



中华人民共和国国家标准

GB/T 18391.4—2009/ISO/IEC 11179-4:2004
代替 GB/T 18391.4—2001

信息技术 元数据注册系统(MDR) 第4部分:数据定义的形成

Information technology—Metadata registries (MDR)—
Part 4: Formulation of data definitions

(ISO/IEC 11179-4:2004, 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
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

2 一致性 1

3 术语和定义 1

4 数据定义要求与建议的概要 2

4.1 要求 2

4.2 建议 3

5 规定 3

5.1 前提 3

5.2 要求 3

5.3 建议 5

参考文献..... 7

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4部分,等同采用 ISO/IEC 11179-4:2004《信息技术 元数据注册系统(MDR) 第4部分:数据定义的形成》(英文版),仅有编辑性修改。

本部分代替 GB/T 18391.4—2001《信息技术 数据元的规范与标准化 第4部分:数据定义的编写规则与指南》。

本部分与 GB/T 18391.4—2001 相比主要变化如下:

- 名称变更为:《信息技术 元数据注册系统(MDR) 第4部分:数据定义的形成》;
- 4.1 和 5.2 中的“规则”改为“要求”;
- 4.2 和 5.3 中的“指南”改为“建议”;
- 取消了 4.1a) 具有唯一性(在出现此定义的任何数据字典中);
- 增加了 4.2h) 适合被定义的元数据项的类型。

为了保持与英文版本的一致性,英文版中斜体表示的内容本部分也使用斜体表示,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归口。

本部分由中国标准化研究院负责起草。

本部分主要起草人:刘植婷、邢立强、李小林、史立武、孙广芝。

本部分于 2001 年首次发布,本次为第一次修订。

引 言

数据处理和电子数据交换很大程度上依赖于数据库中数据记录的准确性、可靠性、可控性和可校验性。数据的正确而恰当的使用与解释的前提是数据的提供者和使用者都对数据的含义和表示有共同的理解。为了促进这种共同的理解,应定义数据的若干特性或属性。这些数据的特性通称为“元数据”,即“描述数据的数据”。GB/T 18391 的本部分给出了基于元数据注册系统的数据定义形成的要求与建议。这些定义的目的在于确定、描述、解释和澄清数据的含义,以促进数据元的标准化和可重复使用,并促进信息系统的数据共享与集成。

元数据注册系统的结构用一个概念数据模型的形式来规定。元数据注册系统被用来保存数据元和有关概念的信息,例如,“数据元概念”、“概念域”和“值域”。通常,它们都被看作是“元数据项”。这样的元数据对于清晰地描述、记录、分析、分类和管理数据是必需的。

GB/T 18391 的本部分规定的定义要求与建议并不总是适用于在词汇和语言字典中出现的术语定义。应用在语言字典和元数据注册系统中的要求之间存在着差异。GB/T 18391 的要求比自然语言字典的要求有更多的限制。例如,语言字典可以有包含术语或字词的多种意义的多种定义,然而为特定语境所研制的数据定义不宜在该语境内具有多种意义。数据定义是用来说明一个概念或多个概念的,这些概念通过数据集、数据值、数据元或其他元数据项来表示。可以确立一个单一的参考定义,其他定义可以声明与其等价(如,可以在一种语言中建立一个参考定义,其他语言的定义可以声明与其等价)。元数据项可以在特定语境内有一个首选定义,不赞成使用其他的定义。

许多数据定义包含着自身需要被定义的术语(例如,“费用”,“津贴”,“递送”)。某些术语在不同的行业部门中可能具有不同的定义。因此,大多数元数据注册系统有必要对定义所用术语建立一个关联词汇表。

信息技术 元数据注册系统(MDR)

第4部分:数据定义的形成

1 范围

GB/T 18391 的本部分规定了构建数据和元数据定义的要求与建议。本部分仅描述了定义的语义方面,而不考虑定义的格式。尤其当用于 GB/T 18391 第3部分所规定的元数据注册系统的内容时,可广泛用于确定数据和元数据的定义。

