

DB4403

深 圳 市 地 方 标 准

DB4403/T XXX—XXXX

专病数据库建设规范

Specification for the construction of specialized disease database

(送审稿)

XXXX-XX-XX 发布

XXXX-XX-XX 实施

目 次

前 言 III

引 言 IV

1 范围 5

2 规范性引用文件 5

3 术语和定义 5

4 缩略语 6

5 建设原则和建设目标 7

 5.1 建设原则 7

 5.2 建设目标 7

6 建设内容与技术要求 8

 6.1 数据范围 8

 6.2 核心功能模块及要求 8

 6.2.1 数据采集模块 8

 6.2.2 数据存储模块 8

 6.2.3 数据治理模块 9

 6.2.4 多模态数据管理模块 9

 6.2.5 数据质控模块 9

 6.2.6 数据检索模块 9

 6.2.7 数据转换、分析与导出模块 9

 6.2.8 数据共享与多中心协作模块 9

 6.2.9 随访管理模块 9

 6.2.10 知识图谱模块 9

 6.2.11 数据安全性与隐私保护模块 9

 6.2.12 区块链数据存证 10

 6.2.13 区块链服务支撑 10

 6.2.14 项目管理模块 10

 6.3 参考技术架构 10

 6.4 设计要求 11

 6.4.1 业务需求驱动的设计 11

 6.4.2 模块化系统架构 11

 6.4.3 数据服务操作要求 11

 6.5 数据要求 11

 6.5.1 术语标准化 11

 6.5.2 元数据管理要求 12

 6.5.3 数据字典建设要求 12

 6.5.4 专病指标集制定要求 12

 6.5.5 数据模型与结构要求 12

- 6.5.6 数据质量控制要求 12
- 6.5.7 数据共享与互操作要求 12
- 6.6 数据安全性与隐私保护要求 13
 - 6.6.1 数据分类分级 13
 - 6.6.2 数据加密 13
 - 6.6.3 数据访问控制 13
 - 6.6.4 数据去标识化与隐私保护 13
 - 6.6.5 数据传输与存储安全 14
 - 6.6.6 数据完整性保护 14
 - 6.6.7 系统日志与审计 14
- 6.7 数据合规要求 15
- 7 建设流程和管理 15
 - 7.1 建设准备阶段 15
 - 7.2 系统部署与配置阶段 15
 - 7.3 测试与验收阶段 16
 - 7.4 专病数据库实施与迁移 16
 - 7.5 数据运维与生命周期管理 16
- 8 专病数据库评价 17
 - 8.1 评价标准结构 17
 - 8.2 满足临床研究的需求评价 17
 - 8.3 映射真实世界场景评价 17
- 附 录 A （规范性） 参考实施流程 19
 - A.1 实施流程阶段划分 19
 - A.2 实施步骤说明 19
 - A.2.1 需求调研与规划 19
 - A.2.2 系统设计与架构搭建 19
 - A.2.3 数据采集与治理 19
 - A.2.4 系统测试与优化 19
 - A.2.5 用户培训与试运行 19
 - A.2.6 验收与正式上线 20
- 附 录 B （规范性） 数据利用策略与方法 21
 - B.1 数据利用的基本原则 21
 - B.2 数据应用场景 21
 - B.3 数据分析与挖掘 21
 - B.4 数据共享与协作 21

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。

本文件由深圳市卫生健康委员会提出。

本文件由深圳市卫生健康委员会归口。

本文件起草单位：深圳市妇幼保健院、深圳市人民医院、深圳市第三人民医院、深圳市第二人民医院、北京大学深圳医院、香港大学深圳医院、深圳市中医院、深圳市儿童医院、中国医学科学院阜外医院深圳医院、中国医学科学院肿瘤医院深圳医院、上海信医科技有限公司

本文件主要起草人：李笑天、耿庆山、卢洪洲、曾晖、周丽萍、徐小平、蔡本辉、麻晓鹏、胡洋、王绿化、文萍、田赋颖、冯东雷、丁万夫、杨川川、滕国召、闪莹、贾亦真、吴嘉、苏琪茹、宗慧、徐文娟、许德俊、徐朗。

引 言

随着医学科技的快速发展，精准医疗、个性化治疗和大规模临床研究日益依赖高质量、结构化的数据支持。专病数据库作为以疾病为核心、患者为对象的全周期数据管理平台，能够实现跨科室、跨机构、跨区域的数据整合与共享，支持临床研究、患者管理和医疗决策。

专病数据库通过系统化采集和管理患者的多维度数据（如诊疗记录、实验室结果、影像资料、基因数据等），涵盖疾病的起因、经过、结果和转归。它不仅为个性化诊疗方案的制定提供数据支持，还通过多中心协作促进大规模临床研究和创新药物、器械的研发，推动科研成果的转化。

本文件旨在为医疗机构及科研单位提供专病数据库的建设与管理规范，保证专病数据库在设计、数据采集与整合、数据存储与管理、数据利用、安全管理、质量控制等方面达到统一的技术标准，具备高效性、可扩展性、互操作性和数据安全性。通过标准化建设，保证不同医疗机构在数据质量、数据共享、隐私保护方面具备一致的技术基础，推动临床研究、患者管理和医疗决策的高效开展。

专病数据库建设规范

1 范围

本文件规定了专病数据库的建设原则和建设目标、建设内容和技术要求、建设流程和管理等关键环节。

本文件适用于医疗机构、科研单位及相关服务提供方开展专病数据库的建设与管理。标准的适用对象包括但不限于单中心和多中心的专病数据库，以及用于临床研究、创新药物和医疗器械研发的专病数据库。

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 22239—2019 信息安全技术 网络安全等级保护基本要求
- GB/T 25069—2022 信息安全技术 术语
- GB/T 35273—2020 信息安全技术 个人信息安全规范
- GB/T 39725—2020 信息安全技术 健康医疗数据安全指南
- GB/T 39786—2021 信息安全技术 信息系统密码应用基本要求
- GB/T 37964—2019 信息安全技术 个人信息去标识化指南
- GB/T 42460—2023 信息安全技术 个人信息去标识化效果评估指南
- GB/T 42384—2023 健康信息学数据交换标准HL7临床文档架构（版本2）
- WS/T 305—2023 卫生健康信息数据元数据标准
- WS/T 500—2016 电子病历共享文档规范
- WS/T 790—2021 区域卫生信息平台交互标准

