

中华人民共和国国家标准

GB/T 23824.3—2009/ISO/IEC TR 20943-3:2004

信息技术 实现元数据注册系统(MDR) 内容一致性的规程 第3部分:值域

Information technology—Procedures for achieving metadata
registry content consistency—
Part 3: Value domains

(ISO/IEC TR 20943-3:2004, IDT)

2009-05-06 发布

2009-11-01 实施

中华人民共和国国家质量监督检验检疫总局
中国国家标准化管理委员会

发布

目 次

前言	I
引言	II
1 范围	1
1.1 背景	1
1.2 目的	1
1.3 限用范围	1
1.4 注册方法——值域和数据元	1
2 规范性引用文件	1
3 术语和定义	2
4 值域的理解	2
4.1 导言	2
4.2 总则	2
4.3 值域的结构	7
4.4 作为值域的代码集	11
4.5 作为值域的分类体系	11
4.6 数据类型和值域	12
4.7 计量单位	13
4.8 维度	15
4.9 对值域的分类	16
5 对值域的注册	17
5.1 导言	17
5.2 注册值域的规则	17
5.3 策略	19
5.4 示例	20
附录 A (规范性附录) 用于值域和概念域的元模型	29
参考文献	30

前 言

GB/T 23824《信息技术 实现元数据注册系统(MDR)内容一致性的规程》分为四个部分:

- 第 1 部分:数据元;
- 第 2 部分:XML 结构化数据;
- 第 3 部分:值域;
- 第 4 部分:综述。

本部分为 GB/T 23824 的第 3 部分,等同采用 ISO/IEC TR 20943-3:2004《信息技术 实现元数据注册系统(MDR)内容一致性规程 第 3 部分:值域》。

本部分的附录 A 是规范性附录。

本部分由中国标准化研究院提出。

本部分由全国电子业务标准化技术委员会归口。

本部分起草单位:中国标准化研究院。

本部分主要起草人:张艳琦,邢立强,陈琳,孙广芝,史立武,岳高峰,娄晓琳,张荫芬,刘植婷。

引 言

基于 GB/T 18391《信息技术 元数据注册系统(MDR)》的元数据注册系统之间的元数据交换,不仅依赖于注册系统软件符合该标准,而且还取决于注册系统之间的元数据内容的可比性。尽管该标准规定了数据的规范和注册,但关于如何将内容注册到系统中还有一些实际问题。基于实施该标准的组织的经验,探究内容问题的技术报告将有助于目前和将来的用户。

元数据注册系统可用于注册多种组织的数据资源集的数据元、值域、其他对象和相关属性。值域用于规定数据元的允许值,元数据注册系统可存储描述值域的信息、标准列表的代码和分类方案。

本部分的依据是 GB/T 18391.3,该系列标准描述了用于管理数据语义学的注册系统的组织方式。该标准以概念模型的形式规定了注册系统的结构。概念模型目的不在于成为计算机系统的一个逻辑或物理数据模型。

GB/T 18391.3 规定了值域和关联的概念域。在生成概念域和值域的过程中,需要规则和关系的概念化和联结。应使值域能够得以再利用,并使之规则化。元素等价域的值之间所具有的关系需要保存在元数据注册系统中。一些概念等价域的值之间也具有关系,它们也需要保存。本部分将描述如何完成该任务。

尽管元数据注册系统可以用来存贮多种元数据项的信息,但本部分只针对值域、概念域、相关属性和关系。本部分的目的在于确保对值域属性的内容有共同的理解,以便元数据可以在不同的注册系统之间得以共享。

信息技术 实现元数据注册系统(MDR)

内容一致性的规程 第3部分:值域

1 范围

1.1 背景

GB/T 18391 的元数据注册系统(MDR)是管理可共享数据的一种工具,是关于数据参考信息的一个全面的、权威的来源。MDR 通过记录和分发数据描述来支持标准化和协同化过程,这有利于数据在组织及用户间的共享。MDR 提供了与涉及特定数据元、值域和分类方案的文件的联系,以及与那些对象所应用的信息系统的联系。当与数据库联合使用时,注册系统能够使用户更好地理解来自该数据库的信息。

注册系统不包含数据本身。它包含了对数据进行清晰描述、编目、分析和分类时所必需的元数据。它给出了关于数据单元含义、表示和标识的理解。该标准标识用于确定系统之间共享数据含义时所需的信息元素。

1.2 目的

GB/T 23824 的本部分旨在描述一组在注册系统中获得值域及其属性一致性注册的规程。本部分不是数据手册,而是用户指南,用于概念化值域和它的组成部分,以便始终建立质量良好的元数据。必要时,组织可以采纳或增加规程。

1.3 限用范围

本部分的范围限定于值域、概念域以及它们的相关属性和关系。整个部分用示例说明所描述的概念。

1.4 注册方法——值域和数据元

在元数据注册系统中,注册值域是可以选择的。一些注册机构将这些集合作为值域,其他机构将这些集合作为数据元。本部分除了明确的说明外,总是将这些集合作为值域。这种选择有助于说明注册许多不同类型值域的方法。

2 规范性引用文件

下列文件中的条款通过本部分的引用而成为 GB/T 23824 的本部分的条款。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部分,然而,鼓励根据本部分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部分。

GB/T 5271.4 信息技术 词汇 第4部分:数据的组织(GB/T 5271.4—2000,eqv ISO/IEC 2382-4:1999)

GB/T 10112 术语工作 原则与方法(GB/T 10112—1999,neq ISO/DIS 704:1997)

GB/T 18391.1 信息技术 元数据注册系统(MDR) 第1部分:框架(GB/T 18391.1—2009,ISO/IEC 11179-1:2004,IDT)

GB/T 18391.2 信息技术 元数据注册系统(MDR) 第2部分:分类(GB/T 18391.2—2009,ISO/IEC 11179-2:2005,IDT)

GB/T 23824.3—2009/ISO/IEC TR 20943-3:2004

GB/T 18391.3 信息技术 元数据注册系统(MDR) 第3部分:注册系统元模型与基本属性(GB/T 18391.3—2009,ISO/IEC 11179-3:2003,IDT)

GB/T 18391.4 信息技术 元数据注册系统(MDR) 第4部分:数据定义的形成(GB/T 18391.4—2009,ISO/IEC 11179-4:2004,IDT)

GB/T 18391.5 信息技术 元数据注册系统(MDR) 第5部分:命名和标识原则(GB/T 18391.5—2009,ISO/IEC 11179-5:2005,IDT)

GB/T 18391.6 信息技术 元数据注册系统(MDR) 第6部分:注册(GB/T 18391.6—2009,ISO/IEC 11179-6:2005,IDT)

GB/T 23824.1—2009 信息技术 实现元数据注册系统(MDR)内容一致性的规程 第1部分:数据元(ISO/IEC TR 20943-1:2003,IDT)

3 术语和定义

GB/T 18391 中确立的及其下列术语和定义适用于 GB/T 23824 的本部分。

3.1

概念等价域 **conceptually equivalent domains**

表示相同概念域的若干值域。

3.2

元素等价域 **elementarily equivalent domains**

两个可枚举值域的允许值之间如果存在一一对应的关系,则为元素等价,这样,给定的任何一对对应允许值,它们的值含义是等价的。

注1:参见4.2.5的示例。

注2:所有元素等价域概念上也是等价的。这是根据这样的事实得出的:元素等价域具有相同的值含义集,因此,它们表示相同的概念域。

注3:元素等价是所有可枚举值域的一种等价关系。因此,任何数量的可枚举值域彼此间可以是元素等价的。参见示例5.4.1。

4 值域的理解

4.1 引言

本章主要描述值域的若干方面:

- a) 关于值域的总则;
- b) 值域中存在的结构或关系;
- c) 作为值域的代码集;
- d) 作为值域的分类方案;
- e) 数据类型和值域的关系;
- f) 计量单位的应用;
- g) 维度的重要性;
- h) 值域分类。

本部分始终用示例来说明所描述的概念。表明所有构件之间关系的详细模型参见附录A。

4.2 总则

4.2.1 引言

值域是一组允许值的集合。允许值是某个值和该值含义的组合。相关含义称为值含义。本部分中,允许值可表示为用尖括号界定的一个有序对: $\langle \text{值}, \text{值含义} \rangle$ 。值域是一个或多个数据元有效值的集合,它用于验证信息系统和数据交换中数据的有效性,同样也用于描述数据元的元数据的一个不可或缺

的部分。特别是,值域可以帮助理解数据元所表示的数据的内容、形式和结构。

值域分为两大类:可枚举和不可枚举。*可枚举值域*是其所有允许值可以明确列出的值域。*可枚举值域*类型的示例有:代码集、标准分级和分类。*不可枚举值域*是其允许值由规则表示的值域,该规则称为*不可枚举值域描述*。因此,没有明确列出允许值。该规则准确规定了哪些值属于该值域,哪些不属于。不可枚举值域类型的示例包括数字区间、字符串和位图。

