

中华人民共和国卫生行业标准

WS/T 370—2022
代替 WS 370-2012

卫生健康信息基本数据集编制标准

Standard for drafting of health information basic dataset

2022 - 10 - 25 发布

2023 - 04 - 01 实施

前 言

本标准代替WS 370—2012《卫生信息基本数据集编制规范》。与WS 370—2012相比，除编辑性修改外，主要技术变化如下：

- 增加了第4章缩略语；
- 修改了第5章内容结构，卫生健康信息基本数据集至少包含的内容由原来的7部分增加为9部分，增加的部分是术语和定义、索引；
- 修改了6.3 b) 数据集类目编码（DCC），二级类目代码由原来的2位数字修改为1位数字；
- 修改了7.3.1 内部标识符，增加子集顺序号，用3位数字表示，代表数据子集在特定数据集中的序号，从001开始顺序编码。子集顺序号与数据集分类编码之间加“.”区分；
- 修改了第8章数据元索引和表示方法，细化了索引项、编排前后顺序的相关内容；
- 修改了附录A，对卫生健康信息基本数据集分类代码的类目名称和代码进行了修改。

本标准由国家卫生健康标准委员会卫生健康信息标准专业委员会负责技术审查和技术咨询，由国家卫生健康委统计信息中心负责协调性和格式审查，由国家卫生健康委规划发展与信息化司负责业务管理、法规司负责统筹管理。

本标准起草单位：国家卫生健康委统计信息中心、华中科技大学、解放军总医院、空军军医大学、上海市疾病预防控制中心。

本标准主要起草人：胡建平、李岳峰、沈丽宁、刘丽华、张黎黎、王霞、夏天、董方杰、张诚。

卫生健康信息基本数据集编制标准

1 范围

本标准规定了卫生健康信息基本数据集的内容结构、数据集元数据、数据元属性、数据元索引和表示方法。

本标准适用于指导卫生健康信息相关数据集的编制与使用。

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标准必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本标准；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本标准。

WS/T 303 卫生信息数据元标准化规则

WS/T 305 卫生信息数据集元数据规范

WS/T 306 卫生信息数据集分类与编码规则

WS 363.1 卫生信息数据元目录 第1部分：总则

WS 364.1 卫生信息数据元值域代码 第1部分：总则

3 术语和定义

WS/T 303、WS/T 305、WS/T 306、WS 363.1、WS 364.1中界定的以及下列术语和定义适用于本标准。

3.1

基本数据集 basic dataset

在特定主题下，为了满足业务信息系统规范化建设和领域内部以及领域间数据交换与共享需求，设计归纳的各个子系统（或者功能模块）所包含的最小数据元素的集合。

4 缩略语

下列缩略语适用于本标准。

DCC 数据集类目编码 (Dataset Classification Coding)

DE 数据元素 (Data Element)

HDS 卫生领域数据集 (Health Data Set)

5 内容结构

卫生健康信息基本数据集至少应由以下部分内容构成：

封面
前言
名称
1 范围
2 规范性引用文件
3 术语和定义
4 数据集元数据
5 数据元目录
5.1 数据元公用属性
5.2 数据元专用属性
索引

6 数据集元数据

6.1 数据集元数据选取

根据WS/T 305中规定的7个元数据子集，本标准选取了核心元数据中的8个“必选（M）”元数据元素，分为2个元数据子集，对卫生信息基本数据集进行描述。WS 305中的数据集核心元数据描述规范适用于本文件。

6.2 数据集描述格式

卫生健康信息基本数据集元数据采用表1的格式进行描述。

表 1 数据集元数据

元数据子集	元数据项	元数据值
标识信息子集	数据集名称	
	数据集标识符	
	数据集发布方—单位名称	
	关键词	
	数据集语种	
	数据集分类—类目名称	
内容信息子集	数据集摘要	
	数据集特征数据元	

6.3 数据集标识符编码规则

数据集标识符采用字母数字混合码，结构为：数据集类目编码（DCC）_版本标识符（VI）。其中：

- a) 版本标识符（VI）：结构由4部分组成，为“V”+“m..m”+“.”+“n..n”。其中，“m..m”和“n..n”为阿拉伯数字构成，在数学上应是具有意义的正整数。“m..m”表示主版本号，“n..n”表示次版本号。如果数据元更新前后可以进行有效的数据交换，则更新后主版本号不变，次版本号等于当前次版本号加1；如果数据元更新前后无法进行有效的数据交换，则更新后主版本号等于当前主版本号加1，次版本号归0。

- b) 数据集类目编码（DCC）：即数据集分类编码。采用长度 8 位的字母数字混合码，含 1 位“.”。按业务领域代码、一级类目代码、二级类目代码、序号从左向右顺序排列。结构见图 1。

