



中华人民共和国国家标准

GB/T 18391.5—2001
idt ISO/IEC 11179-5:1995

信息技术 数据元的规范与标准化 第5部分：数据元的命名和标识原则

Information technology—
Specification and standardization of data elements—
Part 5: Naming and identification principles for data elements



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 5 部分:数据元的命名和标识原则

GB/T 18391.5—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 21 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: 155066·1-18078 定价 12.00 元

网址 www.bzcbs.com

*

科 目 597—526

版权专有 侵权必究
举报电话:(010)68533533



GB/T 18391.5—2001

0597-526

前 言

本标准等同采用国际标准 ISO/IEC 11179-5:1995《信息技术——数据元的规范与标准化——第 5 部分:数据元的命名和标识原则》。

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、附录 B、附录 C 是提示的附录。

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:冯卫、李小林、刘植婷、邢立强。



ISO/IEC 前言

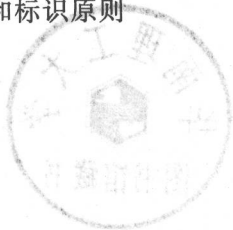
国际标准化组织(ISO)和国际电工委员会(IEC)是世界性的标准化专门机构。作为 ISO 或 IEC 成员的各国家团体机构,通过各类技术委员会参与国际标准的研制。技术委员会由参加各类专业领域技术活动的相关组织组建而成。ISO 和 IEC 的技术委员会在共同感兴趣的领域进行合作。同 ISO 和 IEC 有联系的官方与非官方国际组织,也可以参与该项工作。

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 11179-5 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



引 言

本标准包括原则、规则和指南。原则是建立规则的前提。规则是强制性的,并且是可进行一致性测试的。指南是上述规则在好的示例中的应用。

目 次

前言	I
ISO/IEC 前言	II
引言	III
1 范围	1
2 引用标准	1
3 定义	1
4 数据的标识结构原则	3
5 数据注册标识的规则	4
6 结构化命名约定指南	4
7 主题词表应用指南	6
附录 A(提示的附录) 命名约定示例	7
附录 B(提示的附录) 注册标识示例	8
附录 C(提示的附录) 主题词表示例	9



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5部分:数据元的命名和标识原则

GB/T 18391.5—2001
idt ISO/IEC 11179-5:1995

Information technology—
Specification and standardization of data elements—
Part 5: Naming and identification principles for data elements

1 范围

本标准规定了数据元命名和标识的规则与指南,阐述了数据元标识的组成成分(以下简称成分)和结构。标识,仅被定义为在一个注册簿中建立数据元唯一标识的方法。本标准定义了属性的标识;阐述了每个属性间的关系;包括制定命名约定时可依据的原则,并且描述了命名约定的一个示例。这里所阐述的命名指南也可用于数据元名称以外的数据概念的命名。

本标准应与建立数据元的属性、分类、定义和注册等规则和程序共同使用。

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 18391.1¹⁾ 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
(idt ISO/IEC 11179-1:1999)

GB/T 18391.2²⁾ 信息技术 数据元的规范与标准化 第2部分:数据元的分类
(idt ISO/IEC 11179-2:2000)

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)

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(idt ISO/IEC 11179-4:1995)

GB/T 18391.6—2001 信息技术 数据元的规范与标准化 第6部分:数据元的注册(idt ISO/
IEC 11179-6:1997)

3 定义

下列定义适用于本标准:

3.1 属性 attribute

某个对象或实体的一种特性。(GB/T 18391.3)

注:

1) 该国家标准已报批。

2) 该国家标准在制定中。

3.2 相关环境 context

对使用或产生名称的应用环境或应用规程的指明或描述。(GB/T 18391.3 6.1.6)

3.3 数据元 data element

用一组属性描述其定义、标识、表示和允许值的一个数据单元。(GB/T 18391.3)

3.4 数据元概念 data element concept

能以数据元的形式表示,且与其任何特定的表示法无关的一种概念。(GB/T 18391.3)

3.5 数据标识符 data identifier

由注册机构赋予数据元的标识符(一串字符或图形符号)。

3.6 定义 definition

表述某人、某事基本特性或人、事类别的词或短语: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 年)

