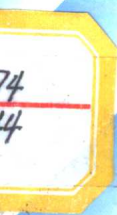


数据元标准化



中国标准出版社

数据元标准化

刘植婷 赵艳华 刘 永 编著

中国标准出版

内 容 提 要

在我国以国民经济信息化为指导,积极筹建信息高速公路、金字工程、EDI、CIMS 等高新技术纷纷启动之时,我们将对离散数据表示领域的最新研究和开发成果介绍给广大读者,以期通过数据元标准化支持全球电子信息的交换。

本书主要阐述了数据元标准化及其协调管理办法,简明扼要地介绍了数据元概念的演变、属性、分类及在信息自动化处理系统中的作用。书中附有大量实例和图解,便于读者加深理解。

数据元标准化

刘植婷 赵艳华 刘永 编著

责任编辑 贾玉勤

*

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: 68522112

永清第一印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

版权专有 不得翻印

*

开本 850×1168 1/32 印张 4 $\frac{1}{2}$ 字数 123 千字

1997年3月第一版 1997年3月第一次印刷

*

ISBN 7-5066-1266-6/TP·033

印数 1—1 500

定价 13.00 元

*

标目 291—07

序

为了积极推进我国数据元标准化的研究工作,使其在“九五”期间更好地配合、支持国民经济信息化工程的开拓,使我国数据元标准的制、修订工作与国家高新技术领域的重点工程紧密结合起来,同时也为了使广大的标准化工作者、信息自动化处理系统的设计人员和各类信息应用者了解国际数据元标准化工作的最新动态,使我国的信息管理工作尽快与国际标准接轨,我们编著了《数据元标准化》一书。

该书较系统地介绍了各类数据元的概念和属性,指出数据元的协调是信息交换和网络建设中的一个关键性问题,并就数据元结构的标准化问题提出了普遍适用的思考方法。为使广大读者便于阅读和理解此书,并通过阅读此书对数据元标准化有一个全面的、清楚的了解,使更多的同志投入到数据元标准化工作中来,我们首先向大家介绍一下数据元标准化的一般概念和国际标准化组织(ISO)从1965~1995年,30年来开展数据元标准化工作的情况,以及我国数据元标准化工作的发展与展望。

一、一般概念

数据元标准化是信息技术标准化的重要组成部分,是信息交换的前提,是信息系统建设的基础和运行的保证,其研究开发水平和推广应用程度,不仅是衡量信息技术标准化水平的标志,而且同信息产业的发展息息相关。

数据(data)一般定义为:事物、概念或指令的一种形式化的表示形式,用以适应人工或自动方式进行通信、解释或处理;数据元(data element)迄今尚没有一个统一的定义。但是,人们从实际工作的需要出发,往往从不同的角度,给予不同的定义。今天,我们便从数据元标准化的需要出发,认可数据元的定义是:表示概念的一类数据,其特

性可由支持信息交换的一组数据元属性来表示。也可以认为:数据元是一组可识别和可定义的数据基本单元。在当今的电子数据交换(EDI)系统中,可以把数据元定义为:在确定的上下文中被认为不可再细分的数据单元;数据元标准化(Standardization of data element)将定义为:对数据元的总则、定义、描述、分类、表示和注册等制定统一的标准,并加以贯彻、实施的过程。

数据元标准化是一项艰巨而复杂的工作,不仅涉及面广、技术性强,而且与具体应用的业务领域密切相关。因此,在制定具体的数据元标准时,必须进行充分的调查研究,收集有关资料和信息,认真分析并确认其存在性。当对一个事物、概念或其属性的可能存在或含义发生分歧时,既要根据已有的知识及其相互间的协调关系,又要考虑到处理对象的唯一性、独立性,然后通过充分协商,取得一致的认识后,对其进行准确的定义或描述。不难看出,人们的互相了解并取得共识,这才是数据元标准化的可靠基础。因此,数据元标准化的成功与否很大程度上取决于广泛的一致性。在国家之间,即国际水平上获得较大范围内的一致性,就需要更多的从事数据元标准化的专家为之努力奋斗,取得共识。

