

地理信息及其元数据标准化

承继成 赵永平

(北京大学遥感与地理信息系统研究所 北京 100871)

摘 要 通过对标准化历史的回顾,阐述了数据标准化在当今人类由工业社会向信息化社会迈进中的作用,并根据全球信息高速公路建设对地学研究的挑战,指出了信息化社会中数据标准对信息采集及信息共享的作用;依据 ISO / TC211 对地理信息数据描述中元数据的建议,通过对 ISO 及 CEN 标准的分析,提出了建立我国元数据标准的基本思路,对我国地理信息标准化的建设具有重要意义。

关键词 ISO, CEN, 地理信息, 数据描述, 元数据, 标准化

标准化是伴随生产力的发展而产生的一种改造客观世界的社会性活动,是对重复性事物和概念所作的统一规定,其本质是管理^[1]。地理信息则是地理学研究内涵的表达,它通过对地理学研究对象进行描象与概念化,用定性或定量方法描述其本质,是一种可用于建立地理对象描述标准的载体,是空间信息标准化研究的主要组成部分。

本文根据标准化发展的历史,通过对地理信息描述规则的分析,指出了地理信息标准化首先应从数据描述中的元数据开始,并依据 ISO 及 CEN 标准的制定规则,对我国元数据标准化的研究进行了探讨,其基本方法对我国信息高速公路建设中地学信息源的流通与共享具有重要意义。

1 标准化综述

标准化是人类生活与生产的重要组成部分,它是人类文明的产物。其历史可追溯到大约 5000 年前的埃及,那时该区部分地方已出现了可互换使用的弓和箭,开始了古代标准化的萌芽。然而,由于由石器时代向现代工业革命的进程经历了几千年的艰辛,古代标准化的发展一直很缓慢,直至 1900 年前,随着枪的批量制造,才从美国诞生了近代工业的标准化。它的发展使人类社会进入了一个崭新的时代,即现代工业化社会。然而对标准化的系统研究却是伴随工业化社会的需求与各种标准团体的成立开始的。

1901 年世界上第一个国家标准化团体英国标准学会诞生,使标准化的研究开始向协调发展方面努力。紧接着,1906 年成立了国际电工委员会 (IEC),1926 年成立了国际标准化协会 (ISA,即国际标准化组织 ISO 的前身,在第二次世界大战中解体),这些组织的成立,为国际标准化发展起了推动作用,也为工业化社会向信息化社会的过渡奠定了基础。

事实上真正对标准化的发展具有影响力的事件还是 1947 年 2 月国际标准化委员会 (ISO) 的成立^[2],特别是 ISO 9000《质量管理和质量保证》系列标准的制定及 ISO 14000 系列环境标准的出台,由于这些行动对整个社会的发展产生了很大的效益,并与人们的生活和生产息息相关,这样不仅扩大了标准化在全球范围内的影响力,而且使 ISO 的标准逐渐成了整个社会遵循的准则。

我国也于 1978 年重新以国家技术监督局的名义进入 ISO,参加国际标准化组织,并开始标准化方面的研究与探讨工作。1980 年成立了中国标准化协会,1989 年颁布了《中华人民共和国标准化法》。这一系列活动标志着我国标准化工作已逐渐走上正规,并纳入了法制管理轨道。

在《中华人民共和国标准化法》中,根据目前国际上标准的制定与实施情况,提出了适合我国国情的标准化政策,并指出在今后我国标准制定中应积极采用国际标准或国外先进标准,同时将我国标准采用国际标准或国外先进标准的程度,分为等同采

用、等效采用和非等效采用 3 种,以便灵活应用。这一政策的实施,为我国标准化的建设起到了推动作用,使得我国迄今已制定的 17000 多个国家标准中,有 43% 左右的采用了国际标准和国外先进标准,并争取到“九五”末期对 ISO、IEC 现行的 12000 多个国际标准除了地理、气候、及基本技术等原因外,都要转化为我国标准,并尽可能等同、等效采用。

2 地理信息标准化的形成

2.1 信息社会化的形成

如果说 1946 年电子计算机的问世是信息革命的序幕,那么 1957 年第一颗人造卫星的升空,便标志着人类进入太空时代,以崭新的角度开始重新认识自己赖以生存的地球,掀起了汹涌澎湃的信息革命的浪潮。

