



中华人民共和国国家标准

GB/T 18391.1—2009/ISO/IEC 11179-1:2004
代替 GB/T 18391.1—2002

信息技术 元数据注册系统(MDR) 第 1 部分:框架

Information technology—Metadata registries (MDR)—
Part 1: Framework

(ISO/IEC 11179-1:2004, IDT)

2009-09-30 发布

2009-12-01 实施

中华人民共和国国家质量监督检验检疫总局 发布
中国国家标准化管理委员会

目 次

前言	III
引言	IV
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
3.1 建模构件的定义	1
3.2 本部分的通用术语	2
3.3 用于元模型的术语	4
3.4 GB/T 18391 的本部分特定术语	7
4 缩略语	7
5 术语学原理	8
6 元数据	8
6.1 概念	8
6.2 数据元的基本模型	8
6.3 数据管理与交换中的数据元	10
6.4 值域的基本模型	10
6.5 分类方案的基本原理	12
7 元数据注册系统	13
7.1 概述	13
7.2 GB/T 18391 MDR 的总体模型	13
7.3 注册概要	14
8 GB/T 18391 综述	14
8.1 各部分介绍	14
8.2 GB/T 18391 第 1~6 部分应用的基本规则	16
9 一致性	16
参考文献	17

前 言

GB/T 18391《信息技术 元数据注册系统(MDR)》分为六个部分:

- 第1部分:框架;
- 第2部分:分类;
- 第3部分:注册系统元模型与基本属性;
- 第4部分:数据定义的形成;
- 第5部分:命名和标识原则;
- 第6部分:注册。

本部分为 GB/T 18391 的第1部分,等同采用 ISO/IEC 11179-1:2004《信息技术 元数据注册系统(MDR) 第1部分:框架》(英文版),仅有编辑性修改。

本部分代替 GB/T 18391.1—2002《信息技术 数据元的规范与标准化 第1部分:数据元的规范与标准化框架》。

本部分与 GB/T 18391.1—2002 相比主要变化如下:

- 标准的名称发生了变更;
- 标准规范的对象进行了扩展,由上一版的数据元扩展为管理项;
- 本部分第5章、第6章、第7章的名称发生了变化;
- 本部分删除了上一版的附录A和附录B;
- 本部分增加了参考文献。

为了保持与英文版本的一致性,英文版中斜体或加粗表示的内容本部分也使用斜体或加粗表示,第3章中术语的英文名称保留英文版本中的大小写格式。

本部分由中国标准化研究院提出。

本部分由全国信息技术标准化技术委员会(SAC/TC 28)归口。

本部分由中国标准化研究院起草。

本部分主要起草人:邢立强、孙广芝、李小林、刘植婷、史立武。

本部分于2002年7月首次发布,本次为第一次修订。

引言

GB/T 18391《信息技术 元数据注册系统(MDR)》描述了数据的语义、数据的表示以及这些数据描述的注册。通过这些描述,可以找到语义的确切理解及数据的有用描述。

本标准的目的在于促进:

- 数据的标准描述;
- 组织内以及组织间对数据的一致理解;
- 跨越时间、空间和应用对数据的重用和标准化;
- 组织内和组织间数据的协同及标准化;
- 数据成分的管理;
- 数据成分的重用。

本标准共有 6 个部分。每部分旨在解决上述需求的一个方面。各部分简述如下:

- 第 1 部分:框架。标准概述和基本概念。
- 第 2 部分:分类。元数据注册系统中分类方案的管理。
- 第 3 部分:注册系统元模型与基本属性。元数据注册系统基本概念模型,包括基本属性和关系。
- 第 4 部分:数据定义的形成。给出了构成高质量数据元及其成分定义的规则与指南。
- 第 5 部分:命名和标识原则。描述了如何为数据元及其成分建立命名约定。
- 第 6 部分:注册。规定了符合 GB/T 18391 的元数据注册系统注册过程的角色和要求。

通常,将描述性数据称为元数据,即,元数据是用于描述其他数据的数据。随着元数据概念的发展,元数据现在一般指用于描述某些其他对象的数据。在本标准中,将元数据范围界定为更为传统的用法,即关于数据的描述。

元数据注册系统是支持注册功能的一个元数据数据库。注册实现三个主要目的:标识、来源和质量监控。标识由赋予(注册系统内)每个注册对象一个唯一的标识符来实现;来源指明元数据及其描述对象的来源;质量监控确保元数据完成其所被赋予的任务。

元数据注册系统用于管理数据的语义。对数据的理解是设计、协调、标准化、应用、重用以及交换数据的基础。设计元数据注册系统的基本模型,目的在于获取数据语义的所有基本成分,而与应用或主题域无关。

元数据注册系统的目标是,各类应用能够确定在现有的元数据注册系统中是否存在合适的对象。如果确认需要一个新对象,鼓励通过适当修改现有描述来派生,以免类似的描述产生不必要的差异。注册也可以辨别两个或多个管理项描述的是否是同一对象,更为重要的是,可以发现在一个或多个方面存在显著差异的管理项是否使用了相似或相同的名称。

在本标准中,数据的基本容器也称为数据元。它可以抽象地存在,也可存在于某个应用系统中,但其描述是相同的。数据元描述既有语义成分,也有表示成分。语义可以进一步细分为语境型和符号型。

语境语义由数据元概念(DEC)描述。数据元概念描述数据所指的对象种类和用于度量该对象的独特特征。符号语义由概念域(CD)描述。概念域是种类的集合,它不必是有限集合。在此,种类表示数据元值域中允许值的含义。

元数据注册系统中,通过与一个特定对象关联的名称、定义、数据类型和相关对象给出该对象的含义。由于名称和定义仅传递一个对象的有限信息,因此,含义的程度是有限的。在一个注册系统中,该对象和其语义相关对象的关系将提供附加信息,而该附加信息与语义相关对象的多少有关。

表示成分是一个数据元可以使用的允许值。每个值对应于概念域中的一个种类。这些允许值的集合称之为值域(VD)。值域可以通过枚举、规则或它们的组合规定所有被允许的值。值所遵循的计算模型由其数据类型给出。

语义和表示成分由 GB/T 18391.3 规定的元数据注册系统的概念模型所包含的属性来描述。符合 GB/T 18391 的元数据注册系统可以描述多种类型的数据。事实上,第 3 部分描述的属性是数据元,并可以注册在一个 GB/T 18391 的元数据注册系统中。而且,描述符或者元数据属性的任何集合都可以理解为数据元,并在元数据注册系统中注册。

