

# T/BIA

团体标准

T/BIA XXXX—202X

## 生物医学本体 第3部分：元数据

Biomedical ontology—  
Part 3: Metadata

(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 录

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
4 本体元数据的描述方法 .....	5
5 本体核心元数据 .....	6
5.1 本体核心元数据组成 .....	6
5.2 本体核心元数据元素 .....	6
5.2.1 本体国际化资源标识符 .....	6
5.2.2 本体名称 .....	6
5.2.3 本体缩略名 .....	6
5.2.4 本体描述 .....	7
5.2.5 创建者 .....	7
5.2.6 联系人电子邮件 .....	7
5.2.7 许可协议 .....	7
5.2.8 本体版本 .....	8
5.2.9 本体版本国际化资源标识符 .....	8
5.2.10 本体表示语言 .....	8
5.2.11 本体语言 .....	8
5.2.13 发布状态 .....	9
5.2.14 发布时间 .....	9
5.2.15 访问地址 .....	10
6 本体扩展元数据 .....	10
6.1 本体扩展元数据组成 .....	10
6.2 本体扩展元数据元素 .....	10
6.2.1 类数量 .....	10
6.2.2 属性数量 .....	10
6.2.3 个体数量 .....	11
6.2.4 复用顶层本体 .....	11
6.2.5 复用参考本体 .....	11
6.2.6 测试 .....	11
6.2.7 评估方法 .....	11
6.2.8 使用示例 .....	11
6.2.9 实体弃用策略 .....	12
6.2.10 版本更新策略 .....	12
6.2.11 分发格式 .....	12
6.2.12 SPARQL 查询端点 .....	13
附录 A (资料性) 本体核心元数据示例 .....	错误! 未定义书签。

附录 B（资料性） 本体扩展元数据示例 ..... 错误！未定义书签。

## 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3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：张敬晨、王哲、杨啸林、周伟、朱彦、邵晨、杨晟、彭苏元、刘丽红、李晓瑛、姚克宇、谢江安、关键、何勇群、程亮、张胜发、罗葳。

## 引言

为促进生物医学本体的可发现性、可获取性与可复用性，规范本体的描述、注册与管理过程，需要统一的本体元数据标准。元数据是本体发布、共享与互操作的基础，对保障本体的质量与可持续性至关重要。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 3 部分，以第 1 部分的开发原则为依据，为第 2 部分开发规范的标准化输出提供元数据模板，共同构成生物医学本体开发的完整规范体系。

# 生物医学本体

## 第3部分：元数据

### 1 范围

本文件规定了生物医学本体元数据的描述方法、组成、元素、实体，以及生物医学本体的核心元数据和扩展元数据。

本文件适用于生物医学本体核心元数据的编目、归档、建库、发布、共享、交换和查询，以及领域扩展等。

### 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 30523-2023 科技资源核心元数据

### 3 术语和定义

下列术语和定义适用于本文件。

#### 3.1

**元数据 metadata**

定义和描述其他数据的数据。

[来源:GB/T 30523-2023,3.2]

#### 3.2

**核心元数据 core metadata**

描述数据最基本信息的元数据最小集合。

[来源:GB/T 30523-2023,3.3]

#### 3.3

**元数据元素 metadata element**

元数据的基本单元。

注：元数据元素在元数据实体中是唯一的。

[来源:GB/T 30523-2023,3.4]

#### 3.4

**元数据实体 metadata entity**

一组说明数据相同特性的元数据元素。

注：可以包含一个或一个以上元数据实体。

[来源:GB/T 30523-2023,3.5]

### 4 本体元数据的描述方法

本体元数据描述方法各部分格式规范设置遵循《科技资源核心元数据》（GB/T 30523-2023）中规

定的元数据描述方法。

## 5 本体核心元数据

### 5.1 本体核心元数据组成

本体核心元数据包括 15 个元数据元素。

- 本体国际化资源标识符；
- 本体名称；
- 本体缩略名；
- 本体描述；
- 创建者；
- 联系人电子邮件；
- 许可协议；
- 本体版本；
- 本体版本国际化资源标识符；
- 本体表示语言；
- 本体语言；
- 学科分类；
- 发布状态；
- 发布时间；
- 访问地址；

