件的个数； B=被评价的数据集中文件的个数
		数据集无效	$X=1-A/B$ 式中： A=有脏数据出现的数据集个数； B=被评价的数据集个数

度量数据丢失或与外部信息关联缺失、未建立关联等。  
完整性维度下的指标定义如表 6 所示。

表 6 完整性评估指标定义

一级指标	定义	二级指标	计算方法
数据元素完整	数据集中应被赋值的数据元素的赋值程度	字段属性完整	$X=A/B$ 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
数据记录完整	数据集中应被赋值的数据记录的赋值程度	记录完整	$X=A/B$ 式中： A=被赋值的数据集中记录的个数； B=预期被赋值的数据集中记录的个数
		数据全面归集	检查业务数据表是否有全量历史业务数据
事项全面覆盖	数据集中应被赋值的数据记录的赋值程度	事项全面覆盖	检查业务数据表是否覆盖全部类型实体

度量跨源、库、表数据信息含义上不一致。一致性维度下的指标定义如表 7 所示。

表 7 一致性评估指标定义

一级指标	定义	二级指标	计算方法
数据内容相同	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	定义一致	$X=A/B$ 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
		结构一致	$X=A/B$

	置的同一数据被同步修改		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
		内容一致	$X=A/B$ 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
关联数据一致	根据一致性约束规则检查关联数据的一致性	关联数据一致	$X=A/B$ 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
准则一致	根据一致性约束规则检查符合准则的一致性	准则一致	$X=A/B$ 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

度量单条数据延迟时点、多条数据延迟时段等。时效性维度下的指标定义如表 8 所示。

表 8 时效性评估指标定义

一级指标	定义	二级指标	计算方法
到达时间	基于日期范围的记录数符合业务需求的程度	时间段记录数符合	$X=A/B$ 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
	基于时间戳的延迟时间符合业务需求的程度。	时间点延迟符合	$X=A/B$ 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
更新频率	单位时间内数据更新频率满足业务需求的程度	更新频率	$X=A/B$ 式中： A=数据集单位时间内更新频次数； B=数据集单位时间内频次总数
时序	数据集中数据元素之间满足一定的时序关系	时序关系符合	$X=A/B$ 式中： A=满足时序关系的记录个数； B=被评价的数据集中记录的个数
更新属性	数据集中更新字段满足业务需求的程度	更新属性的要求	$X=A/B$ 式中： A=满足更新字段要求的记录个数； B=被评价的数据集中记录的个数

度量数据可理解（定义、加工逻辑等）、度量数据结构和内容的变更、溯源、安全等。可审计性维度下的指标定义如表 9 所示。

表 9 可审计性评估指标定义

指标项	定义	指标子项	计算方法
数据可理解	是否提供可解读的元数据文档的度量	数据内容说明	$X=A/B$ 式中： A=包含在元数据文档中数据集元素的个数； B=被评价的数据集中元素的个数
		数据结构说明	$X=A/B$ 式中： A=包含在元数据文档中数据集元素的个数； B=被评价的数据集中元素的个数
		数据加工逻辑说明	$X=A/B$ 式中：

			A=包含在元数据文档中数据集元素的个数; B=被评价的数据集中元素的个数
数据变更	基于数据变更留痕的度量。	元数据变更	$X=A / B$ 式中: A=满足数据变更留痕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
		数据结构变更	$X=A / B$ 式中: A=满足数据变更留痕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
		数据内容变更	$X=A / B$ 式中: A=满足数据变更留痕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
		数据时间变更	$X=A / B$ 式中: A=满足数据变更留痕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
数据溯源	基于数据可溯源的度量	表溯源	$X=A / B$ 式中: A=满足数据溯源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
		行溯源	$X=A / B$ 式中: A=满足数据溯源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
安全规	安全规范是安全和	数据脱敏	$X=A / B$

范	隐私方面的规则，包括数据权限管理，数据脱敏处理等		式中： A=满足安全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
标准依从性	对标准的依从程度	值的标准依从	$X=A/B$ 式中： A=满足标准的数据集中记录的个数； B=被评价的数据集中记录的个数
		格式的标准依从	$X=A/B$ 式中： A=满足标准的数据集中记录的个数； B=被评价的数据集中记录的个数
		元数据的标准依从	$X=A/B$ 式中： A=满足标准的数据集中记录的个数； B=被评价的数据集中记录的个数