二、国际标准化组织(ISO)数据元标准化工作简况

国际标准化组织开展数据元标准化工作,大体经历了三个阶段:第一阶段——起步阶段(1965~1970年)。

这个阶段的工作重点是:对各国数据元编码情况进行调查了解,并着手制定当时数据处理经常遇到的数据元编码标准。

随着电子计算机的广泛应用,应广大计算机用户的要求,国际标准化组织第 97 计算机和信息处理技术委员会(ISO/TC97)于 1965 年 10 月,在日本东京会议上,决定成立研究数据元标准化的工作组(ISO/TC97/WG—K),工作组定名为数据元及其编码表示(Data Elements and their Code Representation)。1966 年,在日内瓦召开了 TC97/WG—K 成立会,会上讨论了工作范围、工作计划、数据元标准化的基本词汇等。

WG—K 工作组的工作计划之一是调查各国以及国际间已应用或准备应用的数据元编码的内容与结构,分商业、企业、个人情况、产业、地区、职业、工种、时间等八个方面。但从 1968 年 4 月发表的调查结果来看,各个国家和国际机构对调查表的答复率很低,这就说明当时许多国家数据编码工作尚未开展。以开展数据处理工作较早的美国来看,也没有达到全国统一编码的水平。

WG—K 工作组除了对数据元编码情况进行调查了解外,还对数据处理中经常遇到的国际单位制(SI)提出了编码标准草案。与此同时,还提出需要拟定对日期的年、月、日、工作周,时刻的时、分、秒等的编码方案。由于当时世界各国对日期的写法习惯不同,执行统一标准难度较大,加之当时不少国家还没有把应用编码标准提到日程上来,更不用说统一推行编码标准了。这一切,都说明数据元标准化的起步十分艰难。

第二阶段——发展阶段(1970~1987 年)。

这个阶段的工作重点是:制定有关描述数据元和数据交换中涉及数据元表示法的国际标准,开始研究数据标准协调一致问题,并着手制定有关指南和规范。

随着社会生产力的发展,信息在自然、社会、经济复合系统中的作用日益重要,信息的商业价值越来越明显,于是各类信息应用系统应运而生。由于数据是各类信息应用系统作用的对象,就象人体里的血流一样,它们的存在和运动决定了系统的生命和活力。因此,研究数据及其标准化的问题就显得格外重要。

为适应新形势对数据元标准化的要求,1970 年,ISO/TC97 在柏林会议上审议通过了美国提出的 WG—K 工作组改变名称及实际工作范围、工作计划的提案,将该组名称改为“数据元表示法”(Representation of Data Elements),其编号由 TC97/WG—K 工作组改为 TC97/SC14 分技术委员会;工作范围扩大了,包括凡与数据交换和数据处理应用有关的,且又经常参与交换的数据元,均属标准化的对象。工作计划中列出数据交换常用的数据元标准化项目,其中包括日期、时间、国别、语言、个人情况,团体组织、会计科目、邮政地址、地理

位置经、纬度表示法、计量单位以及制定有关规范,以供各国协力推行。

该组所以把“编码表示”(代码)更名为“表示法”,是因为“表示法”的使用范围要比“代码”更为广泛,它包括代码、缩写、名称和数值。这就排除了与其他码如字符码、自检码和编程码等在“代码”术语使用上的混乱。

ISO/TC97 于 1972 年 6 月 20 日正式通过了 ISO/TC97/SC14 数据元表示法分技术委员会(SC14)的范围和工作程序。

其范围包括:

用于信息交换和信息处理,一般被交换的数据元表示法的标准
化。

其内容有:

1. 制定有关描述数据元和数据交换涉及到的数据元表示法的
国际标准。

2. 制定有关表示共同关注的数据元的国际标准,包括下述表示
法:

- (1) 日期和时间
- (2) 国家
- (3) 语种
- (4) 个人情况的标识
- (5) 机构的标识
- (6) 账目的标识
- (7) 邮递和海运地址
- (8) 地理位置经、纬度表示法
- (9) 计量单位
- (10) 数值的表示。

