ICS  35.240.80

点击此处添加中国标准文献分类号



中华人民共和国国家标准

GB/T XXXXX—XXXX

|  |
| --- |
|       |

健康信息学 中医药数据集分类与代码

Health Informatics — Traditional Chinese medicine dataset classifying and coding

点击此处添加与国际标准一致性程度的标识

|  |
| --- |
| (征求意见稿) |
|  |

XXXX - XX - XX发布

XXXX - XX - XX实施



目  次

[前  言 II](#_Toc503538087)

[引  言 III](#_Toc503538088)

[1 范围 1](#_Toc503538089)

[2 规范性引用文件 1](#_Toc503538090)

[3 术语和定义 1](#_Toc503538091)

[4 编制原则 2](#_Toc503538095)

[5 分类编码设计 2](#_Toc503538101)

[6 分类与代码 3](#_Toc503538102)

[7 分类与编码的应用. 7](#_Toc503538108)

[参考文献 9](#_Toc503538112)

前  言

本标准按照GB/T 1.1－2009给出的规则起草。

本标准由中国中医科学院中医药信息研究所提出。

本标准由中国标准化研究院归口。

本标准起草单位：中国中医科学院中医药信息研究所、北京英富森软件股份有限公司、中国标准化研究院等。

本标准的主要起草人：XXX，XXX，XXX，XXX，XXX，XXX。

引  言

随着中医药事业的发展，中医药数据资源日益丰富，为更好的发挥中医药数据资源的作用，需要制定有关标准来规范中医药数据资源的管理。数据集是数据资源经人工有序化的结果，也是信息化工作中数据资源存贮的基本形式。中医药数据集分类与编码是实现数据资源管理工作现代化的必要条件，在信息检索、共享、交换中起着至关重要的作用。依据本分类与编码标准将具有某种共同特征的中医药数据集归并在一起，不具有共性的中医药数据集区分开来，并通过设定的代码体系进行编码，使计算机系统或人工能够识别和处理，用以保证中医药数据资源得到有效的管理，并能按类别开发数据集，支持高效率的查询服务。

本标准是在充分参考GB/T 7027—2002《信息分类和编码的基本原则与方法》和GB/T 20001.3—2015《标准编写规则 第3部分：分类编码》的基础上，结合中医药数据集的特点编制的。

本标准中所有类目与代码排列次序及其级别仅表示在本分类体系中的位置，不表示其他含意，与重要程度无关。

健康信息学 中医药数据集分类与代码

1. 范围

本标准规定了中医药数据集的分类与代码。

本标准的分类对象是中医药数据集，不同于学科、专业和行业分类，不能代替文献、情报、图书分类及学术上的各种观点。

本标准适用于中医药领域数据集调查与规划，数据集的组织与描述、整合、汇交、发布和目录查询，以及中医药信息系统的规划与设计。

1. 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅所注日期的版本适用于本文件。凡是不注日期的引用文件，其最新版本（ 所有的修改单）适用于本文件。

GB/T 7027—2002 信息分类和编码的基本原则与方法

GB/T 10113 分类与编码通用术语

1. 术语和定义

下列术语和定义适用于本标准。

* 1.

数据集 data set

具有一定主题，可以标识并能被计算机系统处理的数据集合。[SOURCE: ISO 19115-1:2014, 4.3]

* 1.

中医药 Traditional Chinese medicine

起源于中国的传统医学，以整体论和辨证论治为其特点。[SOURCE: ISO/TS 17938:2014, 2.1]

* 1.

分类 classification

根据事物不同的特性将其分配到一定组织结构中的过程。[SOURCE: ISO 15531-31:2004，3.5.3]

分类代码 classification code

分配给产品的有含义的代码，用于定义数据集中包含何种数据。[SOURCE: ISO 16792:2015, 3.2.1]

面 facet

同类概念的分组。[SOURCE: ISO 25964-1:2011, 2.20]

