



中华人民共和国国家标准

GB/T 18391.1—2002/ISO/IEC 11179-1:1999

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

**Information technology—Specification and standardization
of data elements—Part 1: Framework for the specification
and standardization of data elements**

(ISO/IEC 11179-1:1999, IDT)

2002-07-18 发布

2002-12-01 实施

中华人民共和国
国家质量监督检验检疫总局 发布

目 次

| | |
|------------------------------------|-----|
| 前言 | I |
| ISO/IEC 前言 | II |
| 引言 | III |
| 1 范围 | 1 |
| 2 规范性引用文件 | 1 |
| 3 术语和定义 | 1 |
| 4 缩略语 | 7 |
| 5 方法论无关性 | 7 |
| 6 数据元的基本概念 | 7 |
| 7 数据元与其他数据概念的关系 | 9 |
| 8 GB/T 18391 第 1 部分至第 6 部分综述 | 9 |
| 附录 A(资料性附录) 数据表示和管理的基本概念 | 13 |
| 附录 B(资料性附录) 元数据和数据元 | 23 |

前 言

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。

本部分等同采用国际标准 ISO/IEC 11179-1:1999《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》。

本部分的附录 A 和附录 B 是资料性附录。

本部分由中国标准研究中心提出并归口。

本部分起草单位:中国标准研究中心。

本部分主要起草人:刘植婷、邢立强、史立武。

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票成票。

需要提请注意的是:ISO/IEC 11179 本部分中某些内容可能已经获得了专利权。ISO 和 IEC 并不对全部或部分具有专利权的内容负责。

国际标准 ISO/IEC 11179-1 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 和附录 B 是资料性附录。

引 言

背景

人类通过自然界中事物的特性来认识事物,而这些特性则通过数据来表示。数据的基本单元——数据元的规范,涉及到为每一个数据元的相关特性进行规范化说明,以确保自然界事物表示的一致性与准确性。经过仔细、认真规范和标准化的数据将会大大加强其在系统间及环境间的实用性和共享性。数据共享包括所需数据的查询检索以及与其他用户的数据交换。当数据元按照 GB/T 18391 的要求进行编写,且其文件在一个数据元登记系统得到管理时,则从分散的数据库中查找、检索以及通过电子通信手段发送和接收它们将变得相当简便。

通过自动信息处理系统对于交流的数据元进行识别和标准化,是一项正在进行的重要工作。该项活动的成功及其在世界范围内的应用将对提高政府、企业界及学术界间的交流起到至关重要的作用。

GB/T 18391 的制定致力于解决以下几个最基本的数据共享与标准化问题(但不限于此):

- 缺乏促使全球数据获得与交换的机制,特别是在应用领域;
- 全球唯一的标准数据元的标识符尚不存在;
- 有关数据元特性的文档尚不足以支持数据全面自动化的共享,包括数据的查找、检索与交换;
- 目前尚未形成关于数据元标识、开发和描述的统一指南;
- 在成千上万的数据元中查找和检索某一具体的标准数据元是非常困难或者是几乎不可能的事情;
- 目前尚不存在通用的方法对标准数据元进行有效的组织;
- 尽管有些数据在一个机构内已经标准化,但机构间公用的数据标准却很少;
- 机构间数据的交换导致了专用数据交换表示方法的激增;
- 数据定义和描述的精确程度尚不能支持数据的再利用以及用户多样化的需求;
- 现有用于缩减冗余逻辑数据的目录结构尚不能满足需要;
- 由于标准数据元的缺乏而使全球实施电子数据交换(EDI)受到阻碍;EDI 报文的内容需要标准数据元。

为了推动电子通信的全球化,国际标准界通过细心而持续不断的努力确定了开放系统互连环境(OSIE)。在该环境内,不同的计算机硬件及应用可实现信息共享。开放性信息处理系统所需的四个基本要素(硬件、软件、通信和数据)中的三个(硬件、软件和通信)要素,已经或正在制定为相关标准,而 GB/T 18391 作为数据规范,为开放性信息系统的第四个基本要素,提供了在 OSIE 内实现数据共享的机制。

系统要做到真正的开放,数据必须是易于交换的且可以在各种不同的应用环境内及其相互间进行共享,它可以跨越区域化的网络和分布式的网络。为了数据的共享,数据的使用者与拥有者必须对其含义、表示和标识的理解是一致的。例如,为了理解数据的含义,数据使用者必须在数据元登记系统获得关于数据的描述。数据必须得到充分的描述,并使使用者可以通过很方便的途径得到这些描述。数据元登记系统提供了对数据元内容及表示进行组织的方法,以确保数据描述的一致性,并使数据设计者和使用者较为方便地对其进行定位。数据的统一规范化将为整个软件开发生命周期中的数据检索、交换和一致性的使用提供便利。具有规范意义和形式的信息单元则被称为“已标准化数据元”。

GB/T 18391 的目的

GB/T 18391 描述数据元标准化和注册,使理解和共享数据成为可能。相对于传统的数据管理办法,以 GB/T 18391 来描述数据元标准化及其注册将使一个共享数据环境的产生花费更少的时间和精力。

GB/T 18391 的目的在于如何就离散性数据元的描述及其语义内容(元数据,它以一种一致、标准的方式来表达数据元)的表达与维护提供具体的指南。它还就数据元注册系统的建立提供指南。

GB/T 18391 的初衷是通过电子信息交换以满足国际范围数据开放化交换的需要,它还能:

- 便利数据的获取和注册;
- 加速数据的访问及应用;
- 依据元数据描述的特性,运用智能软件简化数据操作;
- 促成用于 CASE 工具和数据存储的数据表示元模型的开发;
- 方便电子数据的交换与数据的共享。

GB/T 18391 将使以下信息系统和人们之间的数据传递获益:

- 一个机构内;
- 不同的机构间;
- 跨越所有软件、硬件层面和地理、机构及政治的界限。

关于数据元的元数据被存储于数据元注册系统中。数据元注册系统支持数据及其描述的共享。注册是将支持数据共享的元数据文件化的过程。注册应在数据元层面进行,使其含义得到最大程度的理解。GB/T 18391 将使最终用户能自信、准确、无歧义地解释其所想要表达的意思。

GB/T 18391 的用户

GB/T 18391 为数据用户和管理者阐明了共享数据所必需的数据元特性中的一个基本集。并对重要的数据元特性如标识符、定义及分类类型做了特意的强调。GB/T 18391 对数据元注册系统的描述,旨在帮助共享数据的用户对数据元的含义、表示和标识有一个共同的理解。如果已获得数据值,用户可以找到该数据的确切含义。用户还可以通过辨识所需数据的类型从数据库中检索所期望的数据值。

GB/T 18391 为系统分析者和数据管理者提供了能满足需求的数据元的再利用途径,或如果一个数据元不存在,设计新的数据元的途径。在用户从数据库中获得数据元以前,数据服务管理者和系统分析者必须对数据进行逻辑标识和描述,以确保用户不会无意中引进不一致的数据值。如果数据分析师想生成一个共享数据的产品,他们首先应该知道所需特性的数据元是否存在。若存在,他们可以使用它。如果系统分析者复制该数据元,就必须以包含相同信息的相同方式表示数据元。若不存在具有恰好相同特性的数据元,数据管理者就需要设计数据元使其描述适合于软件开发者的需求。GB/T 18391 的目的在于开发对数据元的精确描述。依照这个由多个部分组成的国家标准的原则产生的数据元,可以进行数据交换和检索,而与其所使用的信息处理系统和通信协议无关。

GB/T 18391 为软件开发者提供了确保数据一致的手段。注册系统可以使软件开发者在整个软件开发生命周期(SDLC)中保证数据使用的一致性。注册系统将提供数据元管理的机制并使其在软件开发生命周期(SDLC)的不同阶段具有可追溯性。

GB/T 18391 为数据字典开发者、数据元注册系统、计算机辅助软件工程(CASE)工具和其他数据管理软件提供了用于设计元模型的基础,而元模型又是获取、存储、管理和交换数据元的元数据所必需的。