这样会产生两个主要结果:

- 元数据注册系统可以描述它自身;
- GB/T 18391 没有定义元数据的层次或级别。

因此,GB/T 18391 是一个通用的描述框架,适用于任何类型、任何组织,和任何目的数据。本标准不解决其他的数据管理需要,如数据模型、应用规范、程序代码、项目规划、业务规划和业务政策,这些需要应在别的地方另行解决。

数据处理和电子数据交换应用的增长,在很大程度上取决于数据库中数据的准确、可靠、可控和可校验。正确而恰当地使用数据与解释数据的前提之一,是数据的使用者和拥有者对数据的含义和描述特征(如,表示)有一致的理解。为了达成共识,须定义若干基本属性。

这里规定的基本属性适于对数据字典内容加以定义和规范,也适用于各种管理项集合间的交换或引用。基本属性中的“基本”意味着这些属性对于完整地规范管理项是共同需要的,以确保它们能够适用于各种功能,如:

- 信息处理系统的设计;
- 数据库中数据的检索;
- 用于数据交换的 EDI 报文设计;
- 元数据注册系统的维护;
- 数据管理;
- 字典设计;
- 字典管理;
- 信息处理系统的应用。

“基本”还隐含着它们独立于任何下述内容:

- 应用环境;
- 管理项描述的对象的功能;
- 抽象层;
- 管理项的分组;
- 信息处理系统或数据交换报文设计的方法;
- 元数据注册系统。

“基本”并不意味着 GB/T 18391.3 中规定的属性在所有情况下都是必需的。这些基本属性被分为:必选、条件选、可选几种类型。

信息技术 元数据注册系统(MDR)

第 1 部分:框架

1 范围

GB/T 18391 规定了描述数据所需元数据的种类和质量,以及一个元数据注册系统(MDR)中元数据的管理和使用。本标准用于对数据的表示、概念、含义以及它们之间的关系进行形式化表述,以使人和机器都能理解,而与产生数据的机构无关。本标准不适用于在机器层上以比特和字节对数据进行的物理表示。

GB/T 18391 中的元数据是指对数据的描述,而并非其广义概念。GB/T 18391 的本部分提供了理解各个部分并使之建立联系的途径,它是从概念上理解元数据和元数据注册系统的基础。

2 规范性引用文件

下列文件中的条款通过 GB/T 18391 的本部分的引用而成为本部分的条款。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 15237.1—2000 术语工作 词汇 第 1 部分:理论与应用(eqv ISO 1087-1:2000)

GB/T 18391.2 信息技术 元数据注册系统(MDR) 第 2 部分:分类(GB/T 18391.1—2009, ISO/IEC 11179-2:2005, IDT)

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)

ISO 704:2000 术语工作 原则与方法

3 术语和定义

下列术语和定义适用于 GB/T 18391 的本部分。

3.1 建模构件的定义

本条定义了 GB/T 18391 本部分中使用到的建模构件。

3.1.1

属性 attribute

一个对象或实体的特征。

3.1.2

类 class

具有相同属性、操作、方法、关系和语义的对象集合的描述。

[ISO/IEC 19501-1:2001,定义 2.5.2.9]

3.1.3

(元数据注册系统中)标识符 **identifier**

在一个规定的语境中,能够用来唯一标识与其关联的事物的字符序列。

注:名称不宜用做标识符,因为它是语言相关的。

3.1.4

关系 **relationship**

模型元素之间的联系。

[ISO/IEC 19501-1:2001,定义 2.5.2.36]

3.2 本部分的通用术语

本条定义了 3.1 没有涉及但 GB/T 18391 各部分可能用到的通用术语。

3.2.1

基本属性 **basic attribute**

在元数据项的规范中通常需要的元数据项属性。

3.2.2

特征 **characteristic**

一个对象或对象集特性的抽象结果。

注:特征用于描述概念。

3.2.3

概念 **concept**

通过对特征的独特组合而形成的知识单元。

[GB/T 15237.1—2000,定义 3.2.1]

3.2.4

概念体系 **concept system**

根据概念间相互关系构造的概念集合。

3.2.5

概念数据模型 **conceptual data model**

概念模型 **conceptual model**

用抽象观念表示现实世界的数据模型。

注:一个概念模型表示人们对一个体系的理解。

3.2.6

数据 **data**

信息的可再解释的形式化表示,以适用于通信、解释或处理。

注:数据可以由人工或自动的方式加工、处理。

[GB/T 5271.1—2000,定义 01.01.02]

3.2.7

数据模型 **data model**

数据的图形或文字表示,指明其特性、结构和相互间关系。

3.2.8

定义 **definition**

描述一个概念,并区别于其他相关概念的表述。

[GB/T 15237.1—2000,定义 3.3.1]

3.2.9

指称 designation

指明一个概念的符号表示。

[GB/T 15237.1—2000, 定义 3.4.1]

3.2.10

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,包括事物间的关联。

示例:一个人、对象(物体)、事件、观念、过程等等。

注:实体的存在不依赖于是否有关于它的可用数据。

[ISO/IEC 2382-17:1999, 定义 17.02.05]

