

ICS 11.020

C07

备案号:25592—2009

WS

中华人民共和国卫生行业标准

WS/T 303—2009

卫生信息数据元标准化规则

Rules for data element standardization of health information

2009-01-22 发布

2009-08-01 实施



中华人民共和国卫生部 发布

目 次

前言	III
1 范围	1
2 规范性引用文件	1
3 术语和缩略语	1
3.1 术语	1
3.2 缩略语	4
4 数据元概述	4
4.1 数据元基本模型	4
4.2 值域基本模型	5
4.3 数据元的元数据总体模型	7
4.4 数据元与其他数据概念的关系	7
4.5 卫生信息数据元	8
5 数据元属性	8
5.1 数据元基本属性	8
5.2 数据元附加类属性	10
5.3 数据元属性描述符	10
5.4 数据元属性规范	11
6 卫生信息数据元的命名	14
6.1 命名约定规则	14
6.2 命名约定的实例	15
7 卫生信息数据元的定义	16
7.1 定义规则	16
7.2 定义指南	16
8 卫生信息数据元的分类	17
8.1 关键字	17
8.2 主题词表术语	17
8.3 分类法和本体论分类单元	17
9 卫生信息数据元内容标准编写格式规范	17
9.1 数据元内容标准的格式规范	17
9.2 数据元目录的格式规范	18
参考文献	20

前 言

本标准参考 ISO/IEC 11179:2004《信息技术 元数据注册系统》、GB/T 18391—2001《信息技术 数据元的规范与标准化》和 EN ISO 3166—1:2006《表示国家及其区域名称的代码 国家代码》的基础上,结合卫生信息领域数据元的特性进行了针对性扩展。

本标准由卫生部卫生信息标准专业委员会提出。

本标准由中华人民共和国卫生部批准。

本标准负责起草单位:中国人民解放军总医院。

本标准主要起草人:刘丽华、王才有、金水高、胡建平、王骏、胡凯、张黎黎、饶克勤。

卫生信息数据元标准化规则

1 范围

本标准规定了卫生信息数据元模型、属性、卫生信息数据元的命名、定义、分类以及卫生信息数据元内容标准编写格式规范。

本标准适用于下列活动：卫生信息数据元目录(数据元字典)的研究与制定、卫生信息数据元元数据注册系统的设计与开发、卫生信息标准的研究、教学与交流。

2 规范性引用文件

下列文件中的条款通过本标准的引用而成为本标准的条款。凡是注日期的引用文件，其随后所有的修改单(不包括勘误的内容)或修订版均不适用于本标准。然而，鼓励根据本标准达成协议的各方研究是否可使用这些文件的最新版本。凡是不注日期的引用文件，其最新版本适用于本标准。

GB/T 18391(所有部分) 信息技术 数据元的规范与标准化

3 术语和缩略语

下列术语和缩略语适用于本标准。

3.1 术语

3.1.1 概念 concept

由特征的一个独特组合所生成的知识单元。

[GB/T 15237.1—2000, 3.2.1]

3.1.2 一般概念 general concept

对应两个或两个以上对象的概念，由于共同的特性而形成一组。

[GB/T 15237.1—2000, 3.2.3]

3.1.3 个别概念 individual concept

仅对应一个对象的概念。

[GB/T 15237.1—2000, 3.2.2]

3.1.4 数据元 data element DE

用一组属性规定其定义、标识、表示和允许值的数据单元。

[GB/T 18391.1—2002, 3.14]

3.1.5 数据元概念 data element concept DEC

能以一个数据元的形式表示的概念，其表述与任何特定表示法无关。

[ISO/IEC 11179—1:2004, 3.3.9]

3.1.6 数据元名称 data element name

用于标识数据元的主要手段，由一个或多个词构成的命名。

[GB/T 18391.1—2002, 3.18]

3.1.7 对象类 object class

可以对其界限和含义进行明确的标识，且特性和行为遵循相同规则的观念、抽象概念或现实世界中事务的集合。

[ISO/IEC 11179—1:2004, 3.3.22]

3.1.8 对象类术语 object class term

数据元名称的一个成分,在逻辑数据模型中描述的数据元所属的逻辑数据组,如:“雇员”。

[GB/T 18391.5—2001,3.10]

3.1.9 特性 **property**

一个对象类所有成员所共有的特征。

[ISO/IEC 11179—1:2004,3.3.29]

3.1.10 特性术语 **property term**

数据元名称的一个成分,表示数据元所属的类别。

[GB/T 18391.5—2001,3.11]

3.1.11 限定术语 **qualifier term**

帮助定义和区分数据库中某个名称的一个或几个词。

[GB/T 18391.5—2001,3.12]

3.1.12 表示 **representation**

数据元的值域、数据类型的组合,必要时也包括计量单位或表示类。

[GB/T 18391.1—2002,3.60]

3.1.13 表示类 **representation class**

表示类型的分类。

[ISO/IEC 11179—1:2004,3.3.35]

3.1.14 表示术语 **representation term**

数据元有效值集合的形式。

[GB/T 18391.5—2001,3.15]

3.1.15 计量单位 **unit of measure**

用于计量相关值的实际单位。

[ISO/IEC 11179—1:2004,3.3.36]

3.1.16 值域 **value domain VD**

允许值的集合。

[ISO/IEC 11179—1:2004,3.3.38]

3.1.17 允许值 **permissible value**

在一个特定值域中允许的一个值含义的表达。

[ISO/IEC 11179—1:2004,3.3.28]

3.1.18 值含义 **value meaning**

一个值的含义或语义内容。

[ISO/IEC 11179—1:2004,3.3.39]

3.1.19 维度 **dimensionality**

无单位的计量表述。

注:一个数量是具有一个相关计量单位的值。32°F、0°C、1m 和 100cm 是数量。两个计量单位之间的等价由一个计量单位度量的值与另一个计量单位度量的值之间保持有一个一一对应关系的数量来决定,并独立于相关语境,其特征运算是相同的。等价的计量单位在此意义上具有相同的维度。这里定义的等价在所有计量单位集合中形成了一个等价关系。每类等价对应一个维度。“华氏温度”和“摄氏温度”的计量单位具有相同的维度,因为对于相同的数量,每个由华氏计量的值,同样存在一个由摄氏计量的值,反之亦然。对于每个计量单位的数量可以进行同样的运算。温度一一对应的数量等式有: $C^{\circ} = (5/9) * (F^{\circ} - 32)$ 和 $F^{\circ} = (9/5) * (C^{\circ}) + 32$ 。

