

T/BIA

团体标准

T/BIA XXXX—202X

生物医学本体 第2部分：开发规范

Biomedical ontology—
Part 2: Development specification

(征求意见稿)

202X - XX - XX 发布

202X - XX - XX 实施

北京信息化协会 发布

目次

前 言	III
引 言	4
1 范围	5
2 规范性引用文件	5
3 术语和定义	5
3.1 本体术语	5
3.2 专业术语	5
4 缩略语	6
5 开发规范	6
5.1 确定本体研究目标和范围	6
5.2 开展资源调查	7
5.3 本体设计模式构建	7
5.4 基于设计模式的本体实现	8
5.4.1 使用本体设计模式指导本体结构设计	8
5.4.2 顶层本体：逻辑框架与分类体系	8
5.4.3 中层整合：模式复用与语义对齐	9
5.4.3.1 确定中层本体范围与复用策略	9
5.4.3.2 本体提取与模块化导入	9
5.4.3.3 语义对齐与关系协调	10
5.4.4 语义扩展与约束设计	10
5.4.4.1 新增类	10
5.4.4.2 新增属性	11
5.4.4.3 新增类间关系	11
5.4.5 添加本体元数据描述	11
5.5 本体质量核验	11
附录 A(资料性) 构建生物医学本体示例	13

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。

本文件是《生物医学本体》的第2部分。《生物医学本体》由以下部分组成：

——第1部分：开发基本原则

——第2部分：开发规范

——第3部分：元数据

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。

本文件由北京信息化协会提出并归口。

本文件起草单位：中国医学科学院基础医学研究所、国家人口健康科学数据中心、中国中医科学院中医信息研究所、北京及遇智悦生物科技有限公司、哈尔滨医科大学、重庆邮电大学、中国医学科学院协和医院、中国医学科学院医学信息研究所。

本文件主要起草人：杨啸林、王哲、朱彦、杨晟、张敬晨、邵晨、彭苏元、周伟、李晓瑛、刘丽红、谢江安、关键、程亮、姚克宇、何勇群、张胜发、罗葳。

引言

生物医学本体的开发在实际操作中常面临范围不清、术语重复、结构混乱等问题，影响本体的质量与复用。为系统指导生物医学本体的工程化构建，提升本体的语义一致性与互操作性，特制定本系列文件。

本系列文件由以下三部分构成：

第 1 部分：开发基本原则。明确生物医学本体在开发过程中应当遵循的基本原则，从方法论层面界定本体开发的总体要求与约束。

第 2 部分：开发规范。规定生物医学本体从规划、设计、实现，到验证与发布的完整开发流程，为本体工程实施提供操作性指导。

第 3 部分：元数据。规定生物医学本体元数据的描述方法，明确核心元数据与扩展元数据的组成及其适用范围。

本文件作为第 2 部分，将第 1 部分提出的开发原则转化为可操作的开发规范，涵盖资源调查、概念建模、语义对齐、逻辑验证等关键环节，并为满足第 3 部分的元数据要求提供实施依据。

生物医学本体

第2部分：开发规范

1 范围

本文件规定了生物医学本体在开发过程中应遵循的规范要求，涵盖从目标设定、资源调查、概念建模、语义对齐、术语复用、逻辑验证到发布管理的全流程技术要点。

本文件适用于科研机构、数据平台建设单位、本体开发团队和标准化组织等相关单位的各类生物医学领域的本体工程实践业务中的本体(如疾病、基因、药物、解剖、检验、公共卫生、中医药等)的设计、开发与验证。

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 18391.3-2009 信息技术元数据注册系统(MDR) 第3部分：注册系统元模型与基本属性

GB/T 38371.1-2020 数字内容对象存储、复用与交换规范 第1部分：对象模型

ISO/IEC 21838-1 First edition 2021-08 Information technology Top-level ontologies(TLO) Part 1: Requirements

ISO/IEC 21838-2 First edition 2021-11 Information technology-Top-level ontologies(TLO) Part2: Basic Formal Ontology (BFO)

3 术语和定义

3.1 本体术语

3.1.1

类 class

一个对象集的描述，该对象集具有公共的属性、操作、方法、关系和语义。

[来源：GB/T 18391.3—2009,3.14]

3.1.2

属性 property

一个对象或实体的特征。

[来源：GB/T 18391.3—2009,3.13]

3.1.3

本体设计模式 ontology design pattern

在本体工程中，用于解决反复出现的建模问题的一种可复用语义建模方案。本体设计模式提供了可扩展的实体与关系表示框架，有助于提高本体的结构一致性、可维护性和语义质量。

3.2 专业术语

3.2.1

生物医学研究 biomedical study

生物医学研究是实现研究设计步骤的过程。对一个研究主题进行详细的检查、分析，旨在发现有关它的事实。

3.2.2

样本收集 specimen collection

采集标本的过程。

3.2.3

化验 assay

一种有计划的过程，其目标是通过检查来产生有关重要实体(评估者)的信息。

3.2.4

数据处理 data processing

计算机对数据进行一系列操作，以检索、转换或分类信息。

4 缩略语

下列缩略语适用于本文件。

BFO 基本形式化本体(Basic Formal Ontology)

CHEBI 生物相关化学实体本体(Cheical Entities of Biological Interest)

CL 细胞本体(Cell Ontology)

CSV 逗号分隔值(Comma-Separated Values)

DOID 人类疾病本体(Disease Ontology)

HPO 人类表型本体(Human Phenotype Ontology)

IAO 信息工具体本(Information Artifact Ontology)

ICD 国际疾病分类(International Statistical Classification of Diseases and Related Health Problems)

IRI 国际化资源标识符(Internationalized Resource Identifier)

JSON-LD 关联数据的 JSON 格式(JavaScript Object Notation for Linked Data)

LOINC 观测指标标识符逻辑命名与编码系统(Logical Observation Identifiers Names and Codes)

MeSH 医学主题词表(Medical Subject Headings)

MONDO 统一疾病本体(Mondo Disease Ontology)

NCIt 美国国家癌症研究所词表(National Cancer Institute Thesaurus)

OBI 生物医学调查本体(Ontology for Biomedical Investigations)

OBO 开放生物医学本体格式(Open Biomedical Ontologies Format)

OBO Foundry 开放生物医学本体工厂(Open Biological and Biomedical Ontologies Foundry)

ODP 本体设计模式(Ontology design pattern)

OWL 2 网络本体语言第二版(Web Ontology Language 2)

RO 关系本体(Relation Ontology)

SKOS 简单知识组织系统(Simple Knowledge Organization System)

SNOMED CT 医学系统命名法—临床术语(Systematized Nomenclature of Medicine-Clinical Terms)

TTL 简洁 RDF 三元组语言(Terse RDF Triple Language)

TSV 制表符分隔值(Tab-Separated Values)

UBERON 跨物种解剖结构(UBERON Multispecies Anatomy Ontology)

URI 统一资源标识符(Uniform Resource Identifier)