这些要求与建议适合于形成数据元和其他数据构件类型的定义,这里的数据构件类型可以是实体类型、实体、关系、属性、对象类型(或类)、对象、复合物、代码条目、元数据项,以及由 XML 标记引用的数据。

2 一致性

GB/T 18391 的本部分可以独立地使用,例如,用于定义一个元数据注册系统语境外的数据元。如果在研制定义中已经遵循了这些要求与建议,就可以声明与本部分一致。

在 GB/T 18391 元数据注册系统的语境中,当确立 GB/T 18391.6 中所规定的注册状态时,本部分应构成关于定义的规则。

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

概念域 Conceptual Domain

有效的值含义的一个集合。

注:概念域中的值含义既可枚举也可通过描述表达。

[GB/T 18391.3—2009,定义 3.3.21]

3.2

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]

3.3

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]

3.4

数据元 data element

由一个属性集合规定其定义、标识、表示和允许值的一个数据单元。

[GB/T 18391.3—2009,定义 3.3.36]

3.5

数据元概念 Data Element Concept

以一个数据元的形式来表示的一个概念,其描述与任何特定表示法无关。

[GB/T 18391.3—2009,定义 3.3.38]

3.6

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]

3.7

元数据 metadata

定义和描述其他数据的数据。

[GB/T 15237.1—2000,定义 3.2.18]

3.8

元数据注册(系统) Metadata Registry

用于注册元数据的信息系统。

[GB/T 18391.3—2009,定义 3.2.22]

3.9

元数据项 metadata item

元数据对象的一个实例。

注1: 在 GB/T 18391 各部分中,本术语仅适用于 GB/T 18391.3—2009 第4章由元模型描述的元数据对象的实例。

示例包括数据元、数据元概念、允许值等的实例。

注2: 一个元数据项有相关的属性,适合于其所实例化的元数据对象。

[GB/T 18391.3—2009,定义 3.2.19]

3.10

元数据对象 metadata object

由元模型定义的一个对象类型。

注: 在 GB/T 18391 各部分中,该术语仅适用于 GB/T 18391.3—2009 第4章由元模型描述的元数据对象。示例

包括数据元、数据元概念、允许值等。完全列表见 GB/T 18391.3—2009 的 3.3。

[GB/T 18391.3—2009,定义 3.2.20]

3.11

名称 name

用语言表述的一个对象的指称。

[GB/T 18391.3—2008,定义 3.2.26]

3.12

值域 Value Domain

允许值的集合。

注1: 值域提供表示,但并不隐含与值可能关联的数据元概念,也不表明值的含义。

注2: 允许值可枚举,或可通过一个描述来表达。

[GB/T 18391.3—2009,定义 3.3.140]

4 数据定义要求与建议的概要

为了方便使用者,本章提供不含解释的要求和建议,目的是在理解了这些要求和建议之后使本部分便于使用。对这些要求与建议的说明和示例在第5章给出,以确保对其准确含义的理解。

4.1 要求

数据定义应:

a) 用单数形式阐述;

- b) 要阐述其概念是什么,而不是仅阐述其概念不是什么;
- c) 用描述性的短语或句子阐述;
- d) 仅可使用人们普遍理解的缩略语;
- e) 表述中不应包括其他数据或基本概念的定义。

4.2 建议

数据定义宜:

- a) 阐述概念的基本含义;
- b) 准确而无歧义;
- c) 简练;
- d) 能单独成立;
- e) 表述中不应加入理由、功能用法、领域信息或程序信息;
- f) 避免循环定义;
- g) 对相关定义使用相同的术语和一致的逻辑结构;
- h) 适合被定义的数据项的类型。