3.7 标识符 identifier

见数据标识符。

3.8 词法 lexical

有关一种语言的单词或词汇,而非它的语法和结构。

3.9 名称 name

人们标识事物和概念的基本方式。

3.10 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”。

3.11 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。

3.12 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。

3.13 注册机构 registration authority

经授权对数据元进行注册的组织。

3.14 注册机构标识符 registration authority identifier

经授权对数据元进行注册的任何组织的标识符,即,含有此标识符的属性。

3.15 表示术语 representation term

数据元有效值集合的形式,如:“总额”,“姓名”。

3.16 语义学 semantics

有关词义注释的语言学分支学科(韦氏词典)。

3.17 分隔符 separator

在名称中连接和分隔成分的符号或空格,又称分界符。

3.18 结构设置 structure set

在相关环境中放置对象的方法,用以展示与其他对象的关系,如各种实体关系模型、分类法(taxonomies)和本体论法(ontologies)。

3.19 句法 syntax

字符或字符组之间的各种关系。这些关系与字符或字符组的含义、解释和使用方式无关。它是语言的表示结构及管理语言结构的规则。

3.20 主题词表 thesaurus

按给定顺序排列的受控词汇,其中显示和标识了词汇之间的关系。

3.21 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。

4 数据的标识结构原则

4.1 属性标识

为了区分不同的数据元,用一组相关的五个属性,对每一个数据元进行命名和标识。这些属性是:

- 名称;
- 相关环境;
- 注册机构标识符;
- 数据标识符;
- 版本标识符。

本标准涉及的命名,设定为包括名称和同义名称的属性,这些属性在 GB/T 18391.3(包括了对全部属性的描述)中进行了规定。数据标识符也与 GB/T 18391.3 中的标识符相同。这些属性的原则在下面叙述。

4.2 名称和相关环境

在注册机构中注册的一个数据元至少应当有一个名称,根据该数据元使用的相关环境,可以分配多个名称。每个名称在某个特定的相关环境中有着特殊的作用。为了便于数据管理,可建立严格结构化的名称;使用者可以规定首选名称;在特定的软件环境(如:一个特定程序语言或数据库管理系统)中可以产生简称。

在每个相关环境中可以为许多数据元赋予名称,每个相关环境建立一个命名约定(一般是一组规则),用以详细说明在该相关环境中如何规范名称。一个命名约定应包括相关环境的全部有关方面,这包括:

- 命名约定的范围,如“行业首选名称”;
- 确定名称的机构;
- 指导名称的用词来源和内容的语义规则,如:由数据模型生成的词汇,规范中常用词汇等等;
- 关于用词顺序的语法规则;
- 关于受控单词表、名称长度、字符集、语言词法规则;
- 制定在同一相关环境中名称是否必须具备唯一性的规则。

命名约定的上述内容在第 6 章起草严格结构化命名约定的指南中进行了详述。

4.3 国际注册数据标识符

属性注册机构标识符(RAI)、数据标识符(DI)和版本标识符(VI)构成了国际注册数据标识符(IR-DI)。一个数据元至少需要一个 IRDI。数据标识符由注册机构分配;数据标识符在一个注册机构的范围内必须是唯一的。

由于每个注册机构可以决定其各自的分配方案,因此不能保证用某注册机构的数据标识符能对一个数据元进行唯一标识。例如:如果两个注册机构都用连续的 6 位号码,就会有一组数据元有相同的 DI,然而完全有可能,不同的数据元具有相同的 DI。相反,如果同一个数据元在两处机构注册,它就有两个 DI。因此,一个数据元的标识不仅需要 DI 还需要 RAI。

如果数据元的特定属性改变了,就应产生并注册数据元的新版本,在这种情况下就需要 VI 来完成对数据元的唯一标识。进一步的原则见 GB/T 18391.6。

一个 IRDI 是信息系统、组织或其他希望共享一个特定数据元(但不能利用相同的名称和相关环境)的参与者之间交换数据的关键。当 IRDI 与建立多于一种自然语言的相关环境有关联时,IRDI 也有助于语言的翻译,并且对由不同的注册机构管理的数据元集合之间是一个参考。

GB/T 18391 没有规定 DI 的格式或内容。

对注册机构的要求及有关 IRDI 的讨论见 GB/T 18391.6。

5 数据注册标识的规则

- 5.1 在一个注册机构的注册簿中,每个数据元应有一个唯一的数据标识符。
- 5.2 注册机构标识符、数据标识符和版本标识符的组合,便构成数据元的唯一标识。
- 5.3 需要分配数据标识符的数据元必须是:
 - 按照 GB/T 18391.2 形成的;
 - 按照 GB/T 18391.3 确定属性的;
 - 按照 GB/T 18391.4 定义的;
 - 按照 GB/T 18391.5 命名的;
 - 按照 GB/T 18391.6 注册的。
- 5.4 在一个相关环境中,一个数据元应至少有一个名称。