1975 年,美国国防部的“高级研究项目局”(ADVANCED RESEARCH PROJECT AGENCY,简称 ARPA),通过卫星和无线电传输,形成了连接加州大学、洛杉矶分校、圣芭芭拉分校、犹它州大学和斯坦福大学研究所的 ARPANET 网络,即 internet 的雏形。该网络的建成标志着人类迈入信息化社会的障碍已排除,从而美国政府 1993 年放弃《星球大战》和《新一代高能加速器》计划,推出了“国家信息基础设施”(NI: National Information Infrastructure),即信息高速公路(IH: Information Highway)这一令全世界注目的跨世纪高科技信息基础系统工程,使信息革命的进程进入了实质性阶段。紧接着,世界上许多国家纷纷响应,使信息高速公路建设成为全球研究的热点。

信息革命在全球范围内的兴起,促使信息标准化工作日趋升温。国际标准化组织也意识到为对付信息爆炸和迅速有效地利用其潜力,如果不靠国际标准来有效地引导信息迅猛的发展趋势,必将出现一个电子的混乱局面。于是在 1982 年 ARPANET 采用 TCP/IP 标准解决了网络兼容问题后,ISO 迅速作出反应,并从 TCP/IP 的应用中得到启示,大量吸收 TCP/IP 的先进部分,完善其不足之处,提出了“开放系统互连参考模型(OSI/RM),以求得国际间信息通讯模式的统一。据统计,OSI 方面的正式标准目前已有 120 余项,并且国际上绝大多数厂商都向 OSI 的体系过渡,美国国防部 1986 年也宣布了 TCP/IP 向 OSI 过渡的计划。

OSI 作为世界范围内的标准,涉及的问题十分

复杂,体系也相当庞大,协议程序至今已齐备,在理论上几乎无懈可击,然而在实际应用中,OSI 并没有象 TCP/IP 那样在市场上风行。因此,各国政府又根据本国的实际情况,以 OSI 为基本标准水准,制定了相应的较精练的协议体系,即政府开放系统互连 Profile(GOSIP),以为本国服务。

自 1987 年 GOSIP 问世以来,网络系统互连的研究日趋活跃,并在近年内掀起了又一次高潮,先后推出了美国的“政府 OSI Profile”、英国政府 OSI Profile 和欧洲开放系统指南(EPHOS)等新的体系规范,并且这些体系都以 OSI 为基础,推动着 ISO 标准化的发展。它们的出现,不仅使信息化社会研究的内涵越来越丰富,同时也为信息化社会的到来辅平了道路。

2.2 我国信息研究的现状

信息化社会里,信息力,即一个国家生产、处理、传播和使用各种信息的能力,是综合国力的关键组成部分。因此在人类向信息化社会迈进之时,各国政府对本国信息力的建设都非常重视。我国为了加快国民经济信息化的进程,与国际接轨,也于 1994 年 2 月由国务院负责成立了“国民经济信息化联席会议”(1996 年改组为负有更多职能的“国务院信息化领导小组”),以正式参予全球性信息力的竞争,并与 1997 年 4 月在深圳召开了第一届全国信息化会议,开始了信息建设的全面准备工作。

事实上,信息高速公路时代信息化建设的中心问题便是数据共享的标准问题,其关键也就是建立一套能够被大家普遍接受与采纳的标准,以确保信息的最大共享。我国对这一领域的研究一直比较重视,比如早在 1983 年便在国家科委的支持下开始了有关地理信息(指数字地理信息)标准化的探索工作,另外对信息领域中的其它数据交换标准也给予了高度的重视。

目前我国已建立的与信息技术有关的标准化技术工作组织有 13 个,“九五”期间需要转化为我国国家标准的基础性国际信息技术标准约 800 多项。其中 ISO/IEC JTC1(信息技术)的标准 200 多项,ISO/TC184(工业自动化系统和集成)的标准项目近 60 项,ISO/TC68(金融)的标准项目近 30 项,ITU(国际电信联盟)的标准项目约 200 项,UN/ECE/WP4(国际贸易简化组织),负责 UN/EDIFACT(联合国/用于行政、商业和运输的电子数据交换标准的制定和维护)的项目约 100 项。

“三金工程”是我国信息高速公路建设的重要组成部分,它的标准化工作,将参照国际标准化组织(ISO)的开放系统互联(OSI)的相应标准协议,把国际上已成功应用的分组交换的 CCITT、X.25 推荐标准、适用于公用交换数据网 PSDN 的 X.32 推荐标准、适应国际电子信息交换的信息处理系统(MHS)的 X.400 推荐标准以及电子数据交换 EDI 标准等应用于系统之中,以形成一个开放互联的分布式大系统。按 EDI 标准体系的要求,“九五”期间我国急需制定的 EDI 标准就达 400 余项。其中包括 EDI 基础标准、单证标准、报文标准、通信标准、安全保密标准、编码标准、应用标准及相关标准等。