本体核心元数据的示例见附录A。

### 5.2 本体核心元数据元素

#### 5.2.1 本体国际化资源标识符

**定义：**语义网中用于全局、唯一且持久地标识一个本体的标准化字符串。

**英文名称：**OntologyInternationalizedResourceIdentifier

**数据类型：**字符串

**值域：**自由文本

**短名：**ontIRI

**注解：**必选项，最大出现次数为 1。

#### 5.2.2 本体名称

**定义：**本体的正式名称。

**英文名称：**OntologyTitle

**数据类型：**字符串

**值域：**自由文本

**短名：**onTitle

**注解：**必选项，最大出现次数为1。

#### 5.2.3 本体缩略名

**定义：**本体的缩写名称。

**英文名称：**OntologyAcronym

数据类型：字符串  
 值域：自由文本  
 短名：ontAcr  
 注解：必选项，最大出现次数为1。

#### 5.2.4 本体描述

定义：本体内容的简要介绍，包括内容、目标和适用场景等。  
 英文名称：OntologyDescription  
 数据类型：字符串  
 值域：自由文本  
 短名：ontDes  
 注解：必选项，最大出现次数为1。  
 [来源：GB/T 30523-2023 5.2.3 有改动]

#### 5.2.5 创建者

定义：本体的主要开发者或创建团队  
 英文名称：Creator  
 数据类型：字符串  
 值域：自由文本  
 短名：creator  
 注解：可选项，最大出现次数为N。

#### 5.2.6 联系人电子邮件

定义：用户可联系的问题反馈邮箱。  
 英文名称：ContactEmail  
 数据类型：字符串  
 值域：自由文本  
 短名：conEmail  
 注解：可选项，最大出现次数为N。

#### 5.2.7 许可协议

定义：用户在使用、复制、分发或修改本体资源时应遵守的许可。  
 英文名称：License  
 数据类型：字符串  
 值域：根据代码表填写相应代码，见表1

表1 使用许可代码

代码	许可方式
1	转让
2	传播
3	署名
4	修改
5	汇编

6	收益
7	具体协商，如采用 CC BY4.0 协议

短 名：license

注 解：必选项，最大出现次数为1。

[来源：GB/T 30523-2023 5.2.11 有改动]

#### 5.2.8 本体版本

定 义：本体的当前版本标识，用于区分不同更新。

英文名称：OntologyVersion

数据类型：字符串

值 域：自由文本

短 名：ontVer

注 解：必选项，最大出现次数为1。

#### 5.2.9 本体版本国际化资源标识符

定 义：语义网中用于全局、唯一且持久地标识一个本体版本的标准化字符串。

英文名称：OntologyVersionInternationalizedResourceIdentifier

数据类型：字符串

值 域：自由文本

短 名：ontVerIRI

注 解：必选项，最大出现次数为1。

#### 5.2.10 本体表示语言

定 义：本体开发和表示所采用的表示语言。

英文名称：OntologyRepresentationLanguage

数据类型：字符串

值 域：根据代码表填写相应代码，见表2

表2 本体表示语言代码

代码	本体表示语言	说明
01	OBO	OBO Foundry 早期采用的本体表示语言，结构清晰。
02	OWL	Web Ontology Language，表达能力最强的本体语言，支持逻辑推理，是当前主流推荐格式
03	SKOS	Simple Knowledge Organization System，用于主题词表、分类体系和知识组织系统，不适用于复杂逻辑推理。
04	RDFS	RDF Schema，轻量级本体语言，适合简单分类层级与基本语义关系。
99	其他	供扩展使用

短 名：ontRepLan

注 解：可选项，最大出现次数为1。

#### 5.2.11 本体语言

定 义：描述本体内容的自然语言。例如，中文、英文。

英文名称: OntologyNatureLanguage  
 数据类型: 字符串  
 值域: 自由文本  
 短名: ontLan  
 注解: 可选项, 最大出现次数为1。

### 5.2.12 学科分类

定义: 本体所属领域的学科分类名称全称。  
 英文名称: OntologyDisciplineCategory  
 数据类型: 字符串  
 值域: 限制文本 [按照 GB/T 13745 确定二级学科名称, 格式为: “分类体系的名称(版本号): 具体子类名称”]  
 示例: 《学科分类与代码》(GB/T 13745—2009): 地球科学  
 短名: ontDisCat  
 注解: 必选项, 最大出现次数为 N。  
 [来源: GB/T 30523-2023 5.2.8 有改动]

