

TC609

全国数据标准化技术委员会技术文件

TC609-6-2026-XX

数据基础设施 语义描述和互操作要求

Data infrastructure—Requirements for semantic
description and interoperability

（征求意见稿）

2026-XX-XX 发布

2026-XX-XX 实施

全国数据标准化技术委员会 发布

目 次

前 言	III
引 言	IV
1 范围	5
2 规范性引用文件	5
3 术语和定义	5
4 缩略语	6
5 语义架构要求	6
5.1 总体要求	6
5.2 分级架构	7
5.3 语义内容构成	8
6 语义建模要求	8
6.1 建模通用要求	9
6.2 概念与语义模型构建要求	9
6.3 语义资产维护与发布要求	9
7 语义互操作要求	10
7.1 通用要求	10
7.2 互联互通层要求	11
7.3 语法互操作层要求	11
7.4 语义互操作层要求	11
7.5 智能体互操作层要求	12
8 语义资产质量要求	12
8.1 通用要求	12
8.2 资产内容质量	12
8.3 多模态语义质量	13
8.4 语义互操作质量	13
8.5 应用效果质量	13
8.6 质量评估与持续改进	14
9 安全保障要求	14
9.1 通用要求	14
9.2 基于语义的可信访问控制	14
9.3 权限与使用语义	14
9.4 语义资产安全与可信溯源	14
9.5 语义隐私防护	15
附 录 A （资料性） 语义应用案例	16
参 考 文 献	19

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。

本文件由全国数据标准化技术委员会（SAC/TC609）提出并归口。

本文件起草单位：

引 言

数据基础设施是支撑数据要素市场化、赋能人工智能与智能体应用的关键底座。随着大模型、智能体等新型智能应用的发展，智能体(Agent)逐步成为“用数”主体，数据需进一步做到“可被理解、可被推理、可被可信调用”。现有技术文件覆盖了目录、标识、身份、连接与互通，但尚未覆盖让智能体理解数据含义的“语义层”，导致数据与智能服务在跨节点流通时“传不到、读不懂”，难以支撑智能体的复杂任务规划与长程推理。为补齐这一空白，本文件统一规范数据基础设施语义描述、跨主体语义互通等相关要求。

本文件与数据基础设施相关标准协同：

- a) 与《数据基础设施 数据目录描述要求》的关系：本文件规定目录所承载语义描述项的含义与一致性，目录登记、检索与展示机制由《数据基础设施 数据目录描述要求规定》；
- b) 与《数据基础设施 标识要求》的关系：本文件规定标识的语义解析、映射、多标识对齐机制，《数据基础设施 标识要求》定义全局标识编码规则等内容；
- c) 与《数据基础设施 互联互通基本要求》、《数据基础设施 连接器技术要求》的关系：本文件规定交换内容的语义与语义互操作要求，报文传输与服务通道由《数据基础设施 互联互通基本要求》、《数据基础设施 连接器技术要求》规定；
- d) 与《数据基础设施 用户身份管理和接入要求》的关系：本文件规定用户身份相关语义描述信息，用户身份管理和接入要求由《数据基础设施 用户身份管理和接入要求》规定。

本文件面向数据基础设施中数据资源、数据产品、服务接口及相关对象的语义描述和语义互操作需求，规定语义对象、语义标识、语义建模、语义映射、语义质量和安全保障等要求。智能体、大模型工具调用等可基于本文件机制开展扩展应用。本文件旨在为各功能节点、业务节点开展语义体系建设提供统一指引，构建可被智能体理解、可跨节点互操作的语义底座。

数据基础设施 语义描述和互操作要求

1 范围

本文件规范了数据基础设施建设中语义的架构、建模、互操作、质量管理和安全保障等相关要求。本文件适用于指导各类层级数据基础设施中，各功能节点、业务节点开展语义体系建设。

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。

TC609-5-2025-04 高质量数据集 质量评测规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

语义 semantics

数据所承载的概念含义、关系及业务语境（含规则与约束），是信息理解与知识表达的基础。

3.2

语义描述 semantic description

对数据所承载的概念、关系、规则与约束进行的形式化表达、可被智能体等新型智能应用理解与推理的描述。

3.3

本体 ontology

对业务领域中的概念实体、属性及其相互关系以及相关规则的形式化、可共享的描述，可供智能体等新型智能应用理解与推理。

注：本体可采用W3C RDFS、OWL2等形式化语言表达。

3.4

概念模型 conceptual model

对业务领域中实体、属性、关系、规则与约束进行的、与具体实现无关的抽象表达。

3.5

语义模型 semantic model

基于概念模型，采用形式化语言构建，包含概念实体、属性、关系、约束、规则的机器可读可推理形式化描述，本体是语义模型的标准化实现载体。

3.6

语义标识 semantic identifier

在统一标识基础上，可解析至其所对应概念实体或本体的标识；标识承载语义关联信息，支撑跨域系统之间实体对齐、概念溯源与知识推理。

3.7

语义映射 semantic mapping

不同数据源、目录、本体或多模态数据之间概念实体、关系、口径的对应关系。

3.8

语义对齐 semantic alignment

在跨节点、跨主体场景下，对不同来源语义描述之间的对应关系进行识别与建立的过程。

3.9

语义互操作 semantic interoperability

数据基础设施具备的、支撑智能体等新型智能应用跨节点调用服务与可信使用数据的能力，所交换内容的语义在节点之间能被正确理解与处理。

注：语义互操作以节点间语义描述的可交换、可对齐为基础，支持智能体之间基于一致语义的协作与推理。

3.10

语义资产 semantic asset

以概念模型、语义模型、本体、语义映射、语义规则等形式沉淀的、可被管理、复用和共享的语义化成果。

注：按变更类型，语义资产可分为静态资产与动态资产。静态资产指概念模型、语义模型、语义规则等以模型为主、按版本演进的资产；动态资产指实例数据、语义对齐关系、推理结果等随业务与数据运行持续更新的资产。