1. 编制原则
	1. 科学性原则

自上向下，优先选择最能代表中医药数据集主题的语言、词条定义类目名称，编制数据集分类体系表。

* 1. 系统性原则

综合考虑中医药数据集主题一致性，按其内在联系进行系统化排列，确保类目唯一、结构合理、层次清晰，减少冗余。

* 1. 实用性原则

满足中医药数据集分类编目的简便性、可操作性和通用性需求和数据集查询的一致性理解。

* 1. 稳定性原则

使用稳定的因素作为分类依据，同时提高分类体系的可延展性或兼容性，促进稳定性。

* 1. 可扩展性原则

保证分类体系框架适应中医药数据集不断丰富的内容和日益增长的种类与数量。

1. 分类编码
	1. 分类编码方法
		1. 线分类法

线分类法是将分类对象按选定的若干属性（或特征），逐次地分为若干层级，每个层级又分为若干类目。同一分支的同层类目之间构成并列关系，不同层级类目之间构成隶属关系。

* + 1. 面分类法

选定分类对象的若干属性（或特征），将分类对象按每一属性（或特征）划分成一组独立的类目，每一组类目构成一个“面”。再按一定顺序将各个“面”平行排列。使用时根据需要将有关“面”中的相应类目按“面”的制定排列顺序组配在一起，形成一个新的符合类目。

* 1. 分类编码结构

本标准采用面分类法和线分类法相结合的方法，根据主题进行主要分类，同时在若干相同主题数据集中，又根据其特性，如创建者类型、内容来源等多属性进行分类。

数据集分类编码按中医药领域代码、创建者类型代码、数据来源类型代码、主题类型代码等4部分组成，每部分之间用符号“·”分隔（图1）。中医药数据集类目编码取值范围为阿拉伯数字和英文字母（不区分大小写），且均为半角字符。



图1 中医药数据集类目编码结构

1. 分类与代码
	1. 中医药领域代码

用3位大写英文字母表示。中医药领域统一用TCM表示。

* 1. 创建者类型代码

按照中医药领域业务范围，本标准将数据集创建机构分为教育机构、研究机构、医疗机构、商业机构、政府机构、政府间组织、学会或协会、出版机构、个人及其他10个类型，并为每一类型机构赋予三位字母代码。具体见表1.

表1 数据集创建者类型代码表

|  |  |  |
| --- | --- | --- |
| **代码** | **创建机构类型** | **说明** |
| EDU | 教育机构 | 教育者有[计划](https://baike.baidu.com/item/%E8%AE%A1%E5%88%92)、有[组织](https://baike.baidu.com/item/%E7%BB%84%E7%BB%87/5105529)地对受教育者进行[系统](https://baike.baidu.com/item/%E7%B3%BB%E7%BB%9F)的[教育](https://baike.baidu.com/item/%E6%95%99%E8%82%B2)活动的[组织机构](https://baike.baidu.com/item/%E7%BB%84%E7%BB%87%E6%9C%BA%E6%9E%84) |
| ACA | 研究机构 | 从事中医药领域研究与开发活动的机构 |
| MED | 医疗机构 | 依法定程序设立的从事疾病诊断、治疗活动的卫生机构 |
| COM | 商业机构 | 以生产或买卖商品而获利的机构 |
| GOV | 政府机构 | 国家进行治理和社会管理的机构 |
| INT | 国际组织 | 两个以上国家或其政府、人民、民间团体基于特定目的，以一定协议形式而建立的各种机构。 |
| ASS | 学会/协会 | 由个人、单个组织为达到某种目标，通过签署协议，自愿组成的团体或组织。 |
| PUB | 出版机构 | 进行[图书](https://baike.baidu.com/item/%E5%9B%BE%E4%B9%A6/172039)、[图画](https://baike.baidu.com/item/%E5%9B%BE%E7%94%BB)、[杂志](https://baike.baidu.com/item/%E6%9D%82%E5%BF%97)，[报纸](https://baike.baidu.com/item/%E6%8A%A5%E7%BA%B8/745)和[电子](https://baike.baidu.com/item/%E7%94%B5%E5%AD%90/143051)物品等有版权物品的出版活动的组织。 |
| IND | 个人 | 指个体从业者 |
| OTH | 其他 | 不属于上述类型的机构 |

* 1. 数据来源类型代码

根据数据的来源类型分类，采用两位阿拉伯数字表示。具体见表2。.