3. 制定推荐指南和规范,以保证用规范的方法进行标准化和描
述国际上信息交换涉及到的数据元。

4. 为促使所提供信息交换用的数据标准协调一致,必须加强与
其他机构以及 ISO 委员会的联系。

在此期间,数据元表示法分技术委员会正是根据上述范围和工作内容开展工作的。数据元表示法方面的一些国际标准,就是在这段时间内制定和发布的。

第三阶段——推广应用阶段(1987年~现在)

本阶段的工作重点是:推广应用已制定的数据元表示方面的标准,继续完成有关指南和规范,研究制定数据元协调方面的标准。

随着信息技术的发展,信息产业也随之形成,由于人们对信息技术标准化这一技术基础工作的重视,导致越来越多的组织着手开展这方面的工作。

1987年国际标准化组织(ISO)和国际电工委员会(IEC)两大国际标准化机构,联合成立了信息技术标准化联合技术委员会(Joint Technical Committee, 1),定名为 ISO/IEC JTC1。可以认为,这正是两大国际标准化机构顺应国际信息技术标准化发展的潮流,采取的一个有益的步骤。实际上 ISO/IEC JTC1 是在原计算机和信息处理技术委员会(ISO/TC97)、信息技术设备技术委员会(IEC/TC83)和微处理机系统技术委员会(IEC/TC47B)的基础上改组而成立的,原上述三个组织的各分委员会及其工作组直接进入 ISO/IEC JTC1,数据元表示法(SC14)也不例外。

ISO/IEC JTC1 成立后,JTC1 不仅加强了与 ISO 各委员会的联系,而且更进一步密切了 JTC1 各分委员会之间的协调。如 JTC1/SC14 和 JTC1/WG3〔开放式 EDI 工作组(现为 JTC1/SC30)〕之间确立了以互派技术联络员的方式,进行协调和协同进行标准化的研究工作。JTC1/SC14 派员参加了 ISO 基本语库管理委员会基本语库项目组(BSR)和 JTC1/SC21/WG3 IRDS(信息资源字典结构)起草组的工作。

推广应用数据元标准最为明显的领域是 EDI 系统。数据元标准化的开发对 EDI 在全世界的开发起到了推动作用。这是因为,尽管人们不是用同一种语言,但有了大家一致赞同的数据元概念及其表示方法,人们之间的交换联系就成为可能了。人们只要对信息按共同约定的规则进行统一组织、分类与标识,用同一的概念,并用相同的

表示,就能做到共识,不会产生歧义。这种简化的概念表述,极大减少了人工及计算机的处理差错,有效地避免了数据资料在不同国家、地区的重复输入、翻译等工作,为 EDI 的国际化创造了条件。

从数据元标准化在 EDI 系统中的运用来看,它不仅体现了现代社会对数据元标准化的要求,确定了数据元标准化在 EDI 系统中不可替代的地位,同时也显示出数据元标准化在信息处理越来越电子化过程中的重要地位和作用。

随着人们对数据元标准化的需求,1993 年 JTC1 通过决议,将 SC14 的名称由“数据元表示法”改为“数据元的原则”(Data Element Principles)。工作范围拓广为信息处理系统间交换的数据元(包括总则、定义、描述、分类、表示和注册等)以及有关语法的标准化。根据这一工作范围,SC14 加强了 WG4(数据元标准化协调工作组)的标准化工作。几乎每年都举办 WG4 工作会议。经过几年的努力,WG4 负责制定的有关数据元标准化协调方面的标准,已有三项正式发布实施。

当前 SC14 的工作项目有:

ISO 2955:1983《信息处理——有限字符集系统中使用的国际单位制和其他单位的表示》;

ISO 6523:1984《数据交换——组织标识法的结构》;

ISO 6709:1983《用于地理定位的纬度、经度和高度的标准表示法》;

ISO 5218:1977《信息交换——人的性别表示法》;

ISO/IEC 7826-1:1994《信息技术——代码值交换的一般结构——第一部分:代码表的标识》;

ISO/IEC 7826-2:1994《信息技术——代码值交换的一般结构——第二部分:代码表的注册》;

ISO/IEC DTR 7352《信息技术——在数据交换范围内的数据元分组导则》;