3 术语和定义

GB/T 25069—2022、GB/T 35273—2020、GB/T 39725—2020界定的以及下列术语和定义适用于本文件。

3.1 专病数据库 **disease-specific database**

用于存储、管理和分析特定疾病患者数据的系统化数据库，以及相应的管理系统，旨在支持临床研究、患者管理、随访以及相关科研项目。

3.2 数据治理 **data governance**

通过对数据进行标准化、结构化、归一化等技术处理，保证数据的质量、可用性和一致性，提升数据的管理水平。

3.3 数据元 **data element**

用一组属性规定其定义、标识、表示和允许值的数据单元。

[来源：WS/T 303—2023, 3.1.4]

3.4 元数据 metadata

定义和描述其他数据的数据。

[来源: WS/T 305—2023, 3.1.1]

3.5 数据字典 data dictionary

按照统一标准, 对所有数据元进行结构化、规范化描述的数据集合。

3.6 指标集 indicator set

面向特定业务流程、临床应用或科研目的, 由多个数据元组成的结构化数据集合。

3.7 区块链 blockchain

区块链是一种分布式账本技术, 通过加密和共识机制, 实现数据的透明性、不可篡改性和可追溯性。

3.8 区块链存证 blockchain proof of existence

存证 proof of existence

为了确保存证信息(电子数据)的完整性和真实性, 采用区块链技术实现多节点共识的存证服务。

[来源: T/CESA 1048—2018, 3.1.4]

3.9 审计 audit

安全审计 safety audit

记录受保护医疗数据使用相关的事件, 提供审查追踪。

[来源: WS/T 790.4—2021, 3.1.1]

3.10 角色 role

一组服务于共同目的的活动集合。

[来源: GB/T 32399, 2.1.9]

3.11 区块链节点 blockchain node

区块链网络中负责验证和存储交易数据的计算设备。

3.12 关联性 relevancy

健康和疾病数据与将要开展的临床研究的关联程度。

3.13 稳健性 robustness

稳健性是指系统在面对外部攻击和内部故障时, 通过安全防护、容灾能力和实时监控, 保持稳定运行的能力。

3.14 可靠性 reliability

可靠性是通过标准化的数据模型、人员培训、技术支持以及不干扰临床工作的数据收集方法, 保证数据的完整性、准确性和一致性。

3.15 可及性 availability

保证在专病数据库的管理和使用过程中, 能够及时获取足够的信息以评估关键临床指标。

3.16 数据质量 quality

数据质量是指在数据产生和管理的全生命周期内, 通过系统化的数据治理保证数据的准确性、完整性和一致性。

3.17 去标识化 de-identification

是指通过对个人信息的技术处理, 使其在不借助额外信息的情况下, 无法识别或者关联个人信息主体的过程。

4 缩略语

下列缩略语适用于本文件:

ABAC: 基于属性的访问控制 (Attribute-based Access Control)

ACL: 访问控制列表 (Access Control List)

API: 应用程序接口 (Application program interface)

CDC: 数据变更捕获 (Change Data Capture)

CDISC: 临床数据交换标准组织 (Clinical Data Interchange Standards Consortium)

DICOM: 医学数字成像和通讯协议 (Digital Imaging and Communication in Medicine)

EMPI: 企业级患者主索引 (Enterprise Master Patient Index)

EMR: 电子病历系统 (Electronic Medical Record)

FHIR: 快速医疗互操作性资源 (Fast Healthcare Interoperability Resources)

HIS: 医院信息系统 (Hospital Information System)

ICD-9-CM: 国际疾病分类第九版临床修订本 (International Classification of Diseases, Ninth Revision, Clinical Modification)

ICD-10: 第十版国际疾病分类 (International Classification of Diseases)

LIS: 实验室信息系统 (Laboratory Information System)

MPC: 多方安全计算 (Multi-Party Computation)

MFA: 多因素认证 (Multi-Factor Authentication)

NLP: 自然语言处理 (Natural Language Processing)

PACS: 影像存档和通信系统 (Picture Archiving and Communication System)

RBAC: 基于角色的访问控制 (Role-based Access Control)

SM2: 国密SM2非对称加密算法 (SM2 asymmetric cryptographic algorithm)

SM3: 国密SM3杂凑算法 (SM3 Cryptographic hash algorithm)

SM4: 国密SM4分组密码算法 (SM4 block cipher algorithm)

SNOMED CT: 系统化医学术语 (Systematized Nomenclature of Medicine - Clinical Terms)

5 建设原则和建设目标

5.1 建设原则

专病数据库建设应遵循以下原则:

- a) 规范性原则: 专病数据库数据应采用国际、国家、行业数据标准, 保证数据的可比性和一致性, 支持数据共享与互操作性;
- b) 系统性原则: 统筹专病数据库数据来源, 与现有医疗信息系统无缝对接, 支持跨学科数据整合和利用的需求;
- c) 安全性与数据可信原则: 保证数据的安全与隐私。保证跨机构、跨区域的数据共享与多中心研究中的数据安全、可追溯、完整性、准确性、一致性、可靠性;
- d) 保证质量原则: 采用统一科学的数据治理机制, 保证数据质量;
- e) 易用高效原则: 保证操作简便, 减轻使用者负担, 并通过智能策略和工具, 提升数据利用效率;
- f) 可扩展性原则: 具备良好的扩展性, 能够根据需求扩展功能模块和数据源。

5.2 建设目标

专病数据库的建设旨在通过标准化、系统化地采集、存储、治理和分析特定疾病相关数据, 提升临床研究、患者管理和医疗决策的效率。目标包括:

- a) 支持临床研究: 提供高质量的数据资源, 支持开展临床研究;
- b) 促进科技成果转化: 提供科研成果临床转化的数据支撑, 推动基础研究向临床实践转化;