#### (四) 资讯数据质量评估计算方法

在上述资讯数据质量评估体系的基础上，为度量资讯数据质量的“好”与“坏”，需进一步量化计算综合指标。

首先，针对指标项按满分 100 分进行评分。指标项百分制得分=质量规则符合率\*100。

维度得分= $\sum$ (各指标项百分制得分\*均分权重)，计算结果修约到小数点后 2 位。

质量综合得分= $\sum$ (维度得分\*维度权重)，满分为 100 分，修约到小数点后 2 位。

实际执行过程中，填写资讯数据质量评估打分表，如表 10 所示。

表 10 资讯数据质量评估打分表

维度	一级指标	得分	二级指标	得分	质量规则	公式	取数方式	A 取值	B 取值
准确性	内容正确		内容正确		数据内容正确条数	$A/B*100$			
			勾稽关系正确		证券内码表关联正确数	$A/B*100$			
	格式合规		数据类型		数据库字段与数据字典定义的数据类型相同数	$A/B*100$			
			取值范围		取值范围合理、符合字段枚举的字段数量	$A/B*100$			
			数据长度		数据长度符合的字段数量	$A/B*100$			
			精度		数据精度符合的字段数量	$A/B*100$			
	数据重复		字段重复		字段含义重复	$(1-A/B)*100$			
			记录重复		记录重复	$(1-A/B)*100$			
			文件重复		文件重复	$(1-A/B)*100$			
			数据集重复		数据集包含在其他表中	$(1-A/B)*100$			
	脏数据		字段无效		数据字段中无效字段数量	$(1-A/B)*100$			
			记录无效		数据记录中无效记录数量	$(1-A/B)*100$			
			文件无效		文件无效	$(1-A/B)*100$			
			数据集无效		数据集无效	$(1-A/B)*100$			
准确性维度得分		维度权重							
...		...							
可审计性维度得分		维度权重							
综合得分									
质量等级									

### (五) 资讯数据质量评估等级划分

资讯数据质量评估等级是在量化综合指标的基础上，通过等级划分区间来衡量的质量水平。设定等级划分区间范围需组织数据生产方、加工方、消费方共同研讨，达成一致。

在课题组范围内已形成如表 11 所示等级划分。将资讯数据质量描述为 A、B、C、D、E 五个等级，并给出对应等级的资讯数据质量评估结果说明。

表 11 资讯数据质量评估等级划分

等级	A	B	C	D	E
说明	资讯数据质量评估结果完全达到数据质量要求，且基本不存在数据质量问题。 $90 \leq$ 综合指标	资讯数据质量评估结果基本达到数据质量要求，存在少量的数据质量问题，影响较小。 $80 \leq$ 综合指标 < 90	资讯数据质量评估结果达到数据质量要求，存在较多的数据质量问题，影响一般。 $70 \leq$ 综合指标 < 80	资讯数据质量评估结果不能达到数据质量要求，存在大量的数据质量问题，影响较大。 $60 \leq$ 综合指标 < 70	未开展资讯质量评估，或质量评估结果远不能达到数据质量要求，存在严重的数据质量问题。 综合指标 < 60

#### （六）《证券期货业资讯数据质量评估方法》（标准草案）

以上述研究结果为基础，结合资讯数据的特点，课题组编制了《证券期货业资讯数据质量评估方法》（标准草案），通过标准的方式规定了证券期货业资讯数据质量评估框架、指标、等级划分、评估方法等，适用于资讯数据生存周期各个阶段的数据质量评估。

实际开展资讯数据质量评估过程中，可依据《证券期货业资讯数据质量评估方法》标准，编制配套的《资讯数据质量评估方案》，在评估方案中通过选取评估指标、设定质量

规则、确定维度权重等方式，指导资讯数据质量评估工作的落地实施。同时，评估方案支持按周期灵活调整配置，结合当下行业资讯数据质量管理痛点，实现有针对性地引导行业开展资讯数据质量评估工作。以评促建，以评促改，逐渐提升资讯数据质量，对行业资讯数据应用的持续高质量发展具有重要意义。

## 四、研究实践

在资讯数据中，股指期货行情数据扮演着至关重要的角色，它不仅是投资者决策的重要依据，也是市场参与者进行风险管理和资产配置的关键工具。股指期货行情能够为投资者提供即时的市场动态信息，通过分析股指期货的价格波动，投资者可以迅速捕捉到市场情绪的变化，从而做出相应的投资决策。

课题组以股指期货行情数据质量评估场景为例，实践资讯数据质量评估过程，验证资讯数据质量评估方法的可行性。

### （一）定义评估的目标和范围

通过识别关键数据确定股指期货行情数据质量评价目标为股指期货每日行情数据，范围包括沪深 300 股指期货、中证 500 股指期货、中证 1000 股指期货、上证 50 股指期货全量日行情数据。属性包括合约代码、今开盘、最高价、最低价、成交量、成交金额、持仓量、持仓变化、今收盘、今结算、前结算、涨跌 1、涨跌 2、日期等。

### （二）确定评估数据

确认权威数据来源为中国金融期货交易所官方网站，数据栏目下行情日统计数据。如图 8 所示。

中国金融期货交易所  
China Financial Futures Exchange

输入关键字

首页 | 新闻公告 | 产品 | 数据 | 服务 | 法律法规 | 关于

首页 > 数据 > 行情数据 > 日统计

数据 Data 日统计 期货 期权 全部 日期: 2024-11-07 查询

合约代码	今开盘	最高价	最低价	成交量	成交金额	持仓量	持仓变化	今收盘	今结算	前结算	涨跌1	涨跌2
IF2411	4,000.0	4,172.0	4,000.0	52,134	6,375,049.722	51,158	8	4,170.0	4,135.2	4,033.6	136.4	101.6
IF2412	4,020.0	4,185.0	4,013.4	126,886	15,574,492.536	175,841	13,805	4,183.8	4,147.4	4,042.6	141.2	104.8
IF2503	4,030.0	4,195.2	4,020.4	26,474	3,256,344.240	67,014	2,212	4,195.0	4,157.4	4,051.6	143.4	105.8
IF2506	4,001.8	4,189.0	4,001.8	7,694	943,167.714	14,611	835	4,188.0	4,147.2	4,034.4	153.6	112.8

