

团 体 标 准

T/ACCEM XXXX—2024

健康期望寿命知识库建设指南

Guidelines for construction of healthy life expectancy knowledge base

（征求意见稿）

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

2024 - XX - XX 发布

2024 - XX - XX 实施

中国商业企业管理协会 发 布

目 次

前言 II

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 缩略语 1

5 建设原则 1

6 组成和内容 2

 6.1 知识库组成 2

 6.2 知识库内容 2

 6.3 数据内容 3

7 建设流程 4

 7.1 需求调查和分析 4

 7.2 功能设计 4

 7.3 逻辑设计 5

 7.4 物理设计 5

 7.5 建库 6

8 安全保障 7

 8.1 物理安全 7

 8.2 系统安全 7

 8.3 网络安全 7

9 运行维护 7

 9.1 管理制度 7

 9.2 性能优化 7

 9.3 软硬件维护和升级 7

附录 A（资料性） 健康期望寿命知识库表结构 8

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。

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。

本文件由上海市疾病预防控制中心提出。

本文件由中国商业企业管理协会归口。

本文件起草单位：上海市疾病预防控制中心、XXX、XXX。

本文件主要起草人：XXX、XXX、鲍婕。

健康期望寿命知识库建设指南

1 范围

本文件规定了健康期望寿命知识库建设的建设原则、组成和内容、建设流程、安全保障和运行维护。本文件适用于健康期望寿命知识库的建设、管理和运行维护。

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 2261.1 个人基本信息分类与代码 第1部分:人的性别代码

GB/T 9361 计算机场地安全要求

GB 17859—1999 计算机信息系统 安全保护等级划分准则

WS/T 305 卫生健康信息数据集元数据标准

3 术语和定义

本文件没有需要界定的术语和定义。

4 缩略语

下列缩略语适用于本文件。

DBMS: 数据库管理系统 (Database Management System)

GBD: 全球疾病负担研究 (Global Burden of Disease Study)

5 建设原则

健康期望寿命知识库的建设应遵循以下原则：

- 实用性原则：知识库的建设应集成疾病分类、失能权重等健康期望寿命测算的重要基础数据，形成完整有机标准知识库，实现便捷高效的知识数据应用功能，为健康期望寿命研究提供系统有效的知识服务；
- 规范化原则：知识库的建设应使用一致的命名约定和编码规范来命名和标识数据库中的数据元素，采用标准化的接口和协议，在数据库设计、建立、管理与维护、服务等过程中，符合相关标准规范的要求；
- 安全性原则：在知识库设计、实施、运行和维护等方面采取必要的安全和保密措施，保障整个知识库系统安全、正常和有效地运行及使用；

- d) 系统性原则：知识库建设应在技术指标、标准体系、库体结构、服务方式等方面具有系统性，知识库系统整体上应具有良好的集成性；
- e) 扩展性原则：知识库建设时应考虑后期数据扩展工作，预留足够的存储空间、扩展接口等；
- f) 现势性原则：知识库应定期更新数据库内容，定期进行性能优化，及时发现和解决潜在的性能瓶颈和问题，保证知识库中数据的现势性。

6 组成和内容

6.1 知识库组成

健康期望寿命知识库如图 1 所示，由如下三部分组成：

- a) 数据库：包括现势库和历史库，按数据类型分为疾病分类数据、伤残权重数据、疾病病程分布数据、疾病严重程度分布数据、疾病诊治时间分布数据和疾病负担标准寿命数据；
- b) 支持环境：包括知识库运行所需的硬件、网络和基础软件；
- c) 管理系统：对数据进行储存、管理和运维的软件系统。

