



中华人民共和国国家标准

GB/T 18391.4—2001
idt ISO/IEC 11179-4:1995

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

Information technology—Specification and standardization
of data elements—Part 4: Rules and guidelines
for the formulation of data definitions



2001-07-16 发布

2002-03-01 实施



中华人民共和国
国家质量监督检验检疫总局 发布

中 华 人 民 共 和 国
国 家 标 准
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
GB/T 18391.4—2001

*

中国标准出版社出版
北京复兴门外三里河北街16号
邮政编码:100045

电话:68523946 68517548

中国标准出版社秦皇岛印刷厂印刷
新华书店北京发行所发行 各地新华书店经售

*

开本 880×1230 1/16 印张 1 字数 18 千字
2002年3月第一版 2002年3月第一次印刷
印数 1—2 000

*

书号: 155066·1-18077 定价 12.00 元
网址 www.bzcbs.com

*

科 目 597—525

版权专有 侵权必究
举报电话:(010)68533533



GB/T 18391.4—2001

02-597-525

前 言

本标准等同采用国际标准 ISO/IEC 11179-4:1995《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》。

GB/T 18391 在总标题《信息技术 数据元的规范与标准化》下,包括以下几部分:

- 第 1 部分:数据元的规范与标准化框架;
- 第 2 部分:数据元的分类;
- 第 3 部分:数据元的基本属性;
- 第 4 部分:数据定义的编写规则与指南;
- 第 5 部分:数据元的命名和标识原则;
- 第 6 部分:数据元的注册。

本标准的附录 A 为提示的附录。

本标准由中国标准研究中心提出并归口。

本标准起草单位:中国标准研究中心。

本标准主要起草人:刘植婷、邢立强、冯卫、李小林。

ISO/IEC 前言

国际标准化组织(ISO)和国际电工委员会(IEC)是世界性的标准化专门机构。作为 ISO 或 IEC 成员的各个国家团体机构,通过各类技术委员会参与国际标准的研制。技术委员会由参加各类专业领域技术活动的相关组织组建而成。ISO 和 IEC 的技术委员会在共同感兴趣的领域进行合作。同 ISO 和 IEC 有联系的官方与非官方国际组织,也可以参与该项工作。

在信息技术领域,ISO 和 IEC 已组建了联合技术委员会,即 ISO/IEC JTC1。由联合技术委员会通过的国际标准草案,提交给各国家团体进行表决。一项国际标准的颁布,至少需要 75% 参与表决的国家团体投赞成票。

国际标准 ISO/IEC 11179-4 是由信息技术联合技术委员会(ISO/IEC JTC1)下属的数据管理与交换分委员会(SC32)负责起草的。

ISO/IEC 11179 在总标题《信息技术——数据元的规范与标准化》下,包括以下部分:

- 第 1 部分:数据元的规范与标准化框架;
- 第 2 部分:数据元的分类;
- 第 3 部分:数据元的基本属性;
- 第 4 部分:数据定义的编写规则与指南;
- 第 5 部分:数据元的命名和标识原则;
- 第 6 部分:数据元的注册。

本标准的附录 A 仅供参考。

引 言

数据元定义的目的是给数据元的含义下定义。为了保证质量和一致性,本标准提出了规范数据元定义结构的规则和指南。

准确而清楚的数据元定义是保证数据共享的最重要的因素之一。当双方或多方交换数据时,数据含义的准确一致是至关重要的。传送数据含义的主要载体之一是数据元定义。因此,每个数据元必须具有规范化的定义,这个定义要让每个使用者都能清楚地理解。不规范的数据元定义会引起误解和含义不清,且常常使人们不能成功地进行交流。

本标准包括规则和指南两部分。这些规则和指南适用于信息处理系统和信息交换中所使用的数据元定义的编写。要遵守标准,规则是强制的和可验证的;指南是应遵循的原则。对于规则,能制定客观的试验标准。与指南的一致性可以通过合理性判断进行评价。

本标准中的数据元名称没有遵循特别的句法来命名。“数据元”即是指“数据元类型”,简短的术语更便于使用。

目次

前言 Ⅲ

ISO/IEC 前言 Ⅳ

引言 V

1 范围 1

2 引用标准 1

3 定义 1

4 数据定义规则和指南的概要 2

5 要求 2

附录 A(提示的附录) 参考文献 7



中华人民共和国国家标准

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

Information technology—Specification and standardization
of data elements—Part 4: Rules and guidelines
for the formulation of data definitions