### 5.2.13 发布状态

定义: 本体发布所处的阶段。  
 英文名称: ReleaseStatus  
 数据类型: 字符串  
 值域: 根据代码表填写相应代码, 见表3

表3 发布状态代码

代码	发布状态	说明
01	draft	草拟阶段, 仅用于内部开发, 结构与内容均不稳定。
02	alpha	初步成型, 存在较大结构调整可能, 适合公开征求意见
03	beta	内容较稳定, 可用于试用或社区评估, 但仍可能修改。
04	production	正式发布版本, 可在生产环境使用, 遵循版本控制规则。
05	maintenance	正式版本的后续维护阶段, 包括修正、更新及兼容性保障。
06	deprecated	官方废弃但仍保留, 仅用于历史兼容, 不建议继续使用。
07	retired	完全停止维护, 不再提供支持。
08	superseded	本体已被其他本体或新版本替代, 用于处理继承与语义迁移关系。

短名: relSta  
 注解: 必选项, 最大出现次数为1。

### 5.2.14 发布时间

定义: 本体当前版本的正式发布日期。  
 英文名称: ReleaseDate  
 数据类型: 日期型  
 值域: 日期 (按照 GB/T 7408 执行)  
 短名: relDate  
 注解: 可选项, 最大出现次数为1。  
 [来源: GB/T 30523-2023 5.2.7 有改动]

### 5.2.15 访问地址

定 义：直接访问到本体服务的有效网络地址。

英文名称：accessURL

数据类型：字符串

值 域：自由文本（参见 RFC 2396）

短 名：accURL

注 解：必选项，最大出现次数为1。

[来源：GB/T 30523-2023 5.2.12 有改动]

