

Q/GUOSEN

国信证券股份有限公司企业标准

Q/GUOSEN-ENG-DAG 01-2024

国信证券智能化数据管理工具企业标准

2024-08-29 发布

2024-09-02 实施

国信证券股份有限公司

发布

前 言

本标准于 2024 年 8 月初次起草发布，9 月首次实施，在编制过程中得到各位同仁的大力支持，在此表示感谢！由于编制时间紧促，难免有遗漏之处，恳请广大参阅者多提高贵意见，以供参考校正。

本标准由国信证券股份有限公司提出。

本标准起草单位：国信证券股份有限公司。

本标准主要起草人：李明军、左银康、罗何汇。

本标准于 2024 年 8 月首次发布。

1 范围

本标准规定了智能化数据管理工具在数据质量管理，数据标准管理，元数据管理，数据资产管理，数据安全等领域的标准约束说明，有助于企业规范智能化数据管理工具的建设，提升企业数据管理的品质。

本标准适用于对数据团队的智能化数据管理工具能力进行评价和指导。

2 规范性引用文件

本文件没有规范性引用文件。

3 术语和缩略语

国信证券智能化数据管理工具名词条解释。

表 1 术语说明

术语、缩略语		解释
数据质量管理		数据质量管理是指在数据的整个生命周期中，包括计划、获取、存储、共享、维护、应用和消亡等各个阶段，对可能引发的各类数据质量问题进行识别、度量、监控和预警的一系列管理活动。
数据质量管理	规则智能化配置	规则智能化配置是指根据数据特征和业务需求，利用智能化技术（如规则引擎、人工智能算法等）来实现数据质量规则的灵活管理和自动执行的方法。
	质量自动化监控	质量自动化监控是指利用自动化技术和方法，对数据质量进行实时、连续的监测和控制过程。通过定义的数据质量规则，系统自动地检测、识别和报告数据质量问题，以便及时采取纠正措施，确保数据质量符合预期标准。
	质量智能化分析	利用人工智能技术对数据进行全面、自动化的监控和评估，以及通过智能化方法挖掘和应用数据质量规则，从而提高数据质量管理的效率和效果，为数据质量管理提供洞察力和决策支持。
数据标准管理		数据标准管理是指建立、维护和执行一系列规范化的数据标准，以确保数据在整个组织内一致性、可理解性和可重用性，从而提高数据质量、降低数据管理成本、促进数据驱动决策。
数据标准管理	标准智能化推荐	标准智能化推荐是指利用人工智能和数据分析技术，基于数据特征、业务需求和最佳实践，自动化地推荐适用于特定数据集或业务场景的数据标准和规范，以促进数据一致性、可理解性和可重用性。
	标准管理	数据标准管理是指制定、维护和执行一系列规范化的数据标准，包括数据命名规范、数据格式、数据定义、数据分类等，以确保数据在整个组织内一致性、可理解性和可重用性，从而

		提高数据质量、降低数据管理成本、促进数据驱动决策。
	标准智能落标	标准智能落标是指通过智能化数据管理工具，自动将预设的数据标准应用于实际数据资产中，实现数据资产与数据标准的自动匹配和校验，确保数据质量和规范性。
	落标分析	落标分析是指对现有数据与既定数据标准之间的比对和分析过程，以确保数据的一致性、准确性和标准化。
元数据管理		元数据管理是指对数据的描述信息进行收集、存储、维护和利用的过程，有助于理解、管理和利用数据资产。
元数据管理	元数据质量管理	元数据质量管理指的是对元数据的准确性、完整性、一致性、及时性和可信度等方面进行评估、监控和改进的过程。包括确保元数据与实际数据和业务需求一致，以及确保元数据的更新和维护工作得到有效执行，从而保证元数据对数据管理和决策的可靠性和有效性。
	血缘管理	血缘管理指的是跟踪和记录数据在其生命周期中的来源、变化过程和影响关系的过程。它帮助组织了解数据的流动路径以及数据之间的关联关系，从而支持数据分析、合规性和风险管理等活动。血缘管理可以通过记录数据的源头、数据流转路径、数据转换操作以及数据影响范围等信息来实现。
	元数据分析	元数据分析是指对组织中的元数据进行系统性的分析和解释，以揭示数据资产的特征、关系、使用方式和价值等信息。这种分析可以包括对元数据的结构、内容、质量和血缘等方面进行深入探索，旨在帮助组织更好地理解 and 利用其数据资源，支持数据管理、数据治理和业务决策等活动。
	智能化管理	元数据的智能化管理是指利用先进的技术手段，如人工智能、自然语言处理、大数据分析等，从而增强元数据管理的能力和效率，对元数据进行高效、自动化和智能化的管理。
数据资源管理		数据资源管理是指在整个数据生命周期中，对数据资源进行有效的规划、采集、存储、处理、共享和保护，确保和促进数据资源产生价值的过程。
数据资源管理	资源目录管理	资源目录管理是指建立和维护一个包含组织内所有数据资源的详细清单或目录，其中包括数据集、数据库、数据模型、数据文档、数据字典等信息。资源目录管理旨在提供对数据资产的全面可见性和访问性，以便组织成员可以快速找到所需的数据资源，并了解其基本属性和用途。
	资源智能盘点	资源智能盘点是指利用自动化工具和技术对组织内的数据资源进行全面的扫描、识别和分析，以获取关于数据资源的详细信息，包括其类型、位置、所有者、权限等，从而实现了对数据资产的智能管理和监控。