5 开发规范

5.1 确定本体研究目标和范围

确定研究目标和范围是本体开发的起点，也是后续结构设计、语义对齐和数据互操作的依据。

对于生物学本体，其核心目标是通过语义建模实现知识标准化与可计算化，以支持跨数据库、跨机构的科学数据共享和复用。

在定义目标时应明确：

- 1)目标类型：是用于科研语义标准化与集成、临床数据标准化，还是用于知识图谱构建；
- 2)适用范围：描述的生物医学对象(如疾病、药物、检测指标等)与预期应用场景；
- 3)知识边界：确定建模深度、抽象层次和与上层本体(如 BFO)的对齐程度。

范围界定应遵循满足具体应用、可扩展和可复用的原则，确保初始建模不因过度设计或过窄定义而影响后续集成。

5.2 开展资源调查

资源调查的目的是在建模前全面掌握现有语义资源与标准框架，以避免重复开发、促进本体重利用并提升互操作性。调查应包括以下三个层面：

- 1)知识层：检索并评估可复用的现有本体(如 OBO Foundry、SNOMED CT、LOINC、NCIt、MeSH)。
- 2)数据层：分析现有领域内数据集或数据库的结构、数据元定义及标准，明确与本体的映射关系。
- 3)语义层：识别概念与术语的语义差异和重叠区域，为后续语义对齐提供基础。

调查方法可包括：

- 1)系统性文献与数据库检索；
- 2)专家咨询与社区交流；
- 3)资源比对与差距分析(可使用在生物学领域本体资源库 Onto Portal Alliance、EMBL-EBI Ontology Lookup Service 以及 ROBOT diff、OntoFox 等辅助)。

5.3 本体设计模式构建

本体设计模式是解决本体开发中反复出现的语义建模问题的一种可复用建模方案。在生物学领域，本体设计模式通过定义核心类、关系与约束结构，为知识建模提供统一的语义框架，提升本体的一致性、可扩展性、可维护性和互操作性。

本体设计模式的构建，从识别领域中反复出现的建模问题出发，分析各问题的知识单元实体(entity)、关系类型与语义约束，形成可复用的结构模板，具体步骤如下：

1)问题识别与模式抽象：在明确建模目标与应用场景基础上，识别在目标场景或知识领域中反复出现的语义问题；通过分析现有生物学领域本体和真实数据模型，提炼出具有普遍意义的知识结构；将这些结构抽象为候选模式。

2)确定核心类：列出领域内的核心概念，这些概念将是本体模式图中的主要节点，也是本体构建过程中具体领域术语的锚定点。核心类应能尽可能与顶层或中层本体(如 BFO、OBI、RO)对应，确保可互操作。

3)定义并建立核心关系

依据领域语义特征与知识结构，识别并利用对象属性(object property)形式化描述核心类之间的关系。

区分层级关系(如 is_a, 表示类的继承与分类层次)与关联关系(如 part_of、has_part、participates_in、derives_from 等)。

在本体语义模型中，以类(class)作为节点、以对象属性(object property)作为边，通过公理化表达(如 rdfs:subClassOf、owl:ObjectProperty 等)建立核心类之间的连接，明确其语义关联。

4)定义语义注释

为各核心类与关系添加语义注释信息(annotation properties)，包括标签(rdfs:label)、定义(IAO:0000115)、同义词(oboInOwl:hasExactSynonym)、来源(IAO:0000119)及数据类型说明。

注释应准确反映概念内涵与使用上下文，保证模型的可读性与可追溯性。

5)设定语义约束

为模式中定义的关系和属性添加逻辑约束，包括：

域与范围(Domain/Range)：限定关系适用的类；

基数约束(Cardinality Restrictions)：定义关系的出现次数(如 \forall 、 \exists 、 $=1$)；

类型约束与互斥性(Type and Disjointness Restrictions): 防止逻辑冲突或循环定义。

这些约束应以 OWL 2 描述逻辑形式公理化表达, 以确保语义模型的一致性与推理可计算性。

6) 本体设计模式输出

在完成核心类、关系、注释与约束的定义后, 应将本体设计模式转化为人机双可读的结构化输出, 以支持跨团队复用和后续的自动化语义生成。模式文件应采用标准本体表示格式(如 OWL 2、TTL、OBO 或 JSON-LD), 包含类、属性、公理及逻辑约束。为了后续建模方便, 建议生成基于 CSV/TSV 的语义模板(例如 ROBOT template)便于后续自动生成新术语或模块。也可输出模式结构图, 直观反映核心类的语义关系。

5.4 基于设计模式的本体实现

其目标是依据已构建的本体设计模式, 形成逻辑一致、语义复用、结构分层的可维护本体。设计应遵循“上层约束一致性—中层模式复用—下层语义扩展”的层次化原则, 确保各级本体在逻辑上可对齐、在语义上可继承、在工程上可实现。

实现的具体流程如下:

- 1) ODP 指导本体结构设计: 利用 ODP 组件将抽象语义结构转化为可操作的本体模块。
- 2) 顶层构建: 建议基于 ISO/IEC 21838-2 所定义的基本形式化本体, 形成统一的实体分类框架。
- 3) 中层整合: 将 ODP 与现有领域本体(如 OBI、CL、ChEBI、DOID)进行复用和语义映射。
- 4) 扩展与定制: 基于 ODP 模板向具体应用领域引入新术语、属性与关系的过程。
- 5) 本体元数据描述: 添加本体元数据描述信息。

5.4.1 使用本体设计模式指导本体结构与模块实现

应以已构建的本体设计模式为语义蓝图, 将抽象语义结构逐步实例化为可计算的本体模块。模块之间应共享公共核心类与关系, 实现语义一致与逻辑可组合性。应明确本体设计模式的引用、实例化和组合规则, 保证术语结构与模式定义保持一致。

5.4.2 顶层本体: 逻辑框架与分类体系

顶层设计旨在为领域本体提供统一的逻辑框架与语义约束。推荐采用 ISO/IEC 21838-2 所定义的基本形式化本体作为顶层本体(见图 1), 以保证各模块间的逻辑一致性和跨本体互操作性。