*概念域*是值含义的一个集合。它是其外延是值域的一个汇集的一个概念。概念域也分为两大类:可枚举和不可枚举。*可枚举概念域*的值含义可以明确列出。这类概念域对应于值域中的可枚举类。*不可枚举概念域*的值含义由称之为*不可枚举概念域描述*的规则来表达。因此,值含义不是明确列出的。该规则描述了一个不可枚举值域中允许值的含义。这类概念域对应于值域中不可枚举类。

每个值域表示两类概念:数据元概念(间接的)和概念域(直接的)。*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,也间接表示了数据元概念。然而,值域是与一个概念域直接关联的,因此,表示了该概念,独立于任何数据元。

示例可以有助于表明讨论中的区别,示例如下:

示例:

数据元名称:	雇员性别-代码
数据元概念名称:	雇员性别
数据元概念定义:	一个组织雇员的性别
概念域名称:	人的性别种类
概念域定义:	人的性别列举
值域名称:	人的性别代码(1)
值域定义:	用于人的性别的代码
允许值:	<1,男>
	<2,女>
	<0,未知>

上述值域的代码采自 GB/T 2261.1—2003。标准代码的使用可以确保元数据注册系统和应用系统之间的互操作。但是,一般来说,一个值域代码的选择可以是任意的。在这种情况下,MDR 是获得非标准值域的值及其含义的来源。

此处需要给出关于值域的若干要点。

4.2.2 代码的选择

用于上述值域的代码的选择是任意的。另一个代码集可以这样使用,但该集合是一个不同的值域。使用哪个值域由应用和组织的需要确定。下面的示例是人的性别代码的另一个代码集:

示例:

值域名称:	人的性别代码(2)
值域定义:	用于人的性别的代码
允许值:	<M,男>
	<F,女>
	<U,未知>

4.2.3 允许值的数量

允许值的数量(示例中是 3)也可能是不同的。我们可能需要一个代码表示两性人或一个代码表示变性人。每增加一次新允许值,或减少允许值,将产生一个新值域或值域的新版本。值域的一次改变是否应生成一个新值域还是仅仅产生一个现存值域的新版本由注册机构各自来决定。下例的扩展值域视为一个新值域,而不是 4.2.2 中示例值域的一个版本。

示例:

值域名称:	人的性别代码(3)
值域定义:	用于人的性别的代码
允许值:	<M,男>
	<F,女>
	<H,两性人>
	<T,变性人>
	<U,未知>

4.2.4 概念域的层次

人的性别代码的所有值域可以看作是概念上等价的。概念域中的每个值含义并不要求与一个值关联。然而,一些注册机构可能需要以便更好地区别概念,例如,把人的性别代码划分为5个种类,而不是3个种类,这样必然产生了不同的概念域。下例中,在最高层,所有值域(4.2.1、4.2.2和4.2.3中的示例)表示了人的性别分类这个思想,所以,上位概念域给出了在一个注册系统中需要的、由值域的一个类(例如,人的性别代码)表示的概念。下位概念域提供了将值含义映射到对应值域的列举。

示例:

<u>上位概念域:</u>	(不可枚举的)
概念域名称:	人的性别种类
概念域定义:	人的性别分类类别
<u>下位概念域(1):</u>	(可枚举的)
概念域名称:	人的性别种类:3个值
概念域定义:	用3个种类的人的性别种类的枚举
<u>下位概念域(2):</u>	(可枚举的)
概念域名称:	人的性别种类:5个值
概念域定义:	用5个种类的人的性别种类的枚举

4.2.5 不同允许值之间共享值含义

值含义可以用来连接概念等价域的等价允许值。在元素等价域中,每个值含义连接一个唯一的一对允许值之间的等价代码,每个值域中一个,如下例所示:

示例:

概念域名称:	人的性别种类
概念域定义:	人的性别列举
值域名称:	人的性别代码(1)(见4.2.1的示例)
	人的性别代码(2)(见4.2.2的示例)
值域定义:	用于人的性别的代码

保持这两个可枚举值域之间值含义的一一对应关系定义如下:

HSC(1)	↔	HSC(2)
<1,男>	↔	<M,男>
<2,女>	↔	<F,女>
<0,未知>	↔	<U,未知>

每对对应允许值具有相同的值含义。所以,这两个可枚举值域是元素等价的,因而,也是概念等价的。

上述两个值域中一个值域中的每个允许值和另一个值域的某个允许值共享它的值含义。因此,通过使用值含义便可获得值域之间的值等价。例如,值1和M指男性,或值2和F指女性。这两个值域是元素等价域。

4.2.6 数据元之间的值域共享

雇员性别（雇员按性别分类或区分特征的观念）和学生性别（学生按性别的分类或性别特征的观念）是不同的数据元概念，但它们可以使用同样的值域表示。这样，一个值域[例如，人的性别代码(1)]可以与多个数据元概念关联，从而也和数据元关联。

4.2.7 值域与概念的关联(数据元概念和概念域)

描述相似、而不相同的数据元需要将一个数据元概念与不同的值域联系起来，这些值域被同一个概念域概念化[如示例中，人的性别代码(1)、人的性别代码(2)、人的性别代码(3)]。然而，反之则不成立：同一概念域下的两个值域不需要与同一个数据元概念关联，见下面的示例 1 和示例 2。

示例 1:

概念域名称:	人的性别种类
概念域定义:	人的性别列举
值域名称:	人的性别代码(1)(见 4.2.1 中的示例)
值域定义:	用于人的性别的代码
数据元概念名称:	雇员性别
数据元概念定义:	一个组织雇员的生理性别

示例 2:

概念域名称:	人的性别种类
概念域定义:	人的性别列举
值域名称:	人的性别代码(2)(见 4.2.2 中的示例)
值域定义:	用于人的性别的代码
数据元概念名称:	学生性别
数据元概念定义:	一个教育机构学生的生理性别

4.2.8 未与数据元关联的值域

值域完全不必与一个数据元概念关联。它们可以被独立管理，诸如代码集有时就是如此。例如，作为维护一个标准代码集中代码值的维护机构，即使它不发布使用该值域的数据，代码集是免费提供的。

4.2.9 概念域和数据元概念比较

有两类语义和数据关联：符号语义和语境语义。符号语义指符号也就是值的含义。概念域获得此类语义，因为它是值含义的一个集合。

语境语义解决有关数据的疑问(谁、是什么、何时、何处、为什么和怎么样)。根本上，数据表示某个对象集或对象类的一个成员的一个特性的某个观察记录。观察这个术语通常隐含了人的行为，但机电仪器也可以产生或记录观察结果。

对象类描述谁被观察，特性是被观察对象的显著或描述性特征，而工具(包括人)是如何产生观察。对象类和特性是概念，它们形成与数据关联的基本语境含义。没有它们，就没有数据存在的原因。数据元概念捕捉此含义。

关于仪器(怎么样)和为什么做观察(实验设计)不在 MDR 的范围之内。对各个观察而言何时、何处做观察是具体的。这种个案层面的元数据经常可以在数据库中成套数据元的数据记录中找到。

4.2.10 不可枚举值域

不可枚举值域通过一个规则而不是可枚举的允许值列表来表示或限定数据的。下例是一个不可枚举值域及关联的数据元和数据元概念：

示例:

概念域名称:	行业描述
概念域定义:	描述一个行业的文本
不可枚举概念域描述:	描述一个行业的有限长度的文本
值域名称:	英文文本行业描述
值域定义:	一个行业的文本描述
不可枚举值域描述:	最多可达 60 字符的英文文本
数据元名称:	个人工作行业描述-文本
数据元概念名称:	个人工作的行业描述
数据元概念定义:	个人工作所在行业的描述

4.2.11 具有可枚举和不可枚举组分的值域

一个域既有可枚举也有不可枚举组分的情况尽管很少,但也有可能。这种情况可能发生在值落在一个确定范围,具有一个最小(或最大)值,在最小值以下(或最大值以上)还有一些离散值用于特定的情况。下面的示例 1 说明这种情况:

示例 1:

数据元名称:	家庭每月用水量-加仑
数据元概念名称:	家庭每月用水量
数据元概念定义:	一个家庭每月所使用的水量
概念域名称:	容积(液体)
概念域定义:	用附加特定值对液体容积的计量
不可枚举概念域描述:	液体容积计量
值域名称:	以加仑(美国、液体)×1 000 表示的容积
值域定义:	以千记的美国加仑液体容积
不可枚举值域描述:	大于或等于 0 的整数以及特定值-1 和-2
允许值:	<-1,没报告> <-2,不可计量(例如,用井水或河流)>

另一种具有可枚举和不可枚举部分的值域的情形由示例 2 说明。在此,特定含义附属于一定范围内有限个值。

示例 2:

数据元名称:	下次投票的可能性,数值范围为 0~1
数据元概念名称:	下次投票的可能性
数据元概念定义:	个人在下次投票中可能性的一个计量
概念域名称:	等级测量
概念域定义:	成就、态度、偏好、难度等一个计量
不可枚举概念域描述:	用数量表达的等级测量
值域名称:	偏好等级,0 到 1
值域定义:	用 0 到 1 间数值计量的偏好
不可枚举值域描述:	0 到 1 之间的所有实数
允许值:	<0,绝对不> <0.5,没有偏好> <1,确定的>

注:该示例需要进一步的说明。该数据可以通过读取回答者记号在一个表格中的刻度来获得。值可以通过找到记号到“0”的距离除以整个尺度长度的比率来内插。另外,可枚举和不可枚举品质的使用是这个值域的基础。用相反的偏好尺度表示相同的数据元概念是可能的,那就是,“0”意味着“将参与”,“1”意味着“将不参与”。因此需要与具体允许值关联的值含义来全面地描述这些值域。

4.2.12 值域应用的语义约束

上述示例中使用的值域中的允许值,特别是它们的值含义在目前为止都限制了可能与它们联系的数据元的数量。人的性别代码(1)是一个可枚举值域,仅可用作表示有生命事物的性别的数据元的有效值,因为允许值描述了性别种类。同样,4.2.11中示例2(偏好尺度,0到1),仅可用作表示偏好的数据元的有效值,因为一些允许值的值含义描述了偏好的级别。这两个示例表明,一些值域,通过它们的结构,限制了它们语义的应用。它们仅可表示某些类型的数据元概念。

特别是,可枚举值域,似乎与它们作为有效值的数据元尤其关联,并且所有值域限定在某个方面。例如,下例中的值域不能用来表示数据元*雇员性别-代码*。但是,该示例表明一些值域在许多情形下是适用的:

示例:

数据元名称:	上次选举投票者,是或否
数据元概念名称:	上次选举投票者
数据元概念定义:	指明某人是否在上次选举中投票
概念域名称:	是或否的表示
概念域定义:	是和否的值(for 值为是和否)
值域名称:	是或否代码
值域定义:	回答为是和否的代码
允许值:	<0,否>
	<1,是>

4.2.13 快速变化的可枚举值域(UPC 示例)

以上描述的所有可枚举值域在本质上是固定的。在头脑中它们都被赋予给一个固定的允许值集合,尽管它们可能在某些场合会变化。如果一个可枚举值域要求有更多的允许值,那么一个具有额外允许值的新值域就会产生(参见4.2.1、4.2.2和4.2.3中的示例)。

一些可枚举值域改变(或增长)的很快。一个示例是通用产品代码(UPC)。UPC是一个赋予产品的12位数字码,UPC能唯一地标识产品及其制造商。该代码在数字上表现为12个阿拉伯数字,图形上的形式表现为55个黑白交替的条形。这些条形有4个标准宽度,每组4个连续条形决定一个阿拉伯数字。开始是3个条形,中间是5个条形,这5个条形在6个阿拉伯数字之后,末尾有3个条形。前6个阿拉伯数字标识制造商,接下来的5个阿拉伯数字标识产品,最后一位阿拉伯数字是一个校验码。零售商用这些代码管理存货、确定价格和加快销售点。制造商通过在统一编码委员会(UCC)注册并支付会费。接下来,UPC的厂商协调员管理其所有产品的标识符,退出不再被供给的产品代码,并确保所有产品都被唯一标识。

可能在一个MDR中管理一个快速变化的值域是无意义的。许多变化需要构成、注册管理太多的值域。管理值域可供选择的办法是,定期加以更新,或维护注册系统中值域的参照或指示词。当然,如何管理快速变化的值域取决于注册机构的决策。

另一个选择是,可以将一个快速变化的可枚举值域作为一个不可枚举值域来管理。这样便于维护注册系统中所有元数据。但是,UPC示例表明,这样也有一些明显的缺陷。关于一个有效CPU代码的完整描述是存在的,并且该描述满足不可枚举值域描述的标准。但是,决不可能知道一个特定的UPC代码是否在使用中,知道这类信息具有明显便利。尽管可能有困难,但是,按照可枚举值域维护快速变化值域是了解目前使用那些代码的唯一途径。

4.3 值域的结构

4.3.1 导言

值域间通常具有关联,它们有多种途径来表现。值域之间以及(可能)不同值域的单个值之间具有关联。任何时候,不同值域的单个值之间具有某种关联,那么值域之间也具有某种关联。

值域和允许值之间可以有关联。概念域可以互相关联,因而两个关联值域的共享概念域可能不是单独值域的概念域。一个概念域不必由 MDR 中一个值域来表示。

需要描述的主要关联有两类:值域之间的关系,一个或多个值域间的值的关联。这两类关系由下子条(款)中的实际来详尽阐明。

4.3.2 国际标准行业分类(ISIC)

由联合国统计署维护的国际标准行业分类(ISIC,第三次修订),是用于对国家和国际经济进行分析的行业分类,共4层(又见4.5)。4层都有其名称,它们表示层级的细目。层的名称和每层中项数如下:

- 第一层:门类 17项
- 第二层:大类 60项
- 第三层:中类 159项
- 第四层:小类 259项

前三层中各层的每个项有一个或多个下一层的项。示例如下:

示例 1:	
层级:	国际标准行业分类
门类:	G——批发和零售业;汽车、摩托车和个人、家庭用品维修
大类:	51——批发和代理贸易,汽车和摩托车除外
中类:	513——家庭用品批发
小类:	5131——纺织品、衣服和鞋类批发

批发和零售行业门类内有四层。每层有若干选择和解释性备注帮助用户理解概念。

该示例表明 ISIC 实际上由四个相关值域构成,每层一个。它们关联的原因有两个:1) 每个值域(层)是一个行业的范畴;2) 每个域中的值是该楼层下属项的一个概括,或者,每个域中的项是其上层项的一个具体化。例如,在上例中,项 51 概括了项 513(和项 511、512、514、515 以及 519)。中类层的项是大类层项(51)的具体化。

下例指出了注册 ISIC 所必需的概念域和值域:

示例 2:	
概念域(通用)名称:	行业分类体系
概念域定义:	表示行业范畴的嵌套层级代码
概念域名称:	行业分类体系,第一层
概念域定义:	表示行业范畴的第一层代码
概念域关系:	(行业分类体系的)特化
值域名称:	ISIC 的门类
值域定义:	ISIC 门类(第一)层代码
允许值:	〈A,农业、狩猎业和林业〉 〈B,渔业〉 ... 〈Q,地区外组织和团体〉
概念域名称:	行业分类体系,第二层
概念域定义:	表示行业范畴的第二层代码
概念域关系:	(行业分类体系的)特化
值域名称:	ISIC 大类
值域定义:	ISIC 大类(第二)层代码
允许值:	〈01,农业、狩猎业及相关服务活动〉

	<02,林业、采伐业及相关服务活动>
	<05,渔业、鱼孵化和鱼养殖业务;附属渔业的服务活动>
	...
	<99,地区外组织和团体>
概念域名称:	行业分类体系,第三层
概念域定义:	表示行业范畴的第三层代码
概念域关系:	(行业分类体系的)特化
值域名称:	ISIC 中类
值域定义:	ISIC 中类(第三)层代码
允许值:	<011,谷物种植;市场园林;园艺>
	...
	<020,林业,伐木业和相关服务活动>
	...
	<050,渔业、鱼孵化和鱼养殖业务;附属渔业的服务活动>
	...
	<990,地区外组织和团体>
概念域名称:	行业分类体系,第四层
概念域定义:	表示行业范畴的第四层代码
概念域关系:	(行业分类体系的)特化
值域名称:	ISIC 小类
值域定义:	ISIC 小类(第四)层代码
允许值:	<0111,谷物及其他作物的种植>
	<0112,蔬菜、园艺作物的种植>
	...
	<0200,林业、伐木业和相关服务活动>
	...
	<0500,渔业、鱼孵化和鱼养殖业务;附属渔业的服务活动>
	...
	<9900,地区外组织和团体>

该例中,值域之间的关系通过值域与其概念域之间的关系得以维持,事实上,每个概念域与相同的泛化概念域关联。不同域之间的值的关系通过代码(值)本身来维持。代码的字符数目决定层级。最左边字符是第一层代码,最左边的两个代码是第二层代码,等等。

GB/T 18391 用于元数据注册系统的元模型不支持两个值域值之间的关系类型。如果注册机构确定这是注册系统功能所必需的,则需要增加对元模型的必要扩展。

4.3.3 逻辑观察标识符名称和代码(LOINC)

需要关系的域集合的另一个示例是 LOINC¹⁾。逻辑观察标识符名称和代码(LOINC)提供了用于标识健康领域个体实验室结果、临床观察和诊断研究观察的通用名称和代码的一个标准集。

LOINC 在 *LOINC 用户指南* 中定义了如下一个基本结构:

一个实验结果或临床观察的完整规范的名称包括以下五个或六个主要部分:

——测量成分或(被)分析物的名称(例如,葡萄糖、消炎药如布洛芬)

——观察特性(例如,物品浓度、质量、体积);测量时间安排(例如,时期的或瞬间的)

1) LOINC 见 URL 网 <http://www.regenstrief.org/loinc/loinc.htm>.