3.11

基于语义的可信访问控制 semantics-based trusted access control

以概念、实例、属性及其关系等语义对象为控制对象，对访问主体进行细粒度授权与约束，且决策与执行可审计、可追溯的访问控制方法。

4 缩略语

下列缩略语适用于本文件。

API：应用程序编程接口（Application Programming Interface）

A2A：智能体到智能体协议（Agent to Agent）

JSON-LD：基于JSON的关联数据（JSON for Linking Data）

MCP：模型上下文协议（Model Context Protocol）

OWL2：网络本体语言（OWL 2 Web Ontology Language）

RDF：资源描述框架（Resource Description Framework）

RDFS：资源描述框架模式（RDF Schema）

SPARQL：SPARQL协议和RDF查询语言（SPARQL Protocol and RDF Query Language）

TTL：Turtle（Terse RDF Triple Language）

SWRL：语义网规则语言（Semantic Web Rule Language）

OBAC：基于本体的访问控制（Ontology-Based Access Control）

RBAC：基于角色的访问控制（Role-Based Access Control）

ABAC：基于属性的访问控制（Attribute-Based Access Control）

SHACL：形状约束语言（Shapes Constraint Language）

5 语义架构要求

5.1 总体要求

语义体系架构满足：

- a) 应以概念模型、语义模型、语义映射、语义规则和语义服务为核心，形成连接数据资源、数据产品、智能服务与智能体应用的语义层；语义层向上为智能体、大模型提供语义推理服务，向下对接数据目录、统一标识、身份认证、连接器，提供语义服务、基于语义的可信访问控制等能力。
- b) 应遵循统一规划、分层管理、分域扩展、节点自治、安全可控和开放互操作原则，支撑跨行业、跨节点的语义复用、语义对齐和语义互操作；
- c) 应覆盖需求分析、语义设计、语义实现、验证测试、部署运维、更新维护等语义资产全生命周期管理环节，形成可追溯、可复用、可迁移的管理机制；
- d) 应与数据目录、统一标识、用户身份、连接器、数字合约、使用控制、安全审计等数据基础设施既有能力衔接，语义层负责语义表达、关系解析、推理支撑等，不替代目录发布、身份认证、合约履约和访问控制执行机制；
- e) 宜支持与 RDF、RDFS、OWL2、SPARQL、JSON-LD 等国际通用语义技术兼容，支撑语义资产的机器可读、可校验、可查询和可推理。

5.2 分级架构

语义体系应采用通用级与行业级协同的二级架构，分级架构及控制继承关系见图1：



- b) 行业级：在符合通用级要求的前提下，补充本行业的领域概念、关系、语义标识与映射、行业规则与约束、语义对齐机制等内容；
- c) 各功能节点、业务节点在符合通用级与行业级规定的前提下，开展本节点语义建模、数据映射、语义资产管理、语义服务发布、访问控制绑定和运行审计等内容。

跨级语义关系应满足：

- a) 通用级对所有语义建设具有规则控制、约束和指导作用；当行业级扩展语义与通用级语义冲突时，以通用级语义描述为准
- b) 行业级语义资产应通过继承、扩展、映射等方式声明其与上层语义对象的关系；
- c) 跨节点交换数据或调用智能服务时，应优先引用通用级、行业级已发布语义标识。

5.3 语义内容构成

语义体系应至少包含以下内容：

- a) 概念模型：业务概念、属性、关系、规则与约束的抽象表达；
- b) 语义模型：概念实体、属性、关系、规则与约束的形式化描述；
- c) 语义映射：数据资源、数据产品、服务接口、字段、指标、代码值与语义模型中概念实体、属性、关系之间的对应关系等；
- d) 语义规则：业务规则、推理规则、校验规则、转换规则和口径计算规则与约束等；
- e) 语义资产元数据：语义资产名称、标识、版本、发布状态、责任主体、适用范围、来源依据、变更记录和质量状态等管理信息；