[ISO/IEC 11179—1:2004,3.3.12]

3.1.20 概念域 **conceptual domain CD**

有效的值含义的集合。

- [ISO/IEC 11179—1:2004, 3. 3. 6]
3. 1. 21 **内涵 intension**
 <术语学>组成概念的特征的集合。
 [ISO/IEC 11179—1:2004, 3. 2. 15]
3. 1. 22 **外延 extension**
 <术语学>与一个概念对应的全体对象。
 [ISO/IEC 11179—1:2004, 3. 2. 12]
3. 1. 23 **元数据 metadata**
 定义和描述其他数据的数据。
 [ISO/IEC 11179—1:2004, 3. 2. 16]
3. 1. 24 **元数据项 metadata item**
 一个元数据对象的实例。
 [ISO/IEC 11179—1:2004, 3. 2. 17]
3. 1. 25 **实体 entity**
 任何现存、曾经存在的或可能存在的具体的或抽象的事物,包括事物间的联系。
 [ISO/IEC 11179—1:2004, 3. 2. 10]
3. 1. 26 **属性 attribute**
 某个对象或实体的特征。
 [ISO/IEC 11179—1:2004, 3. 1. 1]
3. 1. 27 **属性值 attribute value**
 某种属性的一个实例表示。
 [GB/T 18391. 3—2001, 3. 2]
3. 1. 28 **分类法 taxonomy**
 依据类及子类间的固有关系进行分类的方法。
 [GB/T 18391. 1—2002, 3. 71]
3. 1. 29 **主题词表 thesaurus**
 按给定顺序排列参照词汇,其中显示和标识了词汇间的关系。
 [GB/T 18391. 5—2001, 3. 20]
3. 1. 30 **注册机构标识符 registration authority identifier RAI**
 经授权对数据元进行注册的任何组织的标识符。
 [GB/T 18391. 5—2001, 3. 14]
3. 1. 31 **数据标识符 data identifier DI**
 由注册机构赋予数据元的标识符。
 [GB/T 18391. 5—2001, 3. 5]
3. 1. 32 **版本标识符 version identifier**
 赋予版本的一个标识符,以此提交和更新数据元注册。
 [GB/T 18391. 5—2001, 3. 21]
3. 1. 33 **管理状态 administrative status**
 注册机构处理注册请求中赋予管理过程的状态的指称。
 [ISO/IEC 11179—1:2004, 3. 3. 3]
3. 1. 34 **数据元字典 data element dictionary**
 列出并定义了所有相关数据元的一种信息资源。
 注:数据元字典可以有各种不同层级,例如:ISO/IEC 委员会级、国际协会级、行业部门级、公司级、应用系统级。

[GB/T 18391.3—2001,3.5]

3.2 缩略语

- CD (conceptual domain)概念域
 DE (data element)数据元
 DEC (data element concept)数据元概念
 VD (value domain)值域
 SO (submitting organization)提交机构
 RO (responsible organization)主管机构
 RA (registration authority)注册机构
 RAI (registration authority identifier)注册机构标识符
 DI (data identifier)数据标识符
 VI (version identifier)版本标识符
 ISO (international organization for standardization)国际标准化组织
 IEC (international electrotechnical commission)国际电工委员会

4 数据元概述

4.1 数据元基本模型

数据元的基本模型是由数据元概念和数据元两部分组成,见图1。

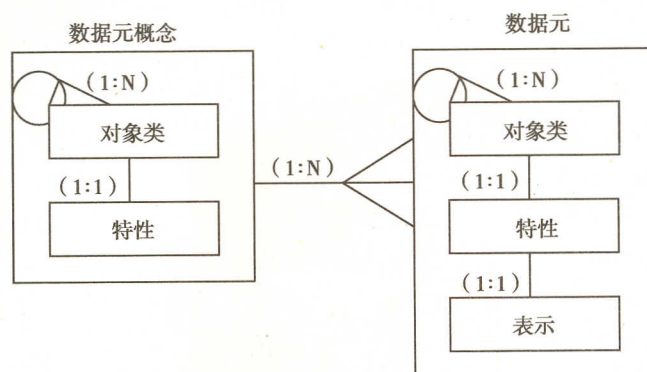


图1 数据元的基本模型

一个数据元概念是由对象类和特性两部分组成,是能以一个数据元形式表示的概念,其描述与任何特定表示法无关。一个数据元是由对象类、特性和表示三部分组成。一个数据元概念对应多个数据元。

对象类是可以对其界限和含义进行明确的标识,且特性和行为遵循相同规则的观念、抽象概念或现实世界中事物的集合。它是我们希望采集和存储数据的事物。对象类是概念,在面向对象的模型中与类相对应,在实体-关系模型中与实体对应,例如,患者、医生、卫生机构等。对象类可能是一般概念。当对象类所对应的对象集有两个或多个元素时,就是一般概念。患者、医生、卫生机构等都是是一般概念。对象类也可以是个别概念。当对象类对应的对象集仅有一个元素时,就是个别概念,例如“北京市医疗机构”就是个别概念。

特性是一个对象类的所有成员所共有的特征。它用来区别和描述对象,是对象类的特征,但不一定是本质特征,它们构成对象类的内涵。特性也是概念,对应于面向对象模型或实体-关系模型中的属性,例如身高、体重、血压、脉搏、血型等。特性也可是一般概念或个别概念。作为个别概念的例子有:病床总数或医疗收入。

表示可包括值域、数据类型、表示类(可选的)和计量单位四部分,其中任何一部分发生变化都成为不同的表示。值域是数据元允许值的集合,例如医疗毛收入这一数据元的值域是用非负实数集(以人民

币为单位)作为它的允许值集合;数据类型是表达数据元允许值的不同值的集合,以这些值的特性和运算为特征,例如患者姓名的数据类型是“字符”;表示类是表示类型的分类,它是可选的,例如性别代码这一值域的表示类是“类别”;计量单位是用于计量相关值的实际单位,例如患者血压测量的计量单位是“mmHg”。

当一个数据元概念与一个表示联系在一起时,就产生了一个数据元。在需要生成概念上相似的数据元时,一个数据元概念可以与不同的表示关联产生不同的数据元。同一概念的表达方法有许多。例如,患者国籍这个数据元概念,可以应用 EN ISO 3166—1 中规定的世界各国 7 种不同的表示,有英文全称、2 位字母码、3 位字母码、地区代码等,每种都包含了一个表示集合,都可以用作与该数据元概念关联的表示。7 种关联就形成了 7 个数据元。