- c) 辅助临床决策：整合多源数据，为个性化诊疗及预后评估提供数据支持；
- d) 推动数据共享与协作：支持跨机构、跨区域的数据共享与协作。

6 建设内容与技术要求

6.1 数据范围

专病数据库的建设应涵盖患者诊疗过程中产生的多维度数据，保证数据的全面性、准确性和时效性。具体数据范围和类型包括：

- a) 患者基本信息：人口学资料（如出生日期、性别、民族、生活地域等）、既往病史、家族病史等；
- b) 疾病诊断与治疗信息：包括疾病诊断、治疗方案、治疗效果、手术记录、用药记录、基因检测等详细数据；
- c) 随访信息：患者的后续治疗、康复数据、预后评估以及长期随访信息，支持随访计划的制定和管理；
- d) 实验室和影像数据：实验室检验数据（如血液、尿液分析等）、病理检查数据、影像数据（如CT、MRI等）；
- e) 遗传信息：患者的遗传数据、基因组数据等；
- f) 诊疗过程数据：患者的症状、病程记录、检查记录等临床数据，通过结构化和标准化处理，用于临床研究和数据分析；
- g) 可穿戴医疗健康数据：以可穿戴医疗健康设备为载体的医疗健康全生命周期数据，包括数据获取、数据传输、数据集成、数据交互、数据反馈过程中产生的数据；
- h) 设备和费用数据：与患者诊疗相关的设备使用信息、医疗费用数据等，支持医疗经济学分析和公共卫生政策制定；
- i) 多模态数据整合：整合包括文本、影像、基因、病理等多种数据源，形成多维度、全面的患者数据集；
- j) 可扩展数据：包括未来可能新产生的临床过程中的数据种类，以及支持临床研究所需的基础数据要素。

6.2 核心功能模块及要求

专病数据库包括以下核心功能模块，以保证数据的采集、管理、分析和共享能够满足科研和临床应用需求：

6.2.1 数据采集模块

数据采集宜使用接口、库表、手工相结合的方式，数据来源宜包括HIS（医院信息系统）、EMR（电子病历系统）、LIS（实验室信息系统）、PACS（影像存档与通信系统）、医疗物联网设备、外部问卷量表等。支持HL7、FHIR、DICOM等国际标准接口，保证数据采集的及时性和全面性。各类数据在进入专病数据库之前，应进行标准化处理，实现多源异构数据的整合，保证数据在各个模块之间流转无误。宜支持数据变更捕获(CDC)技术，及时监控和捕获源数据的变化，实现数据更新一致性。

6.2.2 数据存储模块

应支持结构化和非结构化数据的存储，对于结构化数据（如患者信息），宜使用关系型数据库，对于非结构化数据（如影像、文本病历等），宜采用分布式文件系统或对象存储进行管理；宜采用分布式存储架构，支持适用于不同访问频率的分层存储（热数据、冷数据、归档数据）；系统宜支持数据分片存储及副本复制，保证高可用性和容灾能力，以实现数据的安全性和高效访问。

6.2.3 数据治理模块

应通过对采集的数据进行标准化、清洗、结构化和归一化处理，保证数据的一致性、完整性和可用性；数据标准化处理应符合CDISC、SNOMED CT、ICD-10、ICD-9-CM等标准，保证数据在跨系统、跨机构间的互操作性。

6.2.4 多模态数据管理模块

支持多模态数据的管理和整合，包括文本、影像、基因、病理等不同类型的数据库，保证多维度数据的关联、展示和分析。

6.2.5 数据质控模块

支持通过自动化质控和人工质控相结合的方式，进行实时质控和定期质控，并生成质控报告。

6.2.6 数据检索模块

支持多维度检索，根据需求快速筛选病例和指标，生成研究对象集。

6.2.7 数据转换、分析与导出模块

支持进行初步的数据转换，如重分组、范围筛选等，支持开展统计描述和基础统计推断，帮助用户充分了解数据特征和质量。支持多格式导出（如CSV、Excel、SPSS、R等），或支持与云计算平台的对接。

6.2.8 数据共享与多中心协作模块

保证数据在隐私保护的前提下实现跨机构共享，支持在线授权和跨系统数据下载，提供多中心研究服务组件，促进多中心研究的开展。

6.2.9 随访管理模块

支持患者长期随访数据的采集与管理，保证数据的连续性和时效性，支持随访计划的制定、提醒和数据更新。

6.2.10 知识图谱模块

应支持构建以疾病、症状、药物、检查项等为核心概念的语义网络，提升系统对医疗实体及其关系的结构化表达能力。知识图谱应优先对齐标准化医学术语体系，支持实体与术语的映射、对齐与补充，确保语义一致性与互操作能力，服务于智能检索、临床决策支持及知识发现等应用场景。

6.2.11 数据安全性与隐私保护模块

通过加密、去标识化、访问控制等技术手段，实现数据的隐私保护和安全性，使其符合相关法律法规的要求。系统宜采用SM2、SM3、SM4等加密标准，保证敏感数据的加密存储和传输，支持基于角色的访问控制（RBAC）和多因素认证（MFA）。

6.2.12 区块链数据存证

应集成区块链技术，对关键数据进行存证处理。支持对存证数据的查询和验证，使用者可快速验证数据的真实性和完整性。此外，系统应提供灵活的存证策略配置，满足不同场景下的数据存证需求，为科研数据的安全性和可信度提供保障。

6.2.13 区块链服务支撑

应构建完善的区块链服务支撑体系，包括区块链网络的搭建、智能合约的开发与部署、共识机制的选择与优化。提供区块链节点的管理、监控和维护功能，确保区块链网络的稳定运行。同时，系统应支持区块链与其他系统的集成，实现数据的跨系统、跨机构共享与协作。

6.2.14 项目管理模块

支持开展包括项目计划、任务分配、权限管理、进度跟踪、资源管理、风险管理、文档管理、沟通协作、报告等。

6.3 参考技术架构

专病数据库架构应能保证系统高效、可扩展、安全，支持多数据源采集、存储、处理、分析和共享，应用于临床研究、决策支持、监管应用等场景，具备智能检索、数据分析、随访管理等关键功能。系统通过数据加密、访问控制、隐私保护和区块链技术保障数据安全合规。采用模块化设计，包括应用层、服务层、数据层、安全层和基础层，以实现系统灵活扩展和跨机构数据共享。参考架构图见图1。