2.3 地理信息研究的萌芽

地理信息是伴随地理学的产生而出现的用于描述地理对象的载体,只是随着现代科学技术的进展,被赋予了新的含义,成为地理信息学(或地球信息学)研究范畴的表达。包含了用各种现代化方法采集、量测、分析、存储、管理、显示、传播和应用的与地理和空间分布有关的数据,是测绘学、遥感学、地理信息系统、计算机图形图像学、卫星定位技术、专家系统技术与现代通讯等学科和技术的有机体现^[3]。它在计算机与空间技术冲击下的信息时代,研究内涵已从传统的概念描述向如何用机器语言表达及信息共享方面过渡,成为以数字地理信息为主的信息术语了(所以文中的地理信息指数字地理信息),并在地理信息资源的有效配置与最佳组合状况上开辟着新的研究领域。

自从本世纪 60 年代初加拿大地理学家 Tomlinson 第一次提出建立地理信息系统思想,到 1971 年按 Tomlinson 思想建立的第一个被称为地理信息系统的加拿大地理信息系统(CGIS)投入使用后,地理信息的研究便进入了实质性阶段。特别随着遥感及全球定位系统的发展,有关这三者之间的应用研究及使用软件层出不穷,使全球范围内形成了一种各自为政的局面。面对地理信息研究的这种现状,国际标准化组织已经作出反应,以便形成一种共同的地理信息标准,求得全球信息一体化时代地理信息共享的实现。

早在 ISO 作出反应之前,欧洲标准化组织(CEN)已成立了 TC287 技术委员会,致力于制定地理信息方面的数据标准,推出了 TC287 系列标准草案。面对 CEN 的先行,ISO 于 1994 年 3 月成立了“地理信息/地球信息业(Geographic Information / Geomatics)标

准化技术委员会”,即 ISO / TC211,从事地理信息和地球信息业标准化方面的研究。该委员会与 CEN 达成了共同推进国际标准制定工作的协议,使地理信息的标准化工作也正式纳入了国际标准化轨道。目前 ISO 已有积极成员(P)22 个,观察员(O 成员)13 个。我国 1994 年正式加入该组织,现为 P 成员,并于 1995 年 7 月 25 日由国家技术监督局批准,授权国家测绘局为 ISO / TC 211 国内技术归口主管部门,国家基础地理信息中心为国内技术归口单位,形成了我国的对应组织。并已成立了 5 个工作组(WG),在 20 个工作项目上开展工作,元数据属于其中的第 15 部分。

3 元数据的研究

元数据即 Metadata,在地理界使用的比较早,比如早期各种地图周围的图例等信息便是典型的元数据内容,但它真正在这一领域引起人们的重视却是近年来的事,是随着地理信息标准化发展的需求和相关标准的开始制定引起人们注意的。对于元数据,目前还没有统一的定义,在欧共体的地理信息数据描述草案(CEN / TC 287)中^[4],被简单定义为“关于数据的数据”,而在 ISO / TC 211 中^[5],又被描述为“关于数据集的数据”。事实上,元数据是用于描述数据内容、定义、空间参照、质量及地理数据集管理等方面的数据,它是地理信息标准化的重要组成部分,也是地理信息数据描述的关键技术之一。

在 ISO / TC 211 目前的 442 份文件中,对于地理信息数据描述的草案思想仍有相当一部分来自欧洲,这主要与欧共体在这一领域起步较早有关。对于元数据的描述,目前则已由先前的定性描述向通过 ISO 10303 中 EXPRESS 等模型的定量描述过渡。国内对元数据的标准化研究刚刚起步,目前我们正在根据国际上地理信息标准化制定中的一些建议性草案,针对国内具体情况,对元数据标准中的基本概念、定义及研究范围作较全面的探讨。文中我们首先给出一些基本定义及模型,以便在今后的地理信息及各相应的研究中,能够朝标准化这个国际大方向努力,求得地理信息在进入信息高速公路时代的最大兼容与共享^[6]。

3.1 元数据标准描述中的基本定义

为使文中后面描述较为清楚,首先给出以下相关定义:

- 数据集:** 一个可标识的数据集合。
- 范围:** 数据集包含的水平、垂直地理空间及时间序列。
- 地理区域:** 根据间接参照系(国家或城市)定义的一个数据集所覆盖的空间范围。
- 元数据参考:** 关于元数据的信息。
- 组织:** 一个对数据集关系具有特殊作用的个体或机构。
- 组织角色:** 组织在数据集关系中的作用。
- 连接点:** 单独或一组对数据集关系具有特殊作用的点。
- 连接点角色:** 连接点在数据集关系中的作用。
- 辞典:** 用于描述含多层次结构数据集分类的词典。
- 辞典元素:** 辞典中的条目。
- 必选项(M):** 数据描述中应当提供的元数据元素。
- 可选项(O):** 可有可无的元数据元素。