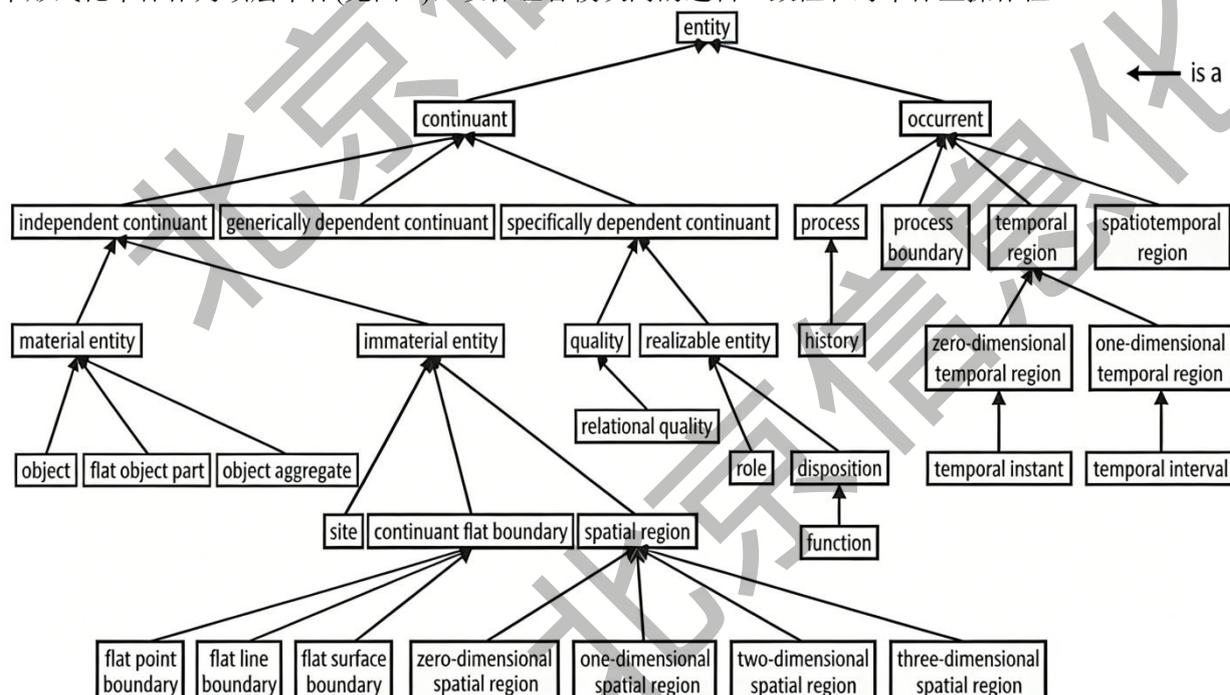


图1 BFO 的系统树形结构

1)明确类别与逻辑定位

应根据领域知识特点，明确本体设计模式中涉及的核心类在 BFO 分类体系中的对应位置，如 *continuant*(常体)、*occurrent*(行体)、*role*(角色)、*quality*(性质)等。核心类的上位类别(不一定是直接上位类)应能映射到 BFO 的相应节点，从而保证逻辑继承关系清晰、语义边界明确。

2)引用与集成方式(Import and Integration)

在构建领域本体时，应通过 OWL import 或等价机制引用 BFO 的正式 OWL 文件，保留其原始 IRI 与命名空间，以确保可追溯性与跨本体一致性。

推荐的引用方式包括：

- 使用 BFO 官方发布版本(如 <http://purl.obolibrary.org/obo/bfo.owl>)；
- 在本体头部声明 owl:imports 指向引用该文件；
- 不应修改或重命名 BFO 类与属性的 IRI，仅可添加本地注释或对齐关系。

3)引用范围与复用策略

本体设计模式构建阶段确定的核心类，应对齐或继承自 BFO 类：

对应于过程类的核心概念，继承自 *bfo:process* 或其子类；

对应于物质实体的核心概念，继承自 *bfo:material entity*；

对应于属性或质量的核心概念，继承自 *bfo:quality*；

对应于角色或功能的概念，继承自 *bfo:role* 或 *bfo:function*。

对于不直接属于 BFO 范畴的领域特有概念，应通过中层本体(如 OBI、IAO 或 RO 等)或高质量领域本体补充表达。

4)版本管理与验证

应明确 BFO 的引用版本，并在本体元数据中记录其版本号、发布日期及访问地址。

在逻辑验证阶段，应确认所有被导入类与属性与 BFO 保持一致，且推理后无逻辑冲突。

5.4.3 中层整合：模式复用与语义对齐

中层整合是在 BFO 顶层框架下，结合具体应用需求与本体设计模式定义的核心语义结构，将本体设计模式与现有高质量领域本体(如 OBI、RO、IAO、CL、CHEBI、DOID 等)进行复用与对齐的过程。

其目标是：在保持逻辑一致性的前提下，实现本体设计模式到领域知识的无缝衔接，避免重复建模，确保语义复用和跨系统互操作。

5.4.3.1 确定中层本体范围与复用策略

应根据本体设计模式的知识域与建模目标，确定中层本体的适用范围及复用策略。

若 ODP 涉及实验、样本、测量等语义，应优先复用 OBI；

若涉及生物结构、解剖层级，建议复用 CL 或 UBERON；

若涉及化学分子或代谢物，建议复用 ChEBI；

若涉及疾病与症状，建议复用 DOID 或 MONDO，ICD 或 SNOMED-CT；

若涉及通用语义关系，应复用 RO；

若涉及数据与注释属性，应复用 IAO。

本体设计模式中的核心类应映射到这些中层本体中的相应概念，形成可继承的语义路径。

5.4.3.2 本体提取与模块化导入

为保证整合的可控性和可维护性，中层本体不宜整体导入，应采用模块化提取(modular extraction)方式复用。

应根据新建本体的语义范围与应用目标，确定抽取范围与深度，使导入的模块既能满足复用需求，又避免冗余依赖。

本体抽取应遵循以下原则：

- 1)最小充分原则：仅抽取与目标领域直接相关的类、属性及其必要的上位类，形成语义闭合的最小单元；
 - 2)结构保持原则：在抽取过程中应尽量保留原有本体的层级结构、关系约束及逻辑依赖，确保导入后语义一致；
 - 3)可追溯原则：抽取与导入过程应记录来源信息、版本号及对应关系，以支持后续验证与维护；
 - 4)一致性原则：抽取后模块应与原本体保持逻辑一致性，不得因局部修改造成语义冲突或断裂。
- 导入的术语应保留其原始 IRI 与注释信息(包括定义、来源、编辑者等)，并在本体元数据中标注其来源本体、版本号、发布日期及引用方式。
- 当导入模块涉及多个来源时，应建立跨来源的引用说明文档，以保障术语对齐的透明性与可复用性。

5.4.3.3 语义对齐与关系协调

语义对齐是确保不同来源本体间逻辑一致与语义可交换的关键步骤。

1) 语义对齐方式：

- 使用 owl:equivalentClass 表达类之间语义等价关系；
- 使用 rdfs:subClassOf 表达类的语义继承或包含关系；
- 使用 owl:sameAs 表示实例级等价关系。

2) 关系协调：

- 生物医学领域的关系应优先复用 RO 定义的标准关系(如 part_of、has_part、participates_in、occurs_in 以及 interacts with 等)；
- 当领域需要更精细的关系时，可通过 owl:propertyChainAxiom 定义复合关系，但须保持可推理性；
- 不建议重复定义 RO 中已有关系或创建本地“近义属性”。

