

中图分类号: K9, G307

文献标识码: A

文章编号: 1672 - 1586(2003)05 - 0001 - 05

国家标准《地理信息 元数据》研制与实现若干问题

蒋景瞳, 刘若梅, 周 旭, 贾云鹏

(国家基础地理信息中心, 北京 100044)

摘 要:通过探讨国家标准《地理信息 元数据》研制和实现的若干问题,包括标准的定位和适用范围、元数据的定义、内容、结构、分级,元数据的扩展,元数据专用标准制定、分层结构的应用与元数据标准实现,以及研制该国家标准策略考虑等问题,以便于对该国家标准能有更加深入地理解和更好地实施该标准。

关键词:元数据;标准;信息共享;标准实现

On the Developing and Implementation of the National Standard “ Geographic Information - Metadata ”

JIAN GJing - tong , LIU Ruo - mei , ZHOU Xu , JIA Yun - peng

(National Geomatics Center of China , Beijing 100044 , China)

Abstract :The paper discusses issues for developing and implementing the national standard , ‘ Geographic Information Metadata ’ , including the tenet and scope of the standard ; the definition , content , structure and level of metadata ; metadata extensions and profile creation ; metadata implementation ; the strategy for developing the national standard . So that the standard could be better understood and implemented .

Key words :metadata ; standard ; information sharing ; standard implementation

0 引言

国际标准化组织经过近 10 年努力,于 2003 年 5 月正式发布了国际标准《地理信息 元数据》(ISO 19115:2003)。我国从 2000 年开始立项研制国家标准《地理信息 元数据》,经过几次调整标准编制思想和策略,最终决定“修改”采用相应地理信息国际标准 ISO 19115。目前,该标准在征求意见的基础上,正处于编写送审稿、提交审查阶段,期望尽早完成审查,报批和发布实施。本文通过探讨地理信息元数据国家标准的定位,元数据的内容、结构与分级,元数据的扩展,对国际标准所做的修改及缘由,领域专用元数据标准制定与元数据标准实现等问题,以便于对该国家标准能有更加深入地理解和更好地实施该标准。

与国际上多数元数据标准相同,国家标准将元数据定义为“关于数据的数据”,是有关数字地理数据标识、覆盖范围、质量、空间和时间模式、空间参照系和分发等方面特征的描述性信息,目的是实现数

据共享;

元数据描述的最基本对象是“数据集”,可以扩展为“数据集系列”,或数据集内的较小的数据分组“要素”和“属性”;

元数据的存储形式为格式化的文本或关系型数据库表;

元数据的使用密切结合计算机网络,在因特网上为用户提供软件工具,以便查询和检索元数据及相关数据库;

元数据的实现一般包括支持元数据内容标准的操作工具软件(含输入、编辑和维护等)的开发、元数据收集、元数据库/文件的建立、元数据库/文件与数据库的链接和元数据查询浏览软件(如超文本浏览器)开发等。

因此,可以说目前元数据已经从简单的描述或索引发展成为用于管理数据、发现数据、使用数据的一种重要的工具与手段。它的作用包括:

数据生产者可以利用元数据对他们的地理数据进行详细的说明;

收稿日期:2003 - 08 - 20

帮助用户了解数据的基本特征,从而决定他们是否使用该数据和能够有效地应用数据;

使数据发现、检索和重复使用变得更加容易。用户能更好地确定地理数据位置、访问、评价、购买和使用地理数据;

1 标准的定位和适用范围

该标准定义的是通用的地理信息元数据,定义了描述地理信息和服务所需要的模式。

适用于:

- 1) 数据集编目、对数据集的完整描述,数据交换网站的信息发布;
- 2) 描述地理数据集、数据集系列、数据集内的较小的数据分组,以及单个地理要素和要素属性。

该标准规定:

- 1) 必选的和条件必选的元数据子集、元数据实体和元数据元素;
- 2) 适合于元数据全部应用范围(数据发现、确定数据的适用程度、数据访问、数据转换和数字数据应用)所需要的最少的元数据集,即所谓核心元数据;
- 3) 可选的元数据元素——在必要时,允许对地理数据进行更详尽的描述;