3.2.11

本质特征 essential characteristic

理解一个概念所必不可少的特征。

3.2.12

外延 extension

〈术语学〉与一个概念对应的全体对象。

注:本术语在 GB/T 18391.3 中有不同的含义。

3.2.13

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。

注:一般概念的示例有“行星”、“城堡”。

3.2.14

个别概念 individual concept

仅对应一个对象的概念。

注:个别概念的示例有“土星”、“坎菲尔镇”。

3.2.15

内涵 intension

〈术语学〉组成概念的特征的集合。

3.2.16

元数据 metadata

定义和描述其他数据的数据。

3.2.17

元数据项 metadata item

元数据对象的一个实例。

3.2.18

元数据对象 metadata object

由元模型定义的一个对象类型。

3.2.19

元数据注册系统 metadata registry; MDR

用于注册元数据的信息系统。

3.2.20

元模型 metamodel

规定一个或多个其他数据模型的数据模型。

3.2.21

名称 name

用语言表述的一个对象的指称。

3.2.22

对象 object

可感知或可想象的任何事物。

注：对象可以是物质的（如，一台发动机、一张纸、一枚宝石）、非物质的（如，转化率、一个项目计划），或假想的（如，一头独角兽）。

3.2.23

注册项 registry item

记录在元数据注册系统中的一个元数据项。

3.2.24

注册系统元模型 registry metamodel

规定一个元数据注册系统的元模型。

3.2.25

术语体系 terminological system

为每个概念赋予指称的概念体系。

3.3 用于元模型的术语

本条定义了用于 GB/T 18391 本部分的术语，它们是 GB/T 18391.3 中规定的元模型中元数据对象的名称。

3.3.1

管理项 administered item

在管理记录中记录其管理信息的注册项。

3.3.2

管理记录 administration record

一个管理项的管理信息的总和。

3.3.3

管理状态 administrative status

注册机构处理注册请求中赋予管理过程状态的指称。

注：注册机构确定管理状态的值及相关含义。参见注册状态。

3.3.4

分类方案 classification scheme

根据对象具有的共同特征将对象进行排列或分组的描述性信息。

3.3.5

分类方案项 classification scheme item; CSI

分类方案的内容项。

注：它可以是分类或本体中的一个节点，叙词表中的一个术语等。

3.3.6

概念域 conceptual domain; CD

有效的值含义的集合。

注：概念域中的值含义既可枚举也可通过描述表达。

3.3.7

语境 context

一个对象被定义或被使用的环境、目的和角度。

注：本术语在 GB/T 18391.3 中有不同的含义。

3.3.8

数据元 data element; DE

由一组属性规定其定义、标识、表示和允许值的数据单元。

3.3.9

数据元概念 data element concept; DEC

能以数据元的形式表示的概念,其描述与任何特定表示法无关。

3.3.10

数据标识符 data identifier; DI

注册机构内一个管理项的唯一标识符。

3.3.11

数据类型 datatype

一些可区分的值的集合,这种区别由这些值的性质以及对这些值的运算所表征。

[GB/T 18221—2000, 定义 4.11]

3.3.12

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 ℉、0 ℃、100 美元和 10 令(纸)是数量。两个计量单位之间的等价由在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。大家熟悉的——对应的数量等式有: $C^{\circ} = 5/9 \times (F^{\circ} - 32)$ 和 $F^{\circ} = 9/5 \times C^{\circ} + 32$ 。

3.3.13

可枚举概念域 enumerated conceptual domain

所有值含义由一个列表进行规定的概念域。

3.3.14

可枚举值域 enumerated value domain

所有允许值由一个列表进行规定的值域。

3.3.15

国际代码分配符 international code designator; ICD

一个组织标识方案的标识符。

注:依据 ISO/IEC 6523-1:1998 的 3.8。

3.3.16

项标识符 item identifier

一个项的标识符。

3.3.17

项注册机构标识符 item registration authority identifier

对项进行注册的注册机构的标识符。

3.3.18

不可枚举概念域 non-enumerated conceptual domain

所有值含义不能由一个列表进行规定的概念域。

3.3.19

不可枚举概念域描述 non-enumerated conceptual domain description

概念域的所有值含义集合的规则、参照或范围的描述或说明。

3.3.20

不可枚举值域 non-enumerated value domain

通过一个描述而不是通过所有允许值列表来规定的值域。

3.3.21

不可枚举值域描述 non-enumerated value domain description

值域所有允许值集合的规则、参照或范围的描述或说明。

3.3.22

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物的集合。

3.3.23

组织 organization

为了某种目的,一个或多个人运作或被指定运作的特定机构架构。

[ISO/IEC 6523-1:1998,定义 3.1]

3.3.24

组织标识符 organization identifier

在一个组织标识方案中赋予一个组织,且在该方案中具有唯一性的标识符。

[ISO/IEC 6523-1:1998,定义 3.10]

3.3.25

组织内部机构 organization part

一个组织内由于信息交换而需要标识的任何部门、分支机构或其他实体。

3.3.26

组织内部机构标识符 organization part identifier; OPI

分配给一个特定组织内部机构的一个标识符。

[ISO/IEC 6523-1:1998,定义 3.11]

3.3.27

OPI 来源 organization part identifier source

组织内部机构标识符的来源。

注:依据 ISO/IEC 6523-1:1998,定义 3.12。

3.3.28

允许值 permissible value

在一个特定值域中允许的一个值含义的表达。

3.3.29

特性 property

一个对象类所有成员所共有的特征。

3.3.30

注册者 registrar

一个注册机构的代表。

3.3.31

注册 registration

一个管理项与其注册机构之间的关系。

3.3.32

注册机构 registration authority; RA

负责维护一个注册库的组织。

3.3.33

注册机构标识符 registration authority identifier

赋予一个注册机构的标识符。

3.3.34

注册状态 registration status

一个管理项在注册生命周期中状态的一个指称。

3.3.35

表示类 representation class

表示的类型分类。

3.3.36

计量单位 unit of measure

用于计量相关值的实际单位。

注：相关概念域的维度必须与指定的计量单位相称。

3.3.37

值 value

数据的值。

3.3.38

值域 value domain; VD

允许值的集合。

3.3.39

值含义 value meaning

一个值的含义或语义内容。

注：给定一个允许值，其值含义的表示应与(且不限定)对应值的表示无关。

3.3.40

值含义描述 value meaning description

一个值含义的描述。

3.3.41

版本 version

管理项唯一的版本标识符。

3.4 GB/T 18391 的本部分特定术语

本条定义了本部分的专用术语，这些术语不适用于本标准的其他部分。

3.4.1

数据构造构件 data construct

一个元数据项描述的对象。

注：个体的数据元、值域、数据元概念、概念域、对象类和特性均为数据构造构件。

3.4.2

数量 quantity

与一个计量单位关联的值。

4 缩略语

本章给出了第3章已定义的一些术语的缩略词。

CD——概念域(Conceptual Domain)

DE——数据元(Data Element)

DEC——数据元概念(Data Element Concept)
DI——数据标识符(Data Identifier)
EDI——电子数据交换(Electronic Data Interchange)
IEC——国际电工委员会(International Electrotechnical Commission)
ICD——国际代码分配符(International Code Designator)
ISO——国际标准化组织(International Organization for Standardization)
JTC1——第1 联合技术委员会(Joint Technical Committee 1)
MDR——元数据注册系统(Metadata Registry)
OPI——组织内部机构标识符(Organization Part Identifier)
RA——注册机构(Registration Authority)
SC32——ISO/IEC JTC1 第 32 分委员会(ISO/IEC JTC1/Sub-committee 32)