3) 对齐验证：

应对语义对齐结果进行系统验证，确保整合后的逻辑一致性与语义完整性。验证工作应包括：

- 一致性检查：确认跨本体映射后无语义冲突、重复定义或孤立节点；
- 完整性检查：确保所有复用类和关系均具备定义、标签及必要的注释信息；
- 对齐记录：应建立“语义映射表(alignment mapping table)”，系统列出本地类、外部类及对应关系类型(如等价关系、从属关系、关联关系等)，并在元数据中记录其来源与验证状态；
- 质量评估：对对齐结果的准确性与可维护性进行人工复核或自动化校验，形成验证报告。

5.4.4 语义扩展与约束设计

语义扩展是在顶层逻辑框架与中层整合的基础上，根据特定应用需求对本体进行领域细化的过程。推荐使用模板分步骤添加类、属性和类间关系，其主要任务是在保证逻辑一致性与语义复用并通过适当的逻辑约束确保模型可计算、可验证。

5.4.4.1 新增类

对于领域特有的概念，应通过新增类的方式实现语义扩展。标准化的类模板可以保证新建类的一致性和规范性，有利于本体开发者之间共享数据和协作开发。表 1 展示了新建类应满足的最小元数据标准。

表 1 新增类最小元数据标准

元数据项名称	RDF/OWL 属性	说明
类 ID *		类唯一标识符，应在命名空间内唯一，具备可解析的 IRI。
标签 *	rdfs:label	类的标准名称，应简洁、明确、符合领域命名规范。必要时可附加多语言标签。
定义 *	IAO:0000115	类定义信息。对类的语义进行精确描述，应能独立表达概念含义。
创建者	dcterms:creator	记录术语来源(文献、标准、上游本体)及责任人(或开发机构)。

*为必填项。

对于多语言本体中的类标签或定义，可为同一实体添加多个带语言标识符。

例如：<rdfs:label xml:lang="zh">血压</rdfs:label>
或者 rdfs:label "血压"@zh

5.4.4.2 新增属性

对于表达新语义关联或数据结构需求的情形，可定义新的对象属性(object property)、数据属性(data property)或注释属性(annotation property)。所有属性应尽可能复用中层关系本体(如 RO、IAO)中的已有关系，仅在必要时新增。

新增属性的命名应遵循语义清晰、结构统一、前缀标准的原则，禁止重复命名或非语义化缩写。表 2 展示了新建属性应满足的最小元数据标准。

表 2 新增属性最小元数据标准

元数据项名称	RDF/OWL 属性	说明
属性标识符*		属性的唯一标识符，应具备可解析的 IRI，并在命名空间中保持唯一。
属性类型*	rdf:type	明确属性类型为对象属性(owl:ObjectProperty)、数据属性(owl:DatatypeProperty)或注释属性(owl:AnnotationProperty)。
标签*	rdfs:label	属性的标准名称，应语义明确、简洁规范。
定义*	IAO:0000115	对属性语义进行精确定义，应能独立说明其逻辑含义与适用场景。
创建者	dcterms:creator	记录属性的创建者或维护机构。
对象属性特征		指定对象属性逻辑特征，如：函数性(Functional)、逆函数性(Inverse functional)、传递性(Transitive)、对称性(Symmetric)、不对称性(Asymmetric)、自反性(Reflexive)、非自反性(Irreflexive)等。
域	rdfs:domain	指定该属性适用的起始类或实体类型。
范围	rdfs:range	指定该属性的目标类、数据类型或值域。
来源		指明属性定义来源(参考本体、标准或文献)。

*为必填项。

5.4.4.3 新增类间关系

新增类建立后，应根据领域知识与语义逻辑，明确定义其与现有类之间的语义关系。类间关系定义是语义扩展的重要组成部分，通过合理建立类间关系，可确保本体语义网络结构完整、逻辑一致、可推理。

类间关系定义要求：

- 1) 应保证每个新增类至少与一个现有类建立逻辑关联(is-a)；
- 2) 类间关系应尽量基于已定义或复用的标准关系(如 is_a、part_of、has_part、derived_from、participates_in 等)，或者前述步骤已经定义的领域特有关系；
- 3) 新增类间关系应保持语义方向清晰、逻辑一致，避免循环依赖或多义关系；
- 4) 在定义类与类之间的关系时，应使用量词限定(Quantifier Restrictions)，来表达类成员对特定关系的存在性或约束性要求。
- 5) 所有类间关系均应可被逻辑推理机识别与验证；
- 6) 为保证建模一致性与可复用性，宜使用统一的类间关系模板(Inter-class Relation Template)定义新增类的逻辑约束。

5.4.5 添加本体元数据描述

向 owl 本体文件添加注释元数据。应提供本体 IRI、版本 IRI、本体标题、描述和许可信息，以支持未来的使用和管理。具体描述可参考“生物医学本体开发最佳实践 第 3 部分：描述规范”。

5.5 本体质量核验

本体质量核验是本体构建后的综合检验环节，旨在验证逻辑一致性、结构完整性、元数据完备性及

领域覆盖性，确保本体满足发布与应用要求。可结合自动化工具与人工复核共同实施。

核验内容包括：

1) 逻辑推理

应基于符合 OWL 2 DL 语义的推理机执行逻辑推理，推导隐含的类层次与关系；推理结束后，应将推理结论显式固化，以形成逻辑闭合模型，并通过规约消除冗余公理，确保本体在语义上完备、在结构上规整。

2) 结构与一致性检查

应对本体的类层次结构、继承路径及导入模块进行一致性检查；应核实外部本体引用的版本信息、国际化资源标识符(IRI)及命名空间配置的正确性；应识别并处理不一致类、孤立节点和重复继承等问题，以确保本体在逻辑关系和结构组织上的完整性与一致性。

3) 覆盖性评估

应依据预先定义的能力问题对本体的语义覆盖范围进行评估；评估可通过执行 SPARQL 查询(如 ASK、SELECT、CONSTRUCT 等)验证本体是否能够正确响应预设的问题；对评估结果与预期目标不一致的部分，应开展模型补充或语义修订，以确保本体的知识范围与设计目标保持一致。

4) 元数据与注释完善

应补充或更新本体的标识信息，包括本体标识符、版本标识符、标题、描述及许可证等内容；应确保元数据记录完整、准确、可追溯，并与本体的版本控制体系保持一致，以支持后续的发布管理和维护工作。

5) 格式转换与报告生成

应根据发布需求，将经验证的本体导出为 OWL、SKOS 等符合标准规范的文件格式；应生成本体质量控制报告，记录核验方法、发现的问题项及相应的修正措施，以确保发布文件的格式统一、内容完整和质量可追溯。

附录 A

(资料性)

A1 构建生物医学本体示例

此部分是应用本标准规范，构建生物医学研究本体 Biological Study Ontology(bso)的示例。示例开发使用的工具为 ROBOT(一个可应用于开放生物医学本体的命令行工具)和 Protégé(本体可视化工具)。实际开发中可不限于这两款工具，读者可自行选择其他工具(如：ontofox、Zooma、NCBO Annotator、SPARQL 等)完成本体开发工作。