表2 数据来源类型代码表

|  |  |  |
| --- | --- | --- |
| **代码** | **数据来源类型名称** | **说明** |
| 01 | 历史文献 | 古代文献中和古代非印刷品中记录的信息。注：中医药领域的古代文献指1911年前的 |
| 02 | 学术出版物 | 印刷品（书籍和期刊）、电子资源（书籍和期刊的电子版本）、以及网站记录上的信息。如标准、技术报告、专利、学术论文、论文、论著、字典等。 |
| 03 | 原始记录 | 在实验、手工记录、录音或任何其他研究行为中的信息源的原始记录。如实验数据、观测数据、病历记录、普查记录、统计数据等。 |
| 09 | 其他 | 未包含在上述三类的内容来源 |

* 1. 主题类型代码

数据集的主题类型可分为三级类目，即一级类、二级类和三级类，共用六位数字代码来表示，其中前两位数字代码表示一级类：中医药事业管理、中医、中药、针灸、民族医药以及其他类，中间两位数字代码表示二级类，最后两位数字代码表示三级类，本标准具体列出了一级和二级主题类目名称和类目说明，三级类目可由用户视使用中的实际情况决定是否进行细分，如不需细分，则最后两位代码均用”00”表示。具体分类和代码见表3。

表3数据集主题类型代码表

|  |  |  |
| --- | --- | --- |
| **代码** | **类目名称** | **说明和举例** |
| **010000** | 中医药事业管理 |  中医药行政管理、机构管理、人员管理、财务管理、物资管理、项目和成果管理、对外交流合作等信息。 |
| 010100 | 行政管理 |  电子政务等信息，如：国家中医药管理局电子政务信息交换系统等。 |
| 010200 | 人力资源管理 |  中医药医疗、教育、产业过程中的人力资源信息，如：岗位设置管理数据库，人员信息管理数据库等。 |
| 010300 | 财务与成本核算管理 |  中医医疗、教育、产业机构财务与成本核算管理信息，如：预算管理数据库，结算管理数据库，财务管理数据库等。 |
| 010400 | 物资管理 |  中医医疗、教育、产业机构的医疗物资管理信息、后勤物资管理信息、企业物资管理信息等，如药品管理数据库、医用耗材管理数据库、被装被服管理数据库、办公物资管理数据库、房屋资产管理数据库等。 |
| 010500 | 项目和成果管理 |  中医药医疗、教学、科研过程汇总的各级各类的课题项目的申报、实施、经费、评审信息等；已通过验收及获得各级各类奖励的中医药项目信息， 研究论文、总结报告、专利等，如：中医药课题申报数据库、中医药在研课题数据库、中医药科研成果数据库、中医药专利数据库等。 |
| 010600 | 中医药对外交流与合作 |  中医药对外交流与合作信息，如：日本中药数据库、中医药国际合作数据库等。 |
| 010099 | 其他 | 未包含在上述6类中的其他中医管理信息。 |
| **020000** | 中医 |  包括中医基础理论、临床实践、养身保健和中西医结合。 |
|  020100 | 基础理论 |  用传统和现代方法研究中医基础理论的信息，包括中医发展过程中的文献研究和基于实验的研究等。如：中医流派数据库，中医病证动物模型数据库，中医医史文献数据库。 |
| 020200 | 临床实践 |  中医方式对疾病的诊断、治疗、预防和其他生理心理障碍的调养等信息，包括在此过程中采用的方法、技术、设备等信息。如：中医肾病诊疗数据库。 |
| 020300 | 养生保健 | 中医保持健康、预防疾病和延长寿命的信息。如：药膳数据库。 |
| 020400 | 中西医结合 | 中西医结合的信息。如：中西医结合治疗中风数据库。 |
| 030099 | 其他 | 未包含在上述4类中的其他中医临床实践相关信息。 |
| **030000** | 中药 |  与中药相关的中药资源、中药材、饮片、中成药、方剂等信息。  |
| 030100 | 中药材和生药 |  中药药用植物、动物和矿物资源以及他们的药用部位信息，包括种植、地理分布、鉴别、采集、初步处理以及涉及到药用植物学、生药学和安全评估的研究等。如：中草药物种基因和转录组数据集，中草药DNA条形码数据集，中草药图片数据库，中草药安全标准数据库 |
| 030200 | 中药药理和化学 |  研究中药药理和中药化学的相关信息，包括中药属性理论、中药药理实验和临床药理实验等。如：中国中药化学成分数据库、中药化学实验数据库。 |
| 030300 | 制药和制备 | 中药材到生药和饮片的物理、化学变化过程中用到的工艺等信息。如：饮片规范数据库。 |
| 030400 | 临床用药 | 在中医理论指导下的药物处方和中药加工方法等信息，包括可直接用于临床实践的中药成品，包括饮片、方剂和中成药等。如：中药古代方剂数据库，中成药副作用数据库。 |
| 030500 | 中药产品贸易 | 中药在市场上进行交易及其过程中的信息。如：中药进出口数据库。 |
| 040099 | 其他 | 未包含在上述9类中的其他中药相关信息。  |
| **040000** | 针灸 | 针灸和经络的相关信息。 |
| 040100 | 经络和腧穴 |  穴位的解剖位置、生理功能；经络的巡行位置、生理功能、经络理论研究信息。如：腧穴解剖标志数据集。 |
| 040200 | 现代机理研究 | 针灸的功能原理和科学内涵的理论研究，主要指使用动物实验进行的研究。如：腧穴文献数据集，针灸有效机制数据集。 |
| 040300 | 针灸器械 | 通过刺激经络和腧穴的来诊断和治疗中使用的器械信息。如：针灸针数据库。 |
| 040400 | 临床应用 |  针灸处方、针刺麻醉、针灸疗法应用于临床和实验的信息，如，针灸处方数据库。 |
| 040099 | 其他 | 未包含在上述4类中的针灸信息。 |
| **050000** | 民族医药 |  我国少数民族医药的信息。 |
| 050100 | 藏族医药 |  藏族医药的信息，如：藏医药数据库。 |
| 050200 | 蒙古族医药 |  蒙古族医药的信息，如：蒙古医药数据库。 |
| 050300 | 维吾尔族医药 |  维吾尔族医药的信息，如：维吾尔医药数据库。 |
| 050400 | 朝鲜族医药 |  朝鲜族医药的信息，如：朝鲜医药数据库。 |
| 050500 | 壮族医药 |  壮族医药的信息，如：壮医药数据库。 |
| 050600 | 回族医药 |  回族医药的信息，如：回医药数据库。 |
| 050700 | 傣族医药 |  傣族医药的信息，如：傣医药数据库。 |
| 050800 | 苗族医药 |  苗族医药的信息，如：苗医药数据库。 |
| 050099 | 其他 |  除去藏、蒙、维、朝、壮、回、傣、苗，其他少数民族医药的信息。 |
| **099000** | 其他 | 未包含在上述5大类中的中医药相关信息。 |

