

《多模态医疗科研数据与共享平台标准》团体标准 编制说明

一、任务来源

《国家数据标准体系建设指南》、《“十四五”卫生健康标准化工作规划》等政策法规明确发展卫生健康领域数据融合应用标准，鼓励发展团体标准，为多模态医疗科研数据治理指明方向；中国人民解放军总医院牵头响应国家要求，为完成“建好平台、形成成果”的业务目标，推动医疗科研数据共享标准体系的建设，提升科研效率与质量，制定多模态医疗科研数据与共享平台标准显得尤为重要。当前，大数据、人工智能等技术的发展为医疗科研数据的处理和分析提供了强大的技术支持。然而，这些技术的应用需要标准化的数据作为支撑。本文件由中国人民解放军总医院、北京大学、浙江大学、中南大学、国家人口健康科学数据中心、北京嘉和美康信息技术有限公司、北京嘉和海森健康科技有限公司、万达信息、北京软件和信息服务业协会共同提出制订。

二、制定规范的必要性和意义

制定《多模态医疗科研数据与共享平台标准》（以下简称本标准）对于推动医疗科研数据共享、促进学术交流与合作具有重要意义，具体表现在解决当前医疗科研数据共享的三大痛点：

1.打破多源异构数据壁垒

医疗科研数据分散于 HIS、LIS、PACS 等系统，格式差异大，导致数据整合成本高、效率低。缺乏统一标准，跨机构协作时数据互操作性差。

2.突破多模态数据质量与知识转化瓶颈

医疗原始数据存在缺失、不一致问题，需规范治理流程提升质量。非结构化文本、影像等数据难以直接用于科研分析，需标准化解析方法。

3.填补安全共享机制缺失

敏感医疗数据共享存在隐私泄露风险，需建立分级授权与审计机制。跨机构数据交换缺乏可信技术支撑。

制定和推行本标准的主要目的是规范明确医疗科研数据应用和共享的范围和要求，包括：

1.构建医疗科研数据全生命周期管理框架

确立多模态医疗科研“数据采集与存储→数据管理与质量控制→数据整合与脱敏→数据分析与知识发现→安全共享与价值反馈”的全生命周期数据应用理念闭环流程，实现数据向知识的转化和共享。

2.统一技术规范

规定多模态数据管理范围、平台基础能力、数据交互协议，解决兼容性问题。

3.建立协同生态

明确多模态医疗科研数据应用各阶段的关联关系：从数据的标准化（第 2-4 部分）到标准数据的知识化（第 5 部分）再到数据和知识的安全共享（第 6 部分）全流程，同时为医学科研单位、医院、医疗 IT 企业提供统一开发与应用基准。

三、主要编制过程

1.第一阶段，成立规范研制起草工作组（以下简称工作组）编制项目立项材料和标准草案。工作组组由 16 名相关专家和专业人员组成，来自中国人民解放军总医院、中国人民解放军火箭军总医院、首都医科大学附属北京佑安医院、国家人口健康科学数据中心、北京大学、清华大学、浙江大学、中南大学、北京嘉和美康信息技术有限公司、北京嘉和海森健康科技有限公司、万达信息、北京软件和信息服务业协会等单位。工作组通过查阅文献，形成理论框架和工作方法，并对已有相关内容基础进行了梳理，多轮研讨沟通，形成立项申请材料和标准草案。工作组于召开立项准备沟通会，为项目立项制定完整方案。

2.第二阶段，项目立项评审。工作组于 2025 年 10 月 22 日在中关村知识产权大厦 A 座，北京软件和信息服务业协会 210 会议室召开了项目立项专家评审会。评审专家包括（姓氏拼音排序）：首都医科大学附属北京佑安医院信息中心原主任崔保丽，中国标准化研究院标准化评估研究所正高级工程师韩冰，中国信息通信研究院人工智能研究所平台部副主任李荪，北京大学大数据分析与应用技术国家工程实验室特聘副研究员王娟，天坛医院信息管理与数据中心主任助理王鑫。标准立项通过专家评审后，信息在全国团体标准信息平台和协会官方网站和北京软协官网发布。

3.第三阶段，编写规范征求意见稿。围绕标准草案及各方意见，工作组开展了进一步调查研究，通过多轮内部研讨和标准内容文件修订，形成征求意见稿。并于 2025 年 11 月 10 日在全国团体标准信息平台和协会官方网站和北京软协官网上进行公示，收集各方建议，并结合各方建议对标准草案进行修改完善，形成标准送审稿。

四、制定规范的原则和依据，与现行法律、法规、标准的关系

制定《多模态医疗科研数据与共享平台标准》是促进医疗科研数据合规共享、释放数据价值、推动精准医学发展的关键举措。其制定工作并非从零开始，而是应在国家现有的法律、法规和标准体系的坚实基础上进行深化与细化。