4.2 值域基本模型

值域的基本模型由概念域和值域两部分组成,一个概念域对应多个值域,见图 2。

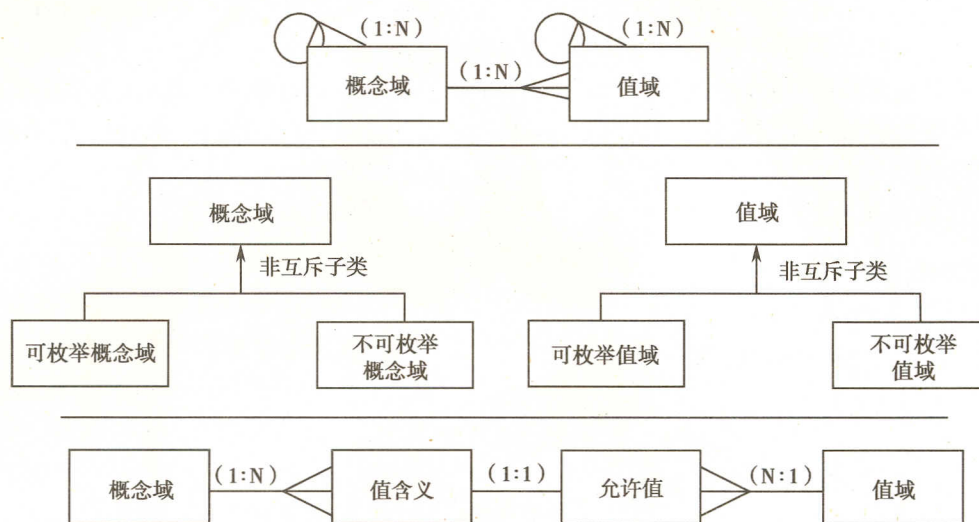


图 2 值域的基本模型

4.2.1 值域

值域是数据元允许值的集合。一个允许值是某个值和该值的含义的组合,值的含义称为值含义。例如,“患者病情状态”数据元的值域是:1 表示危,2 表示重,3 表示一般。1、2 和 3 是值,其值含义分别是危、重和一般。

值域有两种(非互斥的)子类:

- 可枚举值域:由允许值(值和它们的含义)列表规定的值域;
- 不可枚举值域:由描述规定的值域。

一个可枚举值域是包含了它的所有值及值含义的一个列表。例如,“患者血型类别代码”数据元的一个可枚举值域列表见表 1:

表 1 “患者血型类别代码”的值域

值	值含义
1	A 型
2	B 型
3	O 型
4	AB 型
9	不详

一个不可枚举值域是由一个描述来规定。不可枚举值域的描述须准确描述属于该值域的允许值。例如，“传染病死亡率”数据元的值域是大于等于 0 且小于等于 1 的实数。

描述数据有时需要计量单位。例如描述血压的值域记录时，为了准确理解值含义，需要使用计量单位 mmHg 或 kPa；年龄值域记录的计量单位可使用年、月、周、日；血糖值域记录的计量单位可以是 mg/dl 或 mmol/L。所以，计量单位与值域关联。

4.2.2 概念域

概念的外延构成了概念域，一个概念域是一个值含义集合。一个概念域的内涵是它的值含义。概念域也有两种(非互斥的)子类：

- 可枚举概念域：由值含义列表规定的概念域；
- 不可枚举概念域：由描述规定的概念域。

可枚举概念域的值含义可以明确地列举。该类型概念域对应于可枚举类型的值域，见示例 1。不可枚举概念域的值含义由“不可枚举概念域描述规则”来表述。该规则描述了不可枚举值域中允许值的含义。这种类型的概念域对应于不可枚举类型的值域，见示例 2。

如果一个计量单位的任何量可以转化为另一种计量单位下等量的量，则这些计量单位彼此之间是等价的。所有等价的计量单位被认为具有相同的维。例如，长度计量单位(米，厘米等)具有相同的维，年龄计量单位(年、月、周、日)具有相同的维。所以，维与概念域关联。

示例 1: 可枚举概念域

概念域名称: 国别代码
概念域定义: 世界各国名称的表示
值含义: 中国、美国、巴西、法国……
值域名称(1): 国家代码-3 位字母
允许值: <CHN, 中国>; <USA, 美国>; <BRA, 巴西>; <FRA, 法国>; ……
值域名称(2): 国家代码-2 位字母
允许值: <CN, 中国>; <US, 美国>; <BR, 巴西>; <FR, 法国>; ……
值域名称(3): 国家代码-3 位数字
允许值: <156, 中国>; <840, 美国>; <076, 巴西>; <250, 法国>; ……

示例 2: 不可枚举概念域

概念域名称: 体重
概念域定义: 身体所有器官重量的总和
概念域描述规则: 用非负实数表示
值域名称(1): 体重—N5,2
值域描述: 身体所有器官重量的总和, 最大长度 5 位的非负实数, 小数点后保留 2 位数字。
计量单位: 千克
值域名称(2): 体重—N4
值域描述: 身体所有器官重量的总和, 最大长度 4 位的非负整数。
计量单位: 克

4.2.3 关系

每个值域都是概念域的一个元素。多个值域可能是同一个概念域的外延,但一个值域只与一个概念域关联。概念域之间可以存在关系,由此创建概念域的一个概念体系。值域之间也可以存在关系,根据这些关系提供的框架,就能够捕捉相关值域和它们关联概念的结构。

每个值域表示两种概念:数据元概念(间接地)和概念域(直接地)。数据元概念是与一个数据元关联的概念。值域是数据元的表示,因此也间接地表达了数据元概念。但是,值域与一个概念域直接关联,因此,值域对概念的表达与数据元无关。

4.3 数据元的元数据总体模型

数据元的元数据总体模型由概念层和表示层两个部分组成,见图3。概念层包括数据元概念类和概念域类。这两种类都表示概念。表示层包括数据元类和值域类。这两种类都表示数据值的容器。

图3可以给出4个类的若干基本事实:

——一个数据元是一个数据元概念和一个值域的结合体。

——多个数据元可以共享相同数据元概念,这意味着一个数据元概念可以用多种不同方式表示。例如,患者入院时间是一个数据元概念,它可以有多种表示方法,例如用YYYYMMDD表示患者入院的年月日,如果需表示小时和分,则用YYYYMMDDhhmm表示。“患者性别代码-英文全称”、“患者性别代码-1位字母代码”共享同一个数据元概念患者性别。