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

1 范围

GB/T 18391 规定了数据元组成的基本内容,包括元数据。GB/T 18391 的本部分适用于人、机共享数据元的规范化表示和含义,不适用于机器级上数据以位和字节为单位的物理表示。

本部分提供了关联部分的相关环境和从概念上理解数据元的基础。

2 规范性引用文件

下列文件中的条款通过 GB/T 18391 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 5271.4—2000 信息技术 词汇 第4部分:数据的组织(idt ISO/IEC 2382-4:1987)

GB/T 10112—1999 术语工作 原则与方法(neq ISO/DIS 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 18391.2—¹⁾ 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:数据元定义的编写规则与指南(idt ISO/IEC 11179-4:1995)

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

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(ISO/IEC 11179-6:1997)

GB/T 20001.1—2001 标准编写规则 第1部分:术语(ISO 10241:1992,International terminology standards—Preparation and layout,NEQ)

ISO 标准手册 10 数据处理 词汇 1982

3 术语和定义

下表中的术语和定义适用于 GB/T 18391 的各部分。表头中“部分”编号列中的 X 表明该术语在该部分被定义并会在其他章中被使用。以黑体出现于一个术语定义中的词语本身也是一个术语,它将会在该章的其他地方被定义。以普通字体出现的词语采用其被普遍理解的含义。某些词这两种情况都存在。也有的情况是两个或两个以上的术语在一个定义中同时出现,且出现的新术语并未给出定义。在上述情况下,确定术语时是无歧义的。

1) 正在制定中。

| 序号 | 术语 | 定义 | “部分”编号 | | | | | | |
|------|--------------------------------------|--|--------|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 | 5 | 6 | |
| 3.1 | 管理成分 administered component | 汇集了管理属性的成分。 | | × | | | | | |
| 3.2 | 管理状态 administrative status | 注册机构处理注册申请期间所处于各种状态的标示。 | × | | | | | | × |
| 3.3 | 属性 attribute | 某个对象或实体的一种特性。 | × | × | × | × | × | × | × |
| 3.4 | 属性值 attribute value | 某种属性的一个实例表示。 | | | × | | | | |
| 3.5 | 已审核数据元 certified data element | 已登录的数据元。符合 GB/T 18391 规定的质量要求。 | × | | | | | | × |
| 3.6 | 分类方案 classification scheme | 依据对象的共性如:来源、构成、结构、应用和功能等将其排列或分组。 | | × | × | | | | × |
| 3.7 | 分类方案条款 classification scheme item | 分类方案内容的一个成分。这可以是分类法或本体论的一个节点,主题词表中的一个术语等。 | | × | | | | | |
| 3.8 | 分类成分 classified component | 可以依据一个以上的分类方案进行分类的数据元的成分。这些成分包括对象类、特性、表示类、数据元概念、值域和数据元。 | | × | | | | | |
| 3.9 | 备注 comments | 关于数据元的注释。 | | | × | | | | × |
| 3.10 | 概念 concept | 在一组对象共同特性的基础上抽象出来的思想。[GB/T 15237] | × | × | | × | | | |
| 3.11 | 相关环境 context | 对使用名称或产生名称的应用环境或应用规程的指明或描述。 | | | × | | × | × | |
| 3.12 | 数据 data | 对事实、概念或指令的一种形式化表示,适用于以人工或自动方式进行通信、解释或处理。 [GB/T 5271.4—2000] | × | × | | × | | × | |
| 3.13 | 数据字典 data dictionary | 涉及其他数据应用和结构的数据的数据库,即用于存储元数据的数据库[ANSI X3.172—1990]。又见数据元字典。 | × | × | | × | | | |
| 3.14 | 数据元 data element | 用一组属性描述定义、标识、表示和允许值的数据单元。 | × | × | × | × | × | × | × |
| 3.15 | 数据元概念 data concept | 能以数据元形式表示,且与任何特定的表示法无关的一种概念。 | × | × | × | | | × | × |
| 3.16 | 数据元字典 data element dictionary | 列出并定义了所有相关数据元的一种信息资源。又见注册簿。 | | × | × | × | | | |

| 序号 | 术语 | 定义 | “部分”编号 | | | | | | |
|------|---|---|--------|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 | 5 | 6 | |
| 3.17 | 数据元面 data element facet | 易于分类的数据元的任一方面。这包括对象类、特性、表示和数据元概念。 | | × | | | | | |
| 3.18 | 数据元名称 data element name | 用于标识数据元的主要手段,由一个或多个词构成的命名。 | × | | | | | | |
| 3.19 | 数据元注册系统 data element registry | 由注册机构保存的用于描述数据元含义和表示形式的信息资源,包括注册标识符、定义、名称、值域、元数据和管理属性等。又见注册簿。 | × | | | | | | |
| 3.20 | 数据元值 data element value | 数据元允许值集合中的一个值。又见数据值。 | | | × | | | | × |
| 3.21 | 数据标识符(DI) data identifier | 由注册机构分配给数据元的标识符(一串字符或其他图形符号)。 | | | | | | × | × |
| 3.22 | 数据项 data item | 数据元的一个具体值。 | | | | | | | × |
| 3.23 | 数据模型 data model | 以反映信息结构的某种方式对数据组织的描述。 | × | | | | | | |
| 3.24 | 数据管理者 data steward | 对特定的一系列数据资源管理负责的个人或组织。 | × | | | | | | |
| 3.25 | 数据类型 datatype | 由数据元操作决定的用于采集字母、数字和(或)符号的格式,以描述数据元的值。 | × | | | | | | |
| 3.26 | 数据元值的数据类型 datatype of data element value | 表示数据元值的不同值的集合。 | | | × | | | | × |
| 3.27 | 数据值 data value | 值域中的一个元素。 | | × | | | | | |
| 3.28 | 定义 definition | 表述人和事物的基本特性、或者人或事物类别的词或短语;要回答“X是什么?”或“一个X属于什么?”这样的问题;一个词或词组含义的表述[韦氏新世界英语大词典,第三版,1986]。一个数据元基本特性的陈述,并使之区别于其他数据元。 | × | | × | × | × | × | |
| 3.29 | 域 domain | 一种属性的可能数据值的集合。[GB/T 5271]。又见值域。 | × | × | | | × | | |
| 3.30 | 实体 entity | 任何具体的或抽象的事物,包括事物间的联系。[GB/T 5271]。又见对象类。 | × | × | | | | | |

| 序号 | 术语 | 定义 | “部分”编号 | | | | | | |
|------|---|---|--------|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 | 5 | 6 | |
| 3.31 | 穷举域 enumerated domain | 由所有允许值的列表表示的值域。 | × | | | | | | |
| 3.32 | 表示形式 form of representation | 数据元表示形式的名称或描述。如： ‘数量值’、‘代码’、‘文本’、‘图标’。 又见表示术语。 | | | × | | | | × |
| 3.33 | 标识符 identifier | 注册机构内与语言无关的数据元的 唯一标识符。又见数据标识符。给 定相关环境的对象的无歧义的名 称。 | × | × | × | × | × | × | × |
| 3.34 | 信息 information | (信息处理中)关于对象的知识,如: 事实、事件、事物、处理、或想法,包 括在某些相关环境中具有特定含义的 概念。[GB/T 5271]。 | × | | | | | | |
| 3.35 | 信息交换 information interchange | 以某种方式发送和接收数据的过程, 该方式不会使指定给数据的信息 内容或含义在传输中被更改。 | × | | | | | | × |
| 3.36 | 国际注册数据标识符(IR- DI) international registration data identifier | 数据元的一个国际唯一标识符。 | | | | | | | × |
| 3.37 | 关键字 keyword | 用于数据元检索的一个或多个有意 义的字词。 | | × | × | | | | × |
| 3.38 | 表示格式 layout of representation | 在数据元值之内,字符的格式表示 用字符串表现。 | | | | × | | | × |
| 3.39 | 词法 lexical | 有关一种语言的词或词汇,而非它 的语法和结构。 | × | | | | | × | |
| 3.40 | 数据元值的最大长度 max- imum size of data element values | 表示数据元值的(对应数据类型的) 存储单元的最大数值。 | | | | × | | | × |
| 3.41 | 元数据 metadata | 定义和描述其他数据的数据。 | × | | | | | | |
| 3.42 | 数据元值的最小长度 minimum size of data ele- ment values | 表示数据元值的(对应数据类型的) 存储单元的最小数值 | | | | × | | | × |
| 3.43 | 名称 name | 人们标识对象和概念的基本方式。 赋予数据元的单字或多字的名称。 | × | × | × | × | × | × | × |
| 3.44 | 对象 object | 可以想象或感觉的世界的任一部 分。[GB/T 15237]。 | × | | | | | | |