6 结构化命名约定指南

下面是在特定相关环境下用来制定命名约定的指南。附录 A 是一个命名约定的具体示例,它同本章介绍的指南相一致。

指南以通用的词语叙述,并附有示例。规则是由制定名称的原则中引出的,这些规则构成了命名约定。按照这些规则形成的名称可以很容易翻译成其他语言,因为它的句法简单。句法、语义和词法的规则随不同的组织(如企业或为业务部门制定标准的团体)而异。每个组织可以在自己的相关环境中确立名称构成的规则。

如 6.1.1.1 中讨论的,每个数据元是由一组成分构成,它们选自各自相关环境的结构设置。

数据元的名称可以由其成分的名称构成,在一个名称中设置每个成分的含义(语义)和相对或绝对的位置(语法)。它们可以从属于词法规则。它们可以,但不是必须,用一个分隔符来分界。每个成分值的集合或范围应当由一个机构(如一个企业中的数据管理者或者一个国际商务部门命名标准的审定委员会)严格管理。

- 语义规则使含义能够传达;
- 句法规则按一致性和指定顺序来关联成分;
- 词法(词形和词汇)规则减少冗余,增加精确性。

6.1 管理名称语义内容的原则

语义涉及到名称成分的含义,以及对其界定的分隔符。

6.1.1 名称成分语义

成分由离散术语组成。本标准的名称成分由以下术语描述:

- 对象类术语;
- 特性术语;
- 表示术语;
- 限定术语。

6.1.1.1 对象类术语

对象类术语是构成数据元名称的一个成分,它表示某一相关环境中的一项行为或一个对象。建模方法的使用(如实体关系图 ERD 或对象模型)是分别确定所有数据元和与其上一层模型实体的关系位置的一种方法。实体关系模型中实体的属性就等同于数据元,这些数据元通过这种建模方法的进一步应用相互联系起来。在对象模型中,数据元则是通过对象属性表示的。

各种模型为数据元提供了一种分类方案。数据元可以用与之相关模型中的实体来标识,将对象类术语映射为模型中的实体的名称即可。GB/T 18391.1 提供了对象类术语和实体关系模型 ERD、对象模型中的实体之间的映射的示例。

例如,在下面的数据元中:

- 雇员的姓氏
- 费用预算周期合计总额
- 树的高度的测量
- 成员的姓氏

成分雇员、费用、树和成员是对象类术语。

6.1.1.2 特性术语

一组特性术语是由一个特性分类法中的一组名称成分构成的。这个组必须是由离散的(每个定义不能与其他定义相重叠)和完整的(全部的特性,这个组表示了所有说明数据元所需的信息概念)词构成。

例如,在数据元中:

- 雇员的姓氏
- 费用预算周期合计总额
- 成员的姓氏
- 树的高度的测量

成分姓氏、总额和高度是特性术语。

特性术语当然要出现在数据元的定义中。使用两个结构设置中的成分,提供了一种分类的补充方法。用数据元对象类和特性两个成分形成的一个名称,包含了有关数据元极为重要的信息,而且排除了当没有约定使用而被采用的不重要的或不合理的元素。

6.1.1.3 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成的一个成分。每个表示术语由一个受控单词表或一个分类体系形成。表示术语的类别表现形式如下:

- | | | |
|------|------|-------|
| • 名称 | • 总额 | |
| • 测量 | • 数目 | |
| • 数量 | • 文本 | |

这类术语描述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删除一个术语或术语的一部分。这一点可按命名约定中的规则确定。该过程的应用见附录 A。

例如,在数据元中:

- 树的高度的测量
- 雇员的姓氏。

成分测量和姓氏是表示术语。注意姓氏是一个特性术语,为表达的清楚,冗余的字可以删去。

6.1.1.4 限定术语

如果必须对一个数据元进行唯一标识,可以将限定术语附加到对象类术语、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。在命名约定的规则中,建议对限定术语的数量予以限制。

例如,在下面的数据元中:

- 费用预算周期合计总额

成分预算周期是限定术语。

注:对允许的限定术语的形式给予一些限制,有助于减少冗余;通过取消一些同义名,扩大数据的使用范围。这点同样也适用于对象类术语、特性术语和表示术语。用一种机制,如主题词表可以更加便捷(见第 7 章和附录 C)。

6.1.2 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为不同的分隔符。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,在数据元中:

——费用-预算-周期-合计-总额

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。

有些语言,如德语和荷兰语,一般是在一个单词中加入语法因素形成(一个词的含义可能相当于英语或法语由名词或形容词组成的短语)。这些语言用的分隔符不是将各个单词从中间断开,如用连字符、空格和下划线,而是把一个单词中每个名称成分的第一个字母大写。