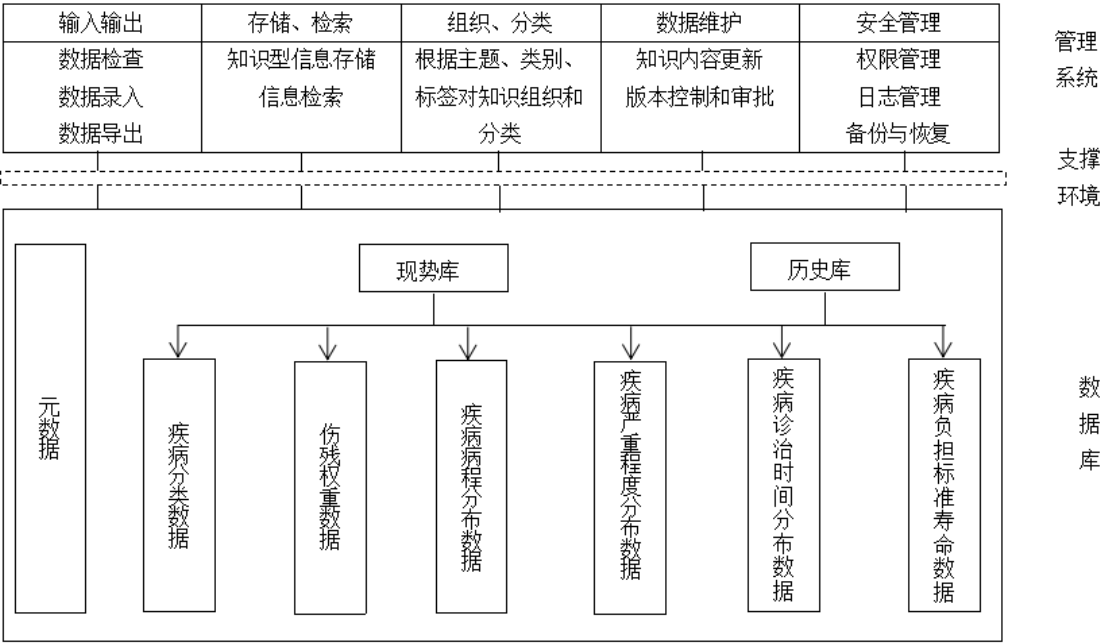


图1 健康期望寿命知识库组成

6.2 知识库内容

6.2.1 术语和定义

包括健康期望寿命相关的术语和指标定义，帮助用户理解和掌握相关知识。

6.2.2 数据集

主要包括疾病分类数据集、伤残权重数据集、疾病病程分布数据集、疾病严重程度分布数据集和疾病诊治时间分布数据集。

6.2.3 图谱和关系

应用知识图谱、关系图等技术，将健康期望寿命相关的知识进行建模和表示，以可视化的方式呈现系统整合的各类知识点和测算指标的关系，以及知识来源信息，实现知识资源全面和充分的开发和知识创新。

6.2.4 实例和案例

包括各种实际案例、应用场景、解决方案等，帮助用户理解和应用知识。

6.3 数据内容

6.3.1 疾病分类数据

6.3.1.1 包括致死性疾病和非致死性疾病信息，其表结构字段名称、字段代码、字段类型及约束条件等参见表 A.1。

6.3.1.2 致死性疾病信息包括：

- a) 一到四级的疾病名称和对应的疾病编码；
- b) 每种疾病的中国疾病分类编码；
- c) 每种疾病的国际疾病分类编码的第九版和第十版信息。

6.3.1.3 非致死性疾病信息包括：

- a) 一到四级的疾病名称和对应的疾病编码；
- b) 每种疾病的中国疾病分类编码；
- c) 每种疾病的国际疾病分类编码的第九版和第十版信息。

6.3.2 伤残权重数据

6.3.2.1 包括各类疾病后遗症的健康状态信息和失能权重信息，其表结构字段名称、字段代码、字段类型及约束条件等参见表 A.1。

6.3.2.2 疾病后遗症的健康状态信息包括但不限于：

- a) 后遗症的名称；
- b) 健康状态名称；
- c) 健康状态描述。

6.3.2.3 失能权重为每种后遗症对应的健康状态的失能权重的点估计和可信区间。

6.3.3 疾病病程分布数据

包括各类疾病的病程信息，包括非致死性疾病名称、疾病编码类别、疾病病程等信息，其表结构字段名称、字段代码、字段类型及约束条件等参见表 A.1。

6.3.4 疾病严重程度分布数据

包括各类疾病不同严重程度的各类后遗症缝补信息，包括疾病名称和编码、后遗症名称、年龄组、性别和严重程度分布等内容，其表结构字段名称、字段代码、字段类型及约束条件等参见表 A.1。