1 确定本体研究目标和范围

本示例构建的生物医学研究本体，其主要目的是整合和标准化生物医学领域的专业知识，以促进不同生物医学数据库和系统之间的数据共享与互操作。在此过程中，本体能够帮助统一术语、定义关键概念并揭示概念之间的关系，从而为跨学科合作提供支持。同时，通过探讨生物医学研究本体的构建过程，提供一个完整的框架，帮助研究人员在开发本体时明确目标、界定范围、选择合适的建模方法，并最终实现跨平台的知识共享与互操作。

2 开展资源调查

本规范探讨的生物医学研究本体是一种用于表示和组织生物医学研究中所涉及的各种概念、设计、实验流程和数据分析的标准化本体。通过开展资源调查，获取的本体资源服务如表 A.1 所示。

表 A.1 本示例中复用的生物医学领域本体资源

资源名称	英文全称	缩写名	描述
基本形式化本体	Basic Formal Ontology	BFO	BFO 是一个领域中立的通用顶层本体。其提供的术语具有高度通用性，如对象、质量、过程、空间和时间区域等，可适用于所有领域。
关系本体	Relation Ontology	RO	RO 本体提供跨多本体共享的关系类型。
信息工具体	Information Artifact Ontology	IAO	IAO 是描述信息实体的本体。
生物医学调查本体	Ontology for Biomedical Investigations	OBI	用于描述实验和临床观测过程。
美国国家癌症研究所词典	National Cancer Institute Thesaurus	NCIt	NCIt 提供了一套统一的、规范化的术语，涵盖了癌症相关的疾病、治疗、临床过程、药物、实验方法、基因等领域。

3 本体设计模式构建

在生物医学研究中，研究设计、实验变量和实验方法是其重要组成部分，其中研究设计是核心，它决定了研究的整体框架、方法和流程。实验变量是基础，它影响实验结果的准确性和可靠性。实验方法是关键，它描述实验的具体操作步骤和技术手段。

对于一项生物医学研究，通常包括以下几个进程：样本采集、实验检测、数据处理等。

样本采集是生物医学研究的起点，也是后续实验检测和数据处理的基础。根据研究目的和实验要求，需要选择合适的样本类型。在生物医学研究中，常见的样本类型包括血液、尿液、组织、细胞等。

实验检测是生物医学研究中的关键环节，它通过对样本的分析和测定，揭示生物体的生理、病理状态或某种物质的性质与含量。在实验过程中，必须严格遵循实验步骤和操作规范，以避免误差和干扰因素的影响。