6 语义建模要求

语义建模建设路径如图2：

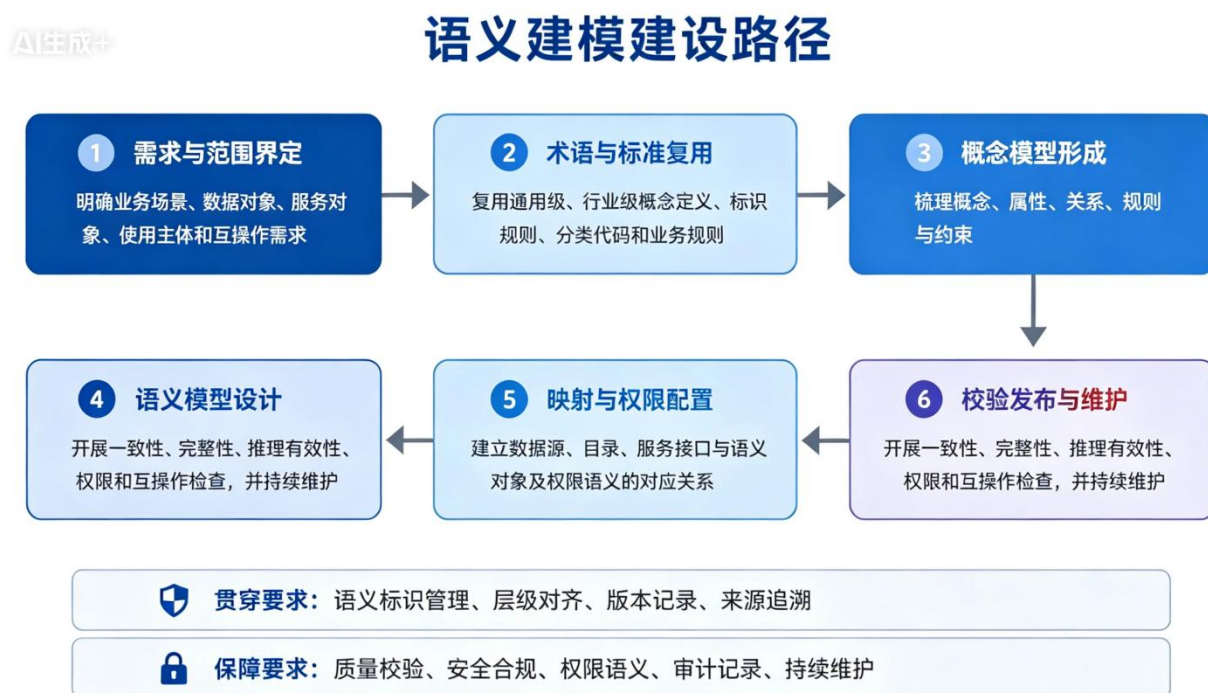


图 2 语义建模建设路径逻辑

6.1 建模通用要求

语义建模应满足：

- a) 应基于业务场景、业务需求、数据资源特征、智能服务能力和跨节点互操作需求开展建模，准确表达概念、属性、关系、口径、约束、规则和上下文；
- b) 应优先复用通用级、行业级已发布的核心概念、标识规则、分类代码和权限语义；确需扩展时，应明确扩展原因、适用范围和上层对齐关系；
- c) 宜按照“需求与范围界定—术语与标准复用—概念模型形成—语义模型设计—映射与权限配置—校验发布与维护”的路径开展语义建模，形成从业务含义到机器可读表达的建设过程；
- d) 语义模型宜采用 W3C RDFS、OWL2 或与之兼容的形式化语言表达，序列化格式宜支持 RDF/XML、Turtle (TTL)、JSON-LD 等；
- e) 语义规则与约束宜采用 SWRL、OWL2 RL、SHACL 或与之兼容的形式化语言表达，语义查询宜支持 SPARQL 或与之兼容的查询方式；
- f) 概念、属性、关系、规则、映射、约束和实例应具备语义标识，语义标识应可解析至对应语义对象及其版本，语义标识规则应符合 NDI—TR—2025—04 的规定；
- g) 建模过程应保留来源依据、责任主体、版本记录和变更记录，保障语义资产可追溯、可审计。

6.2 概念与语义模型构建要求

概念与语义模型构建应满足：

- a) 概念定义应包含名称（含中英文）、语义标识、定义、适用范围、上位概念、下位概念、别名和同义词、示例、来源依据和责任主体等信息；
- b) 语义模型的属性定义应包含属性名称、语义标识、定义、值域、数据类型、计量单位、取值范围、枚举代码、是否必填、敏感级和质量约束等信息；
- c) 关系定义应包含关系名称、语义标识、定义、关联实体、关系类型、谓语类型等信息；
- d) 约束定义应明确适用对象、约束条件、触发条件、校验方式、错误级别和处理建议；
- e) 对于同名不同义、同义不同名等术语歧义问题，应统一标准术语，并通过补充定义、限定条件、唯一语义标识加以区分；应统一业务口径、计量单位与分类编码标准；存量异构系统无法完成标准化整改的，可辅以语义映射实现跨模型兼容；
- f) 当同一对象对应多个异构标识（如国际标准标识、行业标准标识、企业内部标识、监管标识等）时，应通过语义标识与语义映射建立标识间的等价或映射关系，确保可沿不同标识路径解析至同一概念或实例；具体标识体系由行业层结合本文件机制细化。

6.3 语义资产维护与发布要求

语义体系宜支持语义资产的持续维护和规范发布，要求包括：

- a) 宜支持根据业务变化、标准更新、目录变更、服务变更、质量评估结果和安全合规要求，对语义模型、映射、规则和权限语义进行扩展、修订与废止；
- b) 语义资产变更应保留版本记录、变更原因、影响范围、评审结论和版本转换说明，保障语义描述的可追溯性；
- c) 语义资产发布前应完成语法、语义、映射、约束、推理、权限和互操作检查；
- d) 建立语义变更通知机制，语义资产发布后，应同步提醒受影响的语义对齐方、数据目录、连接器服务描述、智能体工具描述、数字合约引用和访问控制策略映射等；