图 8 中国金融期货交易所官方网站

评估数据集来自市场上实际资讯供应商（以下代称为甲、乙、丙），选取资讯供应商甲、乙、丙资讯数据库中，符合评估目标范围的全量日行情数据集。此外，因资讯供应商甲、乙、丙股指期货数据结构差异，设计取数逻辑拉齐方式，处理数据表内容范围不一致、字段命名不统一等问题。

### （三）选取数据质量评估指标

结合资讯数据集-股指期货日行情实际的数据特点，如数据集以数据表形式存储，不涉及存储形式为文件的有关指标等，选取符合数据集特点的评价指标 26 个，如表 12 所示。

表 12 指标选取

维度	一级指标	二级指标
准确性	内容正确	内容正确
		勾稽关系正确
	格式合规	数据类型
		取值范围
		数据长度
		精度
	数据重复	字段重复
		记录重复
		数据集重复

	脏数据	字段无效
		记录无效
		数据集无效
完整性	字段完整	字段完整
	记录完整	记录完整
		数据全面归集
事项完整	事项全面覆盖	
时效性	到达时间 更新频率	时间点延迟时间符合
		更新频率
可审计性	数据可理解	数据内容说明
		数据结构说明
		数据加工逻辑说明
	数据变更	元数据变更
		数据结构变更
		数据内容变更
	数据溯源	表溯源
行溯源		

#### (四) 定义数据质量规则

指标的计算是在规则的基础上进行，因此，在上述确定评估指标的基础上，明确指标下具体规则来评价数据。包括面向表的规则和面向字段的规则，如数据内容正确条数、数据长度符合的字段的数量等。规则分为通用规则和面向数据特点的规则，如加工衍生的字段具有计算逻辑说明的数量、数据记录源头发布时间等。如表 13 所示。

表 13 质量规则

维度	一级指标	二级指标	质量规则
准确性	内容正确	内容正确	数据内容正确条数
		勾稽关系正确	证券内码表关联正确数
	格式合规	数据类型	数据库字段与数据字典定义的数据类型相同数
		取值范围	取值范围合理、符合字段枚举的字段数量
		数据长度	数据长度符合的字段数量

	数据重复	精度	数据精度符合的字段数量
		字段重复	字段含义重复
		记录重复	记录重复
	脏数据	数据集重复	数据集包含在其他表中
		字段无效	数据字段中无效字段数量
		记录无效	数据记录中无效记录数量
完整性	数据集无效	数据集无效	
	字段完整	字段完整	数据字段为空，字段缺失率
	记录完整	记录完整	记录条数
时效性	到达时间更新频率	数据全面归集	历史数据年份覆盖
		事项完整	事项全面覆盖
可审计性	数据可理解	时间点延迟时间符合更新频率	基于权威数据源发布的延迟实时、准实时等
		数据内容说明	说明数据内容、数据范围、数据来源、更新频率、样例数据中具有的数量
		数据结构说明	列名、中文名称、类型、单位、空否、枚举、主键、外键关联中具有的数量
	数据变更	数据加工逻辑说明	加工衍生的字段具有计算逻辑说明的数量
		元数据变更	变更记录情况
		数据结构变更	变更记录情况
	数据溯源	数据内容变更	变更记录情况
		表溯源	数据表来源
		行溯源	数据行来源

### （五）设定质量权重

当前，股指期货日行情数据的准确、完整直接影响了投资者决策的有效性。数据时效方面，源头的数据发布时间稳定，且不涉及复杂的加工处理逻辑，该类数据鲜有发生数据延迟等问题。因此，在实际评估方案中，维度权重划分侧重考虑评估准确性与完整性。设定维度权重：准确性 30%、完整性 30%、时效性 20%、可审计性 20%。同时，为全面刻画资讯的数据质量，选取的评估指标项之间没有倾向性，各指

标项间均分权重。

### (六) 实施质量评估

采用手工、自动等手段校验数据集质量规则的符合程度，将结果记录在表 10 中。评估时间为 2024 年 10 月。截取资讯供应商丙部分评估结果，如表 14 所示。

表 14 资讯供应商丙股指期货日行情数据质量打分

维度	一级指标	得分	二级指标	得分	质量规则	公式	取数方式	A 取值	B 取值
准确性	内容正确	100	内容正确	100	数据内容正确条数	$A/B*100$	抽样 2023 年数 据	6688	6688
			勾稽关系正确	100	证券内码表关联正确数	$A/B*100$	全量	510	510
	格式合规	92.85	数据类型	96.4	数据库字段与数据字典定义的数据类型相同数	$A/B*100$	全量	27	28
			取值范围	100	取值范围合理、符合字段枚举的字段数量	$A/B*100$	全量	22	22
			数据长度	100	数据长度符合的字段数量	$A/B*100$	全量	16	16
			精度	75	数据精度符合的字段数量	$A/B*100$	全量	12	16
	数据重复	98.2	字段重复	96.4	字段含义重复	$(1-A/B)*100$	全量	1	28
			记录重复	100	记录重复	$(1-A/B)*100$	抽样 2020- 2024	0	32400
			文件重复	不涉及	文件重复	$(1-A/B)*100$			
			数据集重复	100	数据集包含在其他表中	$(1-A/B)*100$	全量	0	1
	脏数据	97.4	字段无效	96.4	数据字段中无效字段数量	$(1-A/B)*100$	全量	1	28
			记录无效	95.7	数据记录中无效记录数量	$(1-A/B)*100$	抽样 2020- 2024	33855- 32400	33855
			文件无效	不涉及	文件无效	$(1-A/B)*100$			
			数据集无效	100	数据集无效	$(1-A/B)*100$	全量	0	1