1. 多主题分类编码的应用.

为了满足中医药领域不同业务需要，一个数据集往往包含多个主题或同时涉及不同主题内容，具体分类时应参照以下原则：

* 1. B主题在A主题中应用：

示例1：用方剂和针灸治疗肾病综合征的临床病例报告数据集

这个数据集可以被分配为多分类编码：

TCM.MED.03.020200

TCM.MED.03.040400

* 1. 部分和整体关系：

示例1：中药质量控制规范数据集

这个数据集可以被分配为多分类编码：

TCM.COM.02.030000

TCM.COM.02.030100

TCM.COM.02.030300

* 1. A主题影响B主题：

示例1：针灸有效机制动物实验的文献数据集

这个数据集可以被分配为多分类编码：

TCM.ACA.02.040100

TCM.ACA.02.040200

1. （资料性附录）
用例图及用例分析

# 用例图



# 重点用例分析

## 数据加工/著录

|  |
| --- |
| 用例名称：数据加工/著录 |
| 描述：数据集制作单位使用用例完成元数据库管理、元数据标引（含数据集分类标引）、元数据字段维护等数据加工与著录业务。 |
| 标识符:uc1 |
| 角色: 数据制作单位业务员（datasets creator） |
| 前置条件：数据制作单位业务员已成功登录系统并具有数据集字段维护、数据集分类标引、元数据库管理、元数据标引的权限 |
| 主事件流:1. 数据制作单位业务人员选择登录系统选项，用例开始2. 数据制作单位业务人员输入账号，系统根据规则检查账号的有效性 A1：数据制作单位业务人员账号无效3. 数据制作单位业务人员输入密码，检查密码是否正确A2：密码错误4．显示登录成功提示信息5. 数据制作单位业务人员对元数据库进行管理，系统根据规则判断是否新建元数据库 A3：选择已有元数据库6．新建元数据库：数据制作单位业务人员填写元数据库的基本信息7．数据制作单位业务人员选择元数据库，对元数据库的字段进行管理和维护8. 元数据加工标引9. 完成数据集分类标引，自定义选择相应的中医药领域代码、创建者类型代码、数据来源类型代码、主题类型代码10. 元数据库同步11. 用例结束 |
| 其他事件流：A1: 数据制作单位业务人员账号无效 (1).账号无效的提示信息 (2).返回主事件流第2步A2:密码错误 (1). 系统显示密码错误的提示信息 (2). 返回主事件流第3步A3：选择已有元数据库（1）.系统显示元数据库的具体信息 （2）.跳至主事件流第7步 |
| 后置条件：系统成功将加工著录的元数据库更新至元数据仓储数据库中 |
| 特殊需求：无 |