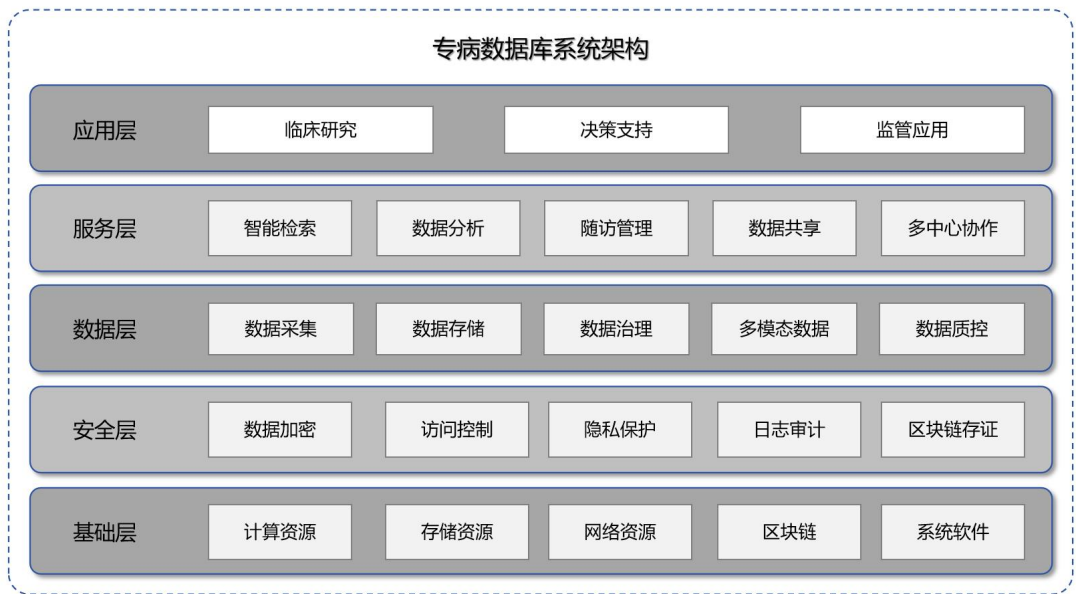


图 1 专病数据库参考架构

系统参考架构的说明如下：

- a) 应用层：展示了专病数据库的具体应用场景，如临床研究、决策支持和监管应用；
- b) 服务层：包括智能检索、数据分析、随访管理、数据共享等核心功能模块，保证数据的高效利用和管理；
- c) 数据层：涵盖数据采集、存储、多模态数据融合和质控，保障数据的完整性和一致性；
- d) 安全层：通过数据加密、访问控制、日志审计和区块链存证，保证数据的安全性和合规性；

- e) 基础层：提供系统运行所需的计算、存储、网络资源及区块链基础设施。

6.4 设计要求

6.4.1 业务需求驱动的设计

系统设计应以业务需求为核心，通过定义专病数据集、梳理数据元素以及明确数据流，保证专病数据库能够有效支持临床和科研需求。具体要求如下：

- a) 专病数据集设计：设计阶段应与业务部门（如临床科室、科研团队）密切合作，明确各类专病数据集的需求，确定核心数据元素（如患者信息、诊疗数据、随访数据等），并保证数据集能够适应临床和科研的多样化需求。
- b) 数据元素定义与标准化：在设计专病数据集时，应根据相关标准定义数据元素，保证数据兼容性和可共享性。设计中应考虑数据的扩展性，允许未来新增或调整数据元素。
- c) 业务流程对接：系统设计应紧密结合临床和科研的业务流程，保证系统能够满足数据采集、分析、随访、科研报告等业务需求。通过设计业务流程与数据流的对接，保证系统在业务层面的高效运作。

6.4.2 模块化系统架构

系统宜采用模块化设计，支持各功能模块独立可扩展。具体要求如下：

- a) 功能模块化设计：系统设计应包括多个核心功能模块，如数据采集模块、数据治理模块、数据质量控制模块、智能检索模块、数据共享模块、随访管理模块和数据安全模块等。每个模块应具备清晰的边界和接口，支持系统的功能完整性和扩展性。
- b) 高内聚、松耦合：各功能模块应相互独立，专注于各自的任务，降低模块间的耦合性，提升系统灵活性。模块之间应通过标准化接口进行数据交换，不同模块能够无缝协作。
- c) 可扩展性与可维护性：系统应支持模块的增减，支持未来能够根据科研和临床需求灵活调整系统功能，每个模块应提供清晰的接口和文档。

6.4.3 数据服务操作要求

数据操作的基本要求如下：

- a) 数据审核与修改：系统应具备数据审核功能，支持数据经过审核后方能进入正式存储，未经审核的数据不得参与分析。数据修改应受到严格权限控制，并且每次修改应经过审核，保证数据的安全性和准确性。
- b) 数据稽查与审计：系统应记录每个数据条目的操作历史，包括数据的创建、修改、删除等操作，实现数据的可追溯性。审计日志宜详细记录操作时间、操作人、操作内容，支持合规性审计。

6.5 数据要求

6.5.1 术语标准化

术语标准化支持专病数据库中的医学术语和编码在不同系统和机构之间保持语义一致，提升数据共享与互操作能力。具体要求如下：

- a) 术语标准化：优先采用国际通用的医学术语标准，对于本地化系统，可通过术语映射逐步引入国际术语体系；
- b) 术语库管理：系统应内置或对接医学术语库，支持术语的新增、更新和维护，保持术语库的完整性和时效性；

- c) 术语映射：系统应支持不同术语系统间的映射与转换，宜实现本地化术语能够自动映射为国际标准化术语，避免因术语不一致造成数据共享障碍。

6.5.2 元数据管理要求

元数据是数据的结构性描述，记录数据背后的来源、创建方式、生命周期属性等。标准推荐遵循 WS/T 305—2023。

6.5.3 数据字典建设要求

系统应参照医学信息相关标准，建立结构清晰、语义一致、可维护的数据字典，以术语标准化和元数据管理为基础，对每一个数据元进行标准化描述，规范字段名称、定义、数据类型、单位、值域、代码系统及适用范围等内容，支撑专病数据库的数据采集、治理、共享与分析。具体要求如下：