数据处理是生物医学研究的最后一个环节，它对实验数据进行整理、分析和解释，从而得出科学结论。

基于上述内容及研究流程，本规范构建了如图 A.1 所示的生物医学研究本体模型。

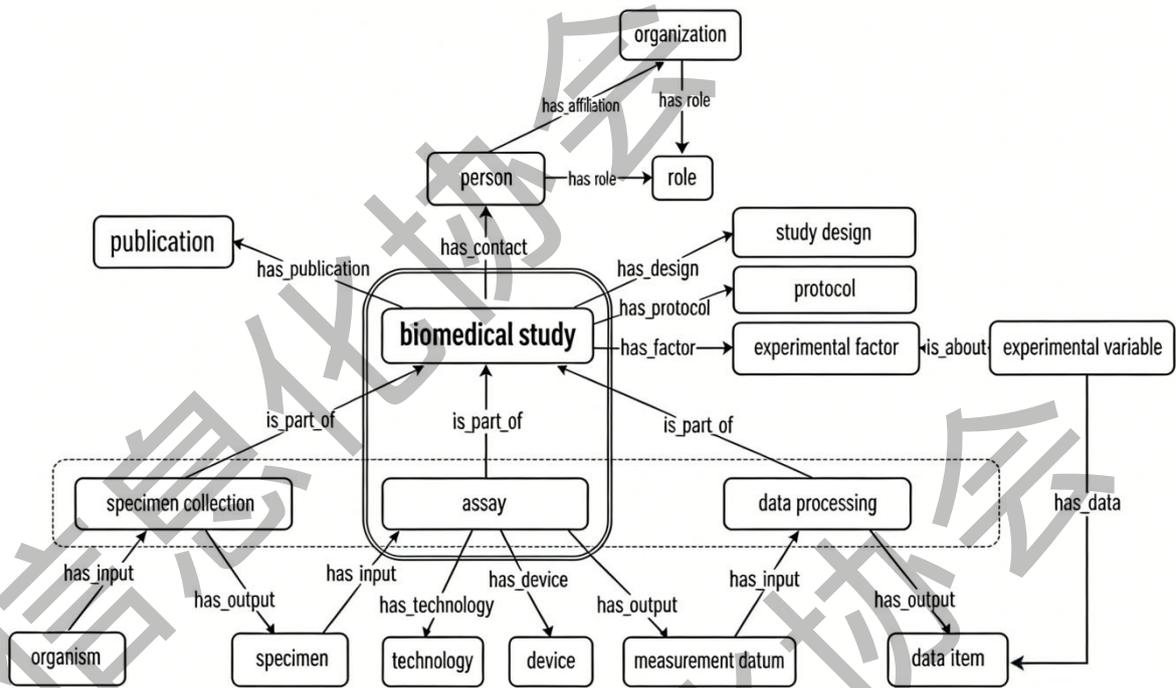


图 A.1 生物医学研究本体模型

4 基于设计模式的本体实现

针对图 2 所示的生物医学研究本体模型,采用本体设计模式的实现流程:顶层构建-> 中层整合-> 扩展与定制,实现的生物医学研究本体整体结构如图 A.2 所示。

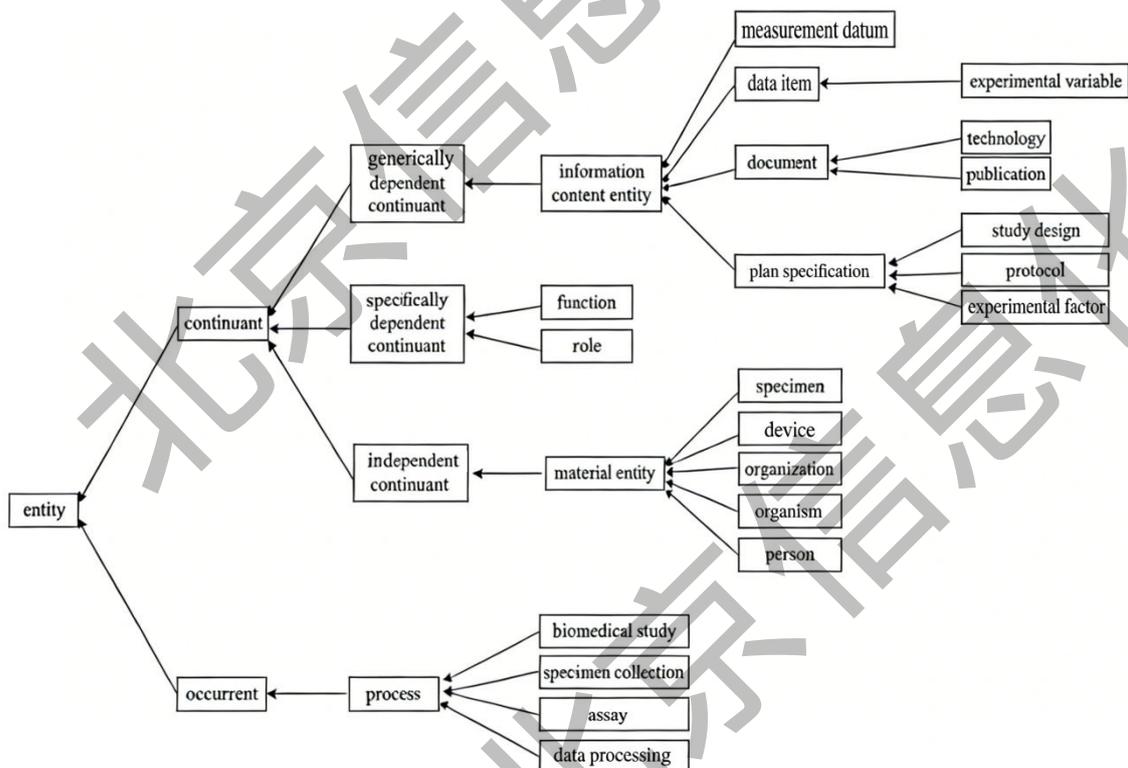


图 A.2 生物医学研究本体核心结构

下面将具体介绍基于本体设计模式实现流程的每一步内容。

4.1 顶层构建

生物医学研究本体采用基本形式化本体(BFO)作为顶层框架(图 1)，这使得待建本体可以继承 BFO 中的类、属性和结构。具体实现时，可使用 `owl:imports` 关键字将其导入。这使得目标本体可以继承 BFO 中的类、属性和结构。示例代码(`bso_dev.owl`)如下：

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <owl:Ontology rdf:about="http://purl.ontochina.org/ontology/bso_dev.owl">
    <owl:imports rdf:resource="file:///D:/bso/backbone/bfo.owl"/>
  </owl:Ontology>
</rdf:RDF>
```

4.2 中层整合

生物医学研究本体采用 IAO、OBI、NCIt 等来源本体作为中层本体进行整合。整合时，需根据“图 3 生物医学研究本体核心结构”中确定的种子术语和具体的抽提算法从来源本体中提取目标本体所需的模块。

4.2.1 本体提取与模块化导入

生物医学研究本体在整合中层本体时，分别以 IAO 中的“`information content entity`”，OBI 中的“`plan specification`”，和 NCIt 中的“`Organism`”、“`Organization`”、“`Person`”等术语作为种子，从来源本体中提取目标本体所需的模块，即为了在应用本体上进行集成而需要提取的实体集，以此来实现术语复用。领域专家可以为给定的应用本体确定正确的种子。本体开发人员可以使用工具来实现基于种子的模块抽提。

下面的示例演示了使用 OBO foundry 的 `robot` 工具从 `iao.owl` 本体文件中以“`information content entity`”为种子提取模块的示例代码。

```
robot extract --method MIREOT --input ./backbone/iao.owl
--upper-term "http://purl.obolibrary.org/obo/IAO_0000030" # information content entity
--lower-term "http://purl.obolibrary.org/obo/IAO_0000027" # data item
--lower-term "http://purl.obolibrary.org/obo/IAO_0000311" # publication
--output ./modules/iao_extract.owl
```

获取到所需模块后，使用 `owl:imports` 关键字将其导入目标本体文件(`bso_dev.owl`)。整合了 IAO、OBI 和 NCIt 等中层本体后的目标本体文件结构如图 A.2 所示。

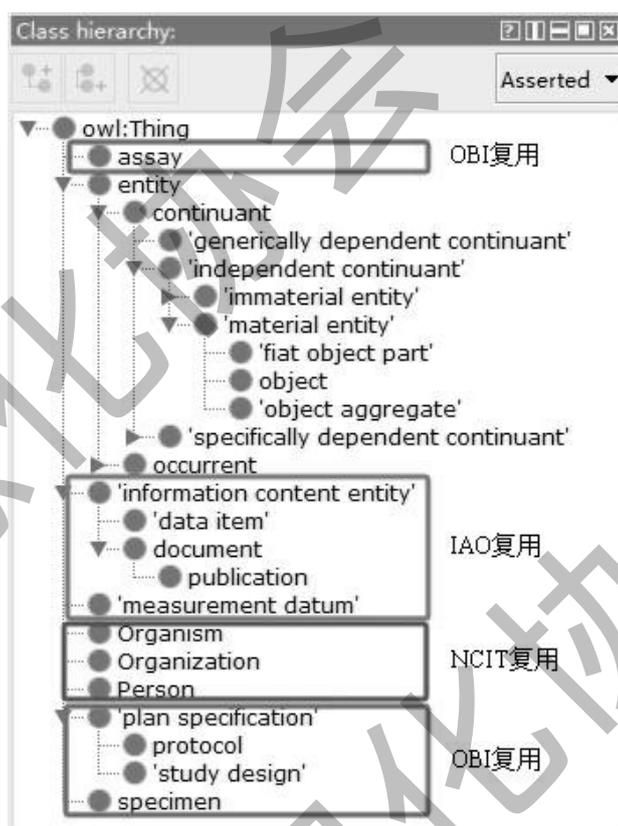


图 A.2 中层本体提取与模块导入后的目标本体结构

除了复用中层本体术语外,在生物医学研究本体中还存在很多关系(如: has role、has input、has output 等),对于这些关系,应优先考虑复用 RO 本体中定义的标准关系。关系复用同术语复用相类似,也可以使用 robot 从 ro.owl 文件中抽提关系模块,然后导入到目标本体中来完成。示例如下:

```
robot extract --method MIREOT --input ./backbone/ro.owl
--upper-term "http://purl.obolibrary.org/obo/RO_0000053" # has characteristic
--lower-term "http://purl.obolibrary.org/obo/RO_0000087" # has role
--upper-term "http://purl.obolibrary.org/obo/RO_0000057" # has participant
--lower-term "http://purl.obolibrary.org/obo/RO_0002233" # has input
--lower-term "http://purl.obolibrary.org/obo/RO_0002234" # has output
--output ./modules/ro_extract.owl
```

将抽提后的关系导入目标本体文件,复用的关系结构如图 A.3 所示:

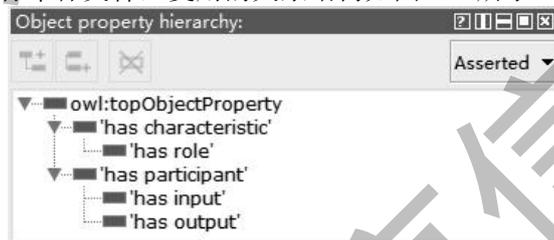


图 A.3 关系提取及复用

4.2.2 语义对齐与关系协调

生物医学研究本体在整合中层本体时,主要通过使用 rdfs:subClassOf 关键字建立术语间的父子关系来进行语义对齐。例如:将 IAO 术语“information content entity”对齐到 BFO 术语“generically dependent continuant”。为此创建一个名为 alignment.owl 的本体文件,在该文件中指明如下的语义对齐关系:

```
<owl:Class rdf:about="http://purl.obolibrary.org/obo/IAO_0000030">
  <!--BFO:generically dependent continuant-->
```

```
<rdfs:subClassOf rdf:resource="http://purl.obolibrary.org/obo/BFO_0000031"/>
<rdfs:label xml:lang="en">information content entity</rdfs:label>
</owl:Class>
```

其他需要对其的来源术语和目标术语如表 A.2 所示。完成语义对齐后，将 alignment.owl 文件导入到目标本体中。整合中层本体并完成语义对齐后的结构如图 A.4 所示。

表 A.2 中层本体语义对齐

来源术语	来源本体	目标术语	目标本体
information content entity	IAO	generically dependent continuant	BFO
measurement datum	IAO	information content entity	IAO
plan specification	OBI	information content entity	IAO
specimen	OBI	material entity	BFO
assay	OBI	process	BFO
organization	NCIt	material entity	BFO
organism	NCIt	material entity	BFO
person	NCIt	material entity	BFO

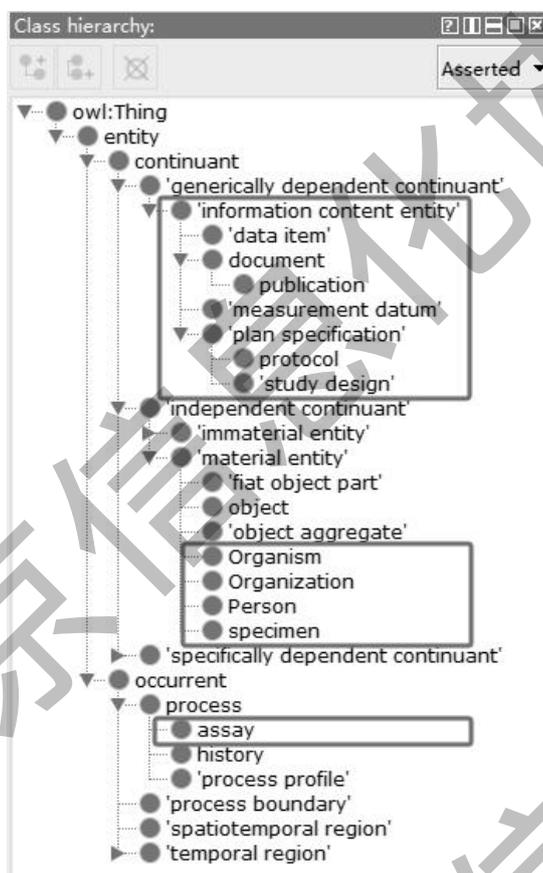


图 A.4 中层本体语义对齐

4.3 语义扩展与约束设计

4.3.1 新增类

对于该示例本体的自有术语(如: biomedical study、specimen collection 等), 应通过创建新类的方式实现语义扩展。标准化的类模板可以保证新建类的一致性和规范性, 有利于本体开发者之间共享数据和协作开发。

ROBOT 提供了一个模板驱动的本体术语生成系统。其模板是一个标签分隔的(CSV)文件, 描述了一组构建本体实体(即类、属性和个体)的模式。结合“表 1 新增类最小信息要求”, 构建的类模板如表 A.3 所示。

表 A.3 新增类模板(bso_classes.tsv)

Class ID	Entity Type	Label	Definition
ID	TYPE	A rdfs:label	A definition SPLIT=
bso:BSO_0000001	class	biomedical study	A biomedical study is a process that realizes the steps of a biomedical study design. A detailed examination, analysis, or critical inspection of a subject designed to discover facts about it.
bso:BSO_0000002	class	specimen collection	A process with the objective of collecting a specimen.
bso:BSO_0000003	class	data processing	A series of operations on data by a computer in order to retrieve or transform or classify information.
bso:BSO_0000004	class	technology	The application of scientific knowledge to practical purposes in any field. It includes methods, techniques, and instrumentation.
bso:BSO_0000005	class	experimental factor	An independent variable manipulated by the experimentalist with the intention to affect biological systems in a way that can be measured by an assay.
bso:BSO_0000006	class	experimental variable	a variable whose values are independent of changes in the values of other variables.
bso:BSO_0000007	class	device	An object intended for use whether alone or in combination for diagnostic, prevention, monitoring, therapeutic, scientific, and/or experimental purposes.

实践中, 使用 robot template 命令可将术语模板文件(bso_classes.tsv)转换为 owl 格式文件:

```
robot template --input ./bso_dev.owl
--template ./templates/bso_classes.tsv
--prefix "bso: http://purl.ontochina.org/ontology/"
--ontology-iri "http://purl.ontochina.org/ontology/bso_classes.owl"
--output ./modules/bso_classes.owl
```

将 bso_classes.owl 导入到目标本体文件中, 添加新类后的本体结构如图 A.5 所示:

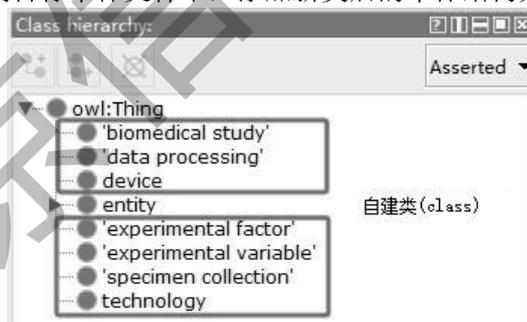


图 A.5 新建类

4.3.2 新增属性

对于表达新语义关联需求的情形, 也可自通过定义新属性来实现。其创建方式同新增类类似。依据“表 2 新增属性最小信息要求”, 设计的模板如表 A.4 所示。

表 A.4 新增属性模板(bso_properties.tsv)

Property ID	Entity Type	Label	Definition	Property type	Property	Characteristic	Domain	Range
ID	TYPE	A rdfs:label	A definition SPLIT=	PROPERTY_TYPE	P %	CHARACTERISTIC	DOMAIN	RANGE
bso:BSO_0000008	object property	is part of	a core relation that holds between a part and its whole					
bso:BSO_0000009	object property	has component	A system, structure, substance, or concept contains one or more specific components or elements.					
bso:BSO_0000010	object property	has publication	Individuals, organizations, or institutions that own or publish some form of publication.	subproperty	has component		biomedical study	publication

bs:BSO_0000011	object property	has design	Individuals, organizations, products, or systems have specific attributes or characteristics shaped by creative activities.	subproperty	has component	functional		
bs:BSO_0000012	object property	has protocol	An individual, organization, system, or process follows or has a set of established rules, procedures, or codes of conduct.	subproperty	has component			

将属性模板文件(bso_properties.tsv)转换为 owl 格式文件

