

团 体 标 准

T/ISC XXX—XXXX

金融领域本体模型构建与应用规范

Specification for construction and application of ontology models in the financial domain

在提交反馈意见时，请将您知道的相关专利与支持性文件一并附上。

（征求意见稿）

XXXX—XX—XX 发布

XXXX—XX—XX 实施

中 国 互 联 网 协 会 发 布

目 次

前 言 III

引 言 V

金融领域本体模型构建与应用规范 1

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 符号和缩略语 2

5 总体框架 2

 5.1 构建原则 2

 5.2 全生命周期建设要求 2

 5.3 角色与职责 3

6 本体建模方法与要素 3

 6.1 建模方法 3

 6.2 实体类型 4

 6.3 关系 4

 6.4 属性 5

 6.5 约束 5

7 术语对齐与映射要求 5

 7.1 术语规范化 5

 7.2 术语对齐 5

 7.3 本体映射 5

8 本体版本演化管理 6

 8.1 版本编号 6

 8.2 变更分类与影响评估 6

 8.3 评审与发布 6

 8.4 废弃与迁移 6

9 本体应用衔接要求 6

 9.1 与知识图谱的衔接 6

 9.2 与检索增强生成（RAG）系统的衔接 7

10 本体质量评价 7

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。

本文件由中国互联网协会归口。

本文件起草单位：

本文件主要起草人：

本文件及其所代替文件的历次版本发布情况为：

——

引 言

随着大模型技术在金融领域的规模化应用，金融机构对高质量、结构化领域知识的需求日益迫切。本体作为对领域概念体系的形式化、规范化描述，是构建金融知识图谱的语义基础，也是提升检索增强生成（RAG）系统知识供给准确性、可解释性与可控性的关键支撑。

当前，金融领域本体建设多由各机构独立开展，建模方法、实体类型划分、关系语义定义、术语使用及版本管理方式差异较大，本体与知识图谱、RAG等下游应用系统的衔接缺乏统一规范，制约了金融知识资源的互联互通、复用共享与规模化应用。

本文件旨在确立金融领域本体模型构建与应用的统一规范，明确金融本体的建模方法与要素、术语对齐与映射要求、本体版本演化管理机制，以及本体与知识图谱、检索增强生成系统的衔接方式，为金融机构、金融科技企业及相关技术服务方开展金融知识工程与大模型应用提供依据。

金融领域本体模型构建与应用规范

1 范围

本文件确立了金融领域本体模型构建的总体框架与基本原则，规定了金融本体建模的方法与要素、术语对齐与映射要求、本体版本演化管理要求，以及本体与知识图谱、检索增强生成（RAG）系统的衔接方式，并给出了本体质量评价要求。

本文件适用于金融领域本体模型的设计、构建、管理与应用，包括银行、证券、保险、基金、期货等金融机构及金融科技企业，也可供金融行业知识基础设施建设参考。

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 7408 数据元和交换格式 信息交换 日期和时间表示法

3 术语和定义

下列术语和定义适用于本文件。

3.1

本体 ontology

对某一领域概念体系的形式化、显式、共享的规范化说明，用于描述该领域的实体类型、关系、属性及约束。

3.2

本体模型 ontology model

本体的形式化表达结果，由实体类型、关系、属性及约束等建模要素构成。

注：本文件中“本体模型”与“本体”在指称构建成果时含义相同。

3.3

实体类型 entity type

对领域中具有共同特征的一类实体的抽象，亦称概念或类。

3.4

知识图谱 knowledge graph

以图形式组织和表示知识的数据结构，由实体、关系及其属性构成。

3.5

检索增强生成 retrieval-augmented generation; RAG

通过检索外部知识源并将检索结果注入生成过程,以增强大模型生成内容准确性与可追溯性的技术架构。

4 符号和缩略语

下列符号和缩略语适用于本文件。

- IRI 国际化资源标识符 (Internationalized Resource Identifier)
- OWL 网络本体语言 (Web Ontology Language)
- RAG 检索增强生成 (Retrieval-Augmented Generation)
- RDF 资源描述框架 (Resource Description Framework)
- SHACL 形状约束语言 (Shapes Constraint Language)
- SKOS 简单知识组织系统 (Simple Knowledge Organization System)
- SPARQL SPARQL协议与RDF查询语言 (SPARQL Protocol and RDF Query Language)