5 术语学原理

GB/T 18391 用到的术语学理论方面的概念应与 ISO 704 和 GB/T 15237.1 保持一致,必要的理论简述如下:

在术语学理论中,一个对象是可以想象或可感知到的事物。概念是特征的独特组合而成的智力构造构件、思想单元或知识单元。概念由称为特征的共同元素来组织或分组。本质特征对于理解一个概念来说是必不可少的。其他特征则是非本质的。构成一个概念的特征总和称为概念的内涵。一个概念所指对象的集合是其外延。

在自然语言中,概念由定义规定的一个唯一的内涵和外延来表述。

一个指称(术语、名称,或符号)表示一个概念。

一个一般概念对应两个或多个对象。一个个别概念仅对应一个对象。也就是说,一个一般概念的外延包括两个或多个对象,一个个别概念的外延仅包括一个对象。

一个概念体系是概念的集合,该集合依据概念间的关系来构造。一个术语体系是一个概念体系,其中每一个概念都具有指称。

6 元数据

6.1 概念

在 GB/T 18391 中,元数据是定义和描述其他数据的数据。这意味着元数据是数据,当数据以这种方式来使用时,就成了元数据。只有在特定的环境下、出于特定的目的或从特定的角度来看,数据才成为元数据。数据用作元数据的环境、目的或角度的集合称为语境。因此,在一定的语境中,元数据就是关于数据的数据。

由于元数据也是数据,因此,元数据可以存储在数据库中,并通过模型来组织。一些模型仅适用于特定的应用,其他一些则更通用。本标准的第 3 部分(注册系统元模型与基本属性)介绍和描述的模型是通用的。该模型表示的是人们对用来描述数据构造的元数据以及元数据之间关系的理解,却不一定涉及在元数据注册系统应用中元数据是如何表示的。这种模型称为概念模型。概念模型用于人类的阅读和理解。

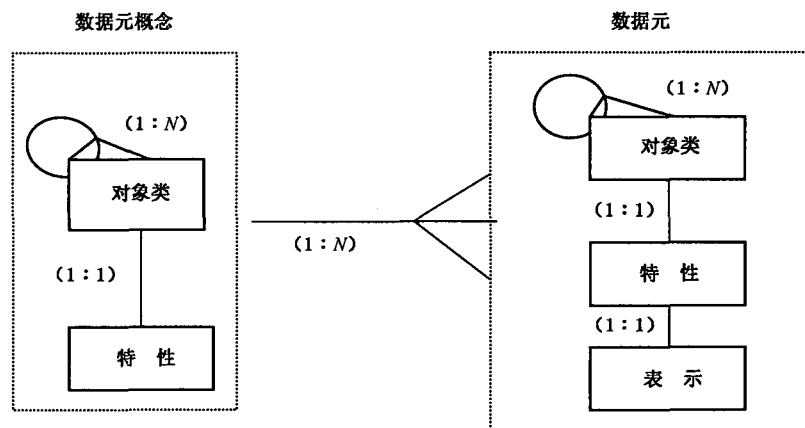
描述元数据的模型常常被称为元模型。在这个意义上,GB/T 18391.3 介绍的概念模型是一个元模型。

6.2 数据元的基本模型

图 1 并非规范性的,但阐明了本条的基本思想。

鉴于 GB/T 18391 的目的,一个数据元由以下两部分组成:

- 数据元概念**:数据元概念是能以一个数据元的形式表示的概念,其描述与任何特定表示法无关。
 - 表示**:表示由值域、数据类型、计量单位(如果需要)、表示类(可选)组成。
- 基于数据建模的视角和 GB/T 18391 的目的,一个数据元概念由两部分组成:
- 对象类**: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。
 - 特性**:一个对象类所有成员所共有的特征。



本图为资料性的,而非规范性的。

图1 数据元的基本模型

对象类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,例如轿车、人、家庭、雇员和订单等。特性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(不包括相关的数据类型),例如颜色、模型、性别、年龄、收入、地址、价格等。

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。上段的例子就是一般概念。记录级数据以这种方式描述。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“美国自然人集合”或“澳大利亚服务行业公司集合”。聚合数据以这种方式描述。特性作为个别概念的例子有:平均收入或总收入。

将一个实际的对象类或特性和它们的名称区分开来是很重要的。这是概念和它们的指称之间的差别。对象类和特性是概念;它们的名称是指称。复杂性起因于人们通过词汇(指称)来传达概念,这样很容易对概念和表示概念的指称产生混淆。例如,绝大多数人看到收入这个词时,一定会确信他们可以无歧义地解释它。但是,收入这一指称传达给所有读者的概念可能并不相同,而且,更为重要的是,收入的每个实例(场合)可能并不指向同一个概念。

也不是所有的想法都能用一种自然语言来简单地表达。例如,“women between the ages of 15 and 45 who have had at least one live birth in the last 12 months”是一个有效的对象类,但却难以用英语命名。某些想法用一种自然语言可能比用另一种自然语言更容易表述。德语单词 *Götterdämmerung* 就没有简单的英语对应词。

当一个表示与一个数据元概念联系在一起时,就产生了一个数据元。表示描述数据的形式,包括值域、数据类型、表示类(可选的),必要时还包括计量单位。值域是数据元允许值的集合。例如,表示家庭年收入的数据元可能用非负整数集(以美元为单位)作为它的一个有效值集。这就是它的值域。

要生成概念上相近的数据元时,如果必要,一个数据元概念可以与不同的值域关联。表示相似事实

的方法有许多,但是,事实反映的概念是相同的。以个人出生国这个数据元概念为例,GB/T 2659 中包含了世界各国 7 种不同的表示法。每种都包含了一个值集合,都可以用作与该数据元概念关联的值域。7 种关联中的每一种就形成一个数据元。对于数据的不同表示法,允许值、数据类型、表示类和可能的计量单位都会随之发生变化。