5 规定

5.1 前提

数据用于特定的目的。使用上的不同,需要某些要求与建议具有不同的表现形式。例如,数据定义在不同的语境中通常要求有不同的层级。5.3a)列举了不同定义需要不同层级的实例。5.3a)“阐述概念的基本含义”的实现取决于语境。传达一个特定定义基本含义的主要特性,因数据的通用或专用层级而异。例如,在航空运输业的语境中“飞机场”这个概念的通用定义也许是适用的,而在商业航空运输业的语境中,定义“飞机场”这个概念所需的主要和基本特性可能是专用的。在元数据注册系统内,由于语言的不同或者群体(如儿童、普通公众或学科领域的专家们)的不同,可以有多个等价定义。有关不同语境中概念之间关系的论述以及如何利用特性来区别不同的概念,见 GB/T 10112—1999 第 5 章。撰写的定义应使用户和共享数据的接收方易于理解。

5.2 要求

为了使有关规范化数据定义结构的要求易于理解,下面给出了有关说明和示例。每项要求之后都附有对其含义的简要说明。说明之后,则是一些示例。每项说明均给出了好的示例,必要时还给出了常见的不好的示例,以说明定义的编写错误。为了说明好的示例与不好的示例之间的差别,在这些示例后附有对原因的说明。应注意,下面是有关数据元定义的示例,并且这些定义是说明性的。

数据定义应:

a) 以单数形式阐述

说明:数据定义所表述的概念应使用单数形式表述(概念本身为复数时除外)。

示例:“物品编号”

1) 好的定义:标识一种物品的一个参考号。

2) 不好的定义:标识物品的参考号。

原因:不好的定义使用了复数的“物品(articles)”,含义不清楚,因为它可能隐含着一个“物品编号”指代多件物品。

b) 要阐述其概念是什么,而不仅是阐述其概念不是什么

说明:编写定义时,仅阐述其概念不是什么并不能对概念作出唯一的定义。

示例:“运费金额”

1) 好的定义:发货人把货物从一个地方运到另一个地方所花费的费用金额。

2) 不好的定义:与包装、提供的文件、装货、卸货和保险无关的各项费用。

原因:不好的定义没有明确说明这个数据的含义。

c) 用描述性的短语或句子阐述(在大多数语言中)

说明:(在大多数语言中)需要使用短语来表述包含概念的基本特性的准确定义。不能使用一个或多个简单的同义词来陈述,也不能以不同的顺序简单地来重复这些名称词。如果需要使用多个描述性短语,则应使用完整的、语法正确的句子。

示例:“代理人名称”

1) 好的定义:被授权代表另一方的那一方的名称。

2) 不好的定义:代表。

原因:“代表”是这个数据元名称的近似同义词,不适宜用作定义。

d) 仅可使用人们普遍理解的缩略语

说明:对包括简称和首字母在内的缩略语含义的理解,通常受到特定环境的限制。环境不同,同一缩写词也许会引起误解或混淆。因此,为了避免歧义,在定义中应使用全称,而不用缩略语。

如果缩略语是人们普遍理解的词,例如“即(i. e.)”和“例如(e. g.)”,或缩略语比合成词的全称更容易理解,而且其本身已被作为一个词采用,例如“雷达(radar)”代表“无线电定位装置(radio detecting and ranging)”,这种缩略语作为例外,可以在定义中使用。

所有简称在第一次出现时,必须予以说明。

示例 1:“潮高”

1) 好的定义:从平均海平面(MSL)到具体潮水平面的垂直距离。

2) 不好的定义:从 MSL 到具体潮水平面的垂直距离。

原因:不好的定义表达的不清楚,因为缩略语 MSL 不为人们普遍理解,一些使用者可能需要参阅其他资料才可确定其含义。因为没有全称,所以在词汇表中很难找到或根本找不到这个术语。

示例 2:“密度计量单位”