6.2 管理名称格式的原则

6.2.1 句法原则

句法原则详细说明了一个名称中各成分的排列。这一排列可能是相对的或者绝对的,或者二者兼有。

a) 相对排列是依据其他成分确定该成分。如在一个约定中的一项规则可能要求一个限定术语必须总是出现在被限定术语的前面。

b) 绝对排列是确定该成分的固定位置。如:一项规则可能要求特性术语总是一个名称的最后成分。

6.2.2 词法原则

这些原则涉及到优选词和非优选词、同义词、缩略语、成分的长度、拼写、容许字符集、大小写区分等等。

7 主题词表应用指南

主题词表是一个非常有用的工具,使用者在其中能够找到各种同义词、近义词和同形异义词作为名称成分。它可以提供名称优选术语和其他术语之间的语义联系。除了同形异义词(单词的拼写一样,表达了不同的概念)的使用指导以外,主题词表可以通过选择以下方式指导使用者;

——同义:从中优选一个词或术语,使其表达的概念比选其他的词或术语更贴切;

——等级:在表示广义和狭义之间的关系时,可用上、下层级来表示;

——相关:把两个或更多相同或不同层级的术语,从语义和概念上联系起来。

一个注册机构可以制定一个名称成分的主题词表,发给感兴趣的团体。另外,还需要制定专业领域的主题词表。

一个主题词表可用文字描述结构,以作为图表的补充。优选术语成为名称的成分;优选术语间的关系表示了成分在结构中的位置。

受控词汇是主题词表的一个优点,可以在一个主题词表中控制描述符的范围,通过等价将同义词和近义词联系起来,并消除同形异义词。这些是使用者使用系统时所需要的所有主题词表的功能。

附 录 A
(提示的附录)
命名约定示例

这些规则出自第 6 章描述的指南。包括若干示例,可供各种相关环境的名称使用,并由专业领域的机构斟酌采用。

A1 语义规则

a) 对象类表示在一个领域内有关的事物,例如:在某领域的一个模型中发现的事物。

示例:费用

b) 需要有一个且仅有一个对象类术语。

c) 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。

示例:合计总额

d) 需要有一个且仅有一个特性术语。

e) 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,即可增加上限定术语。限定术语的顺序并不重要。限定术语是可选的。

示例:预算周期

f) 表示数据元的有效值集合由表示术语来描述。

g) 需要有一个且仅有一个表示术语。

示例:总额

A2 句法规则

a) 对象类术语应处于名称的第一(最左)位置。

b) 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。

c) 特性术语应处于第二位置。

d) 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。

示例:费用 预算 周期 合计 总额

A3 词法规则

a) 名词仅用单数形式,动词(若有的话)为现在时。

b) 名称的各个成分间和多单词术语的单词之间用空格分隔。不允许用特殊字符。

c) 名称中所有单词是组合在一起的。

d) 允许使用缩写词、首字母缩略词和大写首字母。

A4 唯一性规则

同一相关环境的所有名称应是唯一的。

附 录 B
(提示的附录)
注册标识示例

下面是一个数据元标识结构的示例：

数据元 内容				
注册机构标识符 (RAI)	数据标识符 (DI)	版本	名称	相关环境
ISO	848575	1	费用预算周期总额	财务电子数据交换系统
ISO	848575	1	价值-总额	美国-GICS 系统
IEEE	193847	1	运输-费用-总额	工程模型
IEEE	193847	1	我们的-费用-\$	合同报表表头

这是由两个注册机构注册的一个数据元的例子,每个机构都配给它一个数据标识符。在每个机构中该数据元有两个不同的名称。注意,“价值-总额”和“我们的-费用-\$”是应用系统中两个同义名的示例,这些都是没有按照命名约定产生的传统名称,并将继续沿用下去。

这个数据元名称的成分结构(按照第 6 章的原则)如下:



附 录 C
(提示的附录)
主题词表示例

下面是从一个结构术语可能的主题词表中摘选的内容,包括了结构信息,同义词和同形异义词解析。

关键字:

BT—Broader Term(广义术语)

NT—Narrower Term(狭义术语)

UF—Use For(用于)

USE—Use the following term instead(用下列术语替代)

RT—Related Term(相关术语)

SN—Scope Note(范围注释)

主题词表中摘选:

COST

SN——单位支付货物或服务的总额

BT——合同

NT——小额现金

UF——支出

RT——预算

RT——总额

支出

USE——费用

注意,尽管主题词表的条目能显示“费用”同预算和总额的关系,但这种关系的实质是不清楚的。

斜体字用于表示不赞成的词,这是从同义词中区分优选词的方法。

一个或多个层级可以通过列出若干个广义词和狭义词来表示。这是一个主题词表的规定。

在 GB/T 18391.4 中描述了范围注释的定义。