---样品类型(例如,尿、血清)、计量程度(例如,定性的 vs. 定量的)

---测量方法(例如,放射免疫化验、免疫污渍)

这些可以通过下述语法得到形式上的描述:

<被分析物/成分>:<观察属性的种类或计量>:<时间方面>:<系统(样本)>:<尺度>:<方法>

冒号是代码的一部分,用作分隔符。

名称的第一部分可以进一步细分为三个子部分,由~分隔。第一个子部分可以包括多至三层的递增分类指明,由圆点分隔。名称的第三、第四部分(时间方面和系统/样本)也可以用另一个子部分来修饰,用一个 carat 使之与第一部分分开。假如有时间方面,修饰语可以表明观察是一个基于名称规范(最大、最小、平均等等)的选择;假如有系统,如果不是患者(例如,捐血者、胎儿、血液生成物单位),则修饰语标识样本的来源。

给出 LOINC 的详细资料是没有意义的。LOINC 的完整描述是非常复杂的,而且超出了本部分的范围。通过上述的简短描述,很清楚的是,多重代码集和层次用来规定一个 LOINC 的记录。在捕捉不同值域间的所有关系时,一个概念域关系是必要的。下例给出了如何注册 LOINC 结构化值域。时间和衡量维度未予考虑。

示例:

概念域(通用)名称:	逻辑观察标识符名称和代码(LOINC)
概念域定义:	标识健康保健领域个体实验室结果、临床观察和诊断研究观察的内部关联集合
概念域名称:	(被)分析物
概念域定义:	被测量分析物或成分的指明
概念域关系:	{LOINC}的特化
值域名称:	分析物
值域定义:	被测量分析物或成分的名称
允许值:	参见 LOINC 网址
概念域名称:	特性
概念域定义:	被观察或被测量特性的指明
概念域关系:	{LOINC}的特化
值域名称:	特性
值域定义:	被观察或被测量特性的名称
允许值:	参见 LOINC 网址
概念域名称:	样品类型
概念域定义:	样品类型(例如,尿、血清)的指明
概念域关系:	{LOINC}的特化
值域名称:	样品类型
值域定义:	样品类型(例如,尿、血清)的名称
允许值:	参见 LOINC 网址
概念域名称:	方法
概念域定义:	测量方法的指明
概念域关系:	{LOINC}的特化
值域名称:	方法
值域定义:	测量(例如,放射免疫化验、免疫污渍)方法的名称
允许值:	参见 LOINC 网址

4.4 作为值域的代码集

代码集是值域,因为每个代码是一个值,每个代码的含义是一个值含义。如同在性别代码示例中,用代码指定众所周知的含义是一个有用的技术。当人们需要存储关于一个感兴趣对象的一个特征的数据,并用最少的字节这么做时,信息技术领域就开始使用该技术了。代码的使用使得编程变得容易了,因为固定长度的数据比可变长度的数据更易于处理。而且,在一个数据库中,代码比描述性文字更易于维护。

在许多主题领域和标准中,代码集被用作值域。在美国健康保健界,下列引用描述代码集得应用:在 HIPAA(健康保险便携性和可靠性法案)中,一个“代码集”是用于对数据元,诸如术语表格、医学概念、医学诊断代码或医学疗程代码,进行编码的代码的任何集合。用于健康保健业的医学数据代码集包括疾病、缺陷、其他健康相关问题及其症状;导致伤害、疾病、缺陷或其他相关健康问题起因;阻止、诊断、治疗或管理疾病、缺陷或其他相关健康问题所采取的行为;进行这些行为所使用的物品、仪器、设施或其他事物的编码体系。HIPAA 采纳的用于诊断、疗程和药物的管理和财务健康保健业务标准中的数据元需要医学数据的代码集。

在制造业标准化领域(ISO/TC 184/SC 4),标准 ISO 10303《工业自动化系统与集成 产品数据表达与交换》包含了许多将代码集用作数据元值域的示例。依据 21.3.19-si_prefix,该例几乎全文取自 ISO 10303-41:2000。

示例:

一个 si_prefix 是可能与一个 si_单位关联的一个前缀的名称。SI 前缀的定义(值含义)由 ISO 1000:1992《SI 单位及其倍数单位和一些其他单位的应用推荐》来确定。一些示例允许值有:

允许值:	<exa,10 to the 18 th power>
	<peta,10 to the 15 th power>
	<tera,10 to the 12 th power>
	<pico,10 to the -12 th power>
	<femto,10 to the -15 th power>
	<atto,10 to the -18 th power>

国际标准会定义代码集,一些国际标准文件包括多个代码集。例如,ISO 3166-1 中定义的 3 个代码集和其他 4 个可枚举值域基本上是等价的。每个集合中值的含义不必为每个单独的可枚举值域重复。

代码集不必是可枚举值域。代码可以由一个规则生成,这样,由一个规则规定来确定代码的一个代码集是不可枚举值域。另一个事实是,大多数可枚举值域的典型示例是代码集。这可能导致人们错误地认为每个可枚举值域是一个代码集。男婴的名字列表(以及他们名字的含义)是一个值域,但不是一个代码集。

4.5 作为值域的分类体系

一些代码集有置于其中的结构。每个具有结构(例如上述 4.3 的描述)的代码集既是一个分类体系,又是值域。一些人认为每个代码集,甚至每个可枚举值域,就是一个分类体系。这个特性似乎不值得讨论,但是认识到许多分类体系(特别那些也是代码集的)也是值域这一点是很重要的。

美国健康保健界在他们工作中,采用了许多国际分类作为代码集。下面的列表给出一些正在使用的分类法:

国际疾病分类法,第 9 版,临床修改部分(ICD-9-CM),第 1、2 卷,用于下列情况:

- a) 疾病
- b) 损伤
- c) 缺陷
- d) 其他健康相关问题和症状
- e) 损伤、疾病、缺陷或其他健康问题的原因

国际疾病分类法,第9版,临床修改(ICD-9-CM),第3卷,用于下列规程或医院报告对住院患者的疾病、损伤和缺陷采取的行为:

- a) 预防
- b) 诊断
- c) 治疗
- d) 管理

国际统计界已经采纳若干个分类法作为代码集,例如:

国际标准行业分类(ISIC)被统计机构和国际统计组织用来对商业团体进行分类并报告经济统计。在组织内,维护商业团体的注册,它是用于表示每个团体所属行业范畴的数据元的一个代码集。

许多国家和国家集团也有它们自己的标准行业分类(SIC)。ISIC是一个四层体系的SIC。在北美(加拿大、墨西哥和美国),1997年第一次发布了北美行业分类体系(NAICS)。它是一个五层体系的行业分类和代码的SIC。如同ISIC,NAICS也在不同的国家使用,但是每个北美国家为了满足它们组成的需要,增加更多的层级。其他国家也有更加详尽的行业分类。

为了从一个SIC映射到另一个SIC,创立了对照表。一个SIC新版本与它的上一个旧版本之间的映射是特别重要的。这些对照表被分析者用来维护用于跨越由于SIC更新而导致的序列中断的时间序列表格。对照表不限于用于多个SIC。它们被用于所有标准分类或相似分类集合。一个MDR的GB/T 18391元模型不包括一个对照表类。如果一个注册机构认为有必要在其注册系统中包含对照表,那么应对原模型进行本地化的扩展。

除了行业分类外,标准职业分类(SOC)也在应用。许多国家在报告统计数据时采用SOC体系。企业雇用员工的职业是通过使用SOCs来分类、汇总和报告的。人们职位的单个描述也通过SOC和SIC一起使用来分类。

上述示例不会穷尽。它们只是给出了一个分类在跨越主题范围的应用以及用作数据元代码集的广度上的观念。

4.6 数据类型和值域

4.6.1 基础

GB/T 18221定义了一个值空间(亦即值域)以及描述运算和特性的集合构成的数据类型。一个值空间本身不足以规定一个数据类型。

GB/T 18931仅仅针对值域,但是,用一个值域标识一个数据类型,对于共享和理解数据是很重要的。数据类型是在编程语言中声明变量(数据元)的基础,在数据建模的实体和类架构中是隐含的。

示例表明在含义上是相同数据,如何具有不同的数据类型。

示例:	
#1	
值域名称:	容积(整加仑)
值域定义:	以整加仑表示的容积量
不可枚举值域描述:	非负整数
数据类型:	整数
#2	
值域名称:	容积(整加仑)
值域定义:	以整加仑表示的容积量
不可枚举值域描述:	非负整数
数据类型:	字符