| 序号 | 术语 | 定义 | “部分”编号 | | | | | |
|------|---|---|--------|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 | 5 | 6 |
| 3.45 | 对象类 object class | 对象集。现实世界中的想法、抽象概念或事物的集合,有清楚的边界和含义,并且特性和其行为遵循同样的规则而能够加以标识。 | × | | | | | |
| 3.46 | 对象类术语 object class term | 数据元名称的成分,用于表示其所属的对象类;如:“雇员”。 | × | | | | | × |
| 3.47 | 允许的数据元值 permissible data element values | 依据在相应属性中规定的表示形式、布局、数据类型、最大范围和最小范围,表示数据元允许事例的集。该集可以依据名称、参照来源、列举事例的表示或产生事例的规则而加以指定。 | | | × | | | × |
| 3.48 | 特性 property | 对象类的所有个体所共有的某种性质。 | × | | | | | |
| 3.49 | 特性术语 property term | 数据元名称的一个成分,用于表述对象类的特性。(数据元名称的一个成分,表述数据元所属类别。) | × | | | | | × |
| 3.50 | 限定词 qualifier | 帮助定义和呈递唯一性概念的术语。 | × | | | | | |
| 3.51 | 限定术语 qualifier term | 帮助定义和区分数据库中某个名称的词或词组。 | | | | | | × |
| 3.52 | 已登录数据元 recorded data element | 提交的数据元,包含了所有必选属性并已被记录,但其内容可能不满足 GB/T 18391 中其他部分规定的质量要求。 | × | | | | | × |
| 3.53 | 注册簿 register | 一套文件档案(纸质文件、电子文件或复合文件),内有已注册的数据元和相关信息。又见数据元注册系统。 | | | | | | × |
| 3.54 | 注册 registration | 赋予数据元一个明确无二义的标识符以使感兴趣的各方可以获得关于这些数据元的元数据。 | × | | | | | × |
| 3.55 | 注册申请者 registration applicant | 需要从注册机构获得标识符的组织、个人等。 | × | | | | | |
| 3.56 | 注册机构(RA) registration authority | 经授权对数据元或其他对象注册的组织。 | × | | × | | × | × |
| 3.57 | 注册机构标识符(RAI) registration authority identifier | 赋予注册机构的标识符。 | × | | × | | × | × |

| 序号 | 术语 | 定义 | “部分”编号 | | | | | | |
|------|--------------------------------------|--|--------|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 | 5 | 6 | |
| 3.58 | 注册状态 registration status | 数据元在进行注册期间所处各种状态的标示。 | × | | × | | | | × |
| 3.59 | 相关数据参照 related data reference | 数据元与相关数据间的参照。 | | | × | | | | × |
| 3.60 | 表示 representation | 值域、数据类型的组合,必要时也包括度量单位或字符集。 | × | | | | | | |
| 3.61 | 表示类别 representation category | 用于表示数据元的符号、字符或其他表示的类型。 | | | × | | | | × |
| 3.62 | 表示术语 representation term | 数据元名称的成分,用于描述数据元的表示形式。 | × | | | | | × | |
| 3.63 | 主管机构 responsible organization | 对必选属性内容负责的机构或其所属部门。数据元就是由这些必选属性来规定的。 | | | × | | | | × |
| 3.64 | 语义学 semantics | 有关词义注释的语言学分支学科(韦氏)。 | × | | | | | × | |
| 3.65 | 分隔符 separator | 在名称中连接或分隔成分的符号或空格,又称分界符。 | | | | | | × | |
| 3.66 | 已标准化数据元 standardized data element | 在数据元注册系统中可以优先使用的已审核数据元。 | × | | | | | | × |
| 3.67 | 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,分类法和本体论法。 | | | | | | × | |
| 3.68 | 提交机构(SO) submitting organization | 对数据元注册系统的数据元提出增补、变更、取消、删除或撤出的机构或其所属部门。 | | | × | | | | × |
| 3.69 | 同义名称 synonymous name | 与给定名称有区别但表示相同的数据元概念的单字或多字的指称。 | | | × | | | | × |
| 3.70 | 句法 syntax | 语言表述的结构,以及支配语言结构的规则。字符或字符组间的各种关系,这些关系与字符或字符组的含义、解释和使用方式无关。 | × | | | | | × | |
| 3.71 | 分类法 taxonomy | 依据类及其子类间的固有关系进行分类的方法。 | × | × | | | | | |
| 3.72 | 术语 term | 以语言表述的形式对某一特定语言中已经定义的概念的标示。 [GB/T 15237]。 | × | | | | | | |
| 3.73 | 主题词表 thesaurus | 按给定顺序排列参照词汇,其中显示和标识了词汇间的关系。 | × | | | | | × | |

| 序号 | 术语 | 定义 | “部分”编号 | | | | | | |
|------|---------------------------------|--|--------|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 | 5 | 6 | |
| 3.74 | 关系类型 type of relationship | 数据元与相关数据间关系的一种表述。 | | | × | | | | × |
| 3.75 | 值域 value domain | 允许值的集合。 | × | | | | | | |
| 3.76 | 版本 version | 注册机构内,一套逐渐完善的数据元规范中的一个数据元规范发布的标识。 | | | × | | | | × |
| 3.77 | 版本标识符(VI) version identifier | 赋予版本的一个标识符。用于提交或修改数据元注册。 注:在第5部分,等同于版本。 | | | × | | | × | × |

4 缩略语

CASE——计算机辅助软件工程(Computer-Aided Software Engineering)

EDI——电子数据交换(Electronic Data Interchange)

ERD——实体关系图(Entity-relationship Diagram)

IEC——国际电工委员会(International Electrotechnical Commission)

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)

OSIE——开放系统互联环境(Open Systems Interconnection Environment)

RA——注册机构(Registration Authority)

RDBMS——关系数据库管理系统(Relational Database Management System)

SC32——ISO/IEC JTC1 32分委员会(ISO/IEC JTC1/Sub-committee 32)

SDLC——软件开发生命周期(Software Development Life Cycle)

5 方法论无关性

公认的事实是:不同的方法被用于派生出面向应用的数据元。仅数据建模就有许多方法(如:信息工程和面向对象)用于标识和形成数据元。该数据规范标准的六个部分,独立于其他数据元派生方法和技术。因为该标准不仅可应用于所有的数据元,而且可以很好地应用于生成数据元的任何方法。

6 数据元的基本概念

鉴于GB/T 18391的目的,数据元由以下三部分组成:

- 对象类:现实世界中的想法、抽象概念或事物的集合,有清楚的边界和含义,并且特性和其行为遵循同样的规则而能够加以标识;
- 特性:对象类的所有个体所共有的某种性质;
- 表示:值域、数据类型的组合,必要时也包括度量单位或字符集。

对象类是我们希望用于收集和存储数据的事物。对象类的例子有轿车、人、家庭、雇员和定单等。区

分实际的对象类与其名称是很重要的。某些“想法”用一种自然语言可以简单地表述(如英语),而用另一种语言却难以表述(如汉语),反之亦然。例如:“在过去的12个月里至少有一次活产生育的年龄在15~45岁间的妇女”是一个有效对象类,但却难以用汉语简单命名。不过,对象类可以由两个或更多的其他对象类构成。该例可由概念“年龄在15~45岁间的人”和“在过去12个月中至少有一次活产生育的妇女”组合而成。