6.3.5 疾病诊治时间分布数据

包括各类危险因素在不同暴露水平下，与相关疾病发病或死亡的关联程度，包括危险因素名称、暴露剂量、对应的疾病名称和结局指标，以及暴露人群的性别和年龄等，其表结构字段名称、字段代码、字段类型及约束条件等参见表 A.1。

6.3.6 疾病负担标准寿命数据

包括不同年份GBD测算所使用的参考寿命表信息，包括年龄组和期望寿命值，其表结构字段名称、字段代码、字段类型及约束条件等参见表 A.1。

6.3.7 元数据

应符合 WS/T 305 的规定。

7 建设流程

7.1 需求调查和分析

7.1.1 基本要求

在知识库的设计、建库、运行与维护阶段均应进行需求调查和分析：

- a) 设计阶段：了解用户实际需求、现有的数据资源等；
- b) 建设期间：修正系统的建设方案；
- c) 运行阶段：进行数据库软硬件升级和功能、性能调整改进。

7.1.2 需求调查

应选取有代表性的用户进行需求调查，明确调查的内容、制定调查计划并编制调查报告。调查报告的内容应真实、科学，包括但不限于：

- a) 用户概况；
- b) 管理需求；
- c) 应用需求；
- d) 数据需求(使用的数据、生产的数据、需要的数据)；
- e) 安全需求；
- f) 技术装备情况需求(网络、硬件、软件等)。

7.1.3 需求分析

根据需求调查结果进行统计汇总，编写需求分析报告，内容包括但不限于：

- a) 直接用户、潜在用户；
- b) 用户对功能的需求；
- c) 数据覆盖范围、数据类型、格式、内容、数据质量和其他需要明确的对数据库建库有影响的数据情况；
- d) 根据调查分析结果，明确知识库的建设目标和范围，确定数据库所涵盖的内容和数据类型。