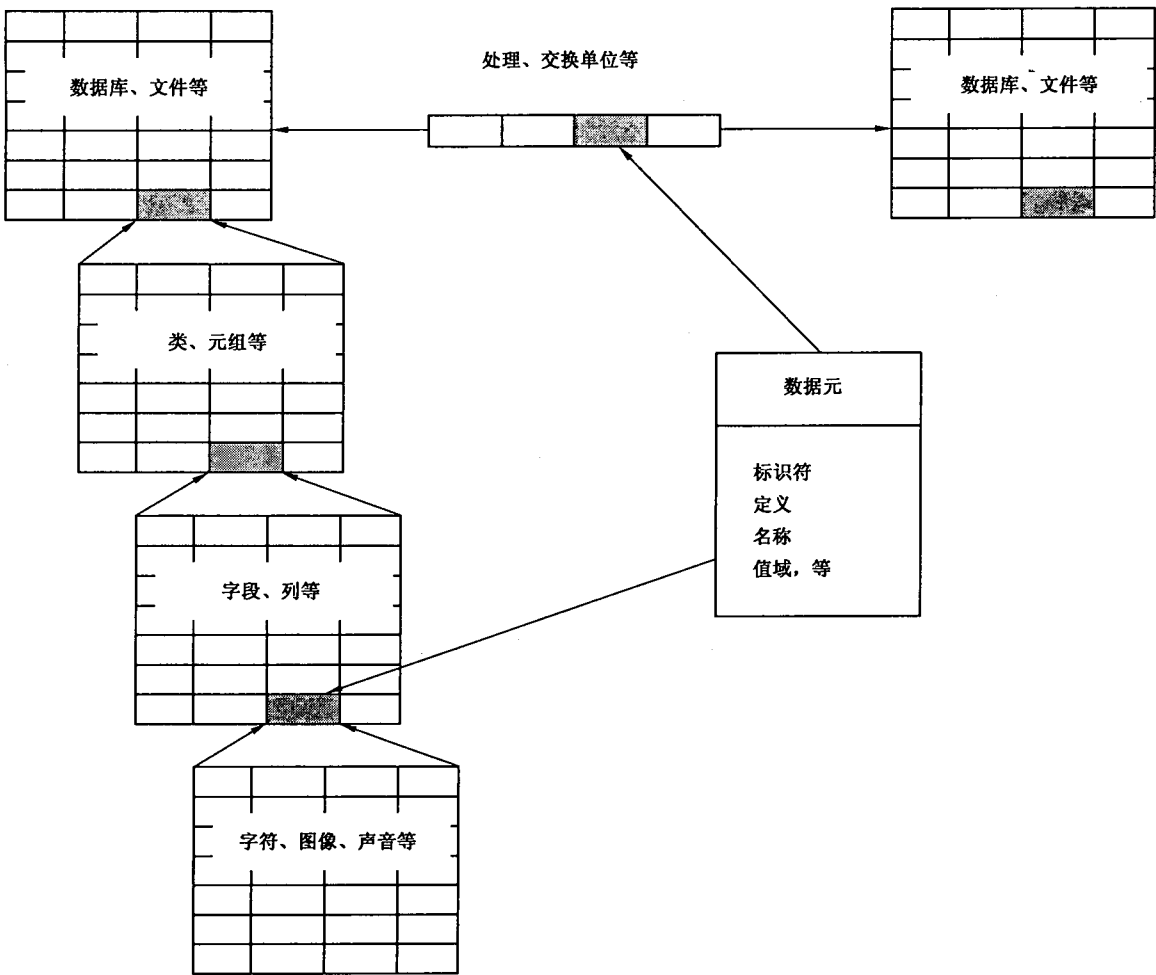
关于数据元描述注册和管理的详细内容参见 ISO/IEC TR 20943-1:2002《获得元数据注册系统内容一致性的规程 第 1 部分:数据元》。

6.3 数据管理与交换中的数据元

图 2 简明刻画了数据元存在的场合。数据元出现在数据库、文档和事务集中。数据元是组织管理数据的基本单位,因此,数据元对于设计组织内的数据库、文档以及组织间用于数据通信的事务集是必不可少的组成部分。

在组织内部,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。数据元本身包含有字符、图像、声音等多类数据。

组织需要将数据传输给其他组织时,数据元是构成事务集的基本单元。事务主要发生于数据库间或文件间,但文件和数据库结构(如记录或元组)在不同组织间并不一定相同。因此,信息(数据加上理解)传输的公共单元就是数据元。



本图为资料性的,而非规范性的。

图 2 数据元和其他数据概念

6.4 值域的基本模型

图 3 本身并非规范性的,但阐明了本条的基本观点。

值域是允许值的集合。一个允许值是某个值和该值的含义的组合。该关联含义称为**值含义**。值域是一个或多个数据元有效值的集合,用于在信息系统和数据交换中确认数据的有效性。它也是描述一个数据元的元数据的必要构成部分。特别是,值域是数据元所表示的数据内容、形式和结构的指南。

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关联含义的一个列表。每对值和含义称为一个**允许值**。每个值的含义称为**值含义**。

一个不可枚举值域由一个描述来规定。不可枚举值域描述准确描述了属于该值域的允许值,例如“大于0小于1的全部实数”。

概念的外延构成了**概念域**,每个值域都是概念域的一个元素。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。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,所以可能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,就能够获得相关值域和它们关联概念的结构。

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枚举。该类型概念域对应于可枚举类型的值域。不可枚举概念域的值含义由称为**不可枚举概念域描述**的规则来表述。这样,值含义并非明确的列出,这个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。参见 ISO/IEC TR 20943-3《元数据注册系统内容一致性的规程 第3部分:值域》。

描述数据有时需要计量单位。如果在数据库中记录温度,为了准确理解值的含义,就必须使用计量单位(如,华氏或摄氏)。再如,火星上岩石的质量由克来计量。但是,计量单位不仅仅限于物理量,货币(如,美元、里拉、英镑)和其他社会经济度量标准也是计量单位。

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,所有的货币具有相同的维。速度的计量,如海里/小时或米/秒,具有相同的维。两个经常被错误认为具有相同维的计量单位是:磅(重量单位)和克。磅是计量力的,而克是计量质量的。

计量单位与值域关联,维与概念域关联。

一些值域中包含了彼此很相似的值。要么是值本身相似,要么是值的含义相同。当出现这些相似时,这些值域也许是同一个概念域的外延。下面的示例表明了本条这个观点和其他观点。