——多个数据元可以共享相同的表示,这意味着一个值域可以被不同数据元重复利用。例如,“住院病人结算费用-元(人民币)”和“门急诊就诊费用-元(人民币)”数据元共享相同的值域“费用-元(人民币)”。

——值域不是必然与一个数据元关联,可以单独管理。

——不同值域所有允许值所对应的值含义都相同时,这些值域在概念上是等价的,因此,对应相同的概念域。例如,前述关于性别的值域:性别代码-1位数字、性别代码-英文全称及性别代码-1位字母代码的允许值所对应的值含义相同,均表示“男性、女性、未知的性别、未说明的性别”。所以这些值域对应同一个概念域“性别”。

——不同值域部分允许值所对应的值含义相同时,这些值域在概念上是相关的,因此,在包含有其各自概念域的概念体系中共享一个由共同的值含义构成的概念域。例如,性别代码-符号的允许值是“♂和♀”,分别表示“男性和女性”。无其他符号表示“未知的性别”和“未说明的性别”,那么性别代码-符号与性别代码-1位数字、性别代码-英文全称及性别代码-1位字母代码值域在概念上相关,共享由共同的值含义“男性和女性”构成的概念域“性别”。

——一个数据元概念仅与一个概念域相关,因此共享相同数据元概念的所有数据元共享概念上相关的表示。例如,“新生儿体重-N5,2(千克)”和“新生儿体重-N4(克)”是两个不同的数据元,共享一个数据元概念新生儿体重。两个数据元的表示在概念上是相关的,共享概念域“体重”。千克和克都是重量的计量单位,称具有相同的维度。因此计量单位与值域相关,维度与概念域相关。

——许多数据元概念可以共享相同的概念域。例如,患者体重和新生儿体重是两个不同的数据元概念,共享一个概念域“体重”。

4.4 数据元与其他数据概念的关系

图4简化表示了数据元所在的不同数据环境。数据元出现在数据库、文件和数据交换集当中。数据元是系统所管理数据的基本单元,因而它必然是其内部进行数据库和文件设计的组成部分,也一定是系统间数据交换集的组成部分。

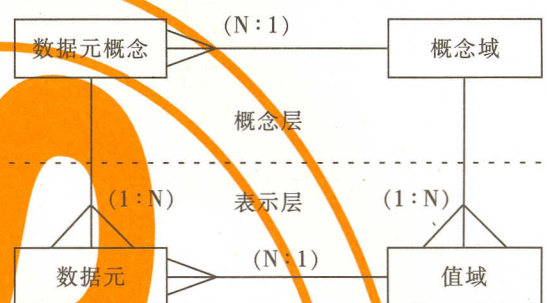


图3 数据元的元数据总体模型

在系统内部,数据库或文件由记录、段和元组等组成,而记录、段和元组则由数据元组成。数据元本身包含各种各样的数据,这些数据包括字符、图像和声音等。

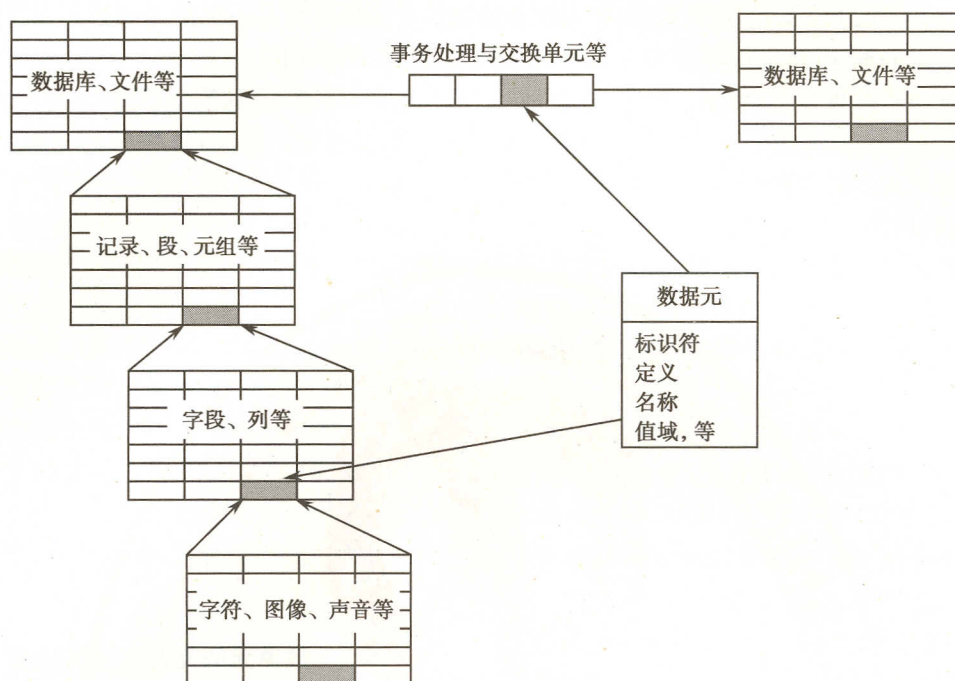


图4 数据元和其他数据概念

当系统间进行数据传输时,数据元是构成交换集的基本单元。数据交换主要发生于数据库间或文件间,但系统间的数据库和文件结构(如记录或元组)并不一定相同,所以进行信息交换的公用单元是数据元。

4.5 卫生信息数据元

卫生信息数据元是卫生这一特定领域的数据元,它的概念和结构遵循通用数据元的概念和结构,但具有自身的特点。医药卫生领域包括基础医学、临床医学、公共卫生、中医药学等多个专业,其数据元的表现形式复杂,例如数字人体、医学影像、基因图谱、中医经络等,除具有通用数据元的属性外,还具有其鲜明的领域特殊性。

5 数据元属性

5.1 数据元基本属性

数据元的基本属性模型见图5。一个数据元规范由一组属性组成,使用了基数型和逻辑相关性两种准则对数据元的基本属性进行分组,分在同一组的属性共同拥有相似的基数和逻辑相关性。

——基数型:每一个数据元规范都可能包含0或1(0:1)、1且仅仅是1(1:1)、0或多(0:n)、1或多(1:n)个列于图5中的属性。

例如:一个数据元规范可能包含0或1个“主管机构”属性,但要求有1且仅仅是1个“定义”属性;可能包含0或多对“相关数据参照”与“关系类型”属性,但要求有1或多个“数据元允许值”属性。