人们用特性来区别和描述对象。特性的例子有颜色、模型、性别、年龄、收入、地址、价格等。另外,特性有可能需要用多个词组加以描述,这要视所用的自然语言而定。

数据元表示部分中最为重要的方面是值域。值域是数据元允许(或有效)值的集。比如,表示家庭年收入¹⁾的数据元可能用一个非负整数集(以美元为单位)作为有效值集。这是一个非穷举域的例子。可供选择的是,有效值可以是事先说明的类别清单,且每个类别都具有某个标识符,如:

- 1 \$ 0 —— \$ 15,000
- 2 \$ 15,001 —— \$ 30,000
- 3 \$ 30,001 —— \$ 60,000
- 4 \$ 60,001 —— +

该值域是穷举域的一个例子。在这两个例子中,都度量了同样的对象类和特性组合——家庭年收入。

对象类和特性的组合是一个数据元概念(DEC)。数据元概念(DEC)是能以数据元形式表示的概念,其描述不包括任何具体的表示。在上面的例子中,家庭年收入¹⁾实际上可以称之为一个数据元概念(DEC),它具有与之相关联的两个可能的表示。因此,数据元由两个部分组成:数据元概念和表示。

图 6.1 阐明了本节的思想。

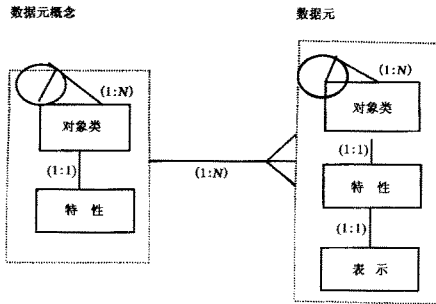


图 6.1 数据元结构

1) GB/T 18391.5 阐明了在一个数据元名下包含了表示术语,因此,家庭年收入不是一个完整的数据元名称。省略是有意的。

7 数据元与其他数据概念的关系

图 7-1 是数据层级的简化表示,表明数据元所在的那些层。数据元出现在数据库、文件和事务集中。数据元是一个组织管理数据的基本单元,因而它必然是组织内部数据库和文件设计,并用于建立与其他组织交流的事务集的组成部分。

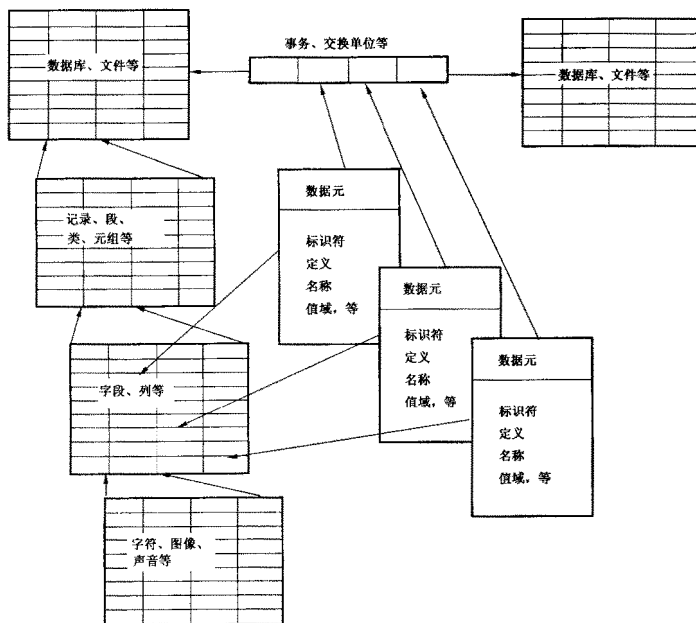


图 7.1 各数据层级的数据元

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。

一个组织需要将数据传输给其他组织时,数据元构成了事务集的基本单元。事务主要发生于数据库间或文件间,但组织间的文件和数据库结构(如记录或元组)并不一定相同。信息(数据加上理解)传输的公共单元是数据元。

8 GB/T 18391 第 1 部分至第 6 部分综述

8.1 各部分介绍

本条介绍了 GB/T 18391 系列标准的每一部分。概述其重点及各部分的重要性。

8.1.1 第一部分

GB/T 18391.1 数据元的规范与标准化框架,介绍并讨论了理解本套标准所必需的数据元的基本概念,并提供了与 GB/T 18391 各单个部分联系的相关环境。

8.1.2 第二部分

GB/T 18391.2 数据元的分类,提供了使数据元概念和数据元与对象类、特性和表示的分类模式相联系的程序和技术。

这些程序和技术将辅助注册机构应用分类方案完成以下内容:

- a) 分析对象类、数据元概念和数据元;
- b) 在以下范畴内进行比较:对象类、数据元概念和数据元;
- c) 减小数据元概念和数据元间的差异;
- d) 确切地定义和标识数据元概念和数据元;
- e) 辅助分析数据元以指定注册状态;
- f) 从数据元注册簿检索数据元概念和数据元;
- g) 识别数据元概念和数据元间的关系。

GB/T 18391.2 制定了一套原则、方法和程序,从分类法/本体论的角度阐明了在描述对象类、特性、表示、数据元概念和数据元(在这里称之为“信息元素”)时,什么是必需的(从最低角度出发)。这包括名称、非智能标识符、定义和其他可以作为属性的事物。GB/T 18391.2 表明如何从分类法到属性以及如何使用 GB/T 18391.3 中阐明的属性和其他所需属性。分类法将在资料性附录中论及,提供建议用于特性信息元素的限定词集和表示信息元素的限定词集。与限定词集一起的还有关于如何解决同义词和同形词问题的说明。GB/T 18391.2 还阐明了信息元素如何使用分类法以及如何使用相关的分类信息描述信息元素。

8.1.3 第三部分

GB/T 18391.3 数据元的基本属性,规定了数据元的属性。但仅限于数据元的基本属性集,而与其在应用系统、数据库和数据交换报文中的应用无关。

数据处理和电子数据交换能否不断地应用、提高将在很大程度上依赖于数据库中准确、可靠、可控制且可检验的数据记录。正确、恰当地使用和解释数据的一个先决条件是:数据的用户和拥有者对数据元的含义和表示有相同的理解。为了确保对数据元的共享,大量的基本属性需定义。

规定的基本属性将应用于下列主要活动中:

- a) 定义和规范数据元字典的内容;
- b) 设计和阐明面向应用的数据模型、数据库和用于数据交换的报文;
- c) 通信和信息处理系统中数据的实际应用;
- d) 不同数据元集间的交换或引用。

“基本”意味着它们在阐明数据元时是完全充分的,且可以确保其可以适用于许多功能,如:

- a) 信息处理系统的设计;
- b) 数据库中数据的检索;
- c) 用于数据交换的 EDI 报文的设计;
- d) 数据字典的维护;
- e) 数据管理;
- f) 字典设计;
- g) 字典管理;
- h) 信息处理系统的应用。

“基本”还意味着它们独立于:

- a) 应用环境;
- b) 数据元的功能(如:限定符、指示符);
- c) 含义抽象的程度(如:像“一个人的姓名”这样一个类概念的表示,或者“一个卡车司机姓名”这样一个具体概念的表示);

- d) 数据元的归类;
- e) 信息处理系统或数据交换报文设计的方法;
- f) 数据元注册系统。

“基本”并不意味着 GB/T 18391.3 中阐明的属性在所有情况下都是必需的。这些基本属性的区别是:

- a) 必选——总是必需的²⁾;
- b) 条件选——在某种指明的条件下是必需的;
- c) 可选——允许但非必需。

8.1.4 第四部分

GB/T 18391.4 数据定义的编写规则与指南,提供了研制准确的数据元定义的指南,列举了许多特定的规则与指南,确切地阐明了如何形成数据元定义。准确的、形式完好的定义是对数据元理解一致的至关重要的要求之一。形式完好的定义对于信息交换也是必要的。只有每一位用户对数据元有共同和准确的理解,其交换才能是毫无困难的。