- e) 宜采用 AI 自动化辅助方法开展概念与关系抽取、映射推荐、冲突发现、质量校验和变更影响分析，保持语义描述与业务变化、标准更新相一致。
- f) 语义资产废止时应明确替代资产、废止时间、影响范围、迁移路径和兼容性说明；

7 语义互操作要求

7.1 通用要求

语义互操作应按图3所示的层次体系构建，自下而上包括互联互通层、语法互操作层、语义互操作层和智能体互操作层。

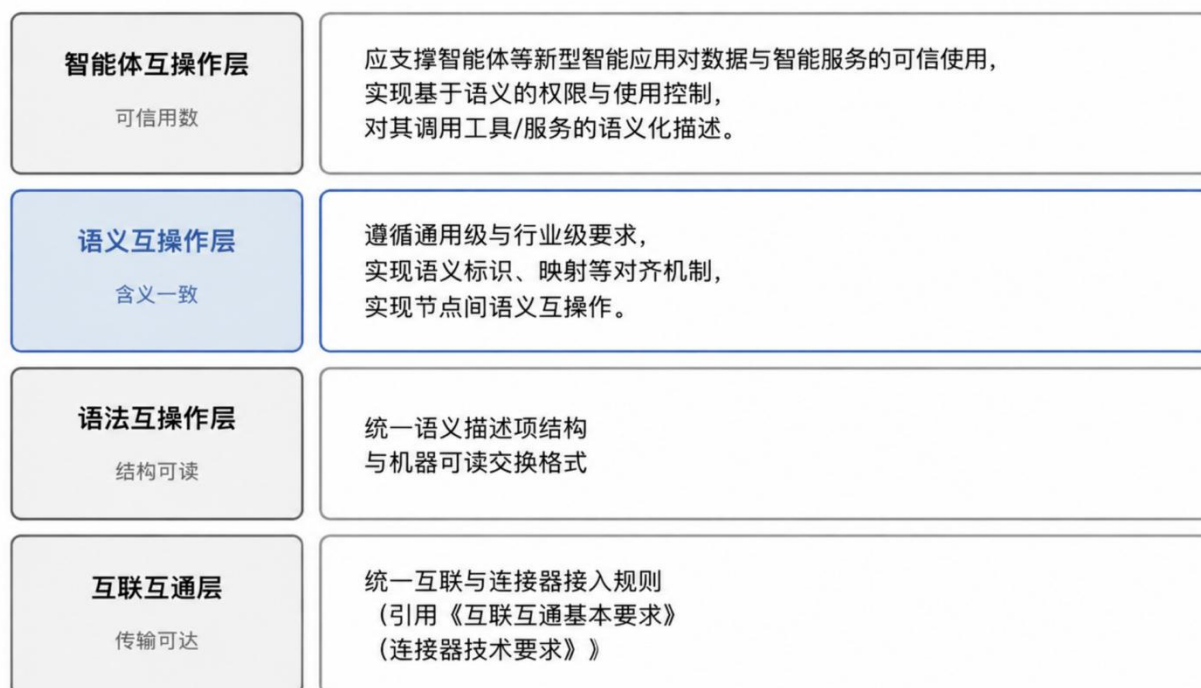


图 3 语义互操作层次体系

语义互操作应满足以下通用要求：

- a) 跨节点交换数据或提供智能服务时，应携带或可解析对应的语义描述；
- b) 不同来源的语义描述应可通过语义对齐建立对应关系，对齐关系宜标注其所依据的语义版本等信息。
- c) 当任一方语义描述发生语义变更（如模型修订、版本升级等）时，应及时识别受影响的语义对齐关系，并对其进行同步更新或重新校验，避免因语义变更导致对齐失效；
- d) 语义对齐关系应保留版本记录，记录其建立时间、所依据的语义版本及变更情况，保障对齐结果的时效性与可追溯性；并适当附带上下文信息，以加强理解。
- e) 语义描述项应符合 NDI—TR—2025—06 的规定，不重复定义；
- f) 各层次互操作应复用通用级、行业级已发布的语义标识与语义资产，不应另建并行语义体系；
- g) 节点间进行语义互操作时，应首先基于通用级建立基础含义共识；当涉及行业特定概念时，应通过行业级语义互操作层进行补充和对齐。

注：本文件规定语义互操作的基础要求；语义自动对齐、一致性校验增强、智能体间协作与推理、策略语义协商等增强能力，在后续版本中规定。

7.2 互联互通层要求

节点互联、报文传输及服务通道（API、MCP、A2A等）由接入连接器承载，应符合NDI—TR—2025—02和NDI—TR—2025—05的规定，本文件不另行规定。

7.3 语法互操作层要求

语法互操作应实现数据交换过程中语义格式和编码规则的一致性与可解析性，确保消息能够被对方系统按约定正确解析和处理，具体要求包括：

- a) 跨节点交换的语义描述应采用机器可读格式，并可被接收方正确解析其结构；
- b) 通用级应统一语义描述项的结构与交换格式；行业级可在其基础上定义行业描述模板与扩展字段，扩展字段不应改变通用级既有结构；
- c) 数据目录检索与详情查询的响应宜包含语义描述项（概念映射、语义关系、口径与计量、语义标识等），其结构与取值应与登记时的语义资产保持一致；
- d) 数据产品交付时，宜随附或可解析其语义描述及所引用语义资产的标识与版本，支撑使用方按一致口径理解和加工数据；
- e) 应支持多版本、多格式报文的语法映射与转换，并建立完善的向后兼容机制，保障语法层面的互通兼容能力。