GB/T 18391.4—2001
idt ISO/IEC 11179-4:1995

1 范围

本标准规定了构建数据元定义的规则和指南。本标准仅介绍数据元定义的语义结构,不对定义的格式作出规定。

这些定义规则和指南适用于数据元,也适用于编写其他类型数据结构的定义,诸如实体类型、实体、关系、属性、对象类型(或类)、对象、段、组合型代码条目和报文。

本标准中的定义规则和指南并不总是适用于词汇表和语言词典中的术语定义。适用于语言词典的规则与适用于数据字典的规则之间存在着差别。例如,语言词典中的术语也许有多个定义,而数据字典的定义在词典中是唯一的,只有一种含义。

许多数据元定义包含有本身需要被定义的术语(如“充电(charge)”、“公差(allowance)”、“交付(delivery)”等)。其中一些术语在不同行业可能有不同的定义。因此,大多数数据字典都需要附带一张供定义用的术语词汇表。在词汇表中标明每个术语的使用领域。

2 引用标准

下列标准所包含的条文,通过在本标准中引用而构成本标准的条文。本标准出版时,所示版本均为有效。所有标准都会被修订,使用本标准的各方应探讨使用下列标准最新版本的可能性。

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

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 5271.4—2000 信息技术 词汇 第4部分:数据的组织(eqv ISO/IEC 2382-4:1987)

GB/T 14805—1993 用于行政、商业和运输业电子数据交换的应用级语法规则(idt ISO 9735:1988)

GB/T 1.6—1997 标准化工作导则 第1单元:标准的起草与表述规则 第6部分:术语标准编写规定(neq ISO 10241:1992)

GB/T 18391.3—2001 信息技术 数据元的规范与标准化 第3部分:数据元的基本属性
(idt ISO/IEC 11179-3:1994)

ANSI X3.172-1990,美国信息系统国家标准词典(ANDIS)

3 定义

本标准应用了下列定义。

3.1 属性 attribute

某个对象或实体的一种特性。

3.2 概念 concept

在一组对象共同特性的基础上抽象出来的思想。

3.3 数据 data

对事实、概念和指令的一种形式化表示,适于人工或自动方式进行通信、解释或处理。

3.4 数据字典 data dictionary

用于数据和其他数据结构的一种数据库;即用于贮存元数据的数据库(ANSI X3.172-1990)。见数据元字典。

3.5 数据元 data element

用一组属性描述定义、标识、表示和允许值的一个数据单元。

3.6 数据元字典 data element dictionary

列出并定义了全部相关数据元的一种信息资源。它与数据字典同义。

注:数据元字典可有不同层级,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。

3.7 定义 definition

表述某人、某事的基本特性,或人和事物类别的词或短语,要回答“X 是什么?”或“X 属于什么?”这样的问题;一个词或词组含义的陈述。[韦氏新世界英语大词典,第三版,1986]。

3.8 域 domain

一种属性的一组可能的数据值(ANSI X3.172-1990)。

3.9 名称 name

人们标识事物和概念的基本方式。

4 数据定义规则和指南的概要

为了方便使用者,本章列出的规则和指南不含说明内容,其目的是使本标准在理解的基础上易于使用。第5章给出了带有说明和示例的规则和指南,以确保人们对其准确含义的理解。

4.1 规则

数据定义应:

- a) 具有唯一性(在出现此定义的任何数据字典中);
- b) 用单数形式阐述;
- c) 要阐述其概念是什么,而不是阐述其概念不是什么;
- d) 用描述性的短语或句子阐述;
- e) 仅可使用人们普遍理解的缩略语;
- f) 表述中不要加入不同的数据元定义或引用下层概念。

4.2 指南

数据定义宜:

- a) 阐述概念的基本含义;
- b) 准确而不含糊;
- c) 简练;
- d) 能单独成立;
- e) 表述中不应加入理论说明、功能说明、范围信息或程序信息;
- f) 避免相互依存;
- g) 相关定义使用相同的术语和一致的逻辑结构。

5 要求

5.1 前提

数据元是为特定的目的而存在和被使用的。使用上的不同,将要求某些规则和指南具有不同的表现形式。例如,不同的相关环境通常要求数据元定义的不同层级特性。5.3a)列举了不同定义需要不同层级特性的实例。5.3a)“阐述概念的基本含义”取决于相关环境。表述某一特定定义的基本含义所必须的主要特性,因数据元的通用化或专业化层级而异。这种主要特性应考虑到与所分析的概念有关的对象类、性质和修饰词的关联。定义诸如像商业航空运输业中“飞机场”这样的概念的主要和基本特性也许是特定的,但是,在不同的相关环境中做为一般的定义也是恰当的。有关不同相关环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。定义的撰写应让使用者和数据共享者易于理解。