属于这些值域的数据在计算机屏幕或打印在纸上时,看起来是相同的,但是在计算机程序中则是完全不同的。传送属于整数值域的数据需要编写一个不同于适于属于字符数据值域的程序接收它。数据类型并不给出完成传送所必需的所有信息,一些其他机器所决定的参数应是已知,但是,它们超出了GB/T 18391的范围。

因此,在数据传送层面,数据类型是一个问题。差异对于正确应用是如此的重要,上例中值域应被视为是不同的。那么,将每个值域与同一个数据元概念关联起来时,需要构成各自不同的数据元。这样,对此数据的正确理解将得到保证。

在GB/T 18391MDR的元模型中,数据类型有4个属性:名称、描述、方案和注解。提供数据类型的名称、它的派生(如果适用)描述、它的本原或派生来源的方案(即GB/T 18221)和其他适当的描述注解就足够了。这样,MDR的用户可以理解所列数据类型的含义,但是MDR不必支持机器来派生数据类型。这一点显得尤为重要,是因为存在不少方案支持数据类型,新的都是定期生成的,而且建立一个模型支持所有这些是不现实的。

4.6.2 具有多于一种数据类型的值域——值含义的限定

有时,一个值域可能有多于一种与之关联的数据类型。在某种意义上,这意味着与允许值关联的值含义是受限定的。在当前的MDR规范(GB/T 18391.3)中,这还意味着值域应分别注册给每种数据类型。

依照GB/T 18221框架,一个值域是值含义与每个值关联的值空间。因此,如果描述的运算或特性随着每个示例而改变,将给这个值域赋予一个以上的数据类型。下例表明这个观念。ISO 3166-1中,两字符字母数字代码和它们的值含义在每种情况下,被认为是允许值。GB/T 18221中定义了数据类型“国家”。

示例:	
#1	
值域名称:	2字母国家代码
值域定义:	用于世界各国的2字母数字代码
允许值:	“字母数字2字符国家代码及含义”
数据类型:	国家(具有国家有形国界法律定义的其他属性)
#2	
值域名称:	2字母国家代码
值域定义:	用于世界各国的2字母数字代码
允许值:	“字母数字2字符国家代码及含义”
数据类型:	国家(具有空间法律定义的其他属性)

在这个示例中,值域和允许值都没有改变,与允许值关联的特性发生了变化。因为允许值没有改变,那么值含义也保持一致。因此,该示例表明,在其结构之外,数据类型与一个值域的应用有关联。这里的应用意味着,在一种情况下,国家被认为是地面上有形国界的意思,在另一种情况下,被认为是地面以上的空间的意思。

在该例中为理解应用差别,数据元概念的定义与值域关联是必要的。空间和地面边界是作为一些对象类的一个特性而与国家关联的概念。这意味着数据类型确实是与一个数据元关联的。下一版本的GB/T 18391.3将阐明这个。

4.7 计量单位

值域所许可的值反映了人类如何量化世界。一个量具有一个与之关联的计量单位。存在多种计量,每种具有多个计量单位的选择(又见4.8),如下所列:

收入 美元,欧元
红血球数量 百万个/立方毫米

质量	克
高度	英寸,厘米
速度	海里/时,米/秒
波长	埃
浓度	百万分浓度
沉淀速度	毫升/时

通常,我们在分析由一个数据元表示的数据时遇到计量单位。但是,在一个 MDR 中使管理的值域之间的计量单位从不同的角度发生联系是有意义的。GB/T 18391.3 描述的 MDR 元模型使计量单位与值域联系起来。ISO 1000:1992 包含了一个更加详尽的 SI 和其他计量单位的描述。又见 G. Schadow 和 C. McDonald 的《计量单位统一代码》(1999)²⁾。

以下列示例考虑可枚举值域:

示例 1:

概念域名称:	年龄序列归类
概念域定义:	表示年龄序列的种类
维度:	时间
值域名称:	年龄种类
值域定义:	由不等范围长度定义的年龄序列。
计量单位:	年
允许值:	<1,0~15 岁> <2,16~29 岁> <3,30~64 岁> <4,65 岁及以上>

上面的可枚举值域使用计量单位年来帮助标记定义组成域的范畴的范围。注意,包含在值含义中的计量单位的使用是理解每个范围的基础。另一个具有相同范围值,但不同值含义的可枚举值域是一个不同的域,因为一个可枚举值域包含的允许值——被定义为值/值含义对。

示例表明一些可枚举值域需要附于其上的计量单位。不需要知道任何关联数据元来决定是否需要计量单位将含义赋予这个域允许值中的值。但是,4.2 中性别代码的示例表明,计量单位是不适于每个可枚举值域的。

对于不可枚举值域,选择是同样的。与许可值关联的计量单位有助于确定域的“值含义”。与相同范围关联的不同计量单位决定了不同的值域。下例中的两个值域阐明了这一点:

示例 2:

#1

概念域名称:	容积
概念域定义:	容积的计量
不可枚举概念域描述:	容积计量
维度:	容积
值域名称:	容积(整数的美国加仑表示)
值域定义:	以整数的美国加仑表示的容积量
不可枚举值域描述:	非负整数
计量单位:	美国加仑

2) G. Schadow 和 C. McDonald(1999)的《计量单位统一代码》,见 URL 网站 <http://aurora.rg.iupui.edu/UCUM/>.

#2

概念域名称：	容积
概念域定义：	容积的计量
不可枚举概念域描述：	容积计量
维度：	容积
值域名称：	容积(整数的升表示)
值域定义：	以整数升表示的容积量
不可枚举值域描述：	非负整数
计量单位：	升

上例值域在概念上是等价的,并共享同一个不可枚举值域描述。但是,每个域中的相同数字(0除外)并意味着相同的事物。一升容积不等同于一加仑。基于这个原因,计量单位的变化将改变域。

4.8 维度

维度是应用计量单位概念部分的名称。诸如米/秒、海里/时、弗隆/两星期等都是计量速度的量度标准。诸如磅、牛顿是计量力的量度标准。上面每个集合中的每个成员属于同一个类,或称之为具有相同的维度。

在数学语言中,维度形成了所有计量单位集合的一个等价类。两个计量单位之间的等价由一个单位集合到另一个集合间存在的可逆变换决定。这意味着,如果存在一种函数使一个计量单位中的值可以映射到另一个计量单位中的值,同时,函数的可逆也能将第二个计量单位中的值映射回为第一个计量单位中的值,那么两个计量单位是同维的。认知这些函数什么时候存在和不存在,是维护一个注册系统很重要的一个方面。下列两个示例说明了这点:

示例 1:华氏—摄氏温度转换

$$F^{\circ} = (9/5)C^{\circ} + 32 \text{ 和 } C^{\circ} = (5/9)(F^{\circ} - 32)$$

这里定义的函数相互间是可逆转换的。

示例 2:英寸—厘米长度转换

$$\text{in} = (2.54) \text{ cm} \text{ 和 } \text{cm} = \text{in} / (2.54)$$

这里定义的函数相互间是可逆转换的。

详细内容参见 F. Olken and J. McCarthy 的《XML 数据类型的计量单位》(1999)³⁾。又见 G. Schadow and C. McDonald 的《计量单位统一代码》(1999)⁴⁾。

了解维度,对于管理一个 MDR 中值域的注册人员和主题专家来说,是非常重要的。注册系统用户需要知道一个计量单位集合提供的数据是否可以转换成另一个计量单位集合中的数据。用于等价概念域的计量单位应具有相同的维度。接下来的两个等价概念不可枚举值域的示例表明了这点:

3) F. Olken and J. McCarthy 的《XML 数据类型的度量单位》(1999)在 URL 的网页 <http://pueblo.lbl.gov/~olken/mendel/w3c/xml.schema.wg/units/syntax.htm>.

4) G. Schadow and C. McDonald 的《计量单位统一代码》on the Web at the URL <http://pueblo.lbl.gov/~olken/mendel/w3c/xml.schema.wg/units/syntax.htm>.