——逻辑相关性:属性除了有相似基数类型外,还可能彼此依赖,也就是说,某种属性在没有其他属性存在的情况下不可能存在。

例如:如果属性“同义名称”和“相关环境”两者有一个存在的话,那么它们两者就都应当存在。类似地,如果属性“相关数据参照”和“关系类型”两者有一个存在的话,那么它们两者就都应当存在。另一方面,即使属性“相关数据参照”和“同义名称”有相同的基数类型(0:n),它们也不能相互依赖而存在,从而它们不能分在同一组。

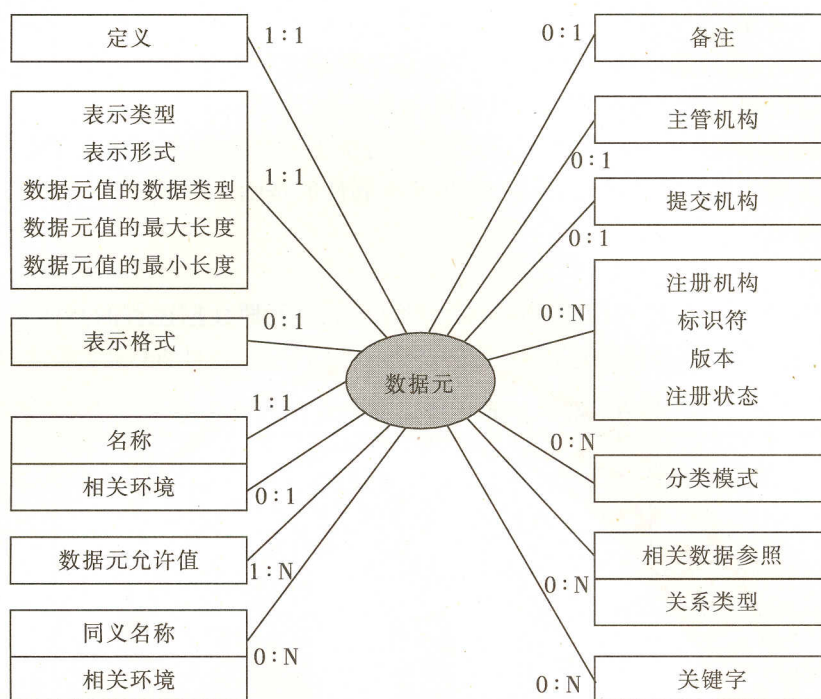


图5 数据元基本属性模型

表2 根据基本属性模型列出了五类 22 项基本属性。纵列“约束”是指在数据元字典中,该属性是“必选(M-mandatory)”,还是“条件选(C-conditional)”,或者是“可选(O-optertional)”。

表2 数据元属性

属性种类	数据元属性名称	约束
标识类	——名称	M
	——标识符	C
	——版本	C
	——注册机构	C
	——同义名称	O
	——相关环境	C
定义类	——定义	M
关系类	——分类模式	O
	——关键字(词)	O
	——相关数据参照	O
	——关系类型	C
表示类	——表示类别	M
	——表示形式	M
	——数据元值的数据类型	M
	——数据元值的最大长度	M
	——数据元值的最小长度	M
	——表示格式	C
	——数据元允许值	M
管理类	——主管机构	O
	——注册状态	C
	——提交机构	O
	——备注	O

5.2 数据元附加类属性

数据元附加类属性是对基本属性的扩展。卫生信息数据元是领域专用数据元,在遵照数据元基本属性的基础上,可以根据本领域数据元的特点和特殊需求进行扩展。

例如,对卫生统计指标数据元的属性进行设置时,可根据卫生统计指标数据元对正负向属性、环节、末端属性、收集方法属性等特殊需求,在五类 22 项基本属性的基础上进行扩展,作为附加类属性。

5.3 数据元属性描述符

5.3.1 属性描述符

属性描述符是对数据元属性进行标准化描述的条目,数据元属性描述符的具体条目和约束条件见表 3。纵列“约束条件”表示该属性描述符条目是“必选(M)”,还是“条件选(C)”,或者是“可选(O)”。

表 3 数据元属性描述符条目与约束条件

属性描述符条目	约束条件	属性描述符条目	约束条件
——名称	M	——数据类型	M
——定义	M	——最大长度	O
——约束	M	——字符集	C
——条件	C	——语言	C
——最多实例数	O	——备注	O

5.3.2 属性描述符条目应用规则

5.3.2.1 名称

“名称”属性描述符是赋予数据元属性的标记。“名称”应当是唯一的,并且应当以字母、汉字、数字式的字符串形式表示。

5.3.2.2 定义

“定义”属性描述符是对数据元属性描述符的概念解释,可使一种数据元属性与其他数据元属性清晰地区别开来。定义以字母数字式的字符串形式表示。

5.3.2.3 约束

“约束”属性描述符是给出一个数据元属性是始终还是有时出现(即含有的值)的限制条件。该描述符可以有如下含义:

- 必选:该数据元属性必须出现;
- 条件选:如果规定的条件存在的话,那么该数据元属性就应当出现;
- 可选:该数据元属性可以出现,也可以不出现。

5.3.2.4 条件

“条件”属性描述符是给出数据元属性应该出现的前提和要求。

5.3.2.5 最多实例数

“最多实例数”属性描述符是在一个数据元规范中规定数据元属性和属性值可出现的最多数目。

注:“最多实例数”可以指属性的重复出现的次数,或指属性出现一次但具有多个赋值(多值属性)。后一种情形需要句法约定,以使属性值相互区别。

5.3.2.6 数据类型

“数据类型”属性描述符是为表达属性值而规定的特定值集合的描述符。属性值的数据类型示例有:“字符”、“序号”、“整数”、“字符串”。

5.3.2.7 最大长度

“最大长度”属性描述符是存储单元最大数目的规格以表达在数据类型中所规定的特定值。

示例:当“数据类型”实例被规定为“整数”并且“最大长度”描述符实例值是“3”时,则表示属性值可以包含最多 3 位整数。

5.3.2.8 备注

“备注”属性描述符是与属性应用有关的注释。

5.4 数据元属性规范

数据元属性规范是依据属性描述符条目和约束条件对数据元的基本属性和附加类属性进行的描述。具体内容如下。

5.4.1 标识类属性

5.4.1.1 名称：名称

定义：赋予数据元的单个或多个字词的指称。

约束：必选

数据类型：字符串

5.4.1.2 名称：标识符

定义：在一个注册机构内与语言无关的一个数据元的唯一标识符。

约束：条件选

条件：如果属性“数据元名称”在一个注册机构内部不是唯一的，那么本属性就是必选的。唯一标识符的分配可作为任何注册机构的注册程序的必选部分。

数据类型：字符串

5.4.1.3 名称：版本

定义：注册机构内，一套数据元规范中的一个数据元发布的标识。

约束：条件选

条件：当对属性进行了更新，并且这种更新满足注册机构制定的新版本的维护规则时，则本属性就是必选的。

数据类型：字符串

5.4.1.4 名称：注册机构

定义：负责维护一个注册库的组织。

约束：条件选

条件：每一个出现的标识符都应指明一个注册机构。

数据类型：字符串

备注：组合属性“注册机构”、“标识符”和“版本”可用来唯一地确定一个数据元。一个数据元可以由多个注册机构来注册。

5.4.1.5 名称：同义名称

定义：与给定名称有区别、但表示相同的数据元概念的单字或多字的指称。

约束：可选

数据类型：字符串

备注：在确定的应用环境中，同义名称常常是熟悉的名称。如果是这种情况的话，要使用属性“相关环境”来指明这个环境。

假如有更多的同义名称存在，则属性“同义名称”与“相关环境”应当作为一对属性来规定。

5.4.1.6 名称：相关环境

定义：对使用或产生名称(或同义名称)的应用环境或应用规程的指明或描述。

约束：条件选

数据类型：字符串

条件：对于属性“同义名称”的每一次出现来说，本属性都是必选的。

当属性“名称”存在于一个信息交换过程中时，本属性是必选的。

- 备注： 为属性“名称”指定属性“相关环境”对于注册机构的注册程序来说是必选的。
- 5.4.2 定义类属性**
- 5.4.2.1 名称：** 定义
- 定义： 表达一个数据元的本质特性并使其区别于所有其他数据元的陈述。
- 约束： 必选
- 数据类型： 字符串
- 5.4.3 关系类属性**
- 5.4.3.1 名称：** 分类模式
- 定义： 根据对象的来源、组成、结构、应用、功能等共同特性，将对象排列或划分成组的模式的分类参照。
- 约束： 可选
- 数据类型： 字符串
- 备注： 分类模式可以用于使数据元与功能上的和(或)组织上的和(或)技术上的分类模式和(或)概念模型相关联。
- 5.4.3.2 名称：** 关键字
- 定义： 用于数据元检索的一个或多个有意义的字词。
- 约束： 可选
- 数据类型： 字符串
- 备注： 本属性可作为记录的关键字(搜索键)与所研究的数据元相关联。
- 5.4.3.3 名称：** 相关数据参照
- 定义： 数据元与相关数据之间的参照。
- 注：参照的数据可以注册在同一数据元的字典中，或者注册在其他的字典、字库中。
- 约束： 可选
- 数据类型： 字符串
- 备注： 1. 当相关数据被其他注册机构所控制时，本属性应当具有唯一的标识符，举例来说：注册机构加上其所分配的有关数据的标识符即构成唯一标识符。
2. 参见数据元的命名规则。
3. 一个数据管理环境应当为应用本属性设置一些规则和约定。
4. 如果本属性存在，则它应当与属性“关系类型”作为一对属性来规定。
- 5.4.3.4 名称：** 关系类型
- 定义： 数据元与相关数据之间关系特性的一种表达。
- 约束： 条件选
- 条件： 若属性“相关数据参照”存在，则本属性就是必选。
- 数据类型： 字符串
- 备注： 1. 本属性应当与属性“相关数据参照”作为一对属性来规定。
2. 关系类型的示例有：“……的限定符”“由……限定”“……的主体”“……的部分”“物理条件”“外部参照”“更高级的标准”“数据元概念”。
3. 一个数据管理环境应当为应用本属性设置一些规则和约定。
- 5.4.4 表示类属性**
- 5.4.4.1 名称：** 表示类别
- 定义： 用于表示数据元的符号、字符或其他表示的类型。
- 约束： 必选
- 数据类型： 字符串

- 备注：表示类别应当由相关的标准来规定。可以用作表示类别的示例有：
 ——字符表示法(ISO/IEC 646)
 ——字符/符号表示法(ISO 第 143 号注册出版物)
 ——条码表示法(EIA-556)
 ——图形表示法
- 5.4.4.2 名称：表示形式
 定义：数据元表示形式的名称或描述，例如：“数值”“代码”“文本”“图标”。
 约束：必选
 数据类型：字符串
 备注：1. 参见 GB/T 18391.2 的有关适用术语(“特性词”或“类别词”)。
 2. 示例 1:对于数据元“患者出生地代码”来说，本属性相当于“代码”。
 3. 示例 2:对于数据元“患者症状描述”来说，本属性相当于“文本”。
 4. 示例 3:对于数据元“患者体重-N5,2(千克)”来说，本属性相当于“数值”。
- 5.4.4.3 名称：数据元值的数据类型
 定义：表示数据元值的不同值的集合。
 约束：必选
 数据类型：字符串
 备注：示例：
 给出两种数据类型，①可能的实例有“字符”、“序数”、“整数”、“实数”、“比例数”、“二进制数”、“有理数”等。②给出扩展的数据类型，如采用 HL7 RIM 的数据类型。
- 5.4.4.4 名称：数据元值的最大长度
 定义：表示数据元值的(与数据类型相对应的)存储单元的最大数目。
 约束：必选
 数据类型：整数
 备注：1. 示例:对于数据元“身份证号码”来说，如果数据类型属性的实例为“字符”，并且数据元值的最大长度属性的实例值为“18”，那么“身份证号码”的数据元值最大不应超过 18 个字符。
 2. “数据元值的最大和最小长度”这两个属性表明数据元值是“不变的”(最大长度与最小长度相等)还是“可变的”(最大长度与最小长度不同)。
- 5.4.4.5 名称：数据元值的最小长度
 定义：表示数据元值的(与数据类型相对应的)存储单元的最小数目。
 约束：必选
 数据类型：整数
 备注：1. 示例:对于数据元“患者症状描述”来说，如果数据类型属性的实例为“字符”，并且数据元值的最小长度属性的实例值为“4”，那么“患者症状描述”的数据元值最小不应少于 4 个字符。
 2. “数据元值的最大和最小长度”这两个属性表明数据元值是“不变的”(最大长度与最小长度相等)还是“可变的”(最大长度与最小长度不同)。
- 5.4.4.6 名称：表示格式
 定义：用字符串表示数据元值的格式。
 约束：条件选
 条件：如果数据元属于“定量数据”类，那么本属性就是必选。如果表示形式属性是“代