准确性维度得分	97.11	维度权重	30%
...	...	...	...
可审计性维度得分	62	维度权重	20%
综合得分	81.68		
质量等级	B		

### （七）数据质量评分及定级

通过对资讯数据质量核检的结果进行汇总评分，数据质量评分采用简单比率法和加权平均法相结合的方法进行计算，评分按照规则级、表级进行数据质量分计算。数据质量等级根据数据质量评分划分为 A、B、C、D、E。计算质量综合得分：甲 79.97、乙 78.48、丙 81.68。质量等级划分：甲 C 级，乙 C 级，丙 B 级。

### （八）数据质量问题分析

通过股指期货数据质量核验，梳理出数据质量问题，并对数据质量问题进行根因分析，找出数据质量问题存在的根本原因。部分问题指标统计情况如图 9 所示。

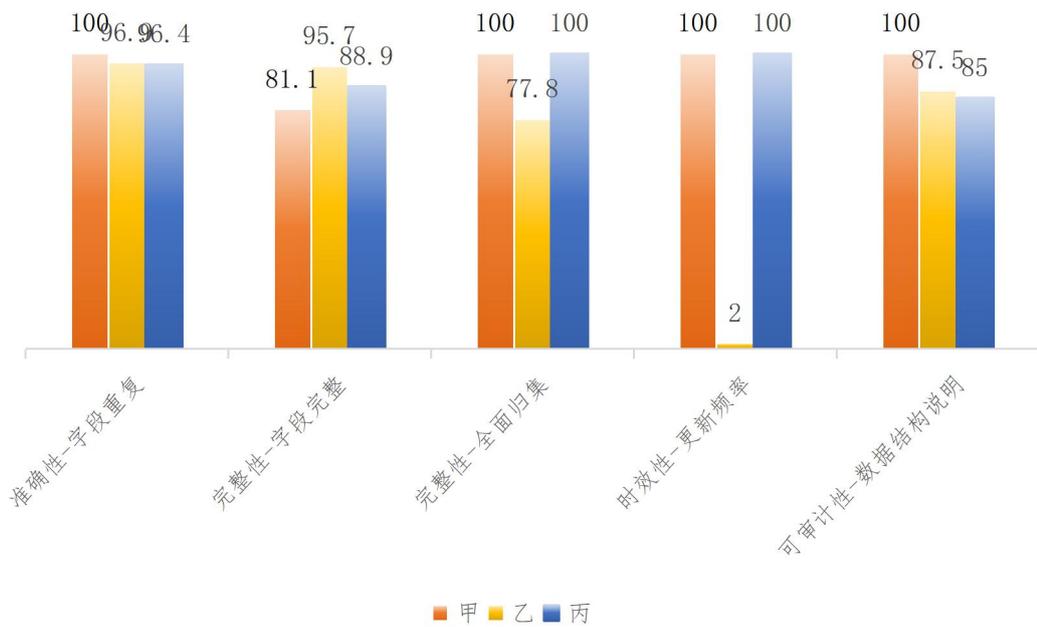


图 9 问题指标统计

识别出的典型问题:

①在准确性-字段重复指标方面，乙存在 ID、JSID 等唯一标识字段，丙存在 TMSTAMP、ENTRYDATE、ENTRYTIME 等更新时间字段，存在字段含义重复问题。

②在完整性-字段完整指标方面，甲缺失持仓量变化、涨跌幅字段，丙存在 12%的空值。

③在完整性-全面归集指标方面，甲、丙覆盖历史全量数据 2006-2024，乙缺失 2010 以前数据。

④在时效性-更新频率指标方面，甲准实时更新、丙半小时更新、乙天更新，与实时更新的质量目标存在差距。

⑤在可审计性-数据结构说明指标方面，乙、丙缺少单位、空否、外键约束等说明。

### (九) 质量结果反馈

最终，将对股指期货日行情数据质量评价结果反馈给数据管理方和数据使用方进行评审和确认，由数据提供方认领资讯的数据质量问题，并制定数据质量问题整改计划，经整改后的数据质量问题应由数据团队进行验证，验证通过后方可继续开发利用，从而对资讯数据质量管理工作进行持续性提升。

## 五、研究结论与建议

（一）通过构建行之有效的资讯数据质量评估体系，为资讯数据质量的提升提供支撑

为进一步推动证券期货业资讯数据质量的提升，课题组深入开展了构建资讯数据质量评估标准的研究工作。通过广泛研究国内外质量维度及相关国家行业标准文件，并结合行业资讯数据的实际应用情况，课题组深入分析了行业已有的资讯数据标准化等规范，设计了一套科学、全面的资讯数据质量评估体系。这一体系填补了行业在资讯数据质量评估领域的空白，为行业资讯数据质量的提升提供了有力支撑。

（二）通过落地评估体系实践应用，充分验证评估体系可行性和实用性

为确保所构建的资讯数据质量评估体系的可操作性和有效性，课题组选取了具有代表性的资讯数据使用场景进行实践应用。通过综合指标计算模型和质量评估体系的实际应用，课题组进行了模拟评估，并验证了评估结果的客观性和公正性。这一实践应用不仅为后续评估体系的改进提供了重要依据，也进一步增强了评估体系的可行性和实用性。