7.4 语义互操作层要求

语义互操作层为本文件规范重点，应实现在数据交换与应用过程中，能够对数据所表达的业务含义形成一致理解，具体要求包括：

- a) 应在统一标识基础上提供语义标识解析，并统一采用领域术语表和命名规范，同一对象在不同节点应解析为一致的概念或者进行映射对齐；
- b) 通用级应统一顶层概念与垂直挂接约束；行业级应将行业概念、口径挂接至通用级顶层概念，不得改变其定义与适用范围；
- c) 跨域语义映射在本版以登记式映射为基础要求：映射需声明发布、版本可溯，并明确等价、上下位、近似等映射类型；映射登记信息应包括映射标识、源实体标识、目标实体标识、映射类型、版本号等信息。
- d) 标准数据元、关键指标的计量单位、统计口径、基准与时区等应可被各方一致理解，数据元需统一规范名称、定义、精度、取值范围，存在差异时应通过语义映射明确对应关系；
- e) 跨节点共享的语义资产应附带语义资产元数据（名称、标识、版本、发布状态、责任主体、适用范围等），并可通过语义标识解析获取；复用方应通过继承、扩展或映射方式声明与被复用资产的关系，不得擅自更改其定义和适用范围；
- f) 被共享语义资产发生修订或废止时，应通知复用方；复用方应评估影响并同步更新相关映射与对齐关系；
- g) 节点宜对外提供语义资产查询服务（如采用 SPARQL 或与之兼容的查询方式），跨节点查询应遵循被查询节点的访问控制与权限语义。

7.5 智能体互操作层要求

智能体互操作层应支撑智能体等新型智能应用对数据与智能服务的可信使用，要求包括：

- a) 应明确“传输协议”与“传输内容语义”相解耦：协议负责传输通道，语义描述负责内容含义；
- b) 节点经 MCP、A2A、API 等协议对外提供工具或智能服务时，宜在协议消息体中携带对应的语义描述，包括概念映射、口径、语义标识及本体引用等；
- c) 经 MCP、Skill 等方式接入的工具集，其工具描述宜采用可被智能体理解的语义化表达，明确输入、输出、前置条件、异常条件与口径说明；
- d) 智能体跨节点取数或调用服务时，应可解析所涉数据与服务的权限与使用语义，并在数字合约与使用控制机制下开展调用；
- e) 通用级应统一权限与使用语义词汇；行业级在符合通用级规则的前提下，细化行业敏感级与使用条件；
- f) 经 A2A 等方式协作的智能体之间，可就所交换内容的语义达成一致，或通过语义对齐解析。

注：多智能体围绕同一任务的语义共享、协作与推理要求，在后续版本中规定。

8 语义资产质量要求

8.1 通用要求

语义资产质量管理应满足：

- a) 质量管理应覆盖语义资产需求分析、设计、实现、验证、发布、运维、更新、废止的全生命周期；
- b) 语义资产质量宜参照 TC609-5-2025-04 的维度框架，从资产内容质量、多模态语义质量、语义互操作质量、应用效果质量四个维度提出要求并开展评测；
- c) 通用级统一制定通用质量基线与评测规则，行业级可在不违背基础通用要求的前提下，结合领域特征补充细化质量要求；
- d) 节点应建立语义资产质量评估与校验机制，明确评估指标、评估频次、问题处置与改进流程；
- e) 质量评估结果应记录于语义资产元数据，作为语义资产发布、复用与共享的依据。

8.2 资产内容质量

语义资产的内容质量要求包括：

- a) 一致性：同一语义概念、术语、口径、约束规则，在同一应用范围内应保持统一；跨应用范围存在差异时，应通过标准语义映射明确对应关系。同模态、多模态同源数据对应的语义描述应逻辑自洽、含义统一，避免语义冲突；
- b) 完整性：语义描述应覆盖实体、属性、关系、约束、上下文等核心要素，匹配数据、服务、智能体工具的业务需求，必备语义项不应缺失；
- c) 准确性：概念定义、业务规则、关联关系如实反映真实业务场景，语义映射、逻辑推理结果准确无误，与来源保持一致；
- d) 规范性：不应存在重复定义、循环定义或相互冲突的公理与规则，语义标识应唯一、可解析；
- e) 机器可处理性：语义资产应机器可读、可校验，其约束与规则应可执行、可验证。语义映射的运行时生效应基于确定性查表或推理规则，不应依赖生成式模型对语义进行即时改写