7.2 功能设计

7.2.1 基本功能

应能满足用户需求和对知识库的管理，应具有数据的输入、输出、存储、处理、查询、更新等基本功能。

7.2.2 安全管理

应包括如下功能：

- a) 权限管理：能根据用户的角色和权限限制对知识内容进行访问和编辑；
- b) 系统日志管理；

c) 备份与恢复。

7.2.3 信息存储与检索

应包括如下功能：

- a) 有效存储大量的健康期望寿命测算所需的知识型信息；
- b) 提供包括但不限于关键词、属性等条件检索方式，查询、检索用户所需的信息。

7.2.4 信息分类

支持根据不同的主题、类别或标签对知识信息进行组织和分类。

7.2.5 可视化表达

支持将健康期望寿命相关的知识进行建模和表示，采用应用知识图谱、关系图等技术将系统整合的各类知识点和测算指标的关系以可视化的方式呈现。

7.2.6 管理与维护

应包括如下功能：

- a) 软硬件升级维护；
- b) 系统功能、性能优化调整；
- c) 数据维护和历史数据管理；
- d) 对于经常更新的知识内容，支持版本控制和审批。

7.3 逻辑设计

7.3.1 数据关联及组织

不同的数据应建立逻辑关联，同类数据间应建立逻辑无缝关联，并采用优化的数据结构和组织方法，减少数据冗余。

7.3.2 图谱数据的组织

应按下列原则分层：

- a) 同一类数据放在同一层；
- b) 相关的数据层可以组成专题；
- c) 有明确关系的数据层应建立关联。

7.3.3 元数据的组织

采用整体化的数据组织方式，并建立与相应数据的关联。

7.3.4 历史数据的组织

应与当前版本的数据组织方式保持一致。数据应正确附加历史年代信息，与现实数据保持可关联性。

7.4 物理设计

7.4.1 软硬件选型

应包括如下内容：

- a) 操作系统；
- b) DBMS；

- c) 数据处理设备，如数据库服务器、文件服务器、网络服务器、Web 服务器、图形工作站等；
- d) 数据存储和备份设备。

7.4.2 数据库结构设计

应包括如下内容：

- a) 基于逻辑设计提出的模型，按照软硬件配置、数据量估算，分配数据库、软件、工作区的物理空间，确定各种数据的目录结构和存储位置；
- b) 各类数据应分别设计各自的数据表结构。其中，矢量数据的每一种地物类型对应一个数据表；
- c) 确定各数据表的数据项名称、类型、宽度和值域范围。

7.4.3 索引结构设计

在数据入库前根据系统软件选型，针对每种数据选定相应的索引关键数据项。数据入库及运行后，根据实际效果对索引结构进行调整。

7.4.4 用户界面设计

根据知识库的各种应用模式，设计体现用户特色的用户界面，界面应简洁、实用。

7.5 建库

7.5.1 数据准备

按照建库设计的要求，收集各种形式的数据和资料，并整理、建档和备份。可搜集的数据和资料包括但不限于：

- a) GBD 网站发布信息；
- b) 发表的文献信息；
- c) 公开数据集。

7.5.2 数据处理

应包括如下内容：

- a) 数据清洗：将各种来源的数据进行清洗、整合，去除重复项、噪音数据和不准确数据；
- b) 数据预处理：采用格式化、标准化、去除不必要的信息等方式对数据进行预处理。形成标准统一的结构化数据集。

7.5.3 数据入库

数据入库应根据所选择的数据组织方式进行，并选择合适的数据库管理系统(DBMS)进行数据存储。可选用手动添加或程序批量入库。数据入库完成后应记录数据入库日志。

7.5.4 数据入库后检查

应包括如下内容：

- a) 完整性检查：检查数据是否有缺失或重复；
- b) 逻辑一致性检查：检查数据是否存放在规定的数据表中，和入库数据是否一致，数据拼接是否无缝；
- c) 属性精度检查：检查属性项名称、类型、长度、顺序、属性值等内容是否正确。

7.5.5 数据建模

对数据进行语义建模，确定数据之间的关联和联系。

7.5.6 数据呈现

构建知识图谱、关系图等形式，展示知识数据之间的结构和关系。

8 安全保障

8.1 物理安全

物理环境应符合 GB/T 9361 的规定，具备防火、防水、防盗、防静电、防雷击、防辐射等安全设施，满足温度、湿度、供电、照明以及设备使用和存放等方面的要求。

8.2 系统安全

应达到 GB 17859—1999 规定的第二级安全保护等级；与公共网络连接应采用逻辑隔离技术，具备防入侵、防病毒等系统安全措施。

8.3 网络安全

应达到 GB/T 22239—2019 规定的第二级安全要求。

9 运行维护

9.1 管理制度

应建立完善的运维管理制度，对知识库、软硬件设备等方面提出明确的运维管理要求，并指定专人负责日常运维工作，保证知识库的正常运行。

9.2 性能优化

应持续对知识库进行性能优化，根据实际使用情况进行参数调整和功能提升，保证知识库的高效运行。

9.3 软硬件维护和升级

- 9.3.1 应定期对软硬件进行巡检，频率宜不低于每月 1 次。
- 9.3.2 发现软硬件故障后，故障恢复时间应小于 24 h。
- 9.3.3 软硬件需更换时，提前做好更换实施预案，更换实施周期应小于 48 h。
- 9.3.4 软硬件的升级应满足知识库安全性要求，并遵循以下原则：
 - a) 兼容性：应能与在用知识库的软硬件设备兼容；
 - b) 可靠性：应选择成熟度高的软件或版本，保持升级后的系统稳定性。