示例1:相似的不可枚举值域

概念域名称:概率

概念域定义:大于0小于1的实数

值域名称(1):概率——2位有效数字

值域描述:精确到2位的所有大于0小于1的实数。

计量单位精确度:小数点后精确到2位

值域名称(2):概率——5位有效数字

值域描述:精确到5位的所有大于0小于1的实数。

计量单位精确度:小数点后精确到5位

示例 2:相似的可枚举值域

概念域名称:世界各国
概念域定义:现有世界各国名称或代码的列表
值域名称(1):国家代码——2 位字母码
允许值:
〈AF,基本地理政治实体“阿富汗民主共和国”〉
〈AL,基本地理政治实体“阿尔巴尼亚人民社会主义共和国”〉
...
〈ZW,基本地理政治实体“津巴布韦共和国”〉
值域名称(2):国家代码——3 位字母码
允许值:
〈AFG,基本地理政治实体“阿富汗民主共和国”〉
〈ALB,基本地理政治实体“阿尔巴尼亚人民社会主义共和国”〉
...
〈ZWE,基本地理政治实体“津巴布韦共和国”〉

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来表示概念,与数据元无关。

关于值域注册和管理的详细示例见 ISO/IEC 20943-3《元数据注册系统内容一致性的规程 第 3 部分:值域》。

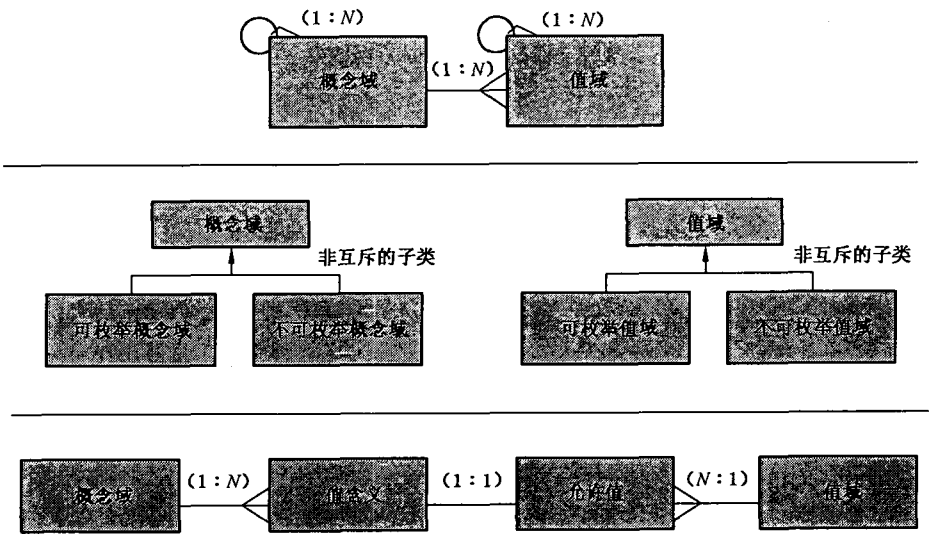


图 3 值域的基本模型

6.5 分类方案的基本原理

鉴于 GB/T 18391 的目的,分类方案是旨在对对象进行分类的一个概念体系。它有特定的结构,限定的内容范围,并将对象指定给体系中已经定义的概念。将概念指定给一个对象的过程称之为分类。概念体系中联系概念的关系联系着其相关概念分类的对象。总之,任何用于对象分类的概念体系,都是一个分类方案。

分类方案的内容范围划定分类方案覆盖的主题范围。分类方案的范围是方案概念体系中最宽泛的

概念,理论上,它决定一个对象是否可以在该方案下进行分类。

概念体系,特别是分类方案可以通过许多途径进行结构化。该结构定义可能存在于概念之间的关系类型,每个分类方案可以使概念与对象关联。在一个特定的分类方案中,关联概念和方案中与关联概念相关的其他概念一起,提供了一个可以用于理解对象含义的概念框架。该框架受分类方案范围的限定。

一个概念体系可以由一个术语体系来表示。指称用于表示体系中的每一个概念,并可以用做关联对象的关键字,以便达到搜索、索引或其他目的。

关系体系是概念体系的一个特殊类型。关系体系中的关系类型就是概念。一个关系类型有 N 个自变量,称为一个 n 元关系类型。“ N 个对象的集合使用 n 元关系类型分类”这个陈述意味着 N 个对象之间存在着一种给定关系类型的关系。

7 元数据注册系统

7.1 概述

元数据也是数据,因此,元数据可以存储在一个数据库中。支持注册功能的一个元数据的数据库就是一个元数据注册系统(MDR)。GB/T 18391.3 给出了一个描述数据的元数据注册系统的概念模型。GB/T 18391.6 描述了 GB/T 18391 注册各方面的要求和规程。对于实际的元数据注册系统,可能还有其他的要求和规程,但这些不在本标准的范围之内。GB/T 18391.4 和 GB/T 18391.5 分别描述了生成良好定义和开发命名协议的规则、指南。GB/T 18391.2 描述了分类的作用。ISO/IEC TR 20943-1 描述了注册数据元时的建议和准则。ISO/IEC TR 20943-3 描述了注册值域时的建议和准则。

一个元数据注册系统包含描述数据构造的元数据。用于描述一个特定数据构造(如,数据元)的所有属性总体上称为一个元数据对象。当具体描述一个特定的数据构造时,实例化后的这些属性称为一个元数据项。元数据项注册之后(如,将元数据登录在元数据注册系统中)成为注册系统项。如果这个注册系统项(如,一个数据元)需要进行管理,也称为管理项。

注:在一般用法上,注册一个描述数据构造的元数据项也被认为是注册该数据构造。实际上,数据构造并不存储在元数据注册系统中,存储的是其描述。这类似于政府跟踪机动车辆的注册系统。每个机动车辆的一个描述登录在注册系统中,而不是车辆本身。但是,人们认为他们注册的是自己的机动车辆,而不是描述。

7.2 GB/T 18391 MDR 的总体模型

GB/T 18391 的概念模型包含两个主要部分:概念层和表示(或句法)层。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。第 6 章给出了图 4 包括的每个类的描述。

