

北京信息化协会团体标准

《生物医学本体 第3部分:元数据》

编制说明

一、工作简况，包括任务来源、协作单位、主要工作过程、研究会团体标准主要起草人及其所做的工作等

1.1 任务来源

随着生物医学本体的广泛应用与快速发展，本体作为一种重要的数据与知识资源，其自身也需要被系统化描述与管理，以支持资源的可发现、可访问、可互操作与可复用。然而，当前国内缺乏统一的本体描述规范，导致本体的元数据信息不完整、结构不一致、语义不明晰，严重影响了本体的跨平台共享与系统集成。为此，北京信息化协会于2025年正式立项《生物医学本体 第3部分：元数据》，作为系列规范的关键组成部分，旨在建立统一的本体元数据描述体系。

1.2 起草单位

本标准起草单位：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。

1.3 主要起草人

本标准主要起草人：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。

主要起草人全面负责标准结构设计、内容撰写、技术论证与多轮修订工作，确保标准的科学性、实用性与前瞻性。

二、立项的必要性，拟解决的问题

生物医学本体作为语义建模与知识表示的核心资源，其描述信息的标准化是实现本体资源FAIR（可发现、可访问、可互操作、可重用）化、支持跨系统高效互操作、提升资源可信度与可用性的关键基础。然而当前该领域存在若干突出问题：本体发布时常因缺乏统一的描述而导致元数据缺失或记录不一致，定义模糊的元数据项进一步削弱了描述的规范性；由于描述不规范，本体在各类数据平台中难以有效注册、检索与复用，造成注册与发现困难；在维护管理方面，普遍缺乏清晰的版本策略、维护计划等关键描述信息，影响本体的长期可持续发展。

针对当前生物医学本体描述信息标准化不足的现状，本部分旨在建立统一的核心元数据描述框架，以系统解决资源注册、发现、维护与互操作中的关键问题。具体而言，通过定义覆盖标识、内容、统计、维护及质量等多个维度的核心元数据项体系，形成标准化的本体描述规范。

三、确定研究会团体标准框架及主要内容，修订研究会团体标准时，应增加新、旧研究会团体标准水平的对比

本标准定位于本体资源的标准化描述，与前两部分共同构成从理念指导、方法实践到输

出规范的全流程闭环。系统构建了覆盖本体核心元数据、统计信息、内容信息、可访问性与维护信息、质量保证等多个维度的描述体系，并提供了统一的数据元定义规则、值域代码表及与国际通用元数据词汇的映射关系。

本标准的制定实现了从无到有的系统性突破。具体表现为：首次建立了针对生物医学领域的专用本体描述规范，统一了元数据项的定义、数据类型与表示格式；创新性地将本体统计信息、内容语义信息、维护策略与质量保证纳入描述体系，提升了本体的透明性、可信度与可持续性。本标准的制定为构建可发现、可访问、可互操作、可复用的生物医学本体生态系统奠定了坚实的描述基础。

四、与现行法律法规、强制性标准等上位标准关系

本标准在起草过程中严格遵循以下法律法规与标准：

文件编号	文件名称	关系说明
GB/T 1.1-2020	标准化工作导则	遵循其起草规则
GB/T 18391.1-2009	信息技术 元数据注册系统(MDR) 第1部分:框架	引用“数据元”“允许值”等核心术语
GB/T30523-2023	科技资源核心元数据	遵循其元数据描述规则与格式规范
ISO/IEC 21838-1/-2	信息技术 顶层本体	引用 BFO 等顶层本体作为内容描述依据

本标准在数据元描述结构上严格对接 WS/T 363 系列标准，在语义描述上映射 DC、FOAF、OMV 等国际通用元数据词汇，确保描述的规范性、可扩展性与国际兼容性。

五、采用国际标准的程度及水平的简要说明

本标准在制定过程中，充分参考并采纳了国际先进的元数据标准理念与实践经验，特别是借鉴了语义网领域通用的资源标识与描述规范，以及国际生物医学本体库（如 OBO Foundry）的元数据管理框架。文件在核心元数据构成、元素定义及代码表设置等方面，实现了与 GB/T 30523-2023《科技资源核心元数据》等国内标准的协调一致，同时兼顾了生物医学领域本体的特殊需求。整体上，本文件的技术内容达到了国内先进水平，为我国生物医学本体的规范化描述、共享与互操作提供了科学依据。

六、重大分歧意见的处理经过和依据

在标准起草过程中，起草组针对以下关键分歧进行了充分讨论与协商：

分歧点	处理方式	最终结论
元数据项粒度与必选性	组织专家研讨会，进行需求论证	分为“核心元数据”与“扩展元数据”
统计信息是否自动生成	调研工具支持度与实际可操作性	明确建议利用 ROBOT 等工具自动生成，减少人工负担
版本策略描述详略	结合开源项目与企业本体维护实践	提供结构化描述模板，支持自由文本补充

所有分歧最终经起草组集体审议达成共识。

七、标准有何先进性或特色之处

本标准填补了生物医学领域在资源描述与标准化方面的空白。其核心贡献在于系统化构建了一个覆盖标识、内容、统计、维护及质量等维度的元数据体系，实现对本体全生命周期的规范描述。通过结构化设计全面贯彻 FAIR 原则，有效提升本体的可发现性、可访问性、互操作性与可复用性。作为系列标准的关键组成，本部分与阐述开发基本原则的第 1 部分、

聚焦开发规范的第 2 部分紧密衔接、相互支撑，共同构成了一个贯穿“理念—实践—描述”的完整规范闭环，为生物医学本体工程的全流程标准化提供了坚实支撑。

八、贯彻研究会团体标准的要求和措施建议

为推动本标准的有效实施与持续完善，将采取多维度并行的路径：首先，通过组织标准宣贯会、针对开发者与平台管理员开展培训，并配套开发元数据填写指南、模板工具与示例库，以降低应用门槛，加强推广普及；其次，推动标准在国家级平台中率先落地；同时，鼓励开发能自动提取、校验与导出本标准元数据的工具插件，构建工具生态；最后，建立公开的反馈渠道并设立定期修订机制，持续收集实践建议并更新元数据项与值域代码，确保规范能适应技术的快速演进与应用场景的不断拓展。

九、其他应予说明的事项。

本标准提供的元数据体系为推荐性规范，可根据具体应用场景进行适度扩展。建议在实施过程中结合本体实际发展阶段（如开发中、测试版、正式发布、已归档）灵活选用元数据项，并鼓励在遵循核心元数据基础上创新扩展，以促进生物医学本体资源的持续丰富与语义互联。