8.1.5 第五部分

GB/T 18391.5 数据元命名与标识原则,就数据元的标识提供了指南。标识是指明、识别特定数据元的一个较为宽泛的术语。标识可以依据标识符的使用,以许多不同的方式来实现。标识包括对人们没有固有含义的数字型标识符的给定;图符(被赋予了含义的图像符号);和与数据元定义和值域有联系的,通常为人们所理解的具有嵌入含义的名称。

名称是赋予数据元的语义的、自然语言的标记,这些标记的变体具有不同的功能。一些名称是为了人们的使用和理解;一些名称是为了在某个特定的系统环境中的应用。名称常常由用户确定,且在不同的用户之间存在差异。本部分标准中的规则描述了名称的不同功能及其如何使用名称。在一个注册机构内每一个数据元有且只能有一个标识符。只要数据元的含义和表示类保持不变,标识符就不会发生变化。标识符仅在一个注册机构内是唯一的。GB/T 18391 并没有规定标识符的格式与内容。

每个数据元可能拥有多个名称——在特定的相关环境中,每个名称都是有意义的。结构化的名称可能被注册用于数据管理。某个首选名称会被其公司加以阐明,许多公用名称会为不同的用户群体所熟悉,缩写名称会被注册用于软件应用等。命名协议将在相关环境内得以执行。相关环境间的协议可以有所不同,相关环境的描述必须与每个数据元名称一起提供。

GB/T 18391.2 中所描述的分类模式可以派生一类结构名称。该名称的构成来源于三种分类法:对象类、特性和表示。每类名称与一些描述性限定词的组将构成一个数据元名称。该名称将有助于用户直观地在分类中查找或放置一个数据元。但是,它不是固定不变的,而且将其作为一个唯一的标识符也是不可靠的。

GB/T 18391.5 论述了数据元的识别与命名。尽管宜使用同样的规则,但它并没有对表示、特性和对象类的名称或标识符进行特别论述。

8.1.6 第六部分

GB/T 18391.6 数据元的注册,提供了注册申请者如何在一个中心注册机构注册数据元,以及如何为每一个数据元分配一个独一无二的标识符的规程和细则。本部分对已注册数据元的维护也做了规定。

已注册数据元的唯一性取决于注册机构标识符和版本的组合,一个注册机构内一个数据元将被分配一个独一无二的标识符。它们也普遍见于许多可用的数据元注册系统中。每一个注册系统由其逻辑和功能上隶属的数据元注册机构来维护。例如:与化学药品有关的数据元很可能在化学药品厂商注册机构下注册。应建立注册簿并为其编写索引,以便于设计应用或报文,例如 EDI,能较为容易地确定是否有合适的数据元已经存在。确定在什么地方建立新的数据元是必要的,并鼓励通过适当的修改从已存在

2) 需要用于记录数据元。GB/T 18391.3 并没有选定记录数据元概念所必需的属性子集的记录。

的条目中派生。这样,可以避免不必要的变更影响类似数据元的构造。注册可以使人们区分发挥相同功能的两个或两个以上的数据元,更为重要的是,它使人们可以区分具有相似或相同名称,但在某个或某几个方面具有巨大差异的数据元。

注册远远复杂于仅仅表明一个数据元是否被注册的双重状态。尽管坚持“好的”数据被注册是诱人的,但却是不实际的。因此,注册数据质量的提高分为三个层次(称作注册状态):已登录数据元、已审核数据元和已标准化数据元。另外,在这些质量层次间存在着管理状态的层次。总之,这些状态层次被称之为管理状态层。它表明在现行注册生命周期中一个已注册数据表示所获得的状态位置点。

8.2 应用 GB/T 18391 第 1~6 部分的基本规则

GB/T 18391 的每部分从数据元规范化的不同方面作了阐述,每部分应与其他部分配合使用。GB/T 18391.1 确立了各部分间关系,并就作为一个总体如何应用给出了指南。GB/T 18391.3 规定了注册申请者应为每一个数据元提供一套必选的元数据项。此外,还提供了潜在可能用到的附加款项列表以及每一个数据元基本属性的详尽特征。由于数据定义和标识这两个必选属性的重要性,分别在 GB/T 18391.4 数据定义的编写规则与指南及 GB/T 18391.5 数据元的命名和标识原则,这两个文件中对其进行了专门而又详尽的论述。

构造数据元的定义应遵循有关定义的文件。数据元的标识应遵循 GB/T 18391.5 中确立的原则。GB/T 18391.2 数据元的分类,规定了一套用于开发数据元及其成分分类模式的属性。当数据元编写完成后,能够在注册机构注册并在该处的注册系统中维护。GB/T 18391.6 数据元的注册,就此步骤提供了指南。

附录 A
(资料性附录)
数据表示和管理的基本概念

本附录介绍了 GB/T 18391 第 2~6 部分所依据的基本思路。

A.1 数据元

数据元是称之为数据的一个广义概念的特殊成员。总而言之,数据是事实、想法或命令的一种表示数据被收集、组织、记录、处理和存放在一个可检索的表中。数据还必须适用于以人工或自动方式进行交换、解释及加工处理。

有许多结构用于数据组织与管理,如数据合成、实体、文件、对象类、对象、记录、关联、关系、行、段、主体域、表以及元组。它们与数据元并不具有相似性,但可以包括或通过一些数据库实现或逻辑建模来等同于数据元的支持。

字节和位也是数据的构件,尽管它们被用于电子媒体中数据元的注册,但并不等同于数据元。在数据库中,数据元可以作为信息组(符号组、域)或字符列来处理。在 Chen 的 ER 数据模型中,它是一个属性(见图 A.5)。在某特定的相关环境中被视为不可分割时,一个数据元则被作为一个单独的数据单位。在自然界中,它是数据的单位,表示关于对象类的单独事实。(如:一个被赋值“M”和“S”的字符码表示了“雇员”这一对象类的婚姻状态的属性)。在其使用范围内,它不可能被分解为更多且具有有用含义的基本信息组。因而,数据元可被定义为在用户论述领域内是与用户相关的。数据元是自然界中对对象类特性以电子或书面形式的表示。

A.1.1 表示

一个特性可由能够被人们解释的某一符号集来体现。一个单独的特性可由几个交替的数据元甚至数据元组(通常称之为数据合成,有时也称之为数据元集或数据元链)来表示。这样,一个特性可以由数据元组构成的数据合成或一个单独的数据元表征。

A.1.2 常用法

所有数据从业者和理论家都会涉及数据元的概念。无论用什么方法或技术,数据元都是整个软件开发生命周期(SDLC)的公用纽带。在 SDLC 的早期阶段,它们曾被视为实体(或对象类)的属性。在 SDLC 的后期,具体的数据值被赋予它们的实例作为符号组或字符列。在 SDLC 的任一阶段,对于软件的生产者和用户来说,数据元是可识别的。

数据元是数据共享和共同持有的最小单元。一些数据元得以共享的信息系统成分有:

- a) 企业信息模型;
- b) 数据模型;
- c) 数据流程图;
- d) 数据库设计(模式、文件、表格);
- e) 接口规范;
- f) 计算机程序。

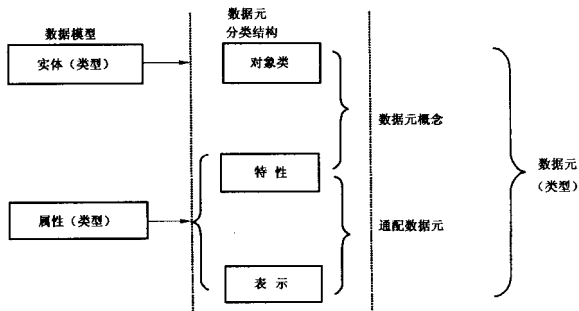


图 A.1 数据元结构

A.1.2.1 数据模型中的数据元

在数据建模出现之前，“数据元”这个术语已被普遍使用。用于表示指导企业商务信息的数据单元传统上被称之为数据元。数据建模的演进就是为捕捉这些数据表示的语义提供一种有效的方法。