（一）标准制定应遵循的核心原则

制定该标准时，应首先确立其核心指导原则，确保标准兼具前瞻性、实用性和安全性。这些原则根植于 GB/T 44109-2024 所倡导的数据治理理念。

1.统筹规划与业务驱动原则

标准制定应服务于国家健康医疗大数据战略和医疗卫生机构的业务发展目标。必须先进行顶层设计，明确平台建设的愿景、目标和实施路线，确保标准工作与业务需求紧密契合。

2.合规性与安全性优先原则

将合规性置于首位，确保数据全生命周期的处理活动严格遵守《个人信息保护法》、《数据安全法》等法律法规。建立覆盖数据分类分级、访问控制、加密传输、脱敏匿名的安全技术和管理体系。

3.互操作性与开放共享原则

通过统一的数据元、代码、接口和元数据标准，打破“数据孤岛”，实现多源异构数据的互联互通。在保障安全的前提下，促进数据的合规、有序开放与共享，最大化数据价值。

4.质量可控与全生命周期管理原则

建立贯穿数据应用全生命周期质量管理体系。明确各环节的质量责任，制定可量化、可监控的数据质量指标，确保科研数据的准确性、完整性和一致性。

5.可持续发展与迭代优化原则

标准应具备一定的灵活性和扩展性，以适应快速发展的医疗技术和数据分析

方法。建立标准的动态评估、反馈和改进机制，形成闭环管理，确保其持续有效。

（二）标准制定的主要依据

标准的制定需有章可循，其主要依据来源于三个层面：

1.法律与行政法规依据

《中华人民共和国个人信息保护法》：是数据处理活动的根本大法，为个人健康信息处理确立了“告知-同意”、目的明确、最小必要等核心原则。

《中华人民共和国数据安全法》：规定了数据分类分级管理、风险评估、监测预警等制度，是平台数据安全体系的顶层设计依据。

《中华人民共和国网络安全法》：明确了网络运营者的安全义务，是平台网络安全建设的基础。

《中华人民共和国人类遗传资源管理条例》：针对涉及人类基因、基因组等遗传资源的采集、保藏、利用和出境活动提供了专项法律规范。

2.国家与行业标准依据

GB/T 44109-2024《信息技术 大数据 数据治理实施指南》：为本标准的整体结构（规划、执行、评价、改进）和核心治理活动（元数据、质量、安全等）提供了方法论基础。GB/T 36073-2018《数据管理能力成熟度评估模型》：为评估和提升平台的数据管理能力提供了详细的评价维度和成熟度等级参考。

WS/T 303-2024《卫生信息数据元目录》、WS/T 305-2024《卫生信息数据元数据规范》等 WS 系列标准：是定义临床数据元、值域、代码的唯一权威来源，确保数据语义的一致性。GB/T 36447-2018《电子健康记录与医疗健康信息共享文档》：定义了临床文档的结构与内容标准，是实现互操作的重要参考。

DICOM（国际标准）、HL7 FHIR（国际标准）：虽为国际标准，但已成为医学影像和临床数据交换的事实全球标准，应在本标准中充分借鉴和引用，确保与国际接轨。

（三）与现行法律、法规、标准的关系

本标准的定位是承上启下的应用层和实施层标准，其与现行体系的关系可概括为“遵循、细化、互补”。

1.对上层法律与法规的遵循与细化

本标准是法律法规在医疗科研数据领域的具体化。法律法规定义了“什么不能做”和“必须遵守的原则”，而本标准则详细规定了“应该如何做”才能符合这些原则。

2.对同级和基础性国家标准的引用与集成

本标准不重复发明基础性标准，而是通过引用和集成的方式，将分散的标准组合成一个适用于多模态医疗科研场景的有机整体。

3.对现有标准空白的补充与创新

针对多模态医疗数据治理中的新兴挑战和空白领域，本标准需发挥补充和创新作用。

五、主要条款的说明，主要技术指标、参数、试验验证的论述

本标准规定了多模态医疗科研数据应用的总体流程应遵循的全生命周期管理理念，该流程体现了数据从原始状态到知识化应用的完整转化过程，各环节相互衔接，构成数据治理活动的核心环节。

本系列标准由以下 6 个部分构成，共同组成完整的多模态医疗科研数据与共享平台标准体系：第 1 部分：总体要求；第 2 部分：心脑血管疾病科研数据集；第 3 部分：平台架构；第 4 部分：样本数据收集、存储及传递；第 5 部分：知识工程；第 6 部分：安全管理。