附 录 A
(资料性)
健康期望寿命知识库表结构

A.1 健康期望寿命知识库结构见表 A.1。

表A.1 健康期望寿命知识库表结构

数据表名称	字段名	数据源名称	定义	字段类型	表示格式	数据源允许值
参考寿命表 (reference life table)	year	年份	参考寿命表制定年份	N	N4	
	Age	年龄组	年龄组	S3	N2	年龄组代码表
	Life Expectancy	期望寿命	各年龄组期望寿命	N	N..10, 6	
伤残权重 (Disability weights)	year	年份	不同年份的伤残权重数据	N	N4	
	Sequela	后遗症名称	各类疾病后遗症的名称	S1	AN..200	
	Health state name	健康状态名称	健康状态的名称	S1	AN..200	
	Health state lay description	健康状态通俗描 述	健康状态的通俗描述	S1	AN..200	
	Mean	均值	权重数据的均值	N	N11, 9	0-1
	Lower	可信区间下限	权重数据的 95%可信区间 下限	N	N11, 9	0-1, <=mean
	Upper	可信区间上限	权重数据的 95%可信区间 上限	N	N11, 9	0-1, >=mean
致死性疾病 分类数据 (Death Causes Mapped to ICD Codes)	Level 1 Cause	疾病名称-第一 级	疾病分类的第一层类别名 称	S1	AN..200	
	Level 1 Cause code	疾病编码-第一 级	疾病分类的第一层类别编 码	S1	AN..5	ICD-10
	Level 2 Cause	疾病名称-第二 级	疾病分类的第二层类别名 称	S1	AN..200	
	Level 2 Cause code	疾病编码-第二 级	疾病分类的第二层类别编 码	S1	AN..5	ICD-10
	Level 3 Cause	疾病名称-第三 级	疾病分类的第三层类别名 称	S1	AN..200	
	Level 3 Cause code	疾病编码-第三 级	疾病分类的第三层类别编 码	S1	AN..5	ICD-10
	Level 4 Cause	疾病名称-第四 级	疾病分类的第四层类别名 称	S1	AN..200	
	Level 4 Cause code	疾病编码-第四 级	疾病分类的第四层类别编 码	S1	AN..5	ICD-10
	CCD	中国疾病分类	中国疾病分类编码	S3	N..3	
	ICD10	国际疾病分类 ICD-10	国际疾病分类代码第十版	S3	AN..5	ICD-10
	ICD9	国际疾病分类 ICD-9	国际疾病分类代码第九版	S3	N..5	ICD-9

表 A.1 健康期望寿命知识库表结构（续）

数据表名称	字段名	数据源名称	定义	字段类型	表示格式	数据源允许值
疾病严重程度分布 (Disease severity distribution)	cause_id	疾病编码	疾病编码	S3	AN..5	ICD-10
	cause_name	疾病名称	疾病名称	S1	AN..200	
	sequela_id_num	后遗症编码	后遗症编码	S3	AN..5	ICD-10
	sequela_name	后遗症名称	后遗症名称	S1	AN..200	
	age_group_name	年龄组	年龄组的类别代码	S3	A2	年龄组代码表
	sex	性别	生理性别的代码	S3	N1	GB/T 2261.1
	severity	严重程度分布	严重程度分布	N	N1,9	
	nonfatal_cause_name	非致死疾病名称	非致死疾病名称	S1	AN..200	
	code_system_id	疾病编码类别	使用疾病编码的类别代码	S2	N1	1. ICD9 2. ICD10
疾病危险因素 (Relative risks)	nfc_duration	疾病病程	疾病病程（天）	N	N..5	
	Rish factors	危险因素名称	暴露的危险因素名称	S1	AN..200	
	Category / Units	暴露剂量	暴露危险因素的水平	S1	AN..200	
	Outcome	对应疾病名称	危险因素暴露对应的疾病名称	S1	AN..200	
	Morbidity / Mortality	疾病结局	危险因素暴露对应疾病的衡量指标	S2	N1	0. 两者 1. 发病 2. 死亡
	Sex	性别	疾病患者的性别	S2	N1	0. 所有人 1. 仅男性 2. 仅女性
	Age_group_name	年龄组	暴露于危险因素的发病/死亡人群的年龄	S3	A2	年龄组代码表