5 总体框架

5.1 构建原则

金融领域本体模型的构建应遵循以下原则:

- a) 业务驱动: 以金融业务场景与应用需求为导向, 优先覆盖核心业务概念;
- b) 标准先行: 优先采用国家标准、行业标准及国际通用技术规范中的术语与建模方法;
- c) 语义明确: 实体类型、关系、属性的定义应清晰、无歧义, 并具有形式化表示;
- d) 可扩展性: 本体结构应支持新业务、新概念的平滑扩展, 保持向后兼容;
- e) 可追溯性: 本体的构建依据、变更历史与版本信息应完整记录并可追溯。

5.2 全生命周期建设要求

金融领域本体建模的生命周期包括需求分析、本体设计、本体实现、验证评估、发布应用、演化维护六个阶段。各阶段的主要活动见表1:

表 1 金融领域本体生命周期各阶段要求

阶段	主要活动	输出物
需求分析	明确应用目标、业务范围与能力问题，识别知识来源。	需求规格说明、能力问题清单
本体设计	确定实体类型、关系、属性与约束，形成本体概念模型。	本体设计文档
本体实现	采用形式化语言编码实现，完成术语对齐与既有本体映射。	本体文件
验证评估	开展一致性校验、质量评价与专家评审。	验证评估报告
发布应用	分配版本号并发布，接入知识图谱、RAG等应用系统。	发布版本、接口文档
演化维护	收集变更需求，实施版本演化管理。	变更记录、新版本

5.3 角色与职责

本体构建与应用过程宜明确领域专家、知识工程师、应用方与治理方四类角色。领域专家负责业务概念与规则的确认；知识工程师负责本体建模与实现；应用方负责提出应用需求并反馈使用效果；治理方负责术语对齐审查、版本评审与发布管理。

6 本体建模方法与要素

6.1 建模方法

6.1.1 建模路径

金融本体建模宜采用自顶向下与自底向上相结合的混合路径：自顶向下路径依据金融业务架构与监管制度抽象核心概念体系；自底向上路径从业务系统数据模式、产品说明书、监管报表等实际知识来源中归纳概念。两条路径的结果应通过评审达成一致。

6.1.2 表示语言

金融本体应采用RDF/RDFS或OWL 2等W3C推荐的形式化语言进行表示；约束规则宜采用SHACL表达；术语与概念体系组织可采用SKOS。本体文件应具备机器可读的序列化形式（如Turtle、RDF/XML、JSON-LD）。

6.1.3 命名规则

- 本体及其元素应使用IRI进行全局唯一标识。命名应遵循以下要求：
- a) 实体类型使用大驼峰命名（如FinancialProduct），关系与属性使用小驼峰命名（如issuedBy）；
 - b) 中文标签（rdfs:label）为必选，英文标签宜同时提供；
 - c) 命名空间宜按业务域划分，并记录于版本说明文档；
 - d) 已被弃用的标识不得复用。

6.2 实体类型

金融本体实体类型应依据金融业务本质特征划分，宜覆盖以下核心类别：金融机构、金融客户、金融产品与服务、账户、合约、交易、金融事件、监管规则、风险要素、凭证与文档等。各类实体类型的定义、划分依据与描述示例见本章示例。