ISO/IEC TR 9789:1994《信息技术——用于数据交换的数据元表示和组织的导则——编码方法和原则》;

ISO/7064:1983《数据处理——校验字符系统》;

ISO/IEC CD 14957《信息技术——数据元值的表示:格式标记》。

ISO/IEC 11179《数据元标准化协调》;该项目分成以下几个部分:

ISO/IEC 11179-1《数据元的生成与标准化框架》;

ISO/IEC 11179-2《定义范围标识概念的分类》;

ISO/IEC 11179-3:1994《信息技术——数据元标准化和规范——第三部分:数据元的基本属性》;

ISO/IEC 11179-4:1995《信息技术——数据元标准化和规范——第四部分:数据定义规范化的规则与指南》;

ISO/IEC 11179-5:1995《信息技术——数据元标准化和规范——第五部分:数据元的命名和识别原则》;

ISO/IEC CD 11179-6《信息技术——数据元标准化和规范——第六部分:数据元的注册》。

三、我国数据元标准化的发展与展望

(一) 我国数据元标准化工作的发展

我国数据元标准化的研究工作始于70年代末期。从1979年全国产品目录统一编码座谈会第一次提出数据元标准化以来,在有关方面的共同努力下,我国的数据元标准化工作发展迅速,在管理机构建设、制度建设、理论研究、标准的制修订、情报交流及国际合作等方面都做了大量的工作,取得了可喜的进步。

1. 首先是编译了一批具有国际组织和工业发达国家有关数据元表示方面的书刊和资料,即《产品、职业、行业分类编码研究报告和译文汇编》、《技术经济信息分类编码统一系统》、《国际标准职业分类》、《国际标准行业分类索引》、《分类编码、文件格式国际国外标准文献汇编》、《国际数据元标准译文集》等。这些书籍和资料的出版发行,为全国数据元标准化科研工作者,各级管理人员提供了一套比较系统的参考资料,从而带动了我国数据元标准化工作的开展。

2. 组织制定了一批数据元方面的标准,如《中华人民共和国行政区划代码》、《国民经济行业分类和代码》、《全国工农业产品(商品、物资)分类与代码》、《信息分类编码的基本原则和方法》、《贸易数据交换——贸易数据元目录》等国家标准,这些标准在全国范围内普遍采用,尤其在历次全国工业普查、人口普查以及全国科技干部普查中得到全面应用和验证,并纳入了我国国家经济信息系统和办公自动化系统及其子系统,现已成为我国 EDI 系统不可分割的组成部分。

3. 完成了一批“七五”、“八五”国家重点科技攻关项目,其中影响面较大的有:75-68-1-8(1-6)《固定资产投资指标体系与代码》、《宏观经济指标体系与代码》。88-02-47-02《全国干部、人事管理信息系统标准化研究》、《学科分类的原则与方法》等重大研究课题。对这些课题富有成效的研究,填补了我国在这些领域的空白,为人工和计算机处理该类信息打下了基础,为国家标准的制定提供了科学依据。

1983年,我国成立了全国计算机与信息处理标准化技术委员会数据元表示分技术委员会(对应 ISO/TC97/SC14),标志着我国有了专门从事数据元标准化的学术组织,并被接纳为 ISO/TC97/SC14 的 P(积极)成员。由于分委员会各成员单位积极参加各项国际标准化活动,包括承担相应的国际义务,对于缩短同国际之间的差距,发挥了十分积极的作用。从 1986 年起,我们每年都派人员参加 JTC1/SC14 年会,并在会上介绍情况,发表论文,受到了各国同行的称赞,有些论文还被纳入委员会的文件。在承担国际标准起草任务方面,我们也做出了努力,国际标准 ISO/IEC CD14957《信息技术——数据元值的表示:格式标记》就是我们和法国联合起草的。1992 年 10 月在我国举办了 JTC1/SC14 第十五届年会与 WG4 工作组会议,会议取得圆满成功,并受到与会各国朋友的称赞。

1984 年 6 月国家经济自动化管理系统一期工程设计任务书列入了组建国家标准局信息分类编码研究所的工作计划。1985 年 1 月国家标准局信息分类编码研究所正式宣布成立,宣告我国数据元标