1) 好的定义:测量物质密度所使用的单位,以单位体积质量(m. p. u.)表示(例如:磅/立方英尺;千克/立方米)。

2) 不好的定义:测量物质密度所使用的单位,以 m. p. u. 表示(例如:磅/立方英尺;千克/立方米)。

原因:m. p. u. 不是常用的缩略语,一些使用者可能不理解它的含义。对该缩略语应写出全称。

e) 表述中不应包括其他数据定义或其他内涵概念

说明:如下面的示例所示,次要的数据元或相关概念的定义不应出现在主要数据元定义中。术语的定义应使用相关术语词汇编写。如果有次要的数据元定义,可用一个注释,附在主要定义正文的末尾或作为词典的分立条目。相关定义可通过关系属性(例如,相互参照)得到。

示例 1:“样品类型码”

1) 好的定义:标识样品种类的代码。

2) 不好的定义:标识被采集的样品种类的代码。样品是一个为试验而采集的少量样本。它既可用做试验的实际样品,又可用做质量控制的代替样品。质量控制样品是用来检验实际样品结果的代替样品。

原因:不好的定义嵌入了两个与当前定义无关的定义。它们是“样品”的定义和“质量控制样品”的定义。

示例 2:“发行银行跟单信用证号码”

1) 好的定义:由发行银行分配给跟单信用证的号码。

2) 不好的定义:由发行银行分配给跟单信用证的号码。跟单信用证是一份文件,在此文件中,银行声明它已签发了跟单信用证,凭此跟单信用证,受益人可以得到支付、接受或按照某些条款和条件进行协商,无须出示担保文件和诸如汇票一类的东西。

原因:不好的定义含有一个应收入词汇表的概念定义。

5.3 建议

数据定义宜：

a) 阐述概念的基本含义

说明：概念中的所有基本特性应出现在语境所对应的相关层级的定义中，避免包含非基本特性。层级的详尽程度取决于系统用户和环境的要求。

示例 1：“托运货物装货序号”（预期语境：任何运输形式）

- 1) 好的定义：指明托运货物在一种运输工具或运输设备中的装载顺序号。
- 2) 不好的定义：指明托运货物在卡车上的装载顺序号。

原因：在预期语境下，托运货物可由各种运输方式实现，如卡车、船只或货运列车。托运货物并不局限于卡车运输。

示例 2：“发票金额”

- 1) 好的定义：一张发票上的收费总计。
- 2) 不好的定义：发票上所列一切收费项目的总计，一方面要考虑减去诸如折扣和让利等数额，另一方面要加上保险费、运输费、装卸费等费用。

原因：不好的定义包含了与主题无关的内容。

b) 准确而不含糊

说明：被定义的概念的确切含义和解释应在定义中一目了然。定义应足够清楚明了，使之只存在一种解释。

示例：“收货日期”

- 1) 好的定义：收货方收到货物的日期。
- 2) 不好的定义：具体货物运抵的日期。

原因：不好的定义没有说明由什么来决定“货物运抵”了。“货物运抵”可理解为产品在预期的目的地卸下的行为，也可理解为客户实际收到这个产品的地点。有可能客户根本没有收到已被卸在货场的产品，也许客户在产品被卸在货场的几天之后才收到它。

c) 简练

说明：定义应简洁明了易于理解，应避免使用诸如“为了本元数据注册系统”、“要描述的术语是”之类无关的限定性短语。

示例：“字符集名称”

- 1) 好的定义：对数据进行编码的表音或表意符号集合的名称。
- 2) 不好的定义：为了元数据注册系统或其他地方所用的系统软硬件能力（处理一个或多个脚本中进行编码的数据）的目的，对数据进行编码的表音或表意符号集合的名称。

原因：在不好的定义中，在“对数据进行编码的表音或表意符号集合的名称”之前的所有短语，都是无关的限定性短语。

d) 能单独成立

说明：概念的含义，从定义本身就能体现出来。理解定义的含义，不需要附加说明和引证。

示例：“学校所在城市名”