3.2 描述地理数据集的元数据

地理数据通常组织成数据集。一个数据集包括一个或多个特征类型事件及它们的相关属性,也可以由其它数据集联合组成。数据集本身及其特征或其它更详细的内容通过元数据表达。

用于表达地理数据集的元数据应通过下面几部分内容给出,其中各标准名词后标有(O)的为可选项,其它均为必选项。

标识信息

简要描述数据集的信息。由数据集标题,可选择标题(O)、缩写标题(O)等组成,是数据集描述中首先要给出的元数据。

数据集概述

用于全面描述数据集的元数据。包括摘要、目的、用途、几何子模式、空间参照系、所用语言、参考文献(O)、示例(O)及相关数据集(O)等。

数据质量信息

质量信息是使数据的用户和生产者能够确定数据贴近规范的程度。所以应当给出描述数据集总体质量的元数据元素,包括数据集的进程历史、专题精度、时态精度、完备性、逻辑一致性、点位精度等内容。

空间参照系信息

地理特征怎样在空间定位的信息通过对空间参照系的描述给出。空间参照系由直接空间坐标系和

间接空间坐标系组成。其中间接空间参照系包含间接空间参照系类型和参照系日期(O);直接空间参照系由水准、椭圆、地图投影及高程参照系等组成。

时空信息

地理数据集在他们的地理层和时态上不尽相同,所以应给出其较详细的信息,包括范围进程信息及数据集完备性、平面范围、垂直范围及时态范围等。

空间数据表示信息

地理特征可能在不同的数据集中用不同的方法定义。元数据应当在给出其特征和定义时加以规定,同时还应描述特征间的相关类型。数据定义内容包括特征类型描述、属性类型描述和相关类型描述(O)3类。其中特征类型描述由特征类型名称、特征类型定义、几何图元、结构图元、特征类型码(O)、事件(O)、特征专题精度(O)、点位精度(O)及完备性(O)组成;属性类型描述由属性类型名称、属性类型定义、属性类型码(O)、属性域值(O)、属性专题精度(O)及时态精度(O)组成;相关类型描述由相关类型名称、相关类型定义、起始特征类型、结束特征类型、基数、约束(O)、相关类型专题精度(O)及逻辑一致性(O)组成。

分类信息

数据集的内容或专题范围可能会通过其特征给出,也可能在分类系统中被描述。如果这样,特征会成为一级或多级结构的组成部分,并且当特征类型规定为一个等级分类时,应通过辞典的方式说明。分类由辞典(O)和辞典元素组成。其中,辞典由辞典名称和辞典管理者组成;辞典元素由术语、定义、同义词、相关术语、广义术语、狭义术语组成,有时也通过图片(O)等加以说明。

分配信息

获取一个数据集时,数据集在哪里、是怎样被保留和怎样被传输的信息是必须的。同样也需要与数据集分配有关的信息,如使用权、版权及特殊数据集的传递媒介等。为满足上述要求,应规定组织和组织角色、连接点和连接点角色及分配等内容。其中,组织和组织角色由组织名称、可选择组织名称(O)、组织名称缩写、组织地址、角色及组织功能(O)组成;连接点及连接点角色由连接点名称、连接点地址及角色组成;分配由约束、版权、价格政策、分配单位、媒介、格式、在线访问、订购及后援服务组成。

元数据参照信息

元数据参照是有关元数据自身的一些信息,包括元数据日期信息、元数据连接信息、元数据标准信息、元数据扩展信息(O)、元数据辅助信息(O)、以及元数据约束信息等内容。

以上各部分为地理数据集描述中元数据的基本组成部分,在数据集描述中,依文中所给顺序依次描述,其它一些细节内容,我们正在制定中。至于元数据的国际标准,将于 1998 年底正式出台,从而形成全球范围内地理信息数据描述中的元数据标准。而从现在到元数据的正式出台,ISO / TC 211 技术委员会对元数据标准草案还要做一定的修改,因此建议国内从事有关地理信息方面的专家,应密切关注 ISO 的进展,努力使自己的研究与国际化接轨。

3.3 元数据应用简析

元数据作为 ISO / TC 211 地理信息 / 地球信息业中的重要组成部分,在未来信息化社会中将成为信息高速公路中地理信息的基本信息流被广泛使用。但在目前实际应用中仍然相当复杂,所以国际标准化组织建议通过 EXPRESS 等语言首先建立一种基本的元数据模型,然后通过该模型构造自己的实际应用模型,以求得从模型建立时就开始规范化、标准化。