- a) 数据字典应完整记录各数据元的属性信息，包括但不限于字段名称、定义、数据类型、单位、可选值、代码系统及适用范围；
- b) 数据字典应与术语体系和元数据标准保持一致，确保语义统一、结构规范，具备良好的扩展性；
- c) 数据字典应具备版本管理能力，支持新增、修改、废止等操作，确保随业务发展及时更新；
- d) 数据字典应支持以结构化、机器可读的方式提供标准字段定义，供系统间配置接口或建立映射，保障跨系统、跨机构的数据结构一致性与语义互操作。

6.5.4 专病指标集制定要求

专病指标集应以数据字典为基础，优先引用其中已定义的数据元进行结构化组合。新增数据元时，应同步纳入数据字典，确保语义一致性与标准统一性。

指标集的制定应由临床专家牵头，联合数据治理和信息化人员协同完成，并经相关评审与一致性校验后发布。制定过程应建立版本管理机制，记录变更内容与生效时间，确保可追溯与持续优化。

6.5.5 数据模型与结构要求

专病数据库中的数据模型和数据结构应具备统一的标准，支持不同系统、不同机构的数据能够无缝集成和共享。具体要求如下：

- a) 数据结构标准化：采用标准化的数据结构模型，保持数据层次清晰，数据元素间的关系明确；
- b) 通用数据模型应用：宜采用通用数据模型或类似标准模型，支持数据的标准化转换和存储；
- c) 数据模型扩展性：数据模型应具备良好的扩展性，支持新增数据项。

6.5.6 数据质量控制要求

数据质量控制是保证专病数据库数据完整性、准确性、一致性、时效性和规范性的关键。具体要求如下：

- a) 质控规则标准化：系统应内置质控规则，涵盖数据的完整性、唯一性和准确性，保证数据质量符合标准。宜随机抽取专病数据库中不少于 1% 的病例数据或不少于 20 例进行数据人工检查（以抽取病例数较多的方式为准）；
- b) 质控流程：系统应支持自动质控流程，定期执行数据质量检查，并生成质控报告，记录质控结果和问题；
- c) 持续质控与改进：专病数据库应具备持续质控能力，应定期进行数据清洗，去除冗余、错误和缺失的数据，维护数据的准确性和可用性，支持数据质量的多轮迭代和改进。

6.5.7 数据共享与互操作要求

系统应支持数据在不同系统、不同机构之间安全、高效地共享和互操作。具体要求如下：

- a) 唯一身份识别：系统应支持企业级患者主索引（EMPI），用于跨机构、跨系统识别患者的唯一身份，实现患者数据的准确整合和关联；
- b) 标准化数据接口：通过标准化 API 接口进行数据交换，实现系统间的数据共享与互操作性；
- c) 跨机构数据共享：系统宜支持跨机构数据共享标准，支持数据去标识化和隐私保护，防止敏感信息泄露；
- d) 数据互操作性：系统应符合国际和国内的数据互操作性规范，支持数据在不同医疗系统间有效交换和使用。对于国内单位，系统应支持本地标准的互操作性。

6.6 数据安全性与隐私保护要求

6.6.1 数据分类分级

专病数据库应依据数据的属性、敏感性及潜在影响，对所采集、存储和处理的数据资源开展分类分级管理。分类分级应符合国家和地方相关标准，结合数据使用场景，明确数据类别与级别，并据此实施差异化的安全控制措施，保障数据的安全性与合规性。具体要求如下：

- a) 数据分类应结合内容特征、业务用途及管理需求，建立清晰、可扩展的分类体系；
- b) 数据分级应综合评估数据泄露、篡改、非法获取等情景下对个人、组织或社会可能造成的影响，合理确定数据等级；
- c) 分类分级结果应作为实施技术防护和管理控制的依据，按照“就高不就低”原则，配置相应的安全保护措施；
- d) 分类分级策略应动态更新，及时反映数据属性、使用范围或外部环境变化。

6.6.2 数据加密

系统应采用加密技术对敏感数据进行加密，并在数据共享和存储过程中进行去标识化处理，保证患者隐私不被泄露。系统的密码应用应符合 GB/T 39786—2021，保证加密算法的安全性，防止未授权访问和数据泄露。要求如下：

- a) 系统应采用国家密码标准进行数据加密，保证数据在存储和传输过程中的安全性与完整性；
- b) 应通过分级权限管理和日志追踪机制，实现只有授权人员能够访问和操作敏感数据，并能够追溯数据操作历史。

6.6.3 数据访问控制

访问控制应基于角色或属性，保证不同用户只能访问其授权范围内的数据。权限管理与访问控制要求如下：

- a) 分级权限管理：系统应根据用户角色和职责分配相应权限；
- b) 访问控制列表（ACL）与角色权限模型：宜通过 ACL 或角色权限模型灵活配置权限，保证敏感数据的安全性；
- c) 基于角色的访问控制和基于属性的访问控制：系统宜同时支持 RBAC 和 ABAC 模型，实现灵活的权限管理；
- d) 多因素认证（MFA）：系统宜使用 MFA 机制，保证用户在访问敏感数据时，提供额外的身份验证层，防止未授权的访问。

6.6.4 数据去标识化与隐私保护

专病数据库中涉及患者的个人信息应严格遵守 GB/T 35273—2020和 GB/T 37964—2019 等相关标准，通过去标识化、加密等技术手段，保证个人隐私得到有效保护。

中小型机构应优先考虑使用基础的去标识化技术和隐私保护工具，实现数据隐私保护的同时，降低技术实现难度。大型机构宜逐步引入更为复杂的隐私计算技术，如差分隐私、多方安全计算和联邦学习，保证在数据共享和分析过程中实现更高水平的隐私保护。

去标识化处理对象包括直接标识符（如姓名、身份证号）和部分间接标识符（如生日、住址、就诊时间等），应在数据共享、出库等环节完成处理。处理策略应结合数据类型、使用场景与机构能力确定，在确保数据可用性的同时降低重识别风险。对高敏感数据，建议参考 GB/T 42460—2023 进行效果评估。