示例 3:

概念域名称:	货币单位表示的数量
概念域定义:	表示货币价值的数量(整数单位)
不可枚举概念域描述:	整数单位表示的货币
维度:	货币单位
值域名称(1):	整数美元的数量
值域定义:	整数美元的货币数量
不可枚举值域描述(1):	非负整数表示的整数美元
计量单位:	美元
值域名称(2):	整数欧元的数量
值域定义:	整数欧元的货币数量。
不可枚举值域描述(2):	非负整数表示的整数欧元
计量单位:	欧元

存在将美元转为欧元及转换回美元的途径,显然,这些是概念等价域。二者表示同样观念,即“货币的数量”。但是,示例也显示出一个问题。为了有些应用,货币转换需要一个函数,对于有些应用,它几乎是连续变化的,例如银行业,因为汇率波动是如此频繁的。如果它们对于注册系统功能很重要,此类问题就需要大量的监控。

下面是(部分)维度方面可能存在的问题,以及管理员和用户在转换单位时面临的困难:

- 货币单位的转换(见上);
- 在给定时间和地点转换温度与转换温度间隔的差异;
- 一个特定事件日期和时间的转换与日期或时间间隔转换的差异;
- 认识到什么时候一些 SI 和英制(或一些其他体系的)单位具有相同的维度。

4.9 对值域的分类

值域由两种主要途径的任意一种来分类,要么通过一个分类方案中项之间的关联,或通过表示类的应用。表示类是一个专用分类方案,表现为一个 MDR 的元模型,用于对值域进行分类。值域、它的属性以及它与数据类型和计量单位的关联构成了一个数据元的表示。因此,表示类是表示的一个分类。它在结构上是非常简单的,本质上是关键字的一个列表。这些关键字是用于描述表示的类的概念的名称。这里,上面若干示例现在将包含一个表示类项:

示例 1:

概念域名称:	人的性别分类
概念域定义:	人的性别概念的列举
值域名称:	人的性别代码(1)
值域定义:	用于人的性别的代码
允许值:	<1,男>
	<2,女>
	<3,未知>
表示类项:	类别

示例 2:

概念域名称:	文本描述
概念域定义:	描述一项行为或事物的文本
不可枚举概念域描述:	有限长度的文本
值域名称:	文本的英文描述
值域定义:	一项行为或事物的文本描述
不可枚举值域描述:	最多可达 60 个字符的英文文本
表示类项:	文本

示例 3:

概念域名称:	容积
概念域定义:	以不同计量单位表示的容积
不可枚举概念域描述:	容积的计量
维度:	容积
值域名称:	(整数升)容积
值域定义:	用整数升计量的容积量
不可枚举值域描述:	非负整数
计量单位:	升
表示类项:	具有计量单位的数量

5 对值域的注册

5.1 引言

本条是对在 MDR 中注册一个值域所需的细节的描述。给出了两个示例,一个是可枚举值域的,另一个是不可枚举值域的。本章的描述将集中在注册值域的特有要求上。

- 5.2 说明注册概念域和值域的规则;
- 5.3 说明注册概念域和值域的策略;
- 5.4 提供基于规则和策略的详细示例。

5.2 注册值域的规则

总体上,模型中的下列类与值域和概念域相关。在一个 MDR 中注册值域和概念域,它们得到增殖:

- 概念域
- 概念域关系(如果必要)
- 可枚举概念域,或不可枚举概念域(可能二者皆有)
- 如果是可枚举——
 - 值含义
- 如果是不可枚举——
 - 描述
- 值域
- 值域关系(如果必要)
- 可枚举值域,或不可枚举值域(可能二者皆有)
- 如果是可枚举——
 - 值

允许值(值和值含义的结合)

如果是不可枚举——

描述

计量单位(如果必要)

数据类型

表示类

附录 A 提供了用于元数据注册系统的元模型中概念域——值域部分的一个 UML 模型。该模型显示了类、它们的名称、属性和关系,并给出了每个属性和关系的要求和基数性。

该模型包含了在一个 MDR 中注册值域和概念域的一些基本规则。这些规则中有的已经在前面的章节中叙述过。其他的在这里讨论。另外,规则的一些基本原理没有在模型的注释中。

- a) 一个概念域自身可以存在,亦即它不需要一个联合的值域用于它的存在,但是,一个值域应有一个关联的概念域。

一些概念域并不直接与值域连接在一起。这样,4.3.2 示例 2 中概念域单独存在表明了这一点。

- b) 应仔细地维护值含义。

值含义是将不同可枚举值域中不同值与相同的含义连结起来的方法;是跟踪一个值的含义随着时间而改变的方法;是在分类法之间建立对应表的方法。相似的概念经常在分类法间交叉表示;这些概念相互间的映射,对那些需要比较不同数据被分类为相似但不相同的分类体系中的用户来说,是很重要的。

- c) 值域的子类,可枚举和不可枚举不是排他的。这意味着一些值域可以有一个可枚举部分和一个不可枚举部分。参见 4.2.11 中的示例 1 和示例 2。

很容易认为子类是排他的,也就是非此即彼,二者不同时存在。几乎所有示例都是如此。但是,特殊情形使得不寻常值域的产生成为必要。

- d) 数据类型是很重要的组分,应认真地文档化,如同 4.6 的描述。

数据类型是数据成功传输所必需的。问题是规定一个数据类型有多个机制。有程序语言、数据库查询语言、数据传输语言、抽象数据定义语言和其他标准。每种都有自己描述和命名其允许数据类型的方法。不是所有这些规范都分配了相同种类的数据类型。不同规范中相同的名称并不意味着相同的事情(名称相同,但实质不同)。不是所有的规范具有相同的初始的数据类型。在不同的规范中,从初始的数据类型生成新的数据类型的方法是不同的。

GB/T 18391.3 并不要求用户使用一个规范来命名和描述数据类型。但它要求用户提供规范名称和数据类型的名称或出处,这样用户可以确定数据类型的内涵。

- e) 计量单位描述数据值,并有助于确定不同来源数据之间的相似性。它们在任何计量文档的所必需的一部分,不论是试验或是一个统计处理。参见 4.7 和 4.8。

计量单位由维度而聚合成类。维度是计量单位的一个概念,和概念域关联。用于概念等价域的计量单位应有相同的维度。

- f) 一个值含义不必与一个允许值关联。一个允许值应有一个相关的值含义。

- g) 一个概念域是值含义的一个集合。(上面的)第 6 条意味着值含义集合不必与用于表示概念域的任何可枚举值域中的允许值一一对应。另外,概念等价可枚举值域不必是基本等价的。非初等等价,但概念等价的域甚至可能有相同数量的允许值。

关于概念等价可枚举值域应有多少允许值是公用的问题尚无定论。一个值域与一个概念域的关联,是主题专家自己的抉择。两个理性的人可能有不同的选择。组织的需求、内部用户的需求、外部用户的需求和标准规定的需求都会有一种影响。

下面的示例给出了三个概念等价、但非基本等价可枚举值域;两个具有相同数量的允许值。

示例:

概念域名称:	学校课程分类
概念域定义:	学校课程概念的列举
值域名称:	学校课程代码(1)
值域定义:	用于学校研习课程的代码
允许值:	〈0, 其他〉 〈1, 数学〉 〈2, 语言〉 〈3, 外语〉 〈4, 历史〉
值域名称:	学校课程代码(2)
值域定义:	用于学校研习课程的代码
允许值:	〈0, 其他〉 〈1, 数学〉 〈2, 语言〉 〈3, 外语〉 〈4, 欧洲历史〉 〈5, 泛美洲历史〉
值域名称:	学校课程代码(3)
值域定义:	用于学校研习课程的代码
允许值:	〈0, 其他〉 〈1, 数学〉 〈2, 语言〉 〈3, 外语〉 〈4, 非洲历史〉 〈5, 亚洲历史〉

5.3 策略

注册值域的策略和步骤可以分为两大类,可枚举值域的和不可枚举值域的。主要的策略由一系列的步骤给出。步骤不必遵循给定的顺序。给定的步骤并不是详尽的。

每个域执行以下步骤:

- 理解什么值是允许的。对于一个不可枚举值域,要求完整理解定义允许值的规则,对于一个可枚举域,要求理解值的结构和含义。起初,不可能知道所有的允许值,但理解典型例子的内容、含义和形式是必要的。
- 确定用于值域的注册系统是否存在一个相配的概念域。如果没有,那么应生成一个新的概念域,并登录在注册系统中。从已存在的若干个概念域中为一个给定的值域确定一个合适的可能是困难的。最好是由一个小组来确定,包括注册员和相关的领域专家。
- 如果一个适当的概念域已经存在,然后确定值域本身是否存在。检验所有表示了选定概念域的概念等价值域。如果给定值域是新的,就应创建并将其登录在注册系统中。
- 如果值域是不可枚举的,应确定规则、计量单位和维度。
- 如果值域是可枚举的,应确定每个允许值的值和值含义。
- 如果概念域事先已经存在,用于概念域的值含义应与用于新的可枚举值域的每个值含义进行对照检查。登录任何事先不存在的值含义。对注册系统中已存在值含义的允许值进行标记。
- 登录允许值,应用连结以前存在的值含义。不必管理值集合。注册系统可能会重复值。