（三）通过评估体系标准化转化，促进行业资讯数据质量水平进一步提升

在充分验证研究成果的基础上，课题组积极寻求与监管部门、行业协会等机构的合作，致力于推动研究成果的标准

化、规范化。通过加强与行业各方的沟通与协作，课题组将努力促进评估体系在更广泛领域内的应用和推广。同时，本研究基于广泛的行业调研，研究团队横跨产学研多个领域，具有较高的科学性和可靠性。因此，将研究成果转化为行业标准后，不仅能够填补行业空白，还能够为提高行业资讯数据的质量起到积极的促进作用，为证券期货业的高质量发展贡献力量。

## 【参考文献】

- [1] 无.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[J].中华人民共和国国务院公报, 2023(1):28-33.
- [2] 中国证监会 国家标准委关于加强证券期货业标准化工作的指导意见
- [3] DAMA International.DAMA 数据管理知识体系指南[M].清华大学出版社,2012.
- [4] 20141203-T-469,信息技术 数据质量评价指标[S].
- [5] 全国金融标准化技术委员会证券分技术委员会 证券期货业数据标准规划（2022-2025）
- [6] JR/T 0275-2023,证券期货业机构内部接口 资讯数据[S].
- [7] Sebastian-Coleman L .Features of the DQAF Measurement Logical Model[M].Elsevier Inc. 2013.
- [8] GB/T 41795-2022,质量技术基础 信息资源数据规范[S].
- [9] NB/T 11083-2023,风电信息管理数据质量评估及治理技术规范[S].
- [10] DB/T 97-2024, 地震观测数据质量评价规范全球导航卫星系统观测》 [S].
- [11] CY/T 239-2021,ISLI 数据质量检测[S].
- [12] 华为数据之道[M].机械工业出版社,2020.

角色	姓名	单位	职务
课题负责人	孟晋津	中信建投证券股份有限公司	信息技术部总监
	韩忠明	北京工商大学	副教授
	李云涛	中信建投证券股份有限公司	研发工程师
课题成员	蔡跃	中信建投证券股份有限公司	研发工程师
	郭城	中信建投证券股份有限公司	数据工程师
	褚丽恒	申万宏源证券有限公司	数据工程师
	侯韩芳	中国标准化研究院	副所长
	王春艳	中国标准化研究院	研究员
	黎峰	证通股份有限公司	副总裁
	吴征	上海大智慧财汇数据科技有限公司	数据开发负责人

# 中华人民共和国 X X X X X 标准

XX/T XXXXX—XXXX

代替 XX/T

## 证券期货业资讯数据质量评估方法

Securities and Futures Industry Information Data Quality Assessment Method

(点击此处添加与国际标准一致性程度的标识)

(工作组讨论稿)

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

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
4 资讯数据质量评估指标框架 .....	1
5 指标说明 .....	3
5.1 指标编码 .....	3
5.2 准确性 .....	3
5.3 完整性 .....	5
5.4 一致性 .....	6
5.5 时效性 .....	7
5.6 可审计性 .....	8
6 资讯质量维度指标评估计算方法 .....	10
7 资讯质量维度评估等级划分 .....	11
附 录 A      (资料性) 质量评估规则 .....	12

## 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××××提出。

本文件由××××归口。

本文件起草单位：

本文件主要起草人：

## 引 言

随着金融科技的快速发展和金融市场的日益复杂化，资讯数据作为基础的数据要素，其共享和使用已成为证券行业不可或缺的重要组成部分。然而，由于数据源头众多、机构间数据交换链路多样、数据加工处理复杂等原因，导致资讯数据质量存在较大差异，出现数据内容错误、数据缺失、数据重复、格式不统一、数据不及时等数据质量问题，资讯数据要素价值释放面临着较大的风险和挑战。因此，针对资讯数据的特点，制定一套数据质量评估方法，以评促建，以评促改，逐渐提升资讯数据质量，对行业资讯数据应用的高质量发展具有重要意义。

# 证券期货业资讯数据质量评估方法

## 1 范围

本文件规定了证券期货业资讯数据质量评估维度、指标、等级划分、质量规则、评估方法等。  
本文件适用于资讯数据生存周期各个阶段的数据质量评估。

## 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

JR/T 0275—2023 证券期货业机构内部接口 资讯数据

## 3 术语和定义

下列术语和定义适用于本文件。

### 3.1

#### **金融资讯 financial information**

影响金融投资有关的市场主体、财政政策、社会环境等公开信息。

注1：金融资讯信息包含但不限于基础信息（公司、证券、市场、行业）、股票信息、基金信息、债券信息、理财产品信息、金融衍生品信息、宏观信息等。

注2：金融资讯信息包含但不限于数据、文本、图片、音频、视频等。

### 3.2

#### **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。

### 3.3

#### **数据集 dataset**

具有一定主题，可以标识并可以被计算机化处理的数据集合。

## 4 资讯数据质量评估指标框架

资讯数据质量评估指标框架包括准确性、完整性、一致性、时效性、可审计性等5个指标维度，内容正确、格式合规、数据重复、脏数据、数据元素完整、更新频率、数据溯源等19个指标项，数据类型、数值范围、数据长度、属性重复、记录无效、字段属性、加工逻辑、更新频率等41个指标子项。