参考方案如下：

- a) 去标识化方案：中小型机构宜采用基础的去标识化技术，如哈希化、假名化等，满足数据处理的基本合规要求；具备技术能力的大型机构，宜采用更为复杂的去标识化方法（如假名化结合蒙特卡洛算法等），进一步提高数据隐私保护的强度。去标识化处理应参考 GB/T 37964—2019，合理选择技术路径，确保数据在可用性与安全性之间取得平衡。
- b) 差分隐私：中小型机构在使用差分隐私技术时，宜借助第三方工具或服务，避免自行开发，确保算法参数设置合理、隐私预算可控。
- c) 多方安全计算（MPC）：MPC 是一种较为先进的隐私计算技术，适用于跨机构协作场景。中小型机构宜选择与第三方服务提供商合作，借助成熟的 MPC 平台实现多个数据持有方的协作分析。
- d) 联邦学习：联邦学习是一种分布式机器学习技术，适用于分布式多中心 AI 建模研究的场景。中小型机构宜优先选择现有的联邦学习平台，通过与其他机构协作开展分布式 AI 建模。

6.6.5 数据传输与存储安全

专病数据库的安全防护应符合GB/T 22239—2019，保证系统具备足够的网络安全防护能力，特别是在数据存储和传输过程中，防止数据泄露与篡改。要求如下：

- a) 加密传输：所有跨地域或跨机构的数据传输应使用加密的传输通道，保证数据在网络传输过程中不被截获或篡改。
- b) 加密存储：专病数据库中的高敏感数据（如基因数据、影像数据）应配备物理和逻辑隔离措施，并采用加密存储。
- c) 分布式存储与备份机制：对于大型机构，宜采用分布式存储方案，并提供异地备份机制，保证数据的高可用性和安全性。

6.6.6 数据完整性保护

系统宜采用基于区块链的数据完整性保护，保证数据在采集、转换、存储、交换、研究计算过程的完整性，保证数据不被篡改、提供安全的数据协作。要求如下：

- a) 数据存证：对数据的采集、转换、存储、交换以及研究计算过程进行全程记录，确保数据的真实性和不可篡改性。应能提供对数据的哈希验证、时间戳、电子签名及分布式存储。
- b) 区块链服务支撑：应能提供对区块链加密技术、共识机制、智能合约、监控与审计等服务功能，支撑应用的数据完整性保护。

6.6.7 系统日志与审计

系统应具备详细的操作日志记录和审计机制，保证所有数据操作的可追溯性和合规性。要求如下：

- a) 操作日志记录：系统应记录所有用户的操作行为，包括数据的访问、修改、删除、导出等操作。日志记录应包括操作时间、操作用户、操作内容等信息，且日志应具备不可篡改性。

- b) 日志审计：系统应定期审计操作日志，保证所有操作符合法规与安全规范。特别是涉及敏感数据的操作（如导出、删除）应进行严格审计。

6.7 数据合规要求

专病数据库的建设和使用应严格遵守国家有关法律法规和伦理，确保数据利用的合法性与合规性。具体要求如下：

- a) 数据合法合规：专病数据库的数据采集、存储、处理和共享应符合《中华人民共和国个人信息保护法》《中华人民共和国数据安全法》《中华人民共和国人类遗传资源管理条例》等相关法律法规。
- b) 遵循伦理要求：专病数据库的数据采集、存储、处理和共享应符合伦理规定，在利用专病数据库开展具体科学研究前，应额外获得伦理审批。
- c) 数据留存与销毁机制：应建立数据安全留存和定期销毁机制，数据在生命周期结束后，应进行安全销毁处理，避免数据滥用和非法利用。
- d) 合规审计与监督：系统应设有合规审计机制，定期检查数据管理行为的合规性，发现问题及时整改，形成闭环管理。

7 建设流程和管理

7.1 建设准备阶段

专病数据库的建设准备阶段宜包括以下流程：

- a) 明确需求：根据具体的学科发展方向、学科优势、创新研究需求、科学问题，参考循证医学证据及专病专家共识，进行研究设计。围绕研究设计，确定数据对象和数据内容。
- b) 制定时间表：制定项目时间表，包括需求分析，数据集确定，系统设计，数据采集、治理与质控，系统测试及优化，专病数据库上线及运维等阶段的时间节点。
- c) 确定数据集，宜包括以下子流程：
 - 1) 确定病例对象：应根据科研或创新发展需求，设定病例对象的纳入标准和排除标准。
 - 2) 确定指标集：应结合现有数据和科研需求，确定专病数据库的通用指标和专病指标，形成指标集。
 - 3) 确定数据来源：宜调研数据的储存位置、格式、接口方式、访问权限、数据质量，综合决定数据采集来源。
 - 4) 数据集评审：应由临床专家、临床流行病学专家、信息技术专家进行数据集评审，保证其符合实际应用需求。

7.2 系统部署与配置阶段

小型机构宜采用单节点部署或虚拟化方案，以满足基础的数据存储和管理需求。随着业务需求的增长，机构宜逐步扩展到分布式架构或云端架构；对于具备较强技术能力的中大型机构，宜采用分布式存储、容器化技术和容灾机制，实现应用程序的快速部署、动态扩展，提供高并发访问和海量数据的管理。部署要求如下：

- a) 覆盖度要求：专病数据库的部署应覆盖所有预定的功能需求，包括数据采集、存储、分析、共享和安全管理。
- b) 部署环境规划：系统应根据业务需求和负载情况进行部署，合理配置硬件（如服务器、存储设备、网络带宽）与软件（如操作系统、数据库系统）。

- c) 高可用性设计：系统宜设计为 7×24 小时连续运行，支持负载均衡，保证高并发访问下的稳定性。
- d) 动态扩展与自动化运维：系统宜配备自动化运维工具，支持动态扩展和自动调整资源配置。
- e) 容灾机制：系统应具备容灾设计，宜通过分布式部署和数据冗余。
- f) 热备与冷备：系统宜支持热备或冷备机制，保证在系统或硬件故障时，能够迅速切换到备份系统，避免数据丢失。
- g) 微服务架构：对于中大型机构或对系统扩展性要求较高的场景，宜采用微服务架构。每个微服务模块应通过标准化 API 进行通信，实现系统的灵活性和扩展性。