准化工作进入了一个新阶段。

从1988年至今八年间,是我国数据元标准化开拓新的应用领域时期。这些新的应用领域包括了建立组织机构代码标识制度和条码的开发应用;推动数据元标准化在EDI、银行电子化、计算机集成制造系统(CIMS)等领域的应用。

组织机构代码标识制度的研究工作,始于1985年,1988年进入了实际工作阶段。在此期间,国务院以国发(1989)75号文转发了国家技术监督局、国家计委、国家科委、民政部、人事部、中国人民银行、国家统计局、国家工商局、国家税务局、国家信息中心等十个部门联合上报国务院的报告。1990年由国家技术监督局牵头成立了由上述十个单位参加的“全国企事业单位和社会团体统一代码标识制度领导小组”,并在中国标准化与信息分类编码研究所内设立了办公室,此后,全国各省市自治区先后成立了相应的领导小组和办公室。

1989年11月经国务院同意,国家技术监督局颁发了GB 11714—89《全国企业、事业单位和社会团体代码编制规则》(参照采用ISO 6523:1984《数据交换——机构标识法的结构》,做为强制性的国家标准贯彻执行。

1993年7月由国家技术监督局、中央机构编制委员会办公室、国家计委等13个部门联合以技监局发(1993)14号文下发了《企事业单位和社会团体代码管理办法》,管理办法规定了采用发放代码证书作为代码的载体,以此作为采集、建立组织机构代码数据库的数据和标识机构的凭证。

1993年3月经人事部批准成立了全国组织机构代码管理中心,全面负责组织制定与建立组织机构代码标识制度有关的国家标准和工作规范,组织实施和协调管理代码的编制、证书发放,组织推广应用;组织协调与国际相关组织的联系。

到目前为止,全国各省、市、县各级代码管理机构所需的计算机设备和代码证书全部配齐到位,在全国范围内已完成近450万个企业、事业、社会团体和行政机构的赋码、数据采集和颁证工作。代码证书的查验制度在统计、银行、税务、公安、交通等管理部门中强制执

行,效果良好。1995年下半年全国组织机构代码管理中心对全国代码工作分东北、华北、西北、西南、华东、中南三片进行了初步验收,全国三十个省除西藏外基本完成颁证工作,省级代码数据库已基本建成。已查验的省级数据库的数据已输入中央数据库,为中央数据库的建立和进行全国代码信息系统联网打下了技术基础。

对条码的研究工作是从1986年开始的,两年以后的1988年12月经国务院批准成立了中国物品编码中心,除继续研究开发和推广应用条码及自动识别技术之外,还由国务院授权负责统一组织、协调管理我国的条码工作,代表国家参加国际物品编码协会。当前我国正在采用的条码前缀号(国别号)690、691、692,就是由国际物品编码协会赋予的。

目前条码中心已承担的国家重点项目——《条码系统研制》和“八五”科技攻关项目《条码印刷品质量监测技术研究》、《交通运输、仓储、物流条码研究》等课题已经完成。到1995年底,参加条码系统的成员已达21500家,有近32万种商品采用了国际通用条码标志,从而保证了这些商品顺利进入国际市场参加平等竞争,为国家的出口创汇做出了贡献。据统计1995年共计有500家商店设置了条码自动扫描系统(POS)。此外,中国物品编码中心还参加了用于行政、商业和运输业的电子数据交换(EDIFACT)的“中国促进EDI应用协调委员会”,担任了中国商用EDI销售组的牵头单位。由于我国条码扫描、打印、检验等设备相继研制成功,并投入批量生产,从而为条码产业的发展奠定了基础。

“八五”期间,国家为使高新技术的开发、应用纳入标准化、规范化的轨道,国家科委相应设立了编制《EDI系统标准化总体规范》、《金融电子化系统标准化总体规范》、《CAD通用技术规范》和《863/CIMS标准化指南》等标准化规范或指南的重点科技攻关项目。在这些规范或指南里均有数据元表示标准化的独立章节,为我国数据元表示的研究、开发、应用、协调提供了依据,极大地促进了成果共享,避免了重复劳动,加速了这些领域的研究成果商品化、产业化的进程。