目前使用的 EXPRESS 语言定义元数据集合的基本模式为:

```

SCHEMA 元数据;
REFERENCE FROM 模式类型;
:
USE FROM 模式类型;
:
TYPE 类型名;
:
END-TYPE;
:
ENTITY 名称;
:
END-ENTITY;
:
END-SCHEMA;

```

上述模块内的内容为 3.2 中提到的各部分及其更详细的分支内容,经格式化后,可通过表格的关系具体反应出。如某一地理面状网中有一电缆线,它的一段线路 A 的特征类型定义如下:

特征类型名称: 线路 A

特征类型定义: 线路中电缆内容保持不变的一段电缆线

几何图无: 线

结构图元: 面状图网中的边

该示例仅仅是一条线路的特征类型描述。事实上,在以地理信息系统为基本平台的地理信息研究中,往往涉及到多内容、多元素、多特征的描述,在数以万条信息描述的标准化过程中,若仍使用传统的基于人工逐条信息定义及描述的方法,难免为数据的标准化带来一定的阻力。所以对元数据的内涵和外延进行全面研究,不仅仅制定出标准,而且通过机器语言实现元数据的自动定义与识别,以形成地理信息元数据标准化的实用性模型,是即将要考虑的首要问题,也是我们正在研究的主要内容。

4 结束语

地理信息标准化,实现信息共享是一个国际潮流和当前地理界研究的热点。目前 ISO / TC 211 正在制订的 20 个标准将于 1999 年全部颁布,而元数据及其标准作为它的重要内容之一,将于 1998 年提前出台,所以该标准的制定与发展目前已到了最关键的时候了。我国虽然对这一领域的研究一直比较重视,并且目前也在做这方面的工作^[6-9],但由于一直没有专门从事这一研究的组织和机构,使得十多年来对地理信息标准化的研究都是从具体标准方面进行,缺乏全面系统的研究,而 ISO / TC 211 则从理论和规律上全面探讨地理信息标准化。所以面对目前全球信息化的发展,应尽快建立我国地理信息标准化方面的专家队伍,总结我国已有的基础,并跟踪国际研究的前沿,使我国地理信息标准化的建设由分散向系统化方面发展,并尽量等同或等效采用国际标准,尽快实现与国际接轨。

参 考 文 献

- 1 陈素枝. 试论标准化的本质. 中国标准化, 1996(1): 11—13.
- 2 程传辉. ISO 四十年来的发展. 中国标准化, 1989(10): 24.
- 3 李德仁. 地理信息学的形成及带来的机遇和挑战. 科技导报, 1995(3): 79.
- 4 CEN / TC, 287, Working Group 2. Geographic Information—Data Description—Metadata, Part 1, July 1995.
- 5 ISO / TC 211 Working Group 3. ISO STANDARD, 15046-15, GEOGRAPHIC INFORMATION—Part 15: METADATA, Version 2. 15, August 1997.

- 6 赵永平, 承继成等. 地理信息数据描述中元数据标准化的研究. 中国标准化, 1997(3): 6—9.
- 7 李军, 陈崇成. 地理科学元数据的研究. 地理研究, 1997, 16(1): 31—38.
- 8 陈常松, 姚绪荣. 地理信息标准化研究. 遥感信息, 1997(2): 19—22.
- 9 蒋景瞳, 刘若梅. 国外GIS标准化进展和我国的对策. 遥感信

息, 1996(3): 19—22.

作者简介

承继成, 教授, 国际欧亚科学院院士。1955年毕业于南京大学地理系, 1960年获原苏联莫斯科大学地理系博士学位, 先后出版了《流域地貌数学模型》等5本著作和50余篇研究论文。

The Normalization Research of Geographic Information and its Metadata

Cheng Jicheng Zhao Yongping

(*Institute of Remote Sensing and GIS, Peking University, Beijing, 100871*)

Abstract With reviewing the history of normalization, the authors have discussed the role of data normalization from the viewpoint of industrialization society to informatization society, also according to the challenge of the global information highway establishment to geoscience research, pointed out the role of data standard to the information collection and information sharing in the informatization society. According to the recommendation of ISO/TC 211 work group to the metadata of geographic information description and the analysis of the ISO and CEN standards, they present a basic idea of how to establish the metadata standard in the country. The research would have a very important meaning for the geographic information normalization establishment.

Key words ISO, CEN, Geographic information, Data description, Metadata, normalization