图 A.1 描述了用于 GB/T 18391 中的数据元结构和术语与一些更为传统数据建模术语的关联。

在一个数据模型中，一个实体(实体类型、对象类等)的某个特性的属性会被企业选择记录为数据。对每个实体，通常有许多属性会引起企业的兴趣。

数据模型和对象模型(面向对象定向范例中)用于识别兴趣体(实体或对象)应用相关环境中的诸多事物。属性提供关于这些实体和对象使用所需的信息。用于整个自动化信息系统环境的数据元是面向对象范例中这些实体或对象以及它们属性的表示。产生于数据模型的数据元的名称的典型形式是实体名称和实体属性名称的合成(图 A.2)。

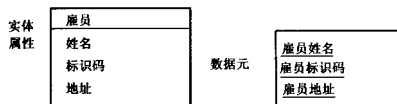


图 A.2 数据模型中的数据元

数据元名称大体上是以前述方式产生的。在对象模型中，类或对象名称与类或对象属性混合使用以形成数据元名称(图 A.3)。对象模型不同于数据模型的地方在于：前者可包含有关对象或类的附加信息，如行为或运行。



图 A.3 对象模型中的数据元

在数据模型或对象模型中，一个属性经常等同于一个数据元(见图 A.5)。它是自然界中对象某个类型单个特性的表示。而另一类思路是将对象类特性的概念与特性的表示分隔开来。由于数据元包含了表示，人们不必依靠其在数据库、屏幕和纸张等中的表示就可以了解其隐含的思想，因而将其隐含思想分隔开来是合理而又明智的。这与那些将属性看作与值域有关，而不是属性具有值域的数据建模技术是

一致的。

在某些情况下,如度量单位,属性与数据元间存在着明显的差异。例如:“日期”是时间连续区上用来度量最近一天的一个点。这样,它可以被看作是一个单独的事实并用一个单独的属性表示在一个数据模型中。但是,有好几种方式来表示一个日期。在美国,最常用的是:月的名称、一月中的天数和年数。它们是三个数据元。然而,同样的日期事实可以用过去某确定日期后所流逝的天数这样一个单独的数据元来表示(如天文学家就使用儒略历)。许多单独事实可以用多于一种度量衡单位来表示,既可用英制表示,也可用公制来表示。

一个数据元概念上的等同体被称之为:属性、特性、数据元概念、逻辑数据元和商务事实。在此,它是作为一个特性来论及的,以区别于人们通常所说的属性(一般包括其表示形式)。这样,特性就成了某个对象类所有成员共有的特征。

特性可以用一个例子来解释。假设一棵树是自然界中我们感兴趣的对象。但是,我们关心的是任一棵树,而不是某棵具体的树。树的特征中我们感兴趣的是其高度。树高就是一个对象类加上一个特性(数据元概念),但还不是一个数据元,因为合适的表示形式尚未阐明。我们可以从多个度量高度的方式中选择一个来表示树高。

相对于特性,人们更倾向于用特性类这个术语来命名数据元那个方面。对象中的一类如:人群没有身高;而每一个对象个体,如:个人,就有身高。因此,对于称之为人群的对象类来说,身高是该对象类的一个特性类。但在该标准中,术语特性的使用遵循常用法则。

有时,数据元可以由几个表示为数据元的要素部分来派生,每个部分又表示为数据元。这些派生物可以有多种形式。一个例子是电话号码由几个要素部分组成。在美国,电话号码由十位数字唯一地描述,这些号码很容易由一个数据元表示。但是,电话公司(和其他公司)需要将电话号码分隔为区号、交换码和线路号,形成三个数据元。将区号、交换码和线路号串联起来(以正确的方式)形成一个数据元以表示整个电话号码。

另一个派生的例子是代数运算。均值(或平均数)的计算需要观测值及其个数,因此,一个人群平均年龄的计算需要表示人群年龄的数据元和表示人群中人数的数据元。这样,就可以用派生中阐明的公式进行显而易见的算术运算。

这些例子是相当简单的,但不难想象那些需要在数据元中搜集的更为复杂的派生类型。与数据元关联的派生为现存的数据元与新需要的数据元之间发生关联提供了一个强有力的手段。

A.1.2.2 表中的数据元

关系型数据库中的数据元以字段名的形式出现于表格中。图 A.4 给出了数据库表格中数据元的一个例子。

记录

属性

数据值

| 雇员 | | | |
|----|----------|--------|--------|
| 号码 | 姓 | 出生日期 | 工资额 |
| 1 | Rood | 47/3/4 | 483.00 |
| 2 | Herden | 48/6/3 | 501.00 |
| 3 | Albright | 51/7/9 | 490.00 |

图 A.4 数据库表格中的数据元

A.1.2.3 数据管理工具中的数据元

图 A.5 标明了数据管理工具中经常与数据元(黑体部分)有关联的术语。

| | | | | |
|--------|-----|-----|-----|-----|
| 传统的 | 文件 | 记录 | 字段 | 字段值 |
| 相关的 | 关系 | 元组 | 属性 | 元素 |
| 对象定向的 | | 类 | 属性 | 实例 |
| RDBMS* | 表 | 行 | 列 | 数据值 |
| ERD** | — | 实体 | 属性 | |
| 替换 ERD | | 实体类 | 属性类 | |
| 术语 | ... | | | |

- * 关系型数据库系统
- ** Chen 格式实体关系图

图 A.5 数据元数据管理术语

A.2 分类

A.2.1 主题词表

主题词表是使相关术语关联的工具,主题词表术语有助于现有数据元的定位。查找名称构件的大量同义同,近义词以及同形异义词使得主题词表成了一个很有用的工具。它能够提供最优先术语和其他术语间语义上的联系。在指导同形异义词(拼写相同而表示不同概念的词)使用的同时,主题词表还可以指导用户通过涉及等同、层次以及关联关系的选择。

一个标准名称构件的主题词表可以由注册员开发并分配给各有关方面;此外,应鼓励主体领域主题词表的开发。

A.2.2 分类法和本体论

分类(如“属”和“种”)至少可以通过两种途径实现。第一种,同时也是最简单的方法是标准化方法,即每当一个新数据元注册时,该方法就会生成一个分类。更理想但非常麻烦的方法是,首先形成一个包括所有可能数据元的完备分类,然后将新注册的数据元放入事先定义好的位置。最为实用的方法是,先形成一个基本的而相对较为简单的分类,在数据元被注册时,允许其在使用严格规则的情况下逐渐完善。

最为普遍的方法是通过词典编纂的形式产生一种分类法。词典编纂过程更加关注语言的词或词汇而并不直接关注语义学。如果人们接受了这样假设:除非一个概念可以用词汇加以描述,否则,人们是无法理解此概念的。这样,词典编纂者的方法就可以服务于数据元的分类要求。即人们可以为其所思考的概念创造词汇。

每一个数据元注册应具有最大满足其用户的特定分类。如此,不同的注册可以选择使用不同的分类。

在一个分类中,每个节点是一个或多个上位类的一个下位类。该节点不仅沿用了上位类的含义,而且其含义同时也受到了上位类的限制。无论一个数据元被定义得如何好,分类对于其中的节点含义的确切描述,无疑具有极大的帮助作用。分类的另一个主要的好处在于有助于一个具体数据元的查找。分类通过大量的数据元描述支持导航查询。

一个基本的分类结构应能有助于数据元的注册、分析和应用。它基于这样的前提:数据元是自然界存在的对象类的特性的表示。数据元的类别由这样三个类组成:

- a) 对象类;
- b) 特性;
- c) 表示形式。

通过查找该类概念的标记,就可以查出所需的数据元。

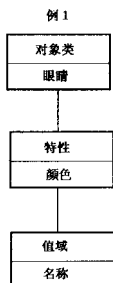
数据元的分类模式最好通过实例进行描述。但是,一个数据元分类模式可以用于多个实例。这主要取决于元数据用户是否为数据的最终用户或是否参与信息资源管理。若是后者,则取决于他们在软件开