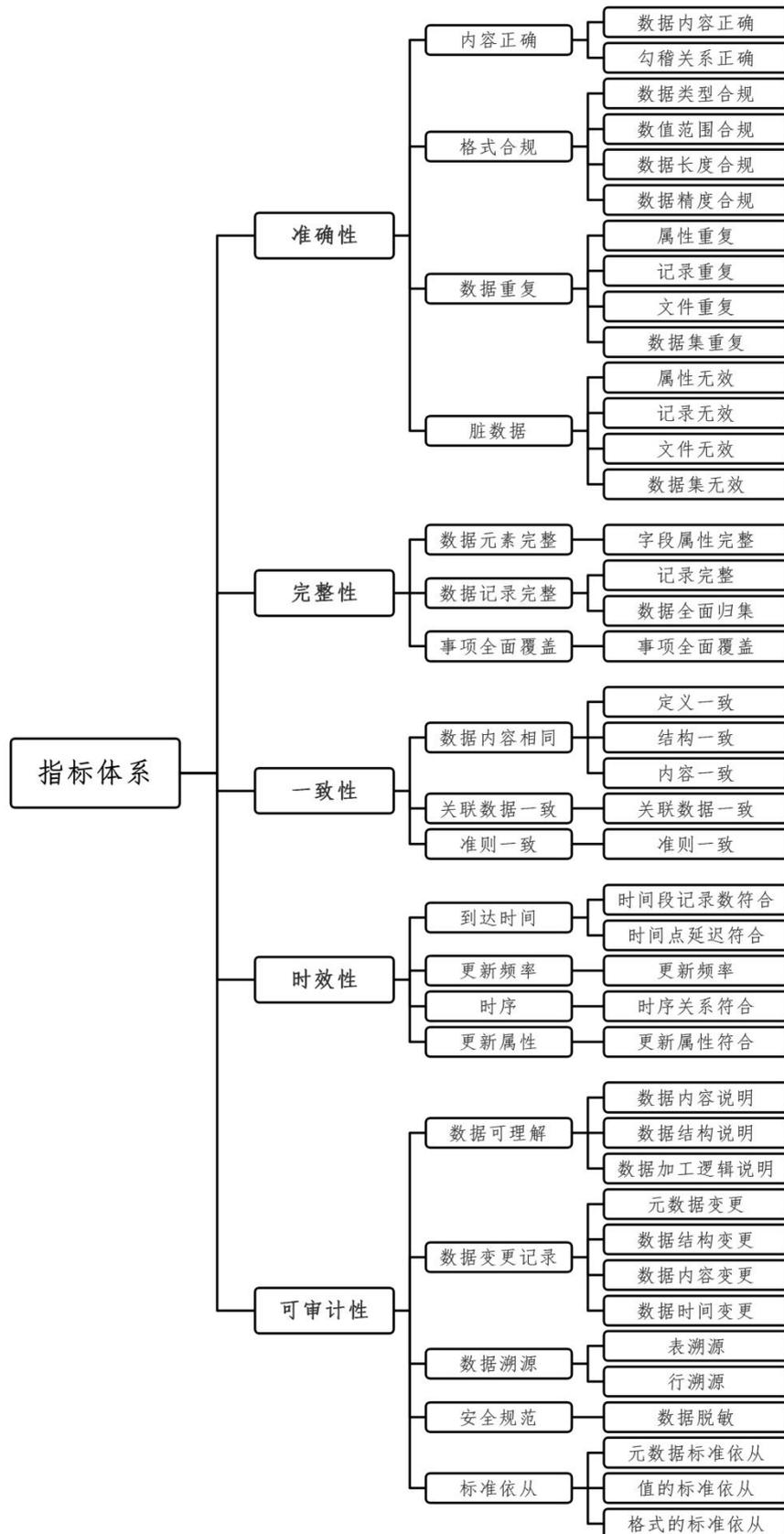


图 1 资讯数据质量评估指标框架

## 5 指标说明

### 5.1 指标编码

XX编码规则。

### 5.2 准确性

度量已有数据内容的准确性，检查人工输入、跨链路复制等错误。度量数据重复或者数据的属性重复。度量数据的定义、值、格式和展现形式符合定义要求。如范围有效性、日期有效性等。准确性评估指标定义见表1。

表 1 准确性评估指标

一级指标	定义	二级指标	计算方法
内容正确	内容是否符合预期	数据内容正确	$X = A / B$ 式中： A=满足数据正确性要求的数据集中数据内容的个数； B=被评价的数据集中数据内容的个数

表1 准确性评估指标（续）

一级指标	定义	二级指标	计算方法
		勾稽关系正确	$X=A/B$ 式中： A=满足数据正确性要求的数据集中关系的个数； B=被评价的数据集中关系的个数
格式合规	数据类型、数值范围、数据长度、精度等是否符合预期	数据类型合规	$X=A/B$ 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
		数值范围合规	$X=A/B$ 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
		数据长度合规	$X=A/B$ 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
		数据精度合规	$X=A/B$ 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
数据重复	特定字段、记录、文件或数据集意外重复的度量	属性重复	$X=1-A/B$ 式中： A=重复的数据集中属性的个数； B=被评价的数据集中属性的个数
		记录重复	$X=1-A/B$ 式中： A=重复的数据集中记录的个数； B=被评价的数据集中记录的个数