7.3 测试与验收阶段

专病数据库测试与验收阶段宜包括以下流程：

- a) 系统测试与性能优化，宜包括以下子流程
 - 1) 功能测试：应对系统的各项功能进行全面测试，包括数据录入、查询、统计、导出等，保证满足用户需求。还应进行边界条件测试，力求系统在极端情况下能稳定运行。
 - 2) 性能测试与优化：通过模拟多用户并发访问场景，进行压力测试，评估系统的负载能力和响应速度。还需通过长时间运行测试，避免内存泄漏、CPU 过高等问题。
 - 3) 安全测试：宜邀请专业的安全团队进行渗透测试，发现潜在漏洞并修复。对敏感数据进行加密存储，并实施严格的访问控制策略。
 - 4) 用户体验测试：通过目标用户试用系统，收集反馈，对系统进行优化。评估系统的易用性，支持用户能方便操作系统。
- b) 系统试运行与验收，宜包括以下步骤
 - 1) 试运行：系统上线试运行，进行数据采集、处理和使用，发现并解决运行中的问题。
 - 2) 问题收集与系统优化：收集试运行期间的问题并进行优化。
 - 3) 项目验收：根据功能、性能和数据质量进行验收，项目正式上线并交付运维。
- c) 培训与课题研究，宜包括以下步骤：
 - 1) 文档编写与培训：应提供详细的技术文档（如安装指南、操作手册等），并对最终用户进行系统操作培训，保证用户能熟练使用系统。
 - 2) 课题研究：使用者利用系统进行课题研究，进行数据分析并验证科研假设。

7.4 专病数据库实施与迁移

专病数据库的实施与迁移宜符合以下要求：

- a) 专病数据库的参考实施流程见附录 A。
- b) 历史数据迁移：新系统上线时，宜将现有历史数据迁移至新的专病数据库中。
- c) 数据同步与及时更新：监控和捕获源数据中的变更，实现专病数据库中的数据能够及时更新，反映最新的患者诊疗状态。
- d) 数据一致性保障：系统应采用事务管理机制，保持数据在迁移过程中的一致性，避免数据丢失或冲突。

7.5 数据运维与生命周期管理

专病数据库建设宜采用以下的数据管理与运维措施：

- a) 自动化运维与监控：系统宜具备自动化运维平台，支持专病数据库的自动化部署、监控、备份和恢复。运维平台宜实时监控系统的状态，并在出现潜在问题时自动发出告警。

- b) 数据审计与追踪：系统宜具备数据审计功能，记录所有数据的访问、修改和删除操作，提供详细的操作日志。
- c) 访问控制与权限管理：系统应支持分级权限管理，实现不同用户根据角色和权限访问相应数据。权限管理宜支持动态调整，实现系统的灵活性和安全性。
- d) 数据归档与分级存储：系统应根据数据使用频率和重要性，制定数据归档策略，将不经常访问的数据转移至冷存储或归档存储，减少存储成本并提高系统效率。
- e) 数据利用：数据的利用策略与方法参见附录 B。
- f) 数据销毁：在数据生命周期结束时，系统应支持安全的数据销毁机制，保证数据在不再使用时能够彻底删除，避免非法访问或滥用。数据销毁应符合相关法规，并生成销毁日志记录。

8 专病数据库评价

8.1 评价标准结构

专病数据库评价标准包括两个方面：一是，评价专病数据库是否满足临床研究的需求，包括对专病数据库与临床研究的相关性和可及性两方面的评价；二是，评价专病数据库是否完整准确映射真实世界场景，包括对专病数据库稳健性、可靠性和数据质量三个方面的评价。专病数据库评价标准结构见图2。

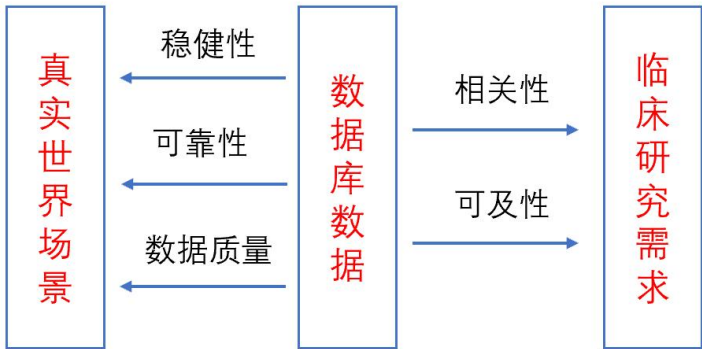


图 2 专病数据库评价标准结构

8.2 满足临床研究的需求评价

评价专病数据库是否满足临床研究的需求，包括：

- a) 相关性（relevancy）：评价专病数据库中的数据与将要开展的临床研究的关联程度。应包括对象（population）、干预/暴露指标（intervention）、对照（comparison）和结局（outcome）四大要素信息；以及评价专病数据库的信息是否能够满足临床和数据监管和质量控制的要求。
- b) 可及性（availability）：评价在实施临床研究过程中，研究人员可以从专病数据库中获取有用的信息，是否足以评价关键临床结局。应包括：是否使用了通用语义定义框架和资料收集的格式、是否有时间序列的资料收集框架、是否根据关键终点设定结局指标以便捕捉关键不良结局、是否可以从其他相关数据库获取有价值信息。另外，可及性还应包括保证授权用户能够便捷、快速地访问所需数据，提升专病数据库的使用效率和用户满意度。

8.3 映射真实世界场景评价

评价专病数据库是否完整准确映射真实世界场景，包括：

- a) 稳健性（robustness）：是指专病数据库收集的数据与临床真实场景的符合程度。评价专病数据库中收集的研究对象是否具有代表性、各类场景的数据在设计 and 采集阶段是否进行标准化、

专病数据库设计阶段是否进行过同行评议、数据是否可校验、是否建立有数据校验模型和数据质量评估标准。

- b) 可靠性 (reliability)：是指在实施过程中数据完整性和准确性的制度保障。评价专病数据库在建设实施过程中是否有相关的制度和流程，包括人员培训和支持系统、通用语义集和术语字典、可靠的数据自动抓取系统和技术、研究对象的代表性策略、避免患者和医务人员的主观偏差策略。
- c) 数据质量 (quality)：是指产生数据后如何采用数据治理，保障数据的质量。评价内容包括数据方面的完整性、准确性、一致性、时效性和规范性。管理方面是否设置风险点的监控和质量审计，安全方面是否有保障系统安全性和数据安全性的措施和条件。