发生命周期中所处的阶段。例如数据建模者应用其定义数据库结构以支持一项具体应用。

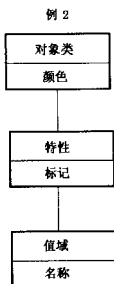
- a) 数据建模者发现需要使用该方法处理一项具体的数据。在本例中,是测量马的高度。
- b) 数据建模者认识到可以称该“特性”为“高度”。为了通过注册,数据元注册被建议使用面向导航的特性分类。在处理尺寸的分类范围内,发现了名为“高度”的这个类。该类名的定义确认了其描述了所指特性。
- c) 数据建模者认识到该对象类可以称之为“马”。在该注册中,对象类分类被建议使用“动物”类下的“马”这一类名。该类名的定义证实这是所需要的。马就成为数据模型使用中的对象类。
- d) 数据元注册表明高度这一特性被视为与马这一对象类相关联。注册则为这一特性是马从其站立的平面到其肩隆的(即:肩部)位距。
- e) 表示分类用于查找用于度量的注册。注册中马的一个标准化的数据元是以公制中米表示的高度。但是,数据建模者知道该数据元不会满足需要,因为最终用户坚持用传统方法——掌宽来测量并记录马的高度。用该分类模式查找注册簿则发现没有这样的数据元被注册。
- f) 数据建模者描述了以掌宽为度量单位测量的关于马高的新数据元的表示形式。该数据元得到了充分描述并被提交注册。只是其表示(即:主要是值域)需要重新描述,因为马这一对象类及其附属特性高度已经被注册过。
- g) 在数据模型的应用中,高度掌宽度量将成为马的一个属性。
- h) 将来,当任何人需要该数据元时,会发现其在注册中的描述使用了三种分类模式。对象分类中马类、特性分类中高度类和表示分类中掌宽度量类的联系,可以使将来的用户直接找到该数据元的此类描述。

在 GB/T 18391.2 数据元的分类中,数据元分类模式得到了详细描述。

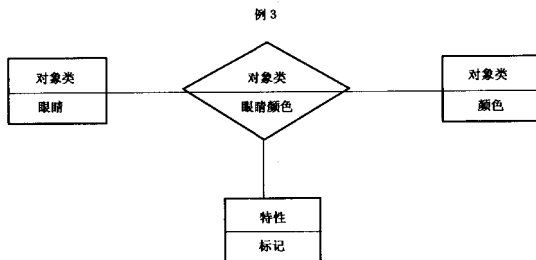
在实际中,特性和对象类间的区别并不常常是绝对的。差异与考虑中所要论述的领域有关,其最好的例子是身体特性。比如,眼睛颜色名称可被视为一个数据元概念,眼睛是其对象类,颜色是其特性。源于该数据元概念的许多可能数据元中的一个数据元可以是眼睛颜色名称。



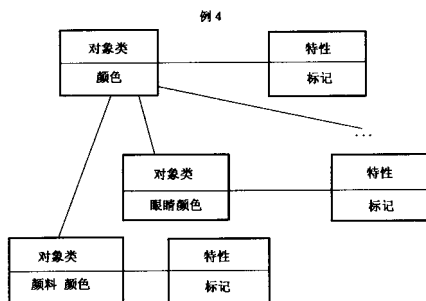
但是,如果某人从事标记各类可观察到的颜色时,颜色将成为对象类,而特性则可被称之为标记。与之关联的一个数据元可以被称之为颜色名称。在此,颜色是一个对象类,而在第一个例子中,它是一个特性。但没有一个可以被视为不正确。



第二个结构可以用于眼睛颜色中,眼睛颜色可以视为眼睛对象类与颜色对象类之间关系的对象类。



另一种描述眼睛颜色的方法是将其作为一个子类。眼睛颜色作为颜色对象类下面的一个子对象类。



例 3 和例 4 结果的实质是一样的,就这两个例子来说,眼睛颜色名称的值域可以是颜色名称值域的子集。

例 5

| |
|---------------|
| 穷举域 |
| 颜色名称 |
| 红色 |
| 灰色 |
| 橙色 |
| 茶色 |
| 米色 |
| · |
| · |
| · |
| 眼睛颜色名称 |
| 棕色 |
| 兰色 |
| 绿色 |
| 淡褐色 |
| 紫罗兰色 |
| · |
| · |
| · |
| 琥珀色 |
| · |
| · |
| · |

关于此情况的另一种思路是：颜色名称值域可以作为 1) 在例 4 中颜色子集，或 2) 在例 3 中与颜色关联所有对象类的一个超集。GB/T 18391 允许此种数据元结构变体的使用。

A.3 数据元的元数据属性

A.3.1 标识

如同由数据元表示的特性和对象类一样，唯一的数字型标识符赋予给每一个已注册的数据元。在有些情况下，唯一的数字型标识符也是一个数据元的给定数值。这些唯一的标识符也被注册在一个数据元注册中。它们使得可以跨越自然语言和应用系统直接解释这些数据元。虽然数据元在不同的自然语言中或在同一种语言的多种版本中会有不同的解释，但具有唯一性的标识符保持不变，并成为由一个版本向另一个版本转化的桥梁。

每一个数据元应获得一个标识符以使该数据元在整个计算空间中得到唯一的标识。而这些计算空间有：应用程序、区域计算系统、分布式的计算系统、组织、企业以及面向所有国家的全球型此类计算空间。标识符不含有任何信息。因此，它们可以由注册员赋予数据元并永久地依附于它们。数据交换用的标识符由注册机构赋予并维护。

在所限定的学科或其主题领域内，除了标识符外，数据元可以被赋予任何一个可替换的名称或图标。一个数据元名称可以是一个注册机构内首选的名称。如同标识符一样，首选名称与数据元之间保持着一一对应的关系。名称通常是数据用户标识数据元并与之相互作用的基本手段。

数据元由不同的注册机构注册。它们中的每一个都首选持有数据用户熟悉的标识符、名称、图符或其他形式的标识。对于可替换标识符，标识符连同其相关环境一起以应用文件的形式表现出来。

一些值域中的每一个数据值也有可能被赋予一个标识符。这对穷举域具有特殊的用处，它将为与数据实例相关的名称国际语言间的翻译提供便利。例如：数据元“眼睛颜色名称”可有穷举域“灰色、兰色、绿色和褐色”，在此，每个数据值可以有一个它们自己唯一的标识符（如，灰色=1 357；兰色=2 468）。

A.3.2 定义

定义是数据元含义的自然语言表述；它的断言。对于数据元开发来说，数据元的定义是极其至关重要的一个方面。为了共享，数据元必须有一个形式上完备、清楚、精确并被普遍理解的定义。GB/T 18391.4包含了对数据元定义的广泛讨论，给出了关于其编写的精确的规则与指南。

A.3.3 表示

数据建模者通常称属性的表示形式为它的“值域”，或简称为“域”。鉴于GB/T 18391的目的，我们称数据元的表示部分为表示。

数据元在商务运作中呈现为值，并在信息交换等功能中得到共享。数据元通常有一个允许值的集合。这个允许值的集合被称之为值域。

数据元从不表示为一个单独的数值，因为它是一个类（如，数据值完整的集合）而不是一个单个事例。比如，雇员标识符是一个数据元，它的值域由一个特定企业中允许值的一个完整列表来描述。这里的数据值仅是雇员标识符所有实例的一个列表。数据元的一个实例只有一个单个数据值并被称之为一个“数据元实例”。

一个特性有定义并隶属一个对象类。相反，一个表示没有定义，但有一个格式类、允许值、最大字符数，如果可以度量，还应有一个度量单位。比如，数据元“月名称”就有以下表示：

- a) 格式类=alpha；
- b) 允许值=January, February, March, April, May, June, July, August, September, October, November, December；
- c) 最大字符数=9；
- d) 度量单位=(不适用，因为不存在度量的问题)。

A.3.3.1 穷举域

尽管数据元的元数据的结构可以从对象类或特性的角度考察，也可从表示的角度考察，这最容易用一个穷举域的例子来表明。穷举域是一个可以由所有允许值列表指定的值域。比如，人们给国家标上名称，就很容易引证它们。包含现今世界上所有国家名称的集合就是穷举域的一个例子。