- 1) 好的定义：学校所在城市的名称。
- 2) 不好的定义：见“学校地点”。

原因：不好的定义不能单独成立，它需要借助另一个定义（学校地点）才能理解本定义的含义。

e) 表述中不应加入理由、功能用法、领域信息或程序信息

说明：尽管它们通常是必要的，但这样的表述不应归入定义本身，因为它们含有与定义无关的信息。如果它们确实有用，可以把这类表述列为不同的元数据属性（见 GB/T 18391.3）。然而，在定义后增加示例是允许的。

- 1) 用于定义的原理不宜作为定义的组成部分（例如：如果一个数据元使用了英里而不用公

里,其理由不宜在定义中说明)。

- 2) 诸如“此数据元不应用于……”的功能用法不宜出现在定义中。
- 3) 关于程序方面的注释,例如,“这个数据元与数据元×××一同连用”这样的话不宜在定义中出现,而应当使用 GB/T 18391.3 规定的“相关数据参照”和“关系类型”属性。

示例:“数据字段标记”

- 1) 好的定义:索引、主题词表、查询、数据库等的数据字段标识。
- 2) 不好的定义:索引、主题词表、查询、数据库等的数据字段标识。数据字段标识用于诸如摘要、表格中的栏目等信息单元。

原因:不好的定义含有功能用法的陈述,以“数据字段标识用于……”开始的陈述应从定义中删去。如果必须有此陈述的话,应将其放在其他属性中。

f) 避免循环定义

说明:两个定义不应相互根据另一定义来为自身定义。一个定义也不应使用另一个概念的定义作为自身的定义。因为这样会导致定义一个概念时,需借助另一个概念,反之,在定义另一个概念时,又需借助原来已给出的那个概念。

示例:带有不好定义的两个数据元

- 1) “雇员 ID 号码”:分配给每一位雇员的号码。
- 2) “雇员”:对应于雇员 ID 号码的人员。

原因:每个定义其含义涉及另一个定义。每个定义都没有指明含义。

g) 对相关定义使用相同的术语和一致的逻辑结构

说明:对相类似的或相关的定义,应使用相同的术语和句法。

示例:下面的示例也阐明了这一观点。两个定义都与相关概念有关,因此应有相同的逻辑结构和相类似的术语。

- 1) 发货日期:供货方发送货物的日期。
- 2) 收货日期:收货方收到货物的日期。

原因:使用相同的术语和句法有利于理解。否则,使用者不知道使用同义的术语和变化了的句法是否意味着有一些不同。

h) 适合被定义的元数据项的类型

说明:元数据注册系统中不同类型的元数据项(例如,数据元概念、数据元、概念域和值域)都起着不同的作用,这些宜反映在定义中。

示例:

数据元概念:“工作等级最高工资额”。

定义:对于相关的工作等级所允许的最高工资。

注:这个数据元概念并不涉及特定的值域。

概念域:“货币金额”。

定义:用货币的单位来表示的总额。

注:此定义涉及货币流通的“幅度”,但不针对特定的货币。

数据元 1:“欧洲工作等级最高工资额”。

定义:对于相关的工作等级所允许的最高工资(用欧元表示)。

数据元 2:“美国工作等级最高工资额”。

定义:对于相关的工作等级所允许的最高工资(用美元表示)。

注:数据元定义可给出明确的值域,因为这样可以完全地将两个数据元区别开来。

参 考 文 献

- [1] ISO 704:2000 术语工作 原则与方法
 - [2] GB/T 15237.1—2000 术语工作 词汇 第1部分:理论与应用
 - [3] GB/T 5271.1—2000 信息技术 词汇 第1部分:基本术语
 - [4] GB/T 18391.3—2009 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
-

中 华 人 民 共 和 国
国 家 标 准
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

*

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 0.75 字数 18 千字
2009年12月第一版 2009年12月第一次印刷

*

书号:155066·1-39524 定价 16.00 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 18391.4-2009