## 6 本体扩展元数据

### 6.1 本体扩展元数据组成

本体扩展元数据包括 12 个元数据元素。

——类数量；

——属性数量；

——个体数量；

——复用顶层本体；

——复用参考本体；

——测试；

——评估方法；

——使用示例；

——实体弃用策略；

——版本更新策略；

——分发格式；

——SPARQL 查询端点；

本体扩展元数据的示例见附录B。

### 6.2 本体扩展元数据元素

#### 6.2.1 类数量

定 义：本体中所包含的类(class)的数量。

英文名称：NumberOfClasses

数据类型：整型

值 域：正整数

短 名：numclass

注 解：可选项，最大出现次数为1。

#### 6.2.2 属性数量

定 义：本体中所包含的属性(property)的数量。

英文名称：NumberOfProperties

数据类型：整型

值 域：正整数

短 名：numPro

注 解：可选项，最大出现次数为1。

### 6.2.3 个体数量

定 义：本体中所包含的个体(Individual)的数量。

英文名称：NumberOfIndividuals

数据类型：整型

值 域：正整数

短 名：numInd

注 解：可选项，最大出现次数为1。

### 6.2.4 复用顶层本体

定 义：在语义建模中所采用或对齐的顶层本体，用于指明该本体的哲学基础和语义一致性框架。

英文名称：TopLevelOntologyReused

数据类型：字符串

值 域：自由文本

短 名：toplevOnt

注 解：可选项，最大出现次数为1。

### 6.2.5 复用参考本体

定 义：本体构建过程中复用或引入的生物学领域参考本体。

英文名称：ReferenceOntologiesReused

数据类型：字符串

值 域：自由文本

短 名：refOnt

注 解：可选项，最大出现次数为N。

### 6.2.6 测试

定 义：判断本体是否达成其声明目标的测试过程。

英文名称：Testing

数据类型：字符串

值 域：自由文本

短 名：test

注 解：可选项，最大出现次数为1。

### 6.2.7 评估方法

定 义：描述本体是否满足其声明的要求、能力问题和目标的方法。

英文名称：EvaluationMethod

数据类型：字符串

值 域：自由文本

短 名：evamet

注 解：可选项，最大出现次数为1。

### 6.2.8 使用示例

定 义：本体被应用到实际使用场景中的情况说明。

英文名称：ExampleOfUse

数据类型：字符串

值 域：自由文本

短 名：ExaUse

注 解：可选项，最大出现次数为N。

#### 6.2.9 实体弃用策略

定 义：实体被删除、拆分、重新定义的策略。

英文名称：EntityDeprecationStrategy

数据类型：字符串

值 域：自由文本

短 名：depStr

注 解：可选项，最大出现次数为1。

#### 6.2.10 版本更新策略

定 义：本体版本更新策略。

英文名称：VersioningUpdatePolicy

数据类型：字符串

值 域：根据代码表填写相应代码，见表4

表4 版本更新策略代码

代码	更新策略
01	每3个月更新一次
02	每6个月更新一次
03	每1年更新一次
04	每2年更新一次
05	每3年更新一次
06	超过3年更新一次
99	不再更新

短 名：verUpdPol

注 解：可选项，最大出现次数为1。

#### 6.2.11 分发格式

定 义：本体发布时采用的数据格式类型。

英文名称：DistributionFormat

数据类型：字符串

值 域：根据代码表填写相应代码，见表5

表5 分发格式代码

代码	分发格式
01	Turtle (.ttl)
02	JSON-LD (.jsonld)
03	RDF/XML (.rdf)

04	OWL/XML (.owl)
05	N-Triples (.nt)
06	OBO(.obo)
99	其他

短 名: disFor

注 解: 可选项, 最大出现次数为1。

#### 6.2.12 SPARQL 查询端点

定 义: 支持标准化语义查询 (SPARQL 1.1) 的接口端点地址。

英文名称: SPARQLEndpoint

数据类型: 字符串

值 域: 自由文本 (参见 RFC 2396)

短 名: spaEnd

注 解: 可选项, 最大出现次数为1。

## 附录 A

(资料性)

## 本体核心元数据示例

## A.1 概述

本附录给出了生物医学本体核心元数据描述的示例。

注：本附录中的示例，仅为说明生物医学本体核心元数据的描述方法，不代表真实情况。

## A.2 示例

生物医学本体的核心元数据描述示例见表 A。

表 A 本体核心元数据

本体核心元数据名称	本体核心元数据内容
本体国际化资源标识符	<a href="https://www.***.***.cn/****/****.owl">https://www.***.***.cn/****/****.owl</a>
本体名称	Biomedical Study Ontology
本体缩略名	BSO
本体描述	生物医学研究本体以构建结构化知识体系为核心，通过对生物医学概念、关系及逻辑的规范化建模，形成机器可读、可推理的语义框架，旨在解决多源生物医学数据的语义异构问题，为跨系统数据整合与智能应用提供标准化语义基础设施。
创建者	国家*****中心
联系人电子邮件	***@***.***.cn
许可协议	7-具体协商，如采用 CC BY4.0 协议
本体版本	2.5.0
本体版本国际化资源标识符	<a href="https://www.***.***.cn/****/releases/2.5.0/****.owl">https://www.***.***.cn/****/releases/2.5.0/****.owl</a>
本体表示语言	02-OWL
本体语言	英文
学科分类	《学科分类与代码》（GB/T 13745—2009）：基础医学
发布状态	04-production
发布时间	2025-06-30
访问地址	<a href="https://www.***.***.cn/****/****.***">https://www.***.***.cn/****/****.***</a>

## 附录 B (资料性)

### 本体扩展元数据示例

#### A.1 概述

本附录给出了生物医学本体扩展元数据的示例。

注：本附录中的示例，仅为说明生物医学本体扩展元数据的描述方法，不代表真实情况。

#### A.2 示例

生物医学本体的扩展元数据描述示例见表 B。

表 B 本体扩展元数据

本体扩展元数据名称	本体扩展元数据内容
类数量	20
属性数量	10
个体数量	5
复用顶层本体	BFO
复用参考本体	复用 DOID 中的术语和 RO 中的关系
测试	基于推理机的一致性检查与 SPARQL 查询验证。
评估方法	本体评估工具（例如：ROBOT report）+ 领域专家评审。
使用示例	例如：HPO 在 clinvar 数据库中使用。
实体弃用策略	采用 oboInOwl:deprecated 标注，并提供替代项说明。
版本更新策略	03-每 1 年更新一次
分发格式	04-OWL/XML (.owl)
SPARQL 查询端点	https://www.***.***.cn/****/****.***