码”，且代码表示需要有具体的结构或格式时，则建议使用本属性。

数据类型： 字符串

备注： 1. 对于定量数据来说，区别整数、十进制标记和浮点计数法是必要的。
2. 对于有具体结构或格式的代码表示来说，代码结构中每个位置的字符类型对于有效意义来说都是重要的。

5.4.4.7 名称： 数据元允许值

定义： 在一个特定值域中允许的一个值含义的表达。

约束： 必选

数据类型： 字符串

备注： 当数据元允许值是编码表示的枚举形式时，每一个数据元值及值含义都应当成对表示。

5.4.5 管理类属性

5.4.5.1 名称： 主管机构

定义： 提供数据元属性权威来源的组织或组织内部机构。

约束： 可选

数据类型： 字符串

备注： 该组织应当被视为数据元的“拥有者”。

5.4.5.2 名称： 注册状态

定义： 一个数据元在注册生命周期中状态的指称。

约束： 条件选

条件： 在注册机构所规定的数据元生存期内，本属性是必选。

数据类型： 字符

备注： 判别注册状态的类型和注册状态的确定应当遵循在数据元注册规程中所阐述的规则。

5.4.5.3 名称： 提交机构

定义： 提出数据元注册请求的组织或组织内部机构。

约束： 可选

数据类型： 字符串

5.4.5.4 名称： 备注

定义： 数据元的注释。

约束： 可选

数据类型： 字符串

5.4.6 附加类属性

卫生信息附加类属性的描述也须遵照数据元属性描述符条目应用规则进行规范。如卫生统计指标数据元扩展的“收集方法”属性，需要描述其定义、约束、数据类型等。

——名称： 收集方法

——定义： 简要阐述该数据元的收集途径。

——约束： 必选

——数据类型： 字符串

6 卫生信息数据元的命名

6.1 命名约定规则

6.1.1 语义规则

——对象类表示在医药卫生领域内有关的事物；

- 需要有一个且仅有一个对象类术语；
- 特性术语应当从特性体系结构设置中产生，并表示出数据的类别；
- 需要有一个且仅有一个特性术语；
- 限定应由专业领域机构决定产生，当需要描述一个数据元并使其在特定的相关环境中唯一时，即可增加限定术语，限定术语的顺序并不重要，限定术语是可选的；
- 表示术语是描述数据元有效值的集合；
- 需要有且仅有一个表示术语。

6.1.2 句法规则

- 对象类术语应处于名称的第一(最左)位置；
- 限定术语应位于被限定成分的前面，限定名称的顺序不应用于区别数据元名称；
- 特性术语应处于第二位置；
- 表示术语应处于最后位置。假如表示术语中有任何字与特性术语中的字重复，则删去冗余词。

6.1.3 针对数据元英文名称的词法规则

- 名词仅用单数形式；动词(若有的话)为现在时；
- 名称的各个成分间和多个单词术语之间用空格分隔，不允许用特殊字符；
- 名称中所有单词应组合在一起；
- 允许使用缩写词，如首字母缩略词和大写首字母。

6.1.4 唯一性规则

同一相关环境的所有名称应是唯一的。

6.2 命名约定的实例

名称成分由以下术语描述：对象类术语，特性术语，表示术语和限定术语。

6.2.1 对象类术语

对象类术语是构成数据元名称的一个成分，表示某一相关环境中的一项行为或一个对象。

例如，有下列的数据元：

- 医生姓名；
- 患者血压；
- 调查对象的身高测量；
- 医疗机构所有制属性。

这里组成数据元各个成分的医生、患者、调查对象和医疗机构等都是对象类术语。

6.2.2 特性术语

特性术语集是由一个特性分类体系中的一个名称成分的集合构成。这个集合应由离散的(术语的定义之间不重叠)和完整的(作为一个整体，该集合包括了规范使用特性的数据元、数据元概念或值域所需的所有信息概念)术语构成。

例如，在上述的数据元中：

- 医生姓名；
- 患者血压；
- 调查对象的身高测量；
- 医疗机构所有制属性。

姓名、血压、身高和所有制是特性术语。

6.2.3 表示术语

表示术语是一个数据元名称中描述数据元表示形式的一个成分。每个表示术语可以从一个受控词表或一个分类体系中得出。表示术语分类的表示形式如下：

- 名称
- 总额

- 测量
- 数量
- 数目
- 文本
-

这类术语描述了数据元有效值集合的形式。通常这类表示术语可能与特性术语有部分重复,此时,可以从结构化名称中将一个术语或术语的一部分删除。

例如,在上述的数据元中:

- 调查对象的身高测量;
- 医生姓名。

成分测量和姓名是表示术语。但由于姓名又是一个特性术语,为了表达清楚,冗余字可以删去。

6.2.4 限定术语

如果必须对一个数据元进行唯一标识,可以将限定术语加到对象类术语、特性术语和表示术语上,这些限定术语也许是从一个相关环境规定的结构设置中产生的。在(确立)命名约定的规则中(时),建议对限定术语的数量予以限定。

例如,在数据元“传染病患者的婚姻状况”中,成分“传染病”是限定术语。

6.2.5 分隔符语义

术语的成分由分隔符来界定。它们可以是:

——没有语义含义。可以用一项命名规则说明分隔符由一个空格或一个确切的特定字符(如一个连字符或下划线)组成,而不管各成分间的语义关系如何。这样的规则简化了名称的生成过程。

——有语义含义。语义含义能由分隔符表达,如将限定术语之间的分隔符和其他成分之间的分隔符设定不同的分隔符。用这种方法,分隔符就把限定术语从名称的其他部分清楚地标识出来了。

例如,在“医疗年人均费用”数据元名称中:

- 医疗-年_人均-费用

在限定术语之间的分隔符是下划线;其他名称成分之间的分隔符是连字符。

7 卫生信息数据元的定义

7.1 定义规则

卫生信息数据元的定义应该满足以下要求:

——具有唯一性(在出现此定义的任何数据字典或数据元目录中)。说明:每个定义必须区别于(字典或目录中的)任何其他定义,以保证专一性。定义中所表述的一个或多个特性必须使被定义的概念与其他概念相区别。