表1 准确性评估指标（续）

一级指标	定义	二级指标	计算方法
		文件重复	$X=1-A/B$ 式中： A=重复的数据集中文件的个数； B=被评价的数据集中文件的个数
		数据集重复	$X=1-A/B$ 式中： A=重复的数据集个数； B=被评价的数据集个数
脏数据	正确字段、记录、文件或数据集之外无效数据的度量	属性无效	$X=1-A/B$ 式中： A=有脏数据出现的数据集中属性的个数； B=被评价的数据集中属性的个数
		记录无效	$X=1-A/B$ 式中： A=有脏数据出现的数据集中记录的个数； B=被评价的数据集中记录的个数
		文件无效	$X=1-A/B$ 式中： A=有脏数据出现的数据集中文件的个数； B=被评价的数据集中文件的个数
		数据集无效	$X=1-A/B$ 式中： A=有脏数据出现的数据集个数； B=被评价的数据集个数

### 5.3 完整性

度量数据丢失或与外部信息关联缺失、未建立关联等。完整性评估指标定义见表2。

表 2 完整性评估指标

一级指标	定义	二级指标	计算方法
数据元素完整	数据集中应被赋值的数据元素的赋值程度	字段属性完整	$X=A/B$ 式中： A=被赋值的数据集中属性的个数； B=预期被赋值的数据集中属性的个数
数据记录完整	数据集中应被赋值的数据记录的赋值程度	记录完整	$X=A/B$ 式中： A=被赋值的数据集中记录的个数； B=预期被赋值的数据集中记录的个数
		数据全面归集	检查业务数据表是否有全量历史业务数据
事项全面覆盖	数据集中应被赋值的数据记录的赋值程度	事项全面覆盖	检查业务数据表是否覆盖全部类型实体

## 5.4 一致性

度量跨源、库、表数据信息含义上不一致。一致性评估指标定义见表3。

表 3 一致性评估指标

一级指标	定义	二级指标	计算方法
数据内容相同	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	定义一致	$X=A/B$ 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
		结构一致	$X=A/B$ 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

表3 一致性评估指标（续）

一级指标	定义	二级指标	计算方法
		内容一致	$X=A/B$ 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
关联数据一致	根据一致性约束规则检查关联数据的一致性	关联数据一致	$X=A/B$ 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
准则一致	根据一致性约束规则检查准则的一致性	准则一致	$X=A/B$ 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

## 5.5 时效性

度量单条数据延迟时点、多条数据延迟时段等。时效性评估指标定义见表4。

表4 时效性评估指标

一级指标	定义	二级指标	计算方法
到达时间	基于日期范围的记录数符合业务需求的程度	时间段记录数符合	$X=A/B$ 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
	基于时间戳的延迟时间符合业务需求的程度。	时间点延迟符合	$X=A/B$ 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

表5 时效性评估指标（续）

一级指标	定义	二级指标	计算方法
更新频率	单位时间内数据更新频率满足业务需求的程度	更新频率	$X=A/B$ 式中： A=数据集单位时间内更新频次数； B=数据集单位时间内频次总数
时序	数据集中数据元素之间满足一定的时序关系	时序关系符合	$X=A/B$ 式中： A=满足时序关系的记录个数； B=被评价的数据集中记录的个数
更新属性	数据集中更新字段满足业务需求的程度	更新属性的要求	$X=A/B$ 式中： A=满足更新字段要求的记录个数； B=被评价的数据集中记录的个数

## 5.6 可审计性

度量数据可理解（定义、加工逻辑等）、度量数据结构和内容的变更、溯源、安全等。可审计性评估指标定义见表5。

表6 可审计性评估指标

一级指标	定义	二级指标	计算方法
数据可理解	是否提供可解读的元数据文档的度量	数据内容说明	$X=A/B$ 式中： A=包含在元数据中数据集元素的个数； B=被评价的数据集中元素的个数
		数据结构说明	$X=A/B$ 式中： A=包含在元数据文档中数据集元素的个数； B=被评价的数据集中元素的个数
		数据加工逻辑说明	$X=A/B$ 式中： A=包含在元数据文档中数据集元素的个数； B=被评价数据集中元素的个数

表7 可审计性评估指标（续）

一级指标	定义	二级指标	计算方法
数据变更	基于数据变更留痕的度量。	元数据变更	$X=A / B$ 式中： A=满足数据变更留痕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
		数据结构变更	$X=A / B$ 式中： A=满足数据变更留痕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
		数据内容变更	$X=A / B$ 式中： A=满足数据变更留痕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
		数据时间变更	$X=A / B$ 式中： A=满足数据变更留痕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

表8 可审计性评估指标（续）

一级指标	定义	二级指标	计算方法
数据溯源	基于数据可溯源的度量	表溯源	$X=A/B$ 式中： A=满足数据溯源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
		行溯源	$X=A/B$ 式中： A=满足数据溯源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
安全规范	安全规范是安全和隐私方面的规则，包括数据权限管理，数据脱敏处理等	数据脱敏	$X=A/B$ 式中： A=满足安全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
标准依从性	对标准的依从程度	值的标准依从	$X=A/B$ 式中： A=满足标准的数据集中记录的个数； B=被评价的数据集中记录的个数
		格式的标准依从	$X=A/B$ 式中： A=满足标准的数据集中记录的个数； B=被评价的数据集中记录的个数
		元数据的标准依从	$X=A/B$ 式中： A=满足标准的数据集中记录的个数； B=被评价的数据集中记录的个数