为满足特殊需要对元数据进行扩展的方法。

该标准原理可以扩展到许多其它形式的地理信息如地图、图表和文本文件,以及非地理数据。

2 元数据的构成

元数据由一个或多个元数据子集(UML 包)构成,元数据子集包含一个或多个元数据实体(UML 类)。

MD-元数据实体是下列实体的聚集:

- 1) 标识信息包含唯一标识数据的信息。标识信息包括有关资源的引用、数据摘要、目的、可信度、状况和联系办法等信息。
- 2) 限制信息包括有关对数据施加的限制的信息。限制实体分为法律限制和/或安全限制。
- 3) 数据质量信息包含数据集质量的总体评价。它包含数据志和数据质量元素。数据质量元素有完整性、逻辑一致性、位置准确度、专题准确度和时间准确度。它们可以进一步细分为若干数据质量量子元素。
- 4) 维护信息包括有关数据更新范围和更新频率的信息。
- 5) 空间表示信息包含数据集中用于表示空间信息

的机制信息,可以定义为栅格空间表示和 MD-矢量空间表示。

- 6) 参照系信息包括数据集使用的空间和时间参照系的说明。参照系可以再分为坐标参照系和椭球体参数。
- 7) 内容信息包含标识使用的要素类目信息和/或描述数据集数据层内容的信息。
- 8) 图示表达类目参照信息包括标识使用的图示表达类目的信息,用于说明数据集使用的图示表达类目。
- 9) 分发信息包括有关资源分发方信息和如何获取资源的信息。数字传送选项、分发方标识和分发格式等元素。数字传送选项包含用于数据集分发的介质、标准订购程序。
- 10) 元数据扩展信息包含用户自行定义的扩展信息,描述扩展的元数据元素,元素的值可以是“代码表”、“枚举”或“代码表元素”

应用模式信息包含有关用于建立数据集的应用模式信息。

元数据数据类型包括:

- 1) 覆盖范围信息描述有关实体的空间和时间覆盖范围,包含有关实体的地理覆盖范围、时间覆盖范围和垂向覆盖范围的信息。其中,地理覆盖范围可以分为边界多边形、地理边界矩形和地理说明三种。
- 2) 引用和负责单位信息提供引用资源(数据集、要素、原始资料、出版物等)的标准方法,以及资源的负责单位信息。负责单位的数据类型包括资源的负责方标识,和/或职务,和/或单位。还说明负责方或单位的地址。

3 地理数据集核心元数据

该标准定义完整的元数据元素集,内容复杂,元数据实体和元素超过 400 个,实际应用中通常仅仅应用它的一个子集。对一个数据集而言,有时只使用基本的最少数量的元数据元素。回答:“特定专题的数据集是否存在(‘什么’?)”、“是否覆盖特定的地区(‘何处’?)”、“数据集特定的日期或时段(‘何时’?)”以及“了解更多情况或订购数据集的联系方式(‘谁’?)”等基本问题。这种标识一个数据集,特别是用于编目所需要的最少量元素称为“核心元数据”元素,它包含的元数据元素除必选元素外,还有部分可选元素能,以求提高互操作能力,所以用户能准确地理解生产者或发布者提供的地理数据和相关的元数据。声称与该标准一致的数据集元数据专用

标准应包含该核心元数据元素。

表1列出描述数据集所需的核心元数据元素(必选的和推荐可选的)。其中“M”表示该元素是必选的,“O”表示该元素是可选的,“C”表示特定条件下该元素是必选的。

表1 地理数据集核心元数据
Tab.1 Core data of geographic data sets

数据集名称 (M)	空间表示类型 (O)
数据集参照日期 (M)	参照系 (O)
数据集负责单位 (O)	数据志 (O)
数据集地理位置 (由4个地理边界坐标或地理标识符确定) (C)	在线资源 (O)
数据集语种 (M)	元数据文件标识符 (O)
数据集字符集 (C)	元数据标准名称 (O)
数据集专题分类 (M)	元数据标准版本 (O)
数据集空间分辨率 (O)	元数据语种 (C)
数据集摘要说明 (M)	元数据字符集 (C)
分发格式 (O)	元数据联系方 (M)
数据集覆盖范围补充信息 (垂向的和时间的) (O)	元数据创建日期 (M)