本次立项的标准为前 3 部分。主要技术内容包括：

第 1 部分：总体要求。规定平台建设的总体框架、基本原则和通用要求，定义标准体系结构和各组成部分间的关系。其中总体框架规定了按照“数据采集与存储→数据管理与质量控制→数据整合与脱敏→数据分析与知识发现→安全共享与价值反馈”全生命周期。通用数据规则规定了：数据应用范围、角色与责任、数据处理流程。数据应用范围中规定了数据来源范围、数据类型范围、数据层级范围；角色与责任中规定了角色定义体系、责任矩阵；数据处理流程中规定了前台科研操作流程、后台管理流程、端到端流程整合。

第 2 部分：心脑血管疾病科研数据集。规定心脑血管疾病科研所需的核心数据元，定义专病数据集结构。定义的数据元属性，包括数据元公用属性、数据元专用属性。其中数据元专用属性根据《卫生健康信息基本数据集编制标准》（WS/T 370—2022）规范内容包含内部标识符、数据元标识符（DE）、数据元名称、定

义、数据元值的数据类型、表示格式、数据元允许值等。

第3部分：平台架构。规定平台技术架构、功能模块和系统要求，定义数据接入、存储、处理和服务的技术规范。技术内容构建了一个以分层模块化（数据层、服务层、应用层）为逻辑核心，以混合存储（文档数据库+分布式检索引擎）和微服务为技术基石，通过标准化接口实现互联互通，并深度融合多层次安全管控（身份认证、RBAC 权限、审计溯源、备份恢复）与全流程数据治理的综合性平台架构，旨在为多模态医疗科研数据的安全共享、高效处理与知识发现提供可靠、可扩展且合规的技术载体。

六、重大意见分歧的处理依据和结果的说明

本规范研究制定过程中未出现重大分歧。

七、采用国际标准或国外先进标准程度的说明，以及国内、外同类产品或标准的对比情况

1.关于多模态医疗科研数据与共享平台的通用国际标准：

（1）HL7 FHIR R4，由 Health Level Seven International 发布，它是医疗保健数据交换标准的升级，用于连接医疗系统，处理外部规范，与现实世界概念关联，并支持数据交换。此项标准覆盖更广，FHIR 侧重通用临床数据，而本标准专注心脑血管科研场景（多模态数据+知识图谱）

（2）DICOM 3.0，由 NEMA/ISO 发布，DICOM 标准支持多种医学成像模态的数据存储，包括 CT、PET、MRI、X 射线和超声波等。它的优势在于是医学领域存储图像的常见标准，能够在单个数据存储中采集和存储图像体和患者数据，并且以 DICOM 格式保存的图像具有高动态范围（高达 16 位）。本标准在 DICOM 基础上要求影像特征解析与语义对齐

（3）ISO 20691:2022 Biotechnology — Requirements for data formatting and description in the life sciences 生物技术 生命科学中数据格式和描述的要求。ISO 20691 通用性强，本标准明确心脑血管生物样本数据规范

（4）OMOPCDM，由 OHDSI 联盟发布，观察性医学结果伙伴关系(OMOP)通用数据模型(CDM)是一个开放的社区数据标准，旨在标准化观察性数据的结构和内容，并实现高效分析，从而产生可靠的证据。OMOP CDM 的核心组成部分是 OHDSI 标准化词汇表。OHDSI 词汇表支持对医学术语进行组织和标准化，以

便在 OMOP 通用数据模型的各个临床领域使用，并支持在构建暴露和结果表型以及其他特征（包括表征、群体层面的效应估计和患者层面的预测研究）时利用知识库进行标准化分析。OMOP 仅定义数据模型，本标准覆盖“采集→处理→知识化→共享”全流程

（二）国内同类标准和技术状况

目前还没有针对多模态医疗科研数据和共享的国家标准或行业标准，相关的标准有：

- 1.GB/T 39725-2020 健康信息共享文档规范（通用框架），未细化到专病科研；
- 2.WS/T 303-2023 医院信息平台技术规范，侧重临床业务而非科研数据治理；
- 3.《人类遗传资源管理条例》，科技部，约束生物样本数据出境，但未提供平台建设标准
- 4.《医疗机构医疗大数据平台建设指南》，是中国医院协会信息专业委员会 CHIMA 为规范和指导我国医疗机构医疗大数据建设及应用工作而编写的。指出了医疗大数据平台的总体架构、技术路线和功能范围，内容原则通用，本标准参照细化。

八、实施标准的措施建议

标准正在征求意见阶段，鼓励相关企业积极参与本标准的意见反馈工作。

九、其他说明

无。

附件：征求意见稿标准意见汇总处理表

《多模态医疗科研数据与共享平台标准》起草工作组

2025 年 11 月 7 日