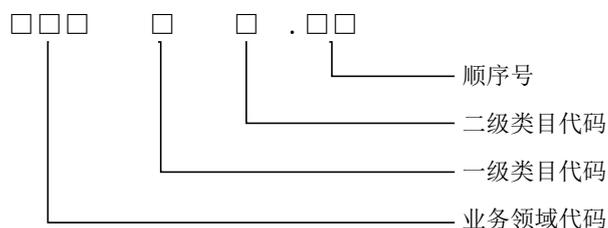


图 1 数据集类目编码（DCC）结构

图1中：

- 业务领域代码：用 3 位大写英文字母表示。卫生健康信息领域统一用 HDS 表示。
 - 一级类目代码：用 1 位大写英文字母表示，从 A 开始顺序编码。
 - 二级类目代码：用 1 位数字表示，数字大小无含义。如有两个以上类目（含两个），从 1 开始顺序编码。如无二级类目，编为 0。
 - 序号：用 2 位数字表示，代表二级类目下数据集的序号，数字大小无含义；从 01 开始顺序编码。序号与二级类目代码之间加“.”区分。
- 数据集分类代码详见附录A。

7 数据元属性

7.1 数据元属性选取

根据WS/T 303中规定的五类22项数据元基本属性中，本标准选取了五类14项数据元基本属性（见表 2），对卫生健康信息基本数据集的数据元进行描述。

表 2 数据元属性列表

序号	属性种类	数据元属性名称	约束
1	标识类	内部标识符	必选
2		数据元标识符	必选
3		数据元名称	必选
4		版本	必选
5		注册机构	必选
6		相关环境	必选
7	定义类	定义	必选
8	关系类	分类模式	必选
9	表示类	数据类型	必选
10		表示格式	必选
11		数据元允许值	必选
12	管理类	主管机构	必选
13		注册状态	必选
14		提交机构	必选

7.2 数据元描述格式

卫生健康信息基本数据集的数据元描述格式采用摘要式。

在同一个数据集中，版本、注册机构、相关环境、分类模式、主管机构、注册状态、提交机构等7个数据元属性的值若完全相同（可称为数据元公用属性），则采用表3的描述格式集中描述，在数据元摘要式目录中可不再分述。

表 3 数据元公用属性描述格式

属性种类	数据元属性名称	属性值
标识类	版 本	
	注册机构	
	相关环境	
关系类	分类模式	
管理类	主管机构	
	注册状态	
	提交机构	

在同一个数据集中，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7个数据元属性的值均不相同（可称为数据元专用属性），采用表4的摘要式目录描述格式进行描述。

表 4 数据元专用属性描述格式

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

7.3 数据元属性描述规则

7.3.1 内部标识符

内部标识符指数据元在某特定数据集中的唯一标识代码。采用长度16位的字母数字混合码，含3位“.”。结构见图2。

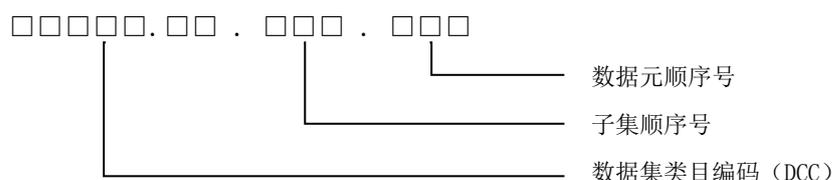


图 2 内部标识符结构

图2中：

- a) 数据集类目编码（DCC）：用 8 位字母数字混合码表示，编码规则见 6.3 b)。
- b) 子集顺序号：用 3 位数字表示，数字大小无含义；从 001 开始顺序编码。如无子集，编为 000。子集顺序号与数据集类目编码之间加“.”区分；
- c) 数据元顺序号：用 3 位数字表示，代表数据元在数据集子集中的序号，数字大小无含义；从 001 开始顺序编码。数据元顺序号与子集顺序号之间加“.”区分。

7.3.2 其他属性

WS 303、WS 363.1、WS 364.1中对数据元属性已做出的规范适用于本文件。

8 数据元索引和表示方法

数据元索引和表示方法如下：

- 以数据集标准中的数据元名称作为索引项，同时给出数据元名称对应的内部标识符；
- 以数据元名称首字的汉语拼音字母顺序编排，为了便于检索在数据元名称首字的汉语拼音首字母相同的索引项之上标出相应的字母；
- 数据元名称为外文字母组成的，其前后顺序按照汉字、拉丁字母、希腊字母、阿拉伯数字的顺序编排。

附录 A

(规范性)

卫生健康信息基本数据集分类代码

卫生健康信息基本数据集分类代码见表A.1

表 A.1 卫生健康信息基本数据集分类代码

一级类目名称	一级类目代码	二级类目名称	二级类目代码
基本信息	A		
公共卫生	B	管理	1
		服务	2
医疗服务	C	管理	1
		服务	2
医疗保障	D	管理	1
		服务	2
药品供应	E	管理	1
		服务	2
卫生综合	F	管理	1
		服务	2
其他	G		