- f) 可溯源性：每一条语义资产（含映射、推理结论）应携带来源标识、置信度与审核状态，作为质量判定与发布的必备元数据；未携带者不应发布。

8.3 多模态语义质量

针对文本、表格、图像、音视频等不同形态的语义资产质量要求包括：

- a) 模态适配性：针对表格类数据，应完整表达嵌套结构、跨表勾稽、内嵌公式等专属语义；针对文本类数据，保障业务条款、实体关系描述连贯完整；针对图像、音视频类数据，确保标注语义与媒体内容正确匹配；
- b) 同源多模态数据语义对齐：文本、表格、图像、音视频等不同形态的同源数据，其所对应的实体、属性、指标、业务关系、约束条件须保持语义统一，不应出现解读偏差、逻辑矛盾；
- c) 跨模态关联完整性：多模态融合场景下，不同数据形态之间的关联语义应完整标注，关联关系清晰可解析，无关联缺失、关联错位问题；
- d) 模态内容联动性：当单一模态原始数据发生变更时，关联的多模态语义描述应同步更新，保持整体内容同步有效。

8.4 语义互操作质量

为保障跨主体、跨系统的数据互通，语义互操作质量要求包括：

- a) 跨主体语义兼容性：跨数据应用主体、跨业务系统开展交互时，本地语义描述、语义映射、语义标识可被正常解析，遵循基础通用统一规则与行业专用扩展要求，不因格式、口径差异造成互操作阻断；
- b) 语义传递完整性：数据、服务跨主体传输过程中，附带的语义标识、概念描述、口径说明、约束规则等关键信息完整传递，无丢失、篡改问题；
- c) 语义对齐有效性：跨主体建立的语义对齐关系准确稳定，语义版本更新、规则调整后，对齐关系可同步迭代，避免对齐失效；语义映射应区分等价与近似关系；近似关系不应在运行时作等价替换使用，需标注为近似召回或经人工确认。
- d) 协议适配质量：基于 API、MCP、A2A 等协议传输的语义内容，符合对应协议规范，语义描述与传输通道解耦，可在不同交互场景下正常复用；
- e) 复用一致性：复用基础通用、行业专用语义资产时，严格继承原有定义与约束，扩展内容不得破坏原有互操作能力。

8.5 应用效果质量

语义资产的应用效果质量要求包括：

- a) 时效性：语义描述应随业务变化、标准更新及时修订，并保留版本记录；
- b) 可用性：语义服务应可正常解析、查询与调用，满足约定的服务可用性要求；
- c) 复用与互操作有效性：基础通用、行业专用语义资产可被正常复用，跨节点语义对齐与互操作结果正确；
- d) 智能应用赋能效果：应评估语义资产对智能体等新型智能应用的理解规划、逻辑推理、标准比对、知识调用等环节的支撑效果。智能体基于本文件语义作出的推理与结论宜可回溯至具体语

义资产，对无法落地的推理应可识别。智能体基于本文件语义作出的推理与结论应可回溯至具体语义资产；对超出语义资产覆盖范围的推理，应可识别并标注为未经确认。

8.6 质量评估与持续改进

配套建立质量评测巡检、版本动态管控机制，具体要求包括：

- a) 评测方式：采用自动化校验结合人工抽检的模式。自动化工具完成格式、约束、一致性批量检测；人工重点核查复杂业务语义、多模态语义、跨主体语义互操作及逻辑推理等内容；发布判断：语义资产发布前应通过质量门限校验，门限应包含：强约束项（如一致性、可溯源性须完全满足）、质量收敛项、人工干预率上限；具体阈值由通用级给出范围、行业级结合领域制定。
- b) 规则设定：评测指标、权重、判定规则由通用级统一规定，行业级按需细化，结合技术发展、业务变化、标准更新，动态优化质量指标与评测方法；
- c) 版本管控：语义资产宜具备完整版本标识、变更记录、操作日志，变更操作全程可审计；
- d) 常态化巡检：各数据应用主体、业务系统应定期开展语义资产质量盘点与复测，结合版本更新、业务迭代开展专项校验。

9 安全保障要求

9.1 通用要求

语义体系的安全保障应符合数据基础设施安全能力的通用要求，在此基础上应针对语义描述、建模、推理及互操作过程中的特有风险，建立语义资产全生命周期的安全防护机制，保障机密性、完整性、可用性及隐私合规性。

9.2 基于语义的可信访问控制

语义体系应支持基于语义的细粒度访问控制：

- a) 应支持以语义资产中的概念、对象、属性及关系为控制对象，进行细粒度授权；
- b) 宜支持将 OBAC（基于本体的访问控制）与 RBAC（基于角色的访问控制）相结合，实现角色到本体对象权限的映射；
- c) 权限控制应精确到概念、实体、实例、属性级别；
- d) 对于高敏感的数据项，应支持与隐私计算、TEE 沙箱等安全能力对接，在可算不可见的情况下进行数据访问控制。

9.3 权限与使用语义

语义描述中的“权限与使用语义”应满足：

- a) 仅对用途、敏感级、使用条件等进行语义化标注；
- b) 用于衔接数字合约与使用控制，不替代其管控机制；
- c) 权限与使用语义由通用级统一定义，行业级在符合通用级规则前提下补充。

9.4 语义资产安全与可信溯源

语义资产的安全保障应满足：

- a) 应保障概念模型、语义模型、语义映射等语义资产在传输与存储过程中的保密性与完整性；
- b) 应防止通过语义推理导致的敏感信息泄露，对于高敏感的数据项应支持通过隐私计算、TEE 沙箱等能力在可算不可见的情况下进行语义推理，必要时对推理结果进行管控；
- c) 语义资产的访问应记录日志，支撑审计与溯源。