	数据资源应用	数据资源应用指的是将组织内的数据资源应用于业务活动、分析过程或决策制定中的过程，包括将数据资源用于支持业务流程优化、市场分析、客户关系管理、产品开发、风险管理等各种业务需求。
数据安全		数据安全是指在整个数据生命周期中，采取一系列措施和策略，以确保数据的保密性、完整性和可用性，防止未经授权的访问、修改或泄露，从而保护数据资产免受安全威胁和风险的影响。
数据安全	自动化分类分级	自动化分类分级是指利用自动化工具和技术对组织内的数据进行识别、分类和分级，以确定其敏感程度和重要性，并根据其特征和规则将其自动分类到适当的安全级别或保护策略中。这种自动化过程可以帮助组织更有效地管理数据安全，降低人为错误和数据泄露的风险。
	权限管理	权限管理是指对数据访问和操作的控制，以确保只有经过授权的用户或实体可以获取特定数据、执行特定操作或访问特定功能。这包括对用户、角色、组织单元或其他实体的权限进行定义、分配和管理，以及对权限的审核、监控和调整。权限管理旨在保护数据免受未经授权的访问、修改或泄露，并确保数据的完整性、保密性和可用性。

4 各能力项要求

一级能力域	二级能力项	功能点	条款	必选/可选项
数据质量管理	规则智能化配置	质量规则生成	支持通过关键字、数据特征、数据标准等，自动化生成数据质量规则的功能。	必选
		规则语义检索	支持对质量规则进行语义理解和意图识别，智能检索相关的质量规则。	可选
	质量自动化监控	异常告警	支持对数据质量进行异常检测和分析，能够预测潜在的数据质量问题并发出告警。	可选
	质量智能化分析	质量问题分析	支持以可视化的形式对数据质量问题进行分析，并形成质量问题报告。	可选
数据标准管理	标准智能化推荐	标准智能推荐	支持基于历史数据标准和业务规则智能聚合推荐数据标准建议，帮助用户可以快速制定和优化标准。	必选
	标准管理	标准标签管理	支持基于标准名称、定义等内容自动化生成标准标签，并分类管理的能力；	可选
		标准语义	支持对标准内容进行语义化检索的能力，用户可以通过关键词、标签分类等	可选

		检索	检索对应的数据标准。	
	标准智能落标	自动核标	支持自动化分析检核字段与所关联的数据标准的匹配程度，确保数据符合相应的数据标准。	必选
		异常识别	支持对数据标准落标问题进行自动检测的能力，并触发告警通知。	可选
	落标分析	落标分析	支持通过可视化形式对落标结果进行分析，并支持数据项差异分析。	必选
元数据管理	元数据质量管理	元数据信息补全	支持基于大模型技术填充缺失的元数据信息的能力。	必选
	血缘管理	血缘解析	支持基于数据解析结果自动化构建数据血缘关系的能力。	必选
		关系图谱构建	支持基于元数据和数据血缘关系及关联关系构建数据关系图谱，以可视化形式展示。	可选
		算子级血缘	支持基于业务、数据关系进行挖掘，实现构建算子级数据血缘的能力	可选
	元数据分析	元数据访问分析	支持元数据访问的分析能力，可视化分析元数据的访问热度、时效性等。	必选
		链路影响分析	支持基于数据上下游的血缘关系，自动化生成变更影响分析报告。	必选
	智能化管理	元数据语义检索	支持对元数据进行语义化检索，提供更加精准的元数据检索能力。	可选
		变更追踪	支持元数据变更自动触发通知告警的能力，并支持比对分析元数据的变更内容。	必选
	资产目录管理	资产语义检索	支持对数据资源进行语义化检索，提供更加精准的数据资源检索能力。	可选
		资产标签化	支持基于数据特征和业务属性自动化生成数据资产标签的能力，并支持分类管理。	必选
	资产盘点	资产分析	支持多维度的数据资产分析能力，支持以可视化形式展示资产发布、访问使用情况等。	可选
		资产推荐	支持基于资产的使用情况、用户应用偏好等推荐相应数据资产的能力。	必选
数据资产管理	数据资产应用	数据智能化应用	支持基于分析需求，智能生成分析报表的能力，并生成数据趋势统计的分析结论。	可选

		智能问答	支持提供基于资产知识图谱或大模型的数据智能问答能力。	可选
数据安全	自动化分类分级	敏感资产盘点	支持自动识别敏感字段的能力，并生成敏感资产标签。	必选
		自动化分类分级	支持基于智能化技术实现自动化分级分类的能力，半自动化或自动化实现数据分类分级。	必选
	权限管理	数据内容访问控制	支持对用户身份和权限对数据内容进行智能化的访问控制和管理。	必选