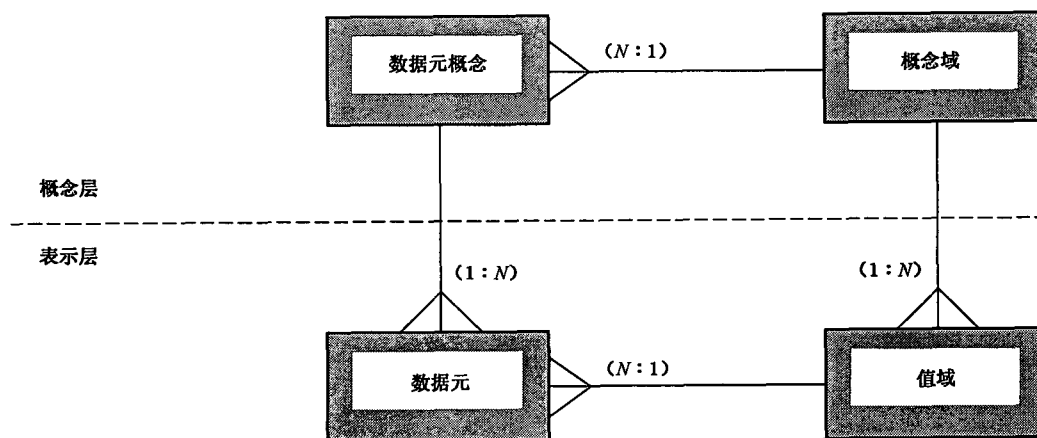


图 4 GB/T 18391 元数据注册系统总体模型

图 4 给出了 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表示(主要是一个值域)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个不同方式表示；
- 数据元可以共享相同的表示,这意味着一个值域可以被其他数据元重复利用；
- 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；
- 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；
- 许多值域可以共享相同的概念域；
- 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。

第 6 章描述了图 4 中没有涉及到的其他事实,在此应提及的有：

- 数据元概念之间的关系可以在一个元数据注册系统中得以维护,这就意味着数据元概念的一个概念体系可以得到维护。
- 概念域之间的关系可以在一个元数据注册系统中得以维护,这就意味着概念域的一个概念体系可以得到维护。

本章随后将描述元数据注册系统中元数据的注册与管理的一些基本内容。

7.3 注册概要

GB/T 18391.6 规定的注册和管理功能将 MDR 与元数据数据库区分开来。这些功能的实现途径是 GB/T 18391.3 规定的元模型设计的一个主要的部分。

注册是应用于 MDR 的一系列的规则、操作和规程。GB/T 18391.6 详细描述了如何遵照 GB/T 18391 进行注册。注册带来的三个最重要的结果是监控元数据质量的能力、起源(元数据的来源)和向 MDR 中每个描述对象赋予标识符。注册还需要一系列的规程来管理注册系统、提交用于注册的对象的数据、维护已提交元数据的主题域。对于一个元数据注册系统的实际实现,可能还有一些其他的要求,但这些不在本标准范围之内。

每个管理项都用一个统一和规定好的方式来维护。标识符、质量措施、主管组织、名称和定义都是一般元数据需要管理的部分。注册是产生或维护用于管理的和用于其他目的的详细元数据的过程。

通过使用注册状态来监控元数据质量。状态记录了质量的级别。GB/T 18391.6 规定了各个级别。每个管理项被赋予一个注册状态,该状态可以随时间而变化。另外,元数据质量是多层面的,也就是说,监控元数据质量有若干个目的。主要的目的有：

- 监控用于每个属性的元数据是否遵守了规则；
- 监控形成定义、生成名称和进行分类时,是否遵守了约定；
- 确定一个管理项是否仍然适用；
- 明确相关管理项之间的相似性,并协调其差异；
- 确定是否可以获得一些管理项更高质量的元数据。

GB/T 18391.6 描述了生成和赋予标识符的规则。MDR 中的每个管理项将被赋予一个唯一的标识符。

注册机构是负责设立规程,管理和维护 MDR 的组织。提交组织负责提请一个新元数据项在注册系统中的注册。主管者负责每个注册项的主题内容。GB/T 18391.6 描述了上述各个角色。

8 GB/T 18391 综述

8.1 各部分介绍

本条介绍了 GB/T 18391 的各部分的要点及重要性。

8.1.1 第1部分

GB/T 18391.1 框架,介绍并讨论了解本套标准所必需的数据元、值域、数据元概念、概念域和分类方案的基本概念,并提供了使 GB/T 18391 各部分关联起来的语境。

8.1.2 第2部分

GB/T 18391.2 分类,提供了管理分类方案这个概念体系的概念模型。这些方案中的概念通过分类过程而与管理项关联。图书馆员、术语学家、语言学家和计算机科学家更精通分类过程,因此本部分不再对此进行描述。从分类中派生出的其他的语义内容才是重点。

将一个对象与来自一个或多个分类方案中的一个或多个概念关联能提供:

- 对对象的更多的理解;
- 相似对象的比较信息;
- 主题域(由分类方案的范围定义)语境中对于一个对象的理解;
- 区别相似概念含义细微差异的能力。

因此,分类方案的管理是最大化 MDR 中信息潜能的重要部分。GB/T 18391.2 为此提供了框架。

8.1.3 第3部分

GB/T 18391.3 注册系统的元模型及基本属性,规定了 MDR 的一个概念模型。该模型的范围仅限于数据元、数据元概念、值域、概念域、分类方案以及其他相关类的基本属性集。GB/T 18391.3—2001 规定的元数据基本属性也包括在此修订版中。

注册系统元模型用 UML 表达,为了方便阅读,将其分为 6 个区。模型中的所有内容都在文本中得到了描述。图表注释框中的若干内容也在文本中得到了描述。

文件中给出了模型中使用到的所有建模构造(类、属性和关系)的列表。这个属性集合就是“基本属性”。第 2、4、5、6 部分描述的属性都包含在注册系统的元模型之中。

注册系统元模型并不是一个组织希望记录的所有元数据的完全描述,因此,设计中,该模型可以根据需要进行扩展。但是,扩展就本质而言,并不是本标准的组成部分。

该部分还用一章描述了符合一致性的条件,一致性表现为严格一致性(符合所有条款)和一致性(符合所有条款,但存在其他条款)。

8.1.4 第4部分

GB/T 18391.4 数据定义的形成,是开发无歧义数据元定义的手册。GB/T 18391.4 给出了大量规则与指南,确切规定了如何形成数据定义。准确、形式完好的定义是获得数据一致理解的一个关键条件。形式完好的定义对于信息的交换也是必要的。只有每一位用户对数据有共同和准确的理解,其交换才能是无障碍的。