## 元数据收割

|  |
| --- |
| 用例名称：元数据收割 |
| 描述：数据中心对数据集制作单位提交的元数据进行收割和同步 |
| 标识符:uc2 |
| 角色: 数据中心业务人员（data center） |
| 前置条件：数据中心业务员已成功登录系统并具有数据管理和发布权限；数据集制作单位已提交数据； |
| 主事件流:1. 中心业务人员选择登录系统选项，用例开始2. 中心业务人员输入账号，系统根据规则检查账号的有效性 A1：业务人员账号无效3. 中心业务人员输入密码，检查密码是否正确A2：密码错误4. 显示登录成功提示信息5. 数据中心业务人员选择数据收割/同步，用例开始6. 选择要读取的源数据库位置7. 选择要写入的目标数据库位置8. 确认同步9. 用例结束 |
| 其他事件流：A1: 业务人员账号无效 (1).系统显示账号无效的提示信息 (2).返回主事件流第2步A2:密码错误 (1). 系统显示密码错误的提示信息 (2). 返回主事件流第3步 |
| 后置条件：无 |
| 特殊需求：无 |

## 索引与发布

|  |
| --- |
| 用例名称：索引与发布用例 |
| 描述：数据中心对收割到的数据集制作单位数据建立索引、并发布到系统前端指定位置。 |
| 标识符:uc3 |
| 角色: 数据中心业务人员（data center） |
| 前置条件：数据中心业务员已成功登录系统并具有数据管理和发布权限 |
| 主事件流:1. 中心业务人员选择登录系统选项，用例开始2. 中心业务人员输入账号，系统根据规则检查账号的有效性 A1：工作人员账号无效1. 中心业务人员输入密码，检查密码是否正确

A2：密码错误4. 显示登录成功提示信息5. 中心业务人员点击更新索引A3：更新索引失败6. 显示索引已完成7. 中心业务人员点击数据发布 8. 选择并指定数据发布的前端位置8. 显示数据发布已完成9. 用例结束 |
| 其他事件流：A1：工作人员无效 (1).系统工作人员无效的提示信息 (2).返回主事件流第2步A2：密码错误 (1). 系统显示密码错误的提示信息 (2). 返回主事件流第3步A3：更新索引失败 (1). 系统显示索引更新失败的提示信息 (2). 返回主事件流第5步 |
| 后置条件：系统成功更新索引，索引文件存入到元数据仓储中。 |
| 特殊需求：无 |

## 分类检索

|  |
| --- |
| 用例名称：分类检索 |
| 描述：数据用户对发布的数据，根据数据集分类进行分类检索。 |
| 标识符：uc4 |
| 角色: 数据用户（datasets user） |
| 前置条件：数据用户已成功登录系统并具有访问权限； |
| 主事件流:1. 用户选择检索功能，用例开始

A1：未完成前置条件2. 用户选择检索条件，输入检索关键字3. 系统根据输入信息查询文件A2：检索失败4. 检索成功5．将检索结果进行统一汇聚与展示6. 用例结束 |
| 其他事件流：A1：未完成前置条件 用户登录认证A2：检索失败返回检索页面 |
| 后置条件：系统成功将已检索的结果返回 |
| 特殊需求：无 |

参 考 文 献

[1]中国图书馆分类法编辑委员会.中国图书馆分类法（第四版），北京:北京图书馆出版社，1999.

[2]吴兰成.中国中医药学主题词表，北京:中医古籍出版社，北京2008.

[3] Medical Subject Headings (National Library of Medicine， U.S.A.

[4] 北京协和医院世界卫生组织疾病分类合作中心.疾病和有关健康问题的国际统计分类第十次修订本（ICD-10），北京:人民卫生出版社，1996.

[5] 王丙义.信息分类与编码，北京:国防工业出版社，2003.

[6] 白殿一，等.标准的编写，北京:中国标准出版社，2009.

[7] WS/T 306—2009卫生信息数据集分类与编码规则.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_