```
robot template --input ./bso_dev.owl
--template ./templates/bso_properties.tsv
--prefix "bso: http://purl.ontochina.org/ontology/"
--ontology-iri "http://purl.ontochina.org/ontology/bso_properties.owl"
--output ./modules/bso_properties.owl
```

将 bso_properties.owl 导入到目标本体文件中，添加新属性后的本体结构如图 A.6 所示：

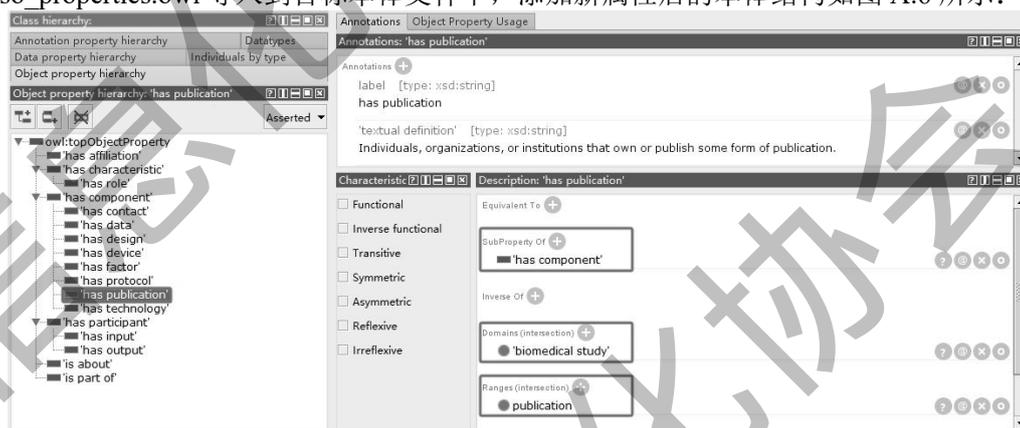


图 A.6 新建属性

4.3.3 新增类间关系

截止目前，新建的类(class)是独立存在的，即：这些新建类之间，新建类和本体复用的术语之间还没有建立关联关系。对于关系的建立，也可以通过模板来实现，如表 A.5 所示。

表 A.5 新增关系模板(bso_relations.tsv)

Class ID	Class Type	Class	is part of	has input	has output
ID	CLASS_TYPE	C %	C 'is part of' some % SPLIT=	C 'has input' some % SPLIT=	C 'has output' some % SPLIT=
bs:BSO_0000001	subclass	process			
bs:BSO_0000002	subclass	process	biomedical study	Organism	specimen

将关系模板文件(bso_relations.tsv)转换为 owl 格式文件

```
robot template --input ./bso_dev.owl
--template ./templates/bso_relations.tsv
--prefix "bso: http://purl.ontochina.org/ontology/"
--ontology-iri "http://purl.ontochina.org/ontology/bso_relations.owl"
--output ./modules/bso_relations.owl
```

将 bso_relations.owl 导入到目标本体文件中，添加新关系后的本体结构如图 A.7 所示：

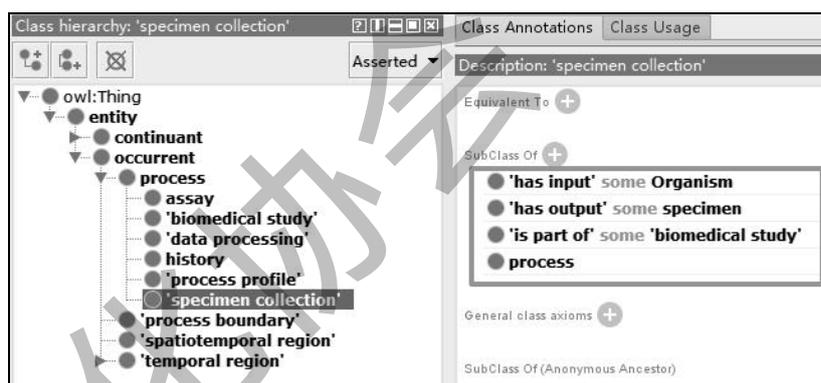


图 A.7 新增类间关系

4.4 添加本体元数据描述

为本体添加描述信息(如标题、作者等)是本体文档化过程中的一部分,有助于明确本体的背景、来源和创建者,提升其可重用性与学术价值。生物医学研究本体元数据信息包括:本体名称、创建者、发布日期、描述等。

```
robot annotate --input ./bso_dev.owl
--annotation dc:title "Biological Study Ontology"
--annotation dc:creator "bmicc"
--annotation dc:date "2025-10-12"
--annotation dc:description "Biomedical Study ontology(BSO) used to demonstrate best practices for ontology
construction."
--output ./bso_dev.owl
```

4.5 本体合并

在目标本体文件(bso_dev.owl)中,使用了大量的 import 关键字来导入各个步骤中生成的代码片段(OWL 片段文件)。为了生成一个完整的本体结构(包括术语、属性、关系等),需要对所有导入的 OWL 文件进行合并。操作如下:

```
robot merge --input ./bso_dev.owl --output ./bso_merged.owl
```

4.6 本体质量检验

使用 HermiT 推理机进行推理,可检查本体是否存在不一致或不可满足类,并自动分类。

```
robot reason --input ./bso_merged.owl --reasoner HermiT --output ./bso_reasoned.owl
```

本体检验用于检查本体中是否存在错误。例如:术语或属性有多个标签或多种定义等。

```
robot report --input ./bso_reasoned.owl --output ./report.tsv
```

表 A.6 展示了目标本体检验报告显示的可能存在的错误信息。

表 A.6 本体检验报告

Level	Rule Name	Subject	Property	Value
ERROR	multiple_definitions	IAO:0000115	IAO:0000115	The official OBI definition, explaining the meaning of a class or property. Shall be Aristotelian, formalized and normalized. Can be augmented with colloquial definitions.@en
ERROR	multiple_definitions	IAO:0000115	IAO:0000115	The official definition, explaining the meaning of a class or property. Shall be Aristotelian, formalized and normalized. Can be augmented with colloquial definitions.@en
ERROR	missing_ontology_license	http://purl.obolibrary.org/obo/bso.owl	dc:license	
ERROR	multiple_labels	IAO:0000111	rdfs:label	editor preferred label@en
ERROR	multiple_labels	IAO:0000111	rdfs:label	editor preferred term

本体开发者可根据提示信息修正错误,直至满足要求为止。

4.7 本体发布

在正式发布本体前，设置目标本体的国际化资源标识符及其版本信息，用来唯一标识该本体。

```
robot annotate --input ./bso_reasoned.owl  
--ontology-iri "http://purl.ontochina.org/ontology/bso.owl"  
--version-iri "http://purl.ontochina.org/ontology/2025-10-12/bso.owl"  
--output ./bso.owl
```

至此，一个完整的，可发布的生物医学本体示例构建完成。