## 6 资讯质量维度指标评估计算方法

相同指标项下的数据质量规则评价目标一致，子指标通过累加进行计算，数据质量规则异常的数量/被评价数据的总量。

指标项取值= $\Sigma$ 子指标数据质量规则异常数量/ $\Sigma$ 被评价数据量

相同维度下，不同指标项间，为避免重要的指标项被“淹没”，采用加权的方式综合考虑维度指标取值。

质量维度综合指标=Σ(权重\*同维度下各指标项取值)

## 7 资讯质量维度评估等级划分

资讯数据质量描述为A、B、C、D、E五个等级，并给出对应等级的资讯数据质量评估结果说明。

表9 资讯数据质量评估等级划分

等级	A	B	C	D	E
说明	资讯数据质量评估结果完全达到数据质量要求，且基本不存在数据质量问题。 90≤综合指标	资讯数据质量评估结果基本达到数据质量要求，存在少量的数据质量问题，影响较小。 80≤综合指标<90	资讯数据质量评估结果达到数据质量要求，存在较多的数据质量问题，影响一般。 70≤综合指标<80	资讯数据质量评估结果不能达到数据质量要求，存在大量的数据质量问题，影响较大。 60≤综合指标<70	未开展资讯质量评估，或质量评估结果远不能达到数据质量要求，存在严重的数据质量问题。 综合指标<60

附 录 A  
(资料性)  
质量评估规则

表A.1给出了资讯信息质量评估规则

表 A.1 质量评估规则

维度	一级指标	二级指标	质量规则	统计方式	
准确性	内容正确	数据内容正确	针对数据集抽样 5%条，与真实数据对比各属性值，计算内容正确率	抽样	
		勾稽关系正确	根据实际表间勾稽关系，设置校验规则	全量	
	格式合规	数据类型合规	根据实际业务规则，设置校验规则	全量	
		数值范围合规	根据实际业务规则，设置校验规则	全量	
		数据长度合规	根据实际业务规则，设置校验规则	全量	
		数据精度合规	根据实际业务规则，设置校验规则	全量	
	数据重复	字段重复	业务表字段重复	全量	
		记录重复	业务表主键值重复	全量	
		文件重复	文件重复	全量	
		数据集重复	数据集重复	全量	
	脏数据	字段无效	数据集中非重复的字段数/应包含字段个数大于 1	全量	
		记录无效	数据集中非重复的记录数/应包含实体个数大于 1	全量	
		文件无效	数据集中非重复的文件数/应包含文件个数大于 1	全量	
		数据集无效	非重复数据集个数/应包含数据集个数大于 1	全量	
	完整性	数据元素完整	字段属性完整	根据实际业务规则，抽样部分字段设置应被赋值校验规则	抽样

表A.1 质量评估规则（续）

维度	一级指标	二级指标	质量规则	统计方式
	数据记录完整	数据记录完整	数据记录完整率=数据集中数据记录数/应包含实体个数（如基本信息表等）； 数据记录完整率=数据集中数据记录数/（特定日期）应包含实体个数（如日行情表等）	全量
		事项全面归集	全面归集完整率=数据集中历史数据年份/应包含全量历史数据年份	全量
	事项全面覆盖	事项全面覆盖	事项全面覆盖率=数据集中包含实体类型数据/应包含全量实体类型数量	全量
一致性	数据内容相同	定义一致	选取同一数据在不同位置存储的多份数据，数据含义应完全一致	全量
		结构一致	选取同一数据在不同位置存储的多份数据，数据结构应完全一致	全量
		内容一致	选取同一数据在不同位置存储的多份数据，数据结构、内容应完全一致	全量
	关联数据一致	关联数据一致	存在关联数据时，关联数据满足“相同数据”质量规则	全量
	准则一致	准则一致	选取同一数据在不同位置存储的多份数据，遵循准则应完全一致	全量
时效性	到达时间	时间段记录数符合	基于日期范围的记录数/符合业务需求的记录数=1	全量
		时间点延迟符合	基于时间点的记录数/符合业务需求的记录数=1	全量
	更新频率	更新频率	实时更新，准实时更新，周期更新	全量
	时序	时序关系符合	数据记录按时序到达	全量
	更新属性	延迟时间符合	业务需求时间戳大于等于数据到达时间戳	全量
可审计性	数据可理解	数据内容说明	提供可解读的数据字典，包含数据内容说明（表、字段注释、枚举等）	全量
		数据结构说明	提供可解读的数据字典，包含数据结构说明、数据类型定义等	全量
		数据加工逻辑说明	提供可解读的数据字典，包含计算公式等加工逻辑	全量

表A.1 质量评估规则（续）

维度	一级指标	二级指标	质量规则	统计方式
	数据变更记录	元数据变更	提供元数据变更留痕信息	全量
		数据结构变更	提供数据结构变更留痕信息	全量
		数据内容变更	提供数据内容变更留痕信息	全量
		数据时间变更	提供数据时间变更留痕信息	全量
	数据溯源	表溯源	提供表级数据来源信息	全量
		行溯源	提供行级数据来源信息	全量
	安全规范	数据脱敏	满足数据脱敏要求	全量
	标准遵从	元数据标准遵从	遵从元数据标准要求	全量
		值的标准遵从	遵从属性值标准要求	全量
		格式的标准遵从	遵从格式标准要求	全量