从数据元标准化国家标准制定情况来看,到目前为止,我们已经编制发布了大约 100 多项数据元国家标准,并在各项工作中发挥着重要的作用。

(二) 我国数据元标准化工作的展望

“九五”期间,我国数据元标准化工作的重点仍然是:继续跟踪国际数据元标准化动态,积极推进我国数据元标准化的研究工作,其主要内容有:

1. 加快进行国际标准转化为国家标准的工作步伐。为了借鉴国际标准,已把现有的国际标准(或草案)译成了中文并编辑成册,定名为《国际数据元标准译文集》,从中选出较成熟的项目列为专题进行研究,为制定国家标准做好技术准备。

2. 深入进行数据元标准化协调的理论研究。包括对数据元的分类、数据元的基本属性、数据元的命名原则和数据元的注册等,要从理论上进行深入的研究,以指导数据元在各个领域的实际应用。

3. 跟踪国内高新技术领域 CIMS、CAD/CAM、EDI、银行电子化、国家“金”字工程的开拓,使数据元标准化工作随动于国内高新技术的进程,使数据元标准的制、修订工作,与国家重点工程(系统)建设紧密结合起来。

4. 做好数据元表示标准的成龙配套工作。同国际数据元标准化的发展进程一样,我国数据元标准化也是从数据元编码标准入手的。按照我国制定的数据元编码,标准体系表确定的 12 个方面的标准系列(参看赵艳华主编《信息分类编码标准化》,中国标准出版社,1989 年版),进行成批、成组的开发,使其填平补齐。近几年内除把重点放在物品分类编码标准的制、修订工作外,应该把劳动力资源——人事信息管理系统指标体系里没有代码的指标项,逐年列入标准制、修订年度计划,使其尽快配套完善。

5. 条码、代码重点项目要限期完成。近期内要全面完成组织机构代码数据采集和代码证书颁发工作,建立起中央数据库和代码信息系统,加强代码工作立法和规章制度建设,强化强制使用的力度,逐步扩大应用范围,为社会各部门提供更多更好的服务。积极开展

“IC 卡”技术标准规范的研究,不断总结“IC 卡”试点工作经验,为国家现代化管理提供确实可靠的技术依据。

条码工作,除继续做好新系统成员的发展外,要对老成员进行复审,推动条码自动扫描商店(POS)系统的建立,积极开展商业电子数据交换规则(EANCOM)的标准化研究,做好商业销售的 EDI 试点工作,推动条码产业的发展。

编 者

1996 年 5 月

目 录

1. 前言	1
2. 概论	3
3. 数据元概念的演变	5
3.1 概念的命名	8
3.2 数据元概念的命名	12
4. 数据元概念的属性	16
4.1 数据元概念属性的类型	16
4.2 数据元概念属性的性质	18
4.3 数据元概念属性的若干例子	19
5. 数据元概念分类	23
5.1 数据元概念分类的层级	23
5.2 数据元概念类别	24
5.3 数据元概念的隶属关系	39
6. 通用数据元	39
7. 通用数据元的属性	41
7.1 通用数据元属性的类型	41
7.2 通用数据元属性的性质	46
7.3 通用数据元属性的实例	47
8. 通用数据元的句法	51
8.1 简单的通用数据元	51
8.2 复合的通用数据元	52
8.3 合成的通用数据元	52
9. 通用数据元字典	53
10. 应用数据元	54
11. 应用数据元的属性	56
11.1 应用数据元属性的类型	56
11.2 应用数据元属性的性质	58

11.3 应用数据元属性的例子	59
12. 应用数据元的句法	62
12.1 简单应用数据元	63
12.2 复合应用数据元	63
12.3 合成应用数据元	63
13. 应用数据元字典	64
14. 不同语言的信息交换	65
15. 离散数据表示的词汇表	67
附录 1 数据元概念分类图	86
附录 2 对特定实体的名词标识的数据元概念矩阵	98
附录 3 要求专有名称的事物表	101
附录 4 信息分类编码国家标准目录	109
附录 5 CIMS 信息分类编码标准体系和标准体系表	117
附录 6 缩略语及有关术语注释	128