4 元数据扩展

该标准提供标准元数据和关联结构,可用于多种多样的地理数据。定义和域值尽可能通用,以满足不同学科对元数据的需求。然而,数据的多样性意味着通用元数据可能适应不了所有的应用。为此,标准提供了定义和应用扩展元数据的规则,以更好地满足特殊用户的需求。

允许扩展的类型包括:增加新元数据子集;建立新的元数据代码表,代替域值为“自由文本”的现有元数据元素的域;创建新元数据代码表元素(扩展代码表);增加新元数据元素;增加新元数据实体;对现有元数据元素施加更加严格的约束条件;对现有元数据元素施加更多限制的域。

为保证扩展的内容与标准严格一致,扩展必须符合以下规则:

- 1) 扩展的元数据元素不应用来改变现有元数据元素的名称、定义或数据类型。
- 2) 扩展的元数据可以定义为实体,可以包含扩展的和现有的元数据元素,作为其组成部分。
- 3) 允许对现有元数据元素施加比该标准要求更加严格的约束条件(如:在该标准中是可选的元数据元素,在扩展后可以是必选的)。
- 4) 允许对元数据元素的域施加比该标准更严格的限制(如:该标准中域为“自由文本”的元数据元素,在专用标准中可以限定为适当值的列表)。

- 5) 允许对该标准允许的域值的使用加以限制(如:在该标准中现有元数据元素域值有5个值,在扩展后可以规定它的域只包含其中3个值,扩展后要求用户从这3个域值中选择1个)。

- 6) 允许对代码表中值的数目进行扩展。

- 7) 不得扩展该标准不允许的任何内容。

在扩展元数据之前,必须仔细地查阅标准中现有的元数据,确认缺少的元数据内容,然后进行扩充。扩展的每一个元数据子集、实体和/或元素,都应定义其名称、缩写名、定义、约束、条件、最大出现次数、数据类型和域值。还要定义其与现有元数据的关系,以便确定结构和模式。

5 专用标准制定

国家标准《地理信息元数据》是基础性标准,一般不直接用于某一领域。各个行业、部门或单位可以以该国家标准为基础,制定“专用标准”,以满足其特定需求。

按照ISO/TC 211制定的相关标准定义,专用标准是一个或多个基础标准的集合,在适用的范围内,为完成一个特定的功能,对所选择的基础标准的章、类、选项和参数的标识[ISO/IEC TR 10000-1:1998]。专用标准是从基础标准和其它专用标准派生的,因此,从定义上讲,与专用标准一致是指与派生该专用标准的基础标准和专用标准的一致。一个简单的专用标准,是基础标准的一个子集。

为制定《地理信息元数据》国家标准的专用标准,应选择全部核心元数据元素,并按需要选择全集元数据中其它元素,特别是必选元素。必要时,还可以按规则扩展基础标准中没有的元数据元素(图1)。

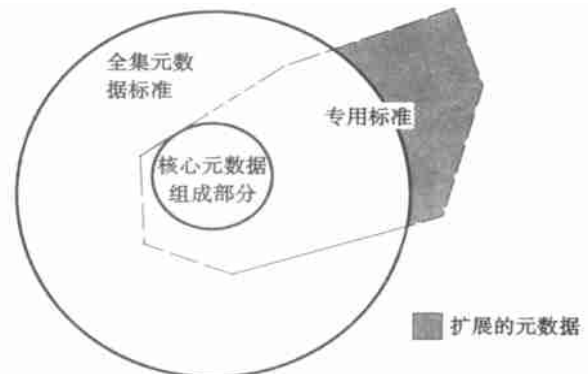


图1 地理信息