——要阐述其概念是什么,而不是阐述其概念不是什么。说明:在阐述数据元的定义时,仅阐述其概念不是什么并不能对概念作出唯一的定义。

——用描述性的短语或句子进行阐述。说明:必须使用短语来形成包含概念的基本特性的准确定义。不能简单地陈述一个或几个同义词,也不能以不同的顺序简单地重复这些名称。如果一个描述性短语不够,则应使用完整的、语法正确的句子。

——仅可使用人们普遍理解的缩略语。说明:这是考虑到对缩略语,包括简称和英文首字母含义的理解,通常会受到特定环境的限制。如某一缩略语在不同环境下也许会引起误解或混淆。则为了避免词义不清,在该定义中就应使用全称,而不用缩略语。所有简称在第一次出现时,必须予以说明。

——表述中不要加入不同的数据元定义或引用下层概念。

7.2 定义指南

卫生信息数据元的定义应尽量做到:

- 阐述概念的基本含义;
- 准确而不含糊;
- 简练;

- 能单独成立；
- 表述中不应加入理论说明、功能说明、范围信息或程序信息；
- 避免相互依存；
- 相关定义使用相同的术语和一致的逻辑结构。

8 卫生信息数据元的分类

对卫生信息数据元进行分类有几个目的：分类可帮助用户从众多的卫生信息数据元中找出某个单一的数据元；方便对卫生信息数据元进行管理和分析；通过继承使原本借助其他属性（如名称和定义）不能完整表述的语义内容得以表达。

数据元的分类模式一般包括：关键字、主题词表、分类法和本体论，其主要作用在于：

- 派生和形成抽象数据元和应用数据元；
- 确保适当属性和属性值的继承；
- 从参照词汇表中派生名称；
- 消除歧义；
- 辨识上位类、同位类和下位类的数据元概念；
- 辨识数据元概念和数据元之间的关系；
- 辅助模块化设计的名称和定义的开发。

8.1 关键字

关键字作为基本属性，可应用于对象类、特性、表示、数据元和数据元概念。关键字的描述如下：

名称：关键字

定义：用于数据元检索的一个或多个有意义的字词。

约束：条件选

数据类型：字符串

8.2 主题词表术语

主题词表术语能够与数据元和数据元概念关联起来。本标准并没有规定主题词表结构。

8.3 分类法和本体论分类单元

分类法是基于概化或特化以及集、子集和集隶属关系这样的数学概念的概念或分类单元的层次结构。本体论是关于分类单元的网状结构，目的在于为自然界某些部分提供模型，由关于对象的类别、对象特性以及自然界中该部分对象间可能联系的理论组成。一个本体论可以包括对分类单元的解释以及对符合语法规则的使用作出限定的正式通则。分类法和本体论中的分类单元可能与下列已分类的数据注册成分相关联：对象类、特性、表示类和数据元概念。本部分并不阐明分类法或本体论的结构。

9 卫生信息数据元内容标准编写格式规范

在医疗卫生各专业领域数据的应用、管理和交换中，卫生信息数据元字典（或数据元目录）的编写是很重要的手段和途径之一。

9.1 数据元内容标准的格式规范

卫生信息各专业领域的数据元标准要求至少由以下部分构成：

封面
目次
前言
引言

名称

- 1 范围
- 2 规范性引用文件
- 3 术语和缩略语
- 4 数据元的描述格式
- 5 数据元的类目分组
- 6 数据元目录
- 7 数据元索引
 - 7.1 按数据元类目分组索引
 - 7.2 按数据元中文名称的顺序索引(汉语拼音字母顺序)

数据元的类目分组可以按照数据元的信息模型表达,也可以按照数据元所属数据集表达。

为了便于数据元的检索,数据元标准宜建立数据元索引,一般分别以数据元类目分组和数据元中文名称汉语拼音字母顺序建立两个索引。在索引中只需列出数据元的中文名称,两种索引的差异在于数据元在索引中的排列顺序不同。

9.2 数据元目录的格式规范

数据元目录是卫生信息数据元标准的核心内容,一般可以有摘要式和字典式两种表达格式。摘要式目录要求列出数据元的所有描述属性,包括必选、可选以及附加属性。字典式目录只需列出几个最为关心的属性,没有固定要求,如只列出必选属性。

摘要式和字典式目录示例分别见图6和图7。

XXXXXXXX	
标识类属性	
名称:	
标识符:	版本:
注册机构:	同义名称:
相关环境:	
定义类属性:	
定义:	
关系类属性:	
分类模式:	关键字:
相关数据参照:	关系类型:
表示类属性:	
表示类别:	表示形式:
数据元值的数据类型:	数据元值的最大长度:
数据元值的最小长度:	表示格式:
数据元允许值:	
管理类属性:	
主管机构:	注册状态:
提交机构:	
备注:	
附加类属性:	

图6 数据元摘要式目录示例

名称	定义	表示类别	表示形式	数据类型	数据元值的 最大长度	数据元值的 最小长度	数据元允许值	……
XXXX1								
XXXX2								
XXXX3								
……								

图7 数据元字典式目录示例

参 考 文 献

- [1] ISO/IEC 20943—3 Information technology-Procedures for achieving metadata registry content consistency-Part 3: Value domains.
- [2] Health Data Standards Committee, Australian Institute of Health and Welfare. National health data dictionary version 12 including version 12 supplement.
- [3] Metadata Online Registry(METeOR). <http://meteor.aihw.gov.au/>.
- [4] The United States Health Information Knowledgebase. <http://ushik.org/registry/x/index.html>.
- [5] The Environment Data Registry. 2006-02-01. <http://www.epa.gov/edr/>.
- [6] 交通部. 交通信息基础数据元管理系统. 2005. <http://jtsjy.moc.gov.cn:88/>.
- [7] 袁满. 数据元概念、表示以及基本数据元之间的关系. <http://www.digitaloilfield.org.cn>.
- [8] 袁满. 数据元技术是解决 IT 行业信息标准化不可忽视的重要技术. <http://www.digitaloilfield.org.cn>.
- [9] 国家质量监督检验检疫总局, 中国国家标准化管理委员会. GB/T 2261. 1—2003 个人基本信息分类与代码 第 1 部分: 人的性别代码[S]. 北京: 中国标准出版社, 2003.
- [10] 国家质量监督检验检疫总局, 中国国家标准化管理委员会. GB/T 2659—2000 世界各国和地区名称代码[S]. 北京: 中国标准出版社, 2000.
- [11] 国家质量监督检验检疫总局, 中国国家标准化管理委员会. GB/T 15237. 1—2000 术语工作词汇 第 1 部分: 理论与应用[S]. 北京: 中国标准出版社, 2000.
-