9.5 语义隐私防护

语义隐私保护应满足：

- a) 支持基于语义的敏感数据隐私防护：发布语义资产时，应对可能暴露敏感数据结构或敏感业务逻辑的对应数据设置脱敏或隐私防护策略；
- b) 防映射关系泄露：应仅发布必要的语义映射、标识信息，对涉及核心商业秘密或敏感个人信息的深层映射关系应加密或实施防护策略。

附 录 A
(资料性)
语义应用案例

A.1 概述

以金融领域“企业经营真实性核验”场景为例，说明本文件各项要求在数据基础设施中的落地方式。金融机构在提供授信等金融服务前，针对企业提交的信息和材料提出核验需求，利用公共数据和公开数据开展多源交叉核验。案例由三类角色协同：授权运营机构（数据基础设施的业务节点，承载城市级可信数据空间与公共数据授权运营平台）提供工商登记、社会保险、住房公积金、合规信用、司法涉诉、资质许可等公共数据，并依法使用相关公开数据；语义平台建设运营机构在业务节点之上提供语义建模、语义服务与语义级访问控制能力，并接入连接器与智能体；金融机构作为功能节点提出核验需求并保留授信等业务决策权。

案例目标：在企业授权、数据不出域和最小必要原则下，通过语义层将企业提交的信息和材料与公共数据、公开数据进行语义对齐和多源交叉核验，形成一致、不一致、查无及需人工复核等核验结果和可追溯证据链，为金融机构授信等服务前的经营真实性审查提供参考；原始公共数据始终留在业务节点内，仅核验结论、风险提示和证据摘要等再加工结果经出域审核后交付。

A.2 语义层的构建

语义层依据本文件第5章、第6章构建，形成可被智能体理解、可跨节点互操作的金融语义资产。

A.2.1 分级架构（见5.2）

按通用级与行业级二级架构组织语义资产：通用级提供跨行业通用的核心概念（如“法人主体”“统一社会信用代码”“时间与金额口径”）、基础关系、语义标识与映射规则、权限与使用语义、质量与安全合规基线；行业级（金融）在符合通用级要求的前提下，补充主体登记与存续、用工与缴存、合规风险、司法涉诉、资质许可等领域概念、关系、规则与约束，并通过继承、扩展、映射声明其与通用级顶层概念的关系；业务节点在符合通用级与行业级规定的前提下，开展本节点语义建模、数据映射、语义资产管理与运行审计。

A.2.2 概念模型与语义建模（见5.3、6）

围绕“企业经营真实性核验”业务目标，按6.1的建模路径构建覆盖主体登记与存续、用工与缴存、合规风险、司法涉诉、资质许可等的概念模型与语义模型，以本体为标准化实现载体；建模全程保留来源依据、责任主体与版本记录，保障语义资产可追溯、可审计。场景所需公共数据和公开数据按核验用途归为五类，经语义映射挂接至概念模型，示例见表A.1。

表 A.1 公共数据与概念模型映射示例

数据类别	主要公共数据	映射到的概念/属性（示例）	在场景中的语义作用
1、主体登记与存续	企业四要素核验-企业基本信息	企业主体、统一社会信用代码、企业名称、法定代表人、身份信息、一致性结果、登记状态	核验主体身份、登记存续状态及企业提交材料的一致性
2、用工与缴存	当月缴纳状态、社保参保人数、近6个月养老保险实缴人数；公积金缴存人数、近一年月均人数、缴至年月、账户状	参保关系、缴纳状态、用工人数及趋势、公积金单位账户、缴存连续性、账户状态	核验企业申报用工规模、人员稳定性与持续经营迹象

数据类别	主要公共数据	映射到的概念/属性（示例）	在场景中的语义作用
	态、封存人数、账户余额、销户状态及日期		
3、合规与风险	是否失信、是否经营异常、是否严重违法失信、是否为股权冻结被执行人、红黑名单类型	失信状态、经营异常状态、严重违法失信状态、股权冻结、守信/失信类型	形成合规状态、信用风险及异常经营的核验旁证
4、司法涉诉	企业破产状态、终本案件（无可供执行财产）、限制高消费（限高令）	破产案件、执行案件、终本状态、限制消费措施及限制事项	核验企业或法定代表人的司法执行状态与履约风险
5、资质许可	许可经营范围、许可事项/资质名称、证照状态、证照有效期	行政许可、许可事项、许可范围、证照状态、有效期起止	核验提交证照真实性、有效性及经营活动与许可范围的一致性

A. 2.3 语义标识与语义映射（见6.1、7.4）

以“统一社会信用代码”作为企业主体的语义标识与跨源关联主键，确保同一企业在不同数据源、不同节点解析为一致的概念实例；采用登记式语义映射（人工建立、声明发布、版本可溯），将各公共数据表及公开数据项的字段、代码值映射至概念模型中的概念、属性、关系，并明确等价、上下位等映射类型，支撑在数据不出域前提下的跨源一致理解与按需取数。

A. 2.4 语义规则（见5.3、6.1）

将经营真实性核验规则表达为可执行、可审计的语义规则，如将企业提供的统一社会信用代码、企业名称、法定代表人等信息与公共数据和公开数据进行比对，形成一致、不一致、查无或需人工复核等结果；依据社会保险、住房公积金缴存等信号核验用工规模与持续经营情况；依据经营异常、严重违法失信、股权冻结、破产案件、终本案件、限制高消费及资质许可状态等识别风险与异常。规则由金融机构（行业级）配置，核验结论关联数据来源与证据链，结果可追溯、可解释。