5.2 规则

为了使有关规范化数据元定义结构的规则易于理解,下面给出了有关说明和示例。每项规则之后都附有对其含义的简要说明。说明之后,则是一些示例。每项说明均配有好的示例,必要时还配有常见的错误示例,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。

数据定义应:

a) 具有唯一性(在出现此定义的任何数据字典中)。

说明:每个定义必须区别于(字典中的)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。

示例:“发货日期”“收货日期”

1) 好的定义:

“发货日期”——供货方发送货物的日期。

“收货日期”——收货方收到货物的日期。

2) 不好的定义:

“发货日期”——交付货物的日期。

“收货日期”——交付货物的日期。

原因:定义“交付货物的日期”不能既用于“发货日期”,又用于“收货日期”这两个数据元。每个定义必须是不同的。

b) 以单数形式阐述

说明:数据定义所表述的概念必须用单数形式表述(概念本身为复数时除外)。

示例:“物品编号”

1) 好的定义:

一个用以标识一种物品的参考号码。(A reference number that identifies an article.)

2) 不好的定义:

标识物品的参考号码。(Reference number identifying articles.)

原因:不好的定义使用了复数的“物品(articles)”,含义不清楚,因为它可理解为一个“物品编号(article number)”指代多件物品。

c) 要阐述其概念是什么,而不是仅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:“运费总额”

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总额。

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。

d) 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)必须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词。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。

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同义词,不适宜用作定义。

e) 仅可使用人们普遍理解的缩略语。

说明:对缩略语,包括简称和首字母含义的理解,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,为了避免词义不清,在该定义中就应使用全称,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。没有全称,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:磅每立方英尺;千克每立方米)。

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:磅每立方英尺;千克每立方米)。

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。

f) 表述中不要加入不同的数据元定义或引用下层概念。

说明:次要的数据元定义或相关概念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小样本。它既可用作试验的实际样品,又可用作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:不好的定义含有另一个概念定义,这个概念定义应收入词汇表。

5.3 指南(指导原则)

数据定义宜:

a) 阐述概念的基本含义。

说明:表述概念的所有基本特性应在该相关环境各相应特性层级的定义中出现。应避免出现非基本特性。是否需要细分层级,取决于系统用户和环境的要求。

示例 1:“托运货物装货序号”(预期环境: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:在预期环境下,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:“发票总额”

1) 好的定义:

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:

具体货物被交付的日期。

原因:不好的定义没有说明由什么来决定“货物交付”了。“货物交付”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“为了本数据字典”、“要描述的术语是”之类无关的限定性短语。

示例:“字符集名”

1) 好的定义:

给一组表音或表意的符号起的名称,数据用这些符号来编码。

2) 不好的定义:

给一组表音或表意的符号起的名称,数据用这些符号来编码。这个数据字典的目的,就像其他地方用的那样,系统硬件和软件的能力就是在多个版本中对编码的数据进行处理。

原因:在不好的定义中,在“数据用这些符号来编码”之后的所有短语,都是无关的限制性短语。

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) 表述中不要加入理论说明、功能说明、范围信息或程序信息。

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的含义无关的信息。如果认为它们确实有用,可以把这类表述列为不同的数据元属性(GB/T 18391.3)。

为列举的定义进行说明,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,其原因不用在定义中说明)。

诸如“此数据元不应用于…”这样的功能说明不宜出现在定义中。

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,而应当按 GB/T 18391.3 的规定,使用“相关数据参照”和“关系类型”。

示例:“数据字段标记”

1) 好的定义:

索引、主题词表、查询、数据库等的数据库标识。

2) 不好的定义:

索引、主题词表、查询、数据库等的数据库标识。数据库标识用于诸如摘要、表格中的栏目等信息单元。

原因:不好的定义含有功能的陈述,以“数据库标识用于…”开始的陈述必须从定义中删去。如果必须有此陈述的话,应将其放在其他属性中。

f) 避免相互依存。

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,附加了另一个概念;反之,在定义另一个概念时,又附加了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) 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,应使用共同的术语和句法。

示例:上述在规则 5.2a) 中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。

附 录 A
(提示的附录)
参 考 文 献

ISO 标准手册 10, 数据处理——词汇汇编, 1982

TRADE/WP. 4/R. 765/Add. 1, 联合国经济与社会理事会, 1991. 7. 30