尽管特性和(表示的)有关的值域可以被考虑用来标识数据元一个潜在有效数据值的集合，对象类通常将宽泛的值域限定为数据值的一个具体的子集，以用于某个具体的数据元。图A.6表明了用不同的对象类对值域的限定。

在图A.6数据元的实例，X公司雇员姓名中：

- a) “名称”的值域是所有可能的名称；
- b) “人的”将其归类为所有可能名称的一个子集；
- c) “被X公司雇用”进一步将名称值域限定为非常具体的一个人名子集：X公司雇员的姓名。

因此，X公司雇员姓名就是由所有可能名称这个值域限定为仅仅是X公司雇员的那些姓名。该数据元所有可能有效的姓名数目就被大大地减少为一些具体姓名。

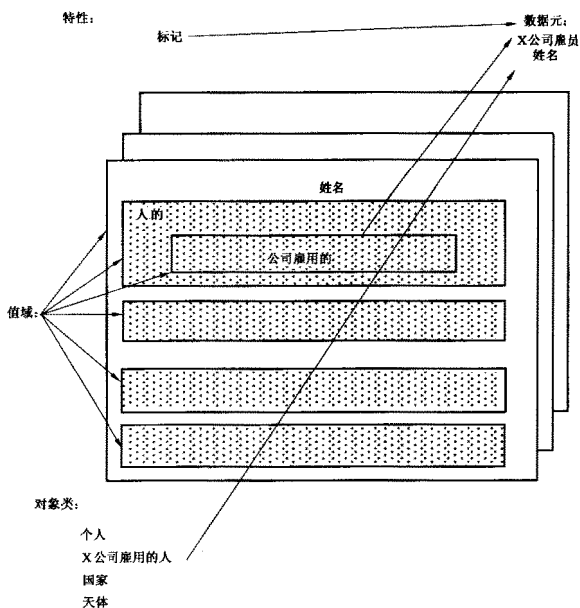


图 A.6 值域限定

一些穷举域可以看作是一个由一些更广泛穷举域构成的交集。图 A.7 表明，在一些情况下，值域是可以重叠的。

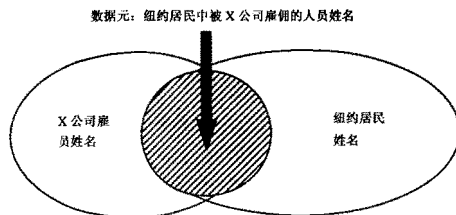


图 A.7 值域交集

A.3.3.2 可再利用域

一些域有被许多数据元使用的潜在可能。比如，所有人类眼睛颜色的名称可被用于雇员、承包商、夫妻和数据库中其他类的人员。

一个严格定义并被指定的域可以被再利用。这些可再利用的域在用于数据元时，具有相同的值域、表示和特性。比如，雇员眼睛颜色或承包商眼睛颜色。我们称这些可再利用域为通配数据元（见图 A.1）。

A.4 数据元注册系统与注册

被授权对数据元注册负责的一个群体被称之为“注册机构”。一个注册机构：

- a) 分配唯一的标识符；
- b) 确保所有所需的元数据得以用文件的形式证明；
- c) 依据元数据的品质和综合程度分配其状态层次。

该注册机构被授权对论述的领域负责。该领域可小至一个公司内的一个分支单位，也可大到整个行业或国家。注册机构对所授权论述领域内所有被认为会被共享的数据元注册负责。

注册机构间存在一定的等级关系，每个机构负责的领域必然大于其下属机构。当需要在低一级的两个或更多注册机构间共享数据元时，可以生成一个高一级的注册机构。低一级的则成为高一级的（具有更广泛的论述领域）注册机构的注册提交机构。

注册机构对其在一个注册系统负责的数据元进行文件描述。

注册机构对注册系统数据元的完整性负责。

附录 B
(资料性附录)
元数据和数据元

该附录讨论元数据和数据元之间的关系。

B.1 什么是元数据

通常,元数据被定义为“关于数据的数据”,但是该定义并不确切,因而会导致歧义和混乱。实际上,元数据是一种信息和文献,它使得数据在经历了时间的推移后,对于用户来说,依然具有可理解性和共享性。它使得在元数据可获得的情况下,数据依然是有用的、可共享以及可理解的。

所用产生数据的组织都有义务提供(生成)必要的元数据,使得数据对于它的内部和外部用户都是可理解的。仅仅使用户可以获得数据,而缺乏理解和解释数据必要的信息显然是不够的。

每一个提供数据的组织还有一个义务就是对元数据的成分及构成,给出确切的定义。其原因包括三个方面:

- a) 在一定的相关环境中,必须对数据和元数据给出明确的界定(尽管在信息系统内,元数据常常被作为另一种数据);
- b) 元数据的整个范围必须被确定,即:充分、全面地描述数据所必需的所有元数据;
- c) 由于每一个组织对元数据的需求和应用都相差甚远,因而无法对元数据给出一个普遍可以接受的定义。

因为元数据是数据(从某种超越某个组织需求的角度来看),元数据也可以像数据库中数据一样,在其储存库中,对元数据进行存储和检索。有序地存储和应用元数据,将使迅速、有效的使用数据成为可能。通过联机数据传播、发送,数据及其元数据就可以同时获得。这样,数据用户就可以通过元数据来了解其需要用的数据的信息。

由于各种组织提供了不同类型的数据,因此对元数据也有不同的需要和要求。但是,元数据宽泛的分类模式使之在不同组织间具有更多的共性。一个普遍确立的方案模式是:

- a) 系统——计算机程序设计及数据库管理所必需的物理及逻辑特征,包括文件的定位、存储介质、记录格式、数据库模式、数据字典等信息;
- b) 应用——数据理解和应用所必需的信息,以及术语定义、搜集程序和工具、数据采集处理等信息;
- c) 管理——成本、时间表、预算及与数据搜集项目、分析管理的相关信息。

综上所述,元数据是数据,并且就形式而言,可以依据数据模型和元模型进行组织。对元数据的全面理解是所有数据用户建立此类模型的首要步骤。这些模型将成为建立元数据库的基础。同时这些模型也成为一些元数据分类的依据,因而,它们也成了分类方案的一部分。

关于元数据的另一种观点则考虑了数据和元数据的可能用户的因素。每一类型的用户会有不同的需求,而且一些用户的需求是互不重叠的。一个程序员也许只需知道系统的元数据,而一个管理人员可能只想看看管理元数据。通常,数据分析员有更为广泛的需求,特别当他们是数据提供组织的外部用户时。总之,不同类型的用户,其需求(基于他们利用数据要回答和解决问题的类型)和水平都会对生成适合于他们的元数据产生影响。

B.2 数据元、元数据、元模型

关于数据元的相关信息是任何一个组织的元数据的一个完整的组成部分。GB/T 18391.2至GB/T 18391.6是这样描述元数据的:一个组织的数据元必须具备元数据。这些元数据将便于用户理解和共享该组织的数据。对于潜在的数据用户来说,分类、基本特性、定义、命名和注册是信息的最主要的几个方面。

将元数据存储于一个库中并使之条理化就需要建模。GB/T 18391.2至GB/T 18391.6是这样描述信息元模型的：元模型是从一个注册系统或库中获取信息所必需的。开发这样的元模型是一件不太容易的事情，但它的开发将大大地增加该标准实施的有用性。

元数据不可能涵盖理解数据元所要表示的数据所必需的所有信息。许多关于数据的内容和管理元数据及组织在该标准中是缺省的，而计算机处理的元数据几乎也是缺省的。例如，一个收集调查资料的组织将会有许多关于如何收集数据的信息。这些诸如样本和问卷设计等常常是理解数据所必要的。但是，样本和问卷设计并不是用来表示和描述数据元的元数据的一部分（源自GB/T 18391）。

确立一个宽泛的元模型将使得额外的元数据的描述成为可能。内容、计算机处理和管理区域均超出了GB/T 18391的范围，它们都需要开发自己的元模型和数据模型。可扩展性将会使新模型拥有自己的位置，以适应于基本的数据元的元模型。这样，就可以建立起一个更加完善的模型。