实体类型的定义应符合以下要求：

- a) 每个实体类型应有唯一的名称、定义及中文标签；
- b) 实体类型之间通过层级关系（subClassOf）组织成分类体系，宜采用单继承为主、多继承审慎使用的原则；
- c) 划分应互斥且尽量穷尽，同级实体类型之间语义边界清晰；
- d) 新增实体类型应说明与既有类型的区别与联系，并附典型实例。

示例：实体类型“金融产品”的描述示例如下：

- 名称：FinancialProduct；中文标签：金融产品；
- 定义：金融机构设计开发并向客户提供的、具有特定收益与风险特征的产品或服务；
- 上位类型：无（顶层实体类型）；
- 下位类型：理财产品、信贷产品、保险产品等；
- 主要属性：产品名称、产品代码、产品类型、风险等级、募集方式；
- 主要关系：发行（金融机构—金融产品）、购买（金融客户—金融产品）；
- 约束：产品代码全局唯一；风险等级取值为R1～R5。

6.3 关系

6.3.1 关系类型

金融本体关系宜包括但不限于以下类型，见表2。

表 2 金融本体关系类型

关系类型	说明
层级关系	实体类型之间的上下位关系。
参与关系	主体对业务活动的参与。
支配关系	所有权、控制权、管理权关系。
时序关系	事件或状态的时间先后关系。
监管关系	监管主体、规则与对象的关联。

6.3.2 关系定义要求

关系的定义应明确以下要素：名称与中文标签、定义域、值域、语义说明、逆关系（如适用）、传递性、对称性、函数性等特性（如适用），以及与既有标准词汇的对应关系。

6.4 属性

属性分为数据属性与对象属性。属性的定义应符合以下要求：

- a) 每个属性应明确所属实体类型或关系、数据类型、取值范围、计量单位（如适用）及是否必填；
- b) 金额类属性应标明币种表示方式，日期类属性应采用GB/T 7408规定的日期时间表示法；
- c) 代码类属性（如机构代码、产品登记编码）应注明所遵循的国家标准、行业标准或监管编码规则；
- d) 枚举型属性应给出完整取值清单及含义说明。

6.5 约束

约束用于保证本体及其实例数据的语义正确性，应包括以下类别，并宜采用SHACL等机器可执行方式表达，见表3。

表3 金融本体约束类别

约束类别	说明	示例
基数约束	限制关系或属性的取值数量。	每份合约至少有一个签约方。
值域约束	限制属性取值的数据类型与范围。	利率取值为非负小数。
互斥约束	规定概念或取值间的互斥关系。	同一账户不能同时为借记账户与贷记账户。
等同约束	规定实例等同或不同的条件。	统一社会信用代码相同的机构为同一实例。
时序约束	限制时间维度上的先后与区间关系。	交易日期不早于账户开立日期。

约束应经过领域专家评审并随本体版本一同发布；实例数据接入时应依据约束进行校验，校验结果应可记录、可追溯。

7 术语对齐与映射要求

7.1 术语规范化

金融本体使用的术语应来源于国家标准、金融行业标准、监管规范性文件及行业通行用法。每个概念应确定唯一的首选术语，同义词、简称、英文对应词宜作为标签管理。术语定义应表述准确、简洁，避免循环定义。

7.2 术语对齐

术语对齐应遵循以下流程：

- a) 收集：系统收集拟对齐的术语及其来源，建立术语清单；
- b) 比对：将本体术语与GB/T、JR/T等相关标准中的术语逐一比对，识别等同、近义、包含、交叉等关系；
- c) 协调：存在差异时，优先采用国家标准术语，其次采用行业标准术语，并保留差异说明；
- d) 记录：对齐结果（对应关系、依据、差异处理决定）应记录并纳入版本说明。

7.3 本体映射

跨机构、跨业务域本体之间的映射应满足以下要求：

表4 本体映射类型及表示方式

映射类型	说明	建议表示
等价映射	两概念语义相同或实质等同。	owl:equivalentClass / skos:exactMatch
上下位映射	一概念为另一概念的上位或下位。	rdfs:subClassOf / skos:broaderMatch
相关映射	两概念存在语义关联但非等同或上下位。	skos:relatedMatch