- h) 确认该概念域是否与其他概念域相关。当将值域的一个集合注册为一个分类时(参见 4.5), 则该概念域与其他概念域相关。分类经常表示为一个层级, 这样, 就会生成一个概念域层级, 以获得每层(值域)表示的概念。
- i) 确认该值域是否与其他值域相关。当将值域的一个集合注册为一个分类(参见 4.5)时, 则该值域与其他值域相关。分类经常表示为一个层级, 这样, 各层间的关系就表示为值域间的关系。
- j) 当前, 元数据注册系统的元模型并不处理允许值之间的关系。关系仅存在于值含义在跨越值域时也是相同的情况下, 而且关系通过连结到一个公共值含义上来实现。关系也存在于层次代码结构中, 如一个分类(参见 4.5), 而且关系通过值的结构, 亦即, 代码本身来实现。

5.4 示例

下列各条中每条都用示例阐明一个值域的完整注册: 一个用于基本等价域的集合; 另一个用于不可枚举值域。两个示例都基于国际标准。示例中所列出的属性都在 GB/T 18391.3 包含和描述的范围之内。

5.4.1 可枚举值域

第一个示例适合于包含在 ISO 3166-1 中可枚举值域的一个关联集合。该部分演示了在一个 MDR 中如何注册可枚举值域(该例中是代码集)。ISO 3166-1 中包含了四个要注册的代码集: 缩写英文名称、2 字符字母码、3 字符字母码和 3 字符数字码。每个代码集中的每个允许值都与相应的定义这些值的值含义关联。这些关联建立了跨越每个值域的元素—元素基础上的一个对应。值含义的集合和一个概念域一起注册。

概念域的注册		
1	概念域	
	概念域语境	ISO 3166-1
	概念域名称	世界上各个国家
	概念域定义	世界上主要的地理政治实体
	概念域类型	可枚举的
	CD 标识符/版本号(DI;VI)	(93-273-8065)1234:1
	维度	代码
2	值含义	
	值含义 ID	10001
	值含义描述	名称为〈阿富汗民主共和国〉的主要地理政治实体
	VM 起始日期	19971001
	VM 终止日期	(不适用)
	值含义 ID	10002
	值含义描述	名称为〈阿尔巴尼亚社会主义人民共和国〉的主要地理政治实体
	VM 起始日期	19971001
	VM 终止日期	(不适用)
	
	值含义 ID	10220

概念域的注册		
	值含义描述	名称为〈津巴布韦共和国〉的主要地理政治实体
	VM 起始日期	19971001
	VM 终止日期	(不适用)
5	其他元数据属性	
	来源	ISO 3166-1:1997
	备注	(不适用)
	提交组织	劳工统计局
	数据管理员	Dan Gillman
6	分类	
	关键字	国家、世界
	群	地理实体、国家标识符
	对象	国家
	提取层	概念域
7	质量控制	
	注册状态	标准的
	管理状态	最终的

#1 值域的注册		
3	值域	
	值域语境	ISO 3166-1:1997
	值域(VD)名称	英文缩写国家名称
	值域名称语境	注册系统
	值域定义	所有国家英文缩写名称
	VD 标识符/版本号 (DI;VI)	(93-273-8065)9876:1
	该值域概念域的 ID	51.41736111
	值域类型	可枚举的
	数据类型	字符串
	数据类型模式	C 编程语言
	最小字符数	4
	最大字符数	44
	格式	%s
	计量单位	(不适用)

# 1 值域的注册		
	精度	(不适用)
	VD 来源(可枚举)	ISO 3166-1:1997
	VD 解释性备注	(不适用)
4	允许值	
	值	阿富汗
	PV 起始日期	19971001
	PV 终止日期	(不适用)
	值含义 ID	10001
	值	阿尔巴尼亚
	PV 起始日期	19971001
	PV 终止日期	(不适用)
	值含义 ID	10002
	
	值	津巴布韦
	PV 起始日期	19971001
	PV 终止日期	(不适用)
	值含义 ID	10220
5	其他元数据属性	
	来源	ISO 3166-1:1997
	备注	(不适用)
	提交组织	劳工统计局
	数据管理员	Dan Gillman
7	质量控制	
	注册状态	标准的
	管理状态	最终的

# 2 值域的注册		
3	值域	
	值域语境	ISO 3166-1:1997
	值域(VD)名称	2 字符的字母数字混合型国家代码
	值域名称语境	Registry 注册系统
	值域定义	所有国家的 2 字符的字母数字混合型国家代码

#2 值域的注册		
	VD 标识符/版本号 (DI:VI)	(93-273-8065)9877:1
	该值域概念域的 ID	51.41736111
	值域类型	可枚举的
	数据类型	长度为 2 的字符串
	数据类型模式	C 编程语言
	最小字符数	(不适用)
	最大字符数	(不适用)
	格式	%2c
	计量单位	(不适用)
	精度	(不适用)
	VD 来源(可枚举)	ISO 3166-1:1997
	VD 解释性备注	(不适用)
4	允许值	
	值	AF
	PV 起始日期	19971001
	PV 终止日期	(不适用)
	值含义 ID	10001
	值	AL
	PV 起始日期	19971001
	PV 终止日期	(不适用)
	值含义 ID	10002
	
	值	ZW
	PV 起始日期	19971001
	PV 终止日期	(不适用)
	值含义 ID	10220
5	其他元数据属性	
	来源	ISO 3166-1:1997
	备注	(不适用)
	提交组织	劳工统计局
	数据管理员	Dan Gillman
7	质量控制	
	注册状态	标准的
	管理状态	最终的

# 3 值域的注册		
3	值域	
	值域语境	ISO 3166-1:1997
	值域(VD)名称	3 字符字母数字混合型国家代码
	值域名称语境	注册系统
	值域定义	所有国家 3 字符字母数字混合型代码
	VD 标识符/版本号 (DI; VI)	(93-273-8065) 9878:1
	该值域概念域的 ID	51.41736111
	值域类型	可枚举的
	数据类型	长度为 3 的字符串
	数据类型模式	C 编程语言
	最小字符数	(不适用)
	最大字符数	(不适用)
	格式	%3c
	计量单位	(不适用)
	精度	(不适用)
	VD 来源(可枚举)	ISO 3166-1:1997
	VD 解释性备注	(不适用)
4	允许值	
	值	AFG
	PV 起始日期	19971001
	PV 终止日期	(不适用)
	值含义 ID	10001
	值	ALB
	PV 起始日期	19971001
	PV 终止日期	(不适用)
	值含义 ID	10002
	
	值	ZWE
	PV 起始日期	19971001
	PV 终止日期	(不适用)
	值含义 ID	10220
5	其他元数据属性	
	来源	ISO 3166-1:1997

#3 值域的注册		
	备注	(不适用)
	提交组织	劳工统计局
	数据管理员	Dan Gillman
7	质量控制	
	注册状态	标准的
	管理状态	最终的

#4 值域的注册		
3	值域	
	值域语境	ISO 3166-1:1997
	值域(VD)名称	3 字符数字型国家代码
	值域名称语境	注册系统
	值域定义	所有国家的 3 字符数字代码
	VD 标识符/版本号(DI; VI)	(93-273-8065)9879:1
	该值域概念域 ID	51.41736111
	值域类型	可枚举的
	数据类型	长度为 3 的数字型字符
	数据类型模式	C 编程语言
	最小字符数	(不适用)
	最大字符数	(不适用)
	格式	%3c
	计量单位	(不适用)
	精度	(不适用)
	VD 来源(可枚举)	ISO 3166-1:1997
	VD 解释性备注	(不适用)
4	允许值	
	值	4
	PV 起始日期	19971001
	PV 终止日期	(不适用)
	值含义 ID	10001
	值	8
	PV 起始日期	19971001
	PV 终止日期	(不适用)

#4 值域的注册		
	值含义 ID	10002
	
	值	716
	PV 起始日期	19971001
	PV 终止日期	(不适用)
	值含义 ID	10220
5	其他元数据属性	
	来源	ISO 3166-1:1997
	备注	(不适用)
	提交组织	劳工统计局
	数据管理员	Dan Gillman
7	质量控制	
	注册状态	标准的
	管理状态	最终的

5.4.2 不可枚举值域

该条提供一个来自 ISO 6709:1983 的具体的关于不可枚举值域的注册。纬度是赤道以北或以南一条子午线的角度距离的计量。该标准给出了纬度计量记录的可变格式,及多于一种的表示法:度和十进制、六十进制(亦即,度和分和弧秒)。该标准还包括关于经度的多于一种的表示法和格式,及高程的一个柔性格式。另外,用于数据转换的一个标准格式也包括在该标准中。

地理信息专家日益喜欢用度和十进制来计量位置坐标,而不少组织仍继续用度、分、弧秒计量纬度和经度。因此,我们选用度、分和弧秒计量的纬度来注册该值域。

概念域的注册		
1	概念域	
	概念域名称	纬度的量度
	概念域名称语境	ISO 6709:1983
	概念域定义	赤道南北点的纬度计量
	概念域定义语境	注册系统
	概念域类型	不可枚举
	不可枚举概念域 描述	纬度计量
	CD 标识符/版本号(DI:VI)	(93-273-8065)2345:1
	维度	角计量
2	值含义	
	值含义 ID	(不适用)