A. 3 与数据基础设施的节点协同

场景运行依赖业务节点、连接器、智能体互操作与语义级访问控制的协同；语义层只规定语义化表达，相关控制与执行由数据基础设施既有机制承担（见5.1）。

A. 3.1 可信数据空间（业务节点）（见5.1、5.2）

授权运营机构的可信数据空间与公共数据授权运营平台构成数据基础设施的业务节点，履行5.2规定的节点职责：本节点语义建模、数据映射、语义资产管理、语义服务发布、访问控制绑定与运行审计。原始数据在节点内被理解、调用与建模；成果出域前经运营机构安全审核与实施机构合法性审核双重把关，构成合规闸门。

A. 3.2 连接器与互联互通（见7.2、7.3）

节点互联、报文传输及服务通道（API、MCP、A2A等）由接入连接器承载，遵循《数据基础设施 互联互通基本要求》《数据基础设施 连接器技术要求》（见7.2）；语法互操作层（见7.3）保障所交换的语义描述采用机器可读格式、可被正确解析。本文件不规定传输机制，仅规定其上承载的语义描述与语义互操作要求。

A. 3.3 智能体互操作与语义级访问控制（见7.5、9.2、9.3）

智能体经MCP、API等接口调用语义服务，调用时携带或可解析概念映射、口径、语义标识与本体引用，实现“传输协议”与“传输内容语义”相解耦（见7.5）；跨节点取数前解析所涉数据与服务的权限与使用语义，并在数字合约与使用控制机制下开展调用（见7.5、9.3）。访问控制以语义资产中的概念、对象、属性、关系为控制对象（OBAC，并与RBAC结合），精确到概念、实体、实例、属性级别（见9.2），按“一场景一审核”授权范围明确用途与边界；对高敏感数据项，结合隐私计算、TEE沙箱等在“可算不可见”下进行访问控制与语义推理（见9.2、9.4）。

A.3.4 端到端协同（见7）

一次企业经营真实性核验从需求发起到成果交付的协同与节点职责见表A.2。

表 A.2 端到端协同流程与节点职责

步骤	主要活动	参与方/节点	对应条款
1 需求发起	金融机构提交企业提供的信息、材料及经营真实性核验需求	金融机构（功能节点）	7.1、7.5
2 语义解析	业务节点内对需求进行语义解析与任务编排	业务节点	6、7.5
3 受控取数	依据权限与使用语义，在授权范围内调用公共数据并获取公开数据	业务节点 / 语义级访问控制	7.4、9.2
4 传输承载	经连接器完成节点互联与报文传输	连接器	7.2、7.3
5 核验分析	依据语义规则开展多源交叉核验，形成核验结论、风险提示与证据链	业务节点	6.1、8
6 出域交付	仅核验结论、风险提示、证据摘要等再加工成果经双重审核后出域	业务节点 / 出域审核	5.1、9.4
7 计量审计与决策	按调用计量、留痕存证；金融机构据此独立作出授信等业务决策	计量审计 / 金融机构	9.4

A.4 标准符合性与价值

本场景在关键要求上落实了本文件，见表A.3。

表 A.3 标准要求与场景落地对应

标准要求	场景落地
分级架构（5.2）	通用级顶层概念骨架+金融行业级概念挂接，继承/扩展/映射声明上下关系
语义建模（6）	以本体为标准化实现载体；统一社会信用代码作为语义标识与关联主键
语义互操作（7）	登记式语义映射；四层互操作；智能体经MCP/API携带语义；连接器承载传输
语义级访问控制（9.2）	概念/属性级OBAC+RBAC，字段/用途级管控，可算不可见
质量与安全（8、9）	多源交叉一致性校验、访问留痕、出域双重审核、敏感映射防护

通过上述落地，场景在数据不出域前提下实现了公共数据和公开数据对智能体的“可被理解、可被推理、可被可信调用”，形成可计量、可审计、可复用的语义资产单元，可向更多机构与场景复制推广。

参 考 文 献

- [1] W3C RDF 1.1 Concepts and Abstract Syntax
 - [2] W3C RDF Schema 1.1 (RDFS)
 - [3] W3C OWL 2 Web Ontology Language
 - [4] W3C SPARQL 1.1 Query Language
 - [5] W3C JSON-LD 1.1
 - [6] NIST SP 800-162 Guide to Attribute Based Access Control (ABAC)
 - [7] Ozgu Can, Murat Osman Unalir. Revisiting Ontology Based Access Control: The Case for Ontology Based Data Access. ICISSP 2022
 - [8] 关于推进行业高质量数据集建设行动的实施方案（国数科基〔2026〕25号）
 - [9] GB/T XXXXX-XXXX 数据基础设施 参考架构
 - [10] GB/T XXXXX-XXXX 数据基础设施 互联互通基本要求
 - [11] GB/T XXXXX-XXXX 数据基础设施 用户身份管理和接入规范
 - [12] GB/T XXXXX-XXXX 数据基础设施 标识要求
 - [13] GB/T XXXXX-XXXX 数据基础设施 连接器技术要求
 - [14] GB/T XXXXX-XXXX 数据基础设施 数据目录描述要求
-