映射条目应标注映射类型、置信度（高/中/低）、建立依据与评审状态。涉及跨机构知识交换的映射应经双方共同评审确认后方可生效，映射的变更纳入本体版本演化管理。

8 本体版本演化管理

8.1 版本编号

本体版本号采用“主版本号.次版本号.修订号”三段式编号（如1.2.0），规则如下：

- 主版本号：发生不兼容变更（删除或改变既有元素语义）时递增；
- 次版本号：发生向后兼容的新增（新增实体类型、关系、属性）时递增；
- 修订号：仅进行编辑性修改（如标签、注释修订）时递增。

8.2 变更分类与影响评估

本体变更分为兼容变更与不兼容变更。实施变更前应开展影响评估，评估范围应覆盖：依赖该本体的知识图谱数据、RAG索引与提示模板、本体映射条目及下游应用接口。不兼容变更应制定迁移方案，并预留不少于一版本的过渡期。

8.3 评审与发布

版本发布前应经过领域评审与技术评审：领域评审确认概念与规则的业务正确性；技术评审确认语法合规性、逻辑一致性及与既有版本的兼容性。发布物应包括本体文件、版本说明（变更清单）、映射文件及校验规则集。

8.4 废弃与迁移

拟废弃的实体类型、关系或属性应先标记为弃用（如owl:deprecated），并在版本说明中给出替代元素与迁移指引；历史版本本体文件应长期保存并可获取，以支撑存量数据与系统的追溯。

9 本体应用衔接要求

9.1 与知识图谱的衔接

本体与知识图谱的衔接应符合以下要求：

- 模式映射：知识图谱的模式（schema）应由本体定义或与本保持显式映射，图谱中的节点类型、边类型与属性应对应到本体的实体类型、关系与属性；
- 实例挂载：实例数据接入图谱时应声明其所属实体类型，并通过本体定义的约束进行校验；

- c) 标识符对齐：图谱实例标识宜采用与本体一致的命名空间，跨图谱融合时依据本体映射进行标识对齐；
- d) 查询接口：宜提供基于SPARQL或图查询语言的语义查询接口，支持按本体概念进行查询与推理；
- e) 同步更新：本体版本升级后，图谱模式与存量实例数据应按迁移方案同步更新，并记录对应的本体版本号。

9.2 与检索增强生成（RAG）系统的衔接

本体与RAG系统的衔接应符合以下要求，典型衔接流程见附录B：

- a) 本体增强索引：宜利用本体的层级关系与同义词扩展检索索引，支持按概念进行文档切分、标注与聚合；
- b) 查询理解：宜利用本体术语对用户查询进行规范化改写与同义扩展，提升检索召回率与准确性；
- c) 知识注入：检索结果注入大模型提示上下文时，应保留与本体的关联标识，宜同时注入相关概念的定义与约束，约束生成内容的语义边界；
- d) 来源追溯：生成内容所依据的知识应可追溯至本体概念及其版本，支撑结果核验与合规审计；
- e) 更新同步：本体版本更新后，RAG系统的索引、术语词典与提示模板应在规定时限内同步更新，并验证一致性。

10 本体质量评价

金融本体发布前及版本演化过程中应开展质量评价，评价维度与指标见表5。

表 5 金融本体质量评价维度

评价维度	说明	典型指标
完整性	对目标业务范围的覆盖程度。	能力问题覆盖率、核心概念覆盖率
一致性	逻辑与语义的无矛盾程度。	推理一致性校验通过率、约束冲突数
准确性	概念与规则的业务正确性。	领域评审通过率、定义合规率
可用性	支撑下游应用的实际效果。	图谱实例校验通过率、RAG检索命中率

评价结果应形成报告，作为版本发布评审的输入；未达到质量要求的版本不得发布。