概念域的注册		
	值含义描述	(不适用)
	VM 起始日期	(不适用)
	VM 终止日期	(不适用)
5	其他元数据属性	
	来源	ISO 6709:1983
	备注	(不适用)
	提交组织	劳工统计局
	Data Steward	Dan Gillman
6	分类	
	关键字	平面坐标、空间的、纬度
	群	地理点位置
	对象	纬度
	提取的层	概念域
7	质量控制	
	注册状态	标准的
	管理状态	最终的

值域的注册		
3	值域	
	值域(VD)名称	纬度的六十进制计量
	值域名称语境	注册系统
	值域定义	地球表面上一个点、地球中点和给定点赤道正北或正南的一个点所形成的角度的六十进制计量。
	值域定义语境	Registry
	VD 标识符/版本号(DI:VI)	(93-273-8065)8765:1
	该值域概念域 ID	97.70902778
	值域类型	不可枚举
	不可枚举值域描述	赤道及赤道以北纬度为正值;赤道以南纬度为负值。度数的取值范围为(DD 格式)0~90;分的取值范围(MM 格式)为 0~59;弧秒的取值范围(SS 或 SS.SSSSS 格式)是 0~59 或 0~59.99999
	数据类型	自由字符

值域的注册		
	数据类型模式	ANSI/ISO SQL
	最小字符数	7
	最大字符数	13
	格式	+/-DDMMSS 到 +/-DDMMSS.SSSSS
	计量单位	度、分、弧秒
	精度	(不适用)
	VD 来源	ISO 6709:1983
	VD 解释性备注	(不适用)
4	允许值	
	值	(不适用)
	PV 起始日期	(不适用)
	PV 终止日期	(不适用)
	值含义 ID	(不适用)
5	其他元数据属性	
	来源	ISO 6709:1983
	备注	(不适用)
	提交组织	劳工统计局
	数据管理员	Dan Gillman
7	质量控制	
	注册状态	标准的
	管理状态	最终的

附录 A
(规范性附录)
用于值域和概念域的元模型

该附录 A 包括 GB/T 18391.3 中描述值域、概念域、它们的属性和它们之间关系的元模型部分(见图 A.1)。这里的元模型为了便于参考。完整的元模型描述和用于描述元模型的 UML 包含在 GB/T 18391.3 中。

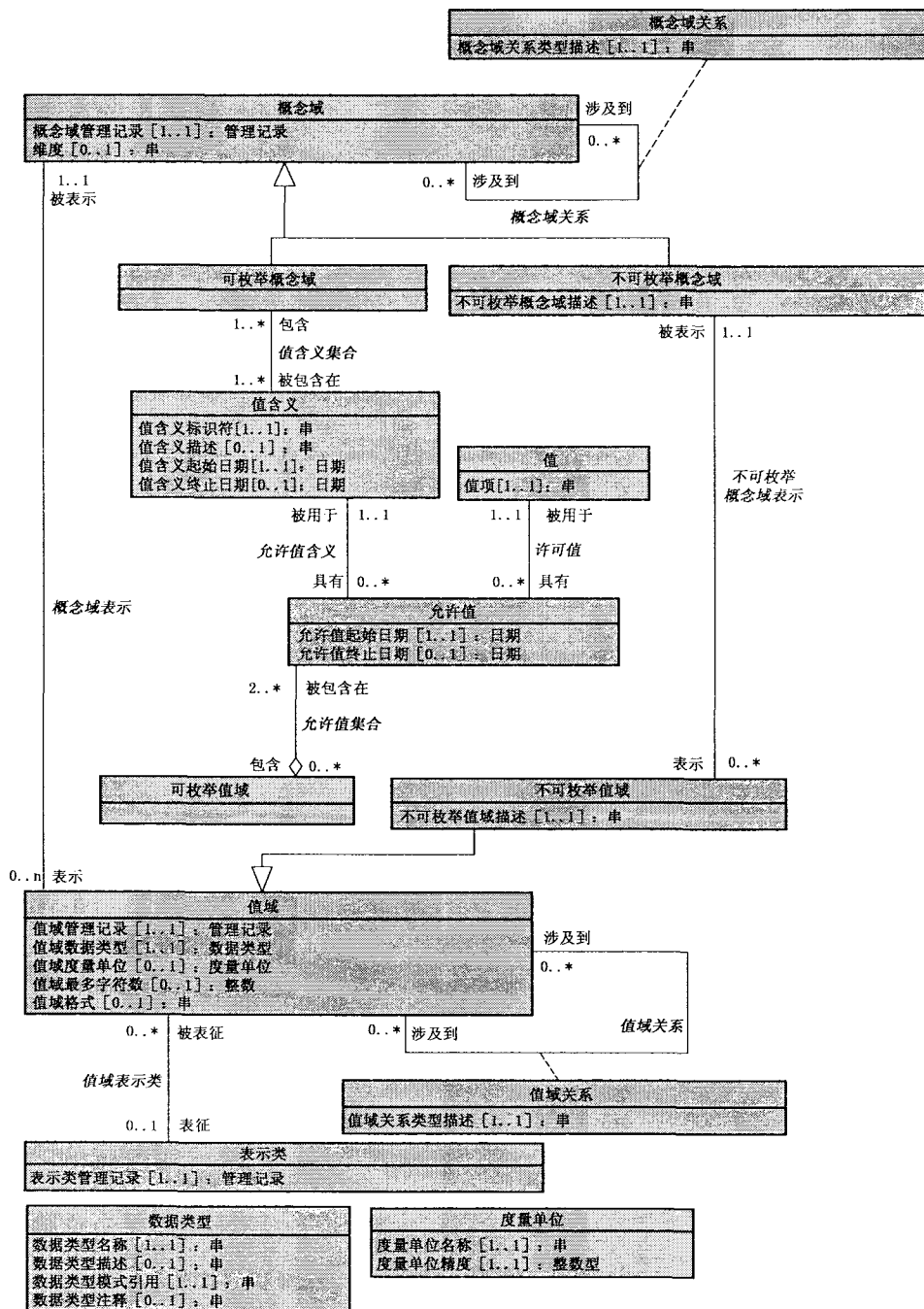


图 A.1 概念域和值域元模型

参 考 文 献

- [1] ISIC—International Standard Industrial Classification—See the URL [http://esa.un.org/unsd/cr/registry/regcst.asp? Cl=2&Lg=1](http://esa.un.org/unsd/cr/registry/regcst.asp?Cl=2&Lg=1))
- [2] ISO 1000:1992, SI units and recommendations for the use of their multiples and of certain other units
- [3] 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
- [4] ISO 6709:1983, Standard representation of latitude, longitude and altitude for geographic point locations
- [5] ISO 3166-1:1997, Codes for the representation of names of countries and their subdivisions Part 1: Country codes
- [6] ISO 10303-41:2000, Industrial automation systems and integration Product data representation and exchange Part 41: Integrated generic resource: Fundamentals of product description and support
- [7] GB/T 18221—2000 信息技术 程序设计语言、环境与系统软件接口 独立于语言的数据类型(idt ISO/IEC 11404:1996)
- [8] ISO/IEC TR 15452:2000, Information technology Specification of data value domains
- [9] LOINC—Logical Observation Identifiers Names and Codes See the URL <http://www.regenstrief.org/loinc/loinc.htm>
- [10] F. Olken and J. McCarthy (1999), Measurement Units in XML Datatypes, See the URL <http://pueblo.lbl.gov/~olken/mendel/w3c/xml.schema.wg/units/syntax.htm>
- [11] G. Schadow and C. McDonald (1999), Unified Code for Units of Measure, See the URL <http://aurora.rg.iupui.edu/UCUM/>
-

中 华 人 民 共 和 国
国 家 标 准
信息技术 实现元数据注册系统(MDR)
内容一致性的规程 第3部分:值域
GB/T 23824.3--2009/ISO/IEC TR 20943-3:2004

*

中国标准出版社出版发行
北京复兴门外三里河北街16号
邮政编码:100045

网址 www.spc.net.cn

电话:68523946 68517548

中国标准出版社秦皇岛印刷厂印刷
各地新华书店经销

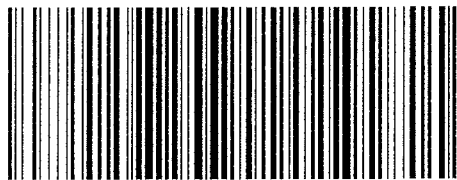
*

开本 880×1230 1/16 印张 2.25 字数 60 千字
2009年11月第一版 2009年11月第一次印刷

*

书号:155066·1-38934 定价 33.00 元

如有印装差错 由本社发行中心调换
版权专有 侵权必究
举报电话:(010)68533533



GB/T 23824.3-2009