附录 A
(规范性)
参考实施流程

A.1 实施流程阶段划分

专病数据库的实施流程分为六个主要步骤，每个阶段环环相扣，保障系统的有效建设、部署和运行，见图A.1。

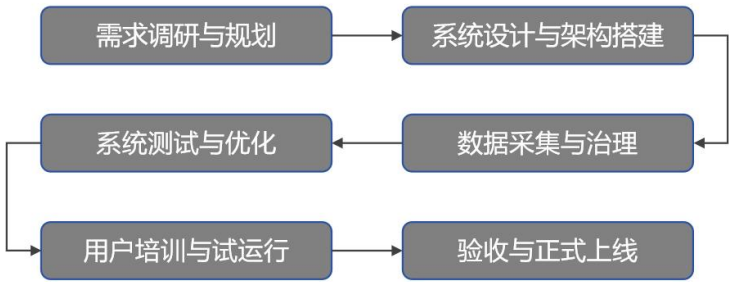


图 A.1 实施流程图

A.2 实施步骤说明

A.2.1 需求调研与规划

在这一阶段，项目团队与临床医生、科研人员等利益相关方共同确定系统的建设目标、应用场景、病种及观察对象，明确数据集的范围和类型。项目组根据这些需求制定建设目标，明确系统功能，并确定项目的时间计划、预算和资源分配。此阶段为后续系统设计及实施奠定基础。

A.2.2 系统设计与架构搭建

根据前期需求调研结果，设计系统架构，明确数据采集、治理、存储、分析及共享等模块的技术方案。开发并使用数据采集接口、库表和手工相结合的方式，从HIS、EMR、LIS、PACS等系统及时协同采集数据。系统架构设计完成后，系统的硬件和软件环境也需搭建与配置，为接下来的数据处理做好准备。

A.2.3 数据采集与治理

启动数据采集接口，导入历史数据。数据从HIS、EMR、LIS、PACS等系统进入专病数据库后，需对其进行清洗，去除冗余数据，修复错误数据，并进行标准化处理，保证所采集的数据符合预期的标准。此时，数据质量控制机制也会被应用，保证数据的准确性、完整性和一致性。

A.2.4 系统测试与优化

在这一阶段，系统的各项功能模块将接受全面测试，保证其能够满足用户需求。系统还需经历性能压力测试，评估在高并发访问情况下的表现，并进行安全测试，保证数据安全性。根据测试反馈，项目团队将对系统进行优化和调试，保证系统的稳定性和高效性。

A.2.5 用户培训与试运行

项目团队为科研人员、临床医生及系统管理员提供培训，保证他们能够熟练掌握系统的操作和功能。随后，系统进入试运行阶段，项目团队收集实际数据运行时的反馈，解决运行中的问题，为最终上线做准备。

A. 2. 6 验收与正式上线

系统试运行完成后，项目团队与用户进行最终验收。验收通过后，系统正式上线，投入实际使用。上线后，系统进入日常运维阶段，项目组会对系统的运行状态进行持续监控，定期优化，保证系统稳定、可靠运行。

附 录 B

(规范性)

数据利用策略与方法

B.1 数据利用的基本原则

数据利用应遵循以下基本原则：

- a) 合法合规性：数据利用应符合相关法律法规，保证患者隐私不被泄露，并符合伦理审查的要求。
- b) 科研与临床支持：专病数据库应支持临床研究、统计分析和患者管理，推动科研成果的临床转化，提升临床决策的科学性。
- c) 高效利用与多维分析：数据利用应支持多维度、多层次的分析需求，宜利用智能化工具（如机器学习、自然语言处理等）提升数据利用效率。
- d) 安全与隐私保护：在数据使用过程中，系统应进行数据脱敏和去标识化处理，并通过严格的权限管理和访问控制，保证数据安全。

B.2 数据应用场景

专病数据库的数据支持科研人员和临床医生进行不同类型的研究和分析利用，应支持以下应用场景：

- a) 单中心临床研究：数据采集与管理来自一个医疗机构，支持内部科研和临床应用。专病数据库支持使用者筛选患者，生成科研队列，并进行回顾性和前瞻性分析，支持疾病发病机制、治疗效果和预后的研究。
- b) 多中心协同研究：支持跨机构多中心临床研究，通过标准化接口实现数据无缝整合与共享。专病数据库支持跨机构数据共享，提升多中心研究的普适性和分析结果的可靠性。
- c) 创新药物与医疗器械研发：基于患者数据支持药物和医疗器械临床研究和安全性评估。
- d) 随访管理与长期健康监测：支持患者的长期随访管理和慢性病健康监测。记录随访期间的健康状况和治疗效果，支持随访计划的制定和管理。
- e) 统计分析 with 报表生成：支持通过描述性统计、回归分析等方式进行历史数据分析，并生成多维度的统计报表。
- f) 临床决策支持：通过整合患者的多维度数据（如实验室数据、影像数据等），为医生提供精准的诊疗信息，辅助个性化治疗方案的制定。

B.3 数据分析与挖掘

数据分析与挖掘要求如下：

- a) 统计分析：专病数据库宜支持多种统计分析方法（如描述性统计、多因素回归分析、生存分析等），为使用者提供数据分析支持。
- b) 机器学习与人工智能：宜支持基于机器学习和人工智能的数据分析，利用历史数据进行疾病预测、治疗效果分析等。
- c) 自然语言处理（NLP）：专病数据库应支持 NLP 技术，处理非结构化数据（如病历文本、影像报告），将其转化为可分析的结构化数据。
- d) 数据可视化：系统应提供数据可视化工具，通过图表等形式展示分析结果，帮助科研人员和医生直观理解数据规律。

B.4 数据共享与协作

专病数据库的数据共享应遵循 GB/T 42384—2023和 WS/T 500—2016，实现不同机构之间的数据互操作性。具体要求包括：

- a) 标准化数据接口：系统应提供标准化 API 接口，支持跨系统、跨机构的数据共享，实现数据的互操作性。
- b) 多中心协作机制：系统宜支持多中心协同研究，通过联邦学习等技术，在不共享数据的前提下，进行模型训练和协同分析。
- c) 数据共享与协作中的数据安全与隐私保护，具体见 7.8 数据安全与隐私保护要求。
- d) 专病数据库在跨区域、跨机构的数据共享时，应遵守 WS/T 790—2021。