定义的有效性是元数据质量的一个方面,遵循第 4 部分的规则与指南可以保证定义的有效性。

8.1.5 第5部分

GB/T 18391.5 命名与标识原则,是赋予管理项指称的手册。指称是命名和标识一个特定数据构造的宽泛术语。

通过命名约定产生数据构造的名称。命名约定是特定语境中生成名称的算法,形成一个命名约定需要语义、句法和词法规则。名称仅是提供数据构造某些语义的一种简单手段,但语义并不完整。句法和词法解决成分(如,允许的字符)、格式以及其他方面的问题。

可以赋予数据构造多个名称,但可以将其中一个作为首选的名称。通常,每个名称在产生该名称的语境范围内使用。

标识符是在元数据的管理和交换中用于表示数据构造的指称。注册机构分配给每个数据构造一个唯一的标识符,该标识符在注册系统的语境中是独一无二的。数据构造的标识符可以用来无歧义地定位数据构造。

8.1.6 第6部分

GB/T 18391.6 注册,提供了为每个数据构造分配一个独一无二的标识符以及注册申请者如何在一个注册机构注册数据构造的操作指南。本部分对已注册管理项的维护也做了规定。注册主要解决MDR中元数据的标识、质量和出处等问题。

管理项的标识符由注册机构的标识符、注册机构内分配给管理项的唯一标识符和版本组合而成。每个注册系统由数据构造逻辑上和功能上隶属的注册机构来维护。例如,有关化学的数据构造很有可能注册在化学品厂商注册机构下。

注册远比仅仅表明一个元数据项是否已被注册来的复杂。尽管倾向于坚持只注册“好的”元数据,但很难实际操作。因此,将管理项质量的改进划分为不同水平,称作注册状态。另外,在这些质量水平间还存在着管理状态的层次。整体上,这些状态层次被称为管理状态。它们表明在现行注册生命周期中一个管理项所处的位置。

MDR管理元数据的出处和责任链,描述了注册机构、数据主管者、主管组织和提交组织的任务和角色,并给出了注册流程的框架。

注册既是过程也是目标。标识符、质量状态、生命周期状态的制定以及出处的描述都是目标。实现这些目标的规则是过程。

8.2 GB/T 18391 第1~6部分应用的基本规则

GB/T 18391的每部分有助于元数据产生、组织和注册的不同侧面。每部分应与其他部分配合使用。GB/T 18391.1确立了各部分间的关系,并就作为一个总体如何应用给出了指导。GB/T 18391.3规定了注册申请者应为每一个注册对象提供的元数据项,并给出了每个基本属性的详尽特征。由于名称、定义和标识在管理描述数据构造的元数据中很重要,在以下两个部分进行了专门而又广泛的论述,构造数据元的定义应遵循GB/T 18391.4,标识和命名应遵循GB/T 18391.5确立的原则。GB/T 18391.2规定了一组用于分类方案及其成分注册和管理的属性。元数据项可以作为注册系统项注册在MDR中,并作为管理项得到管理,GB/T 18391.6就这些步骤提供了指南。

9 一致性

本部分并没有特定的一致性标准。GB/T 18391.1是将各部分标准联结起来的一个框架。因此,本部分不存在一致性的问题。本标准的其他各部分都有自己的一致性条款。

参 考 文 献

- [1] GB/T 1988 信息技术 信息交换用七位编码字符集
- [2] GB/T 2659—2000 世界各国和地区名称代码
- [3] GB/T 4880(所有部分) 语种名称代码
- [4] GB/T 5271 信息技术 词汇
- [5] GB/T 7408 数据元和交换格式 信息交换 日期和时间表示法
- [6] GB/T 9387.1 信息技术 开放系统互连 基本参考模型 第1部分:基本模型
- [7] GB/T 11383 信息技术 信息交换用八位代码结构和编码规则
- [8] GB/T 15273 信息处理 八位单字节编码图形字符集
- [9] GB/T 15417 文献 多语种叙词表编制规则
- [10] GB/T 16262 信息技术 抽象语法记法—(ASN.1)
- [11] GB/T 16647 信息技术 信息资源词典系统(IRDS)框架
- [12] GB/Z 18219 信息技术 数据管理参考模型
- [13] 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
- [14] GB/T 20000.1—2002 标准化工作指南 第1部分:标准化和相关活动的通用词汇
- [15] GB/T 20001.1—2001 标准编写规则 第1部分:术语
- [16] Barr, Avron & Feigenbaum, Edward A. *The Handbook of Artificial Intelligence* (3 Volumes) William Kaufman, Inc., 1981
- [17] Gosh, Sakti P. *Data Base organization for Data Management*, Second Edition. New York: Academic Press, Inc., 1986
- [18] Herman G T. *Theory Algorithms*. Ralston, Anthony & Reilly, Jr., Edwin D. *Encyclopedia of Computer Science and Engineering*, Second Edition. Van Nostrand Reinhold Company Inc., 1983
- [19] ISO/IEC 2375 *Information technology—Procedure for registration of escape sequences and coded character sets*
- [20] ISO 2788 *Documentation—Guidelines for the establishment and development of monolingual thesauri*
- [21] ISO 6093 *Information processing—Representation of numerical values in character strings for information interchange*
- [22] ISO 6862 *Information and documentation—Mathematical coded character set for bibliographic information interchange*
- [23] ISO/IEC 6937 *Information technology—Coded graphic character set for text communication—Latin alphabet*
- [24] ISO TR 9007 *Information processing systems—Concepts and terminology for the conceptual schema and the information base*
- [25] ISO/IEC TR 9789 *Information technology—Guidelines for the organization and representation of data elements for data interchange—Coding methods and principles*
- [26] ISO/IEC 10646 *Information technology—Universal Multiple—Octet Coded Character Set (UCS)*
- [27] ISO/IEC TR 20943 (all parts) *Information technology—Procedures for achieving meta-data registry content consistency*

[28] Judith J Newton, Daniel C Wahl. *Manual for Data Administration*. National Institute of Standards and Technology, 1993

[29] Tasker, Dan. *Fourth Generation Data*. Prentice Hall, 1989

中 华 人 民 共 和 国
国 家 标 准
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

*

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 1.75 字数 43 千字
2009年12月第一版 2009年12月第一次印刷

*

书号:155066·1-39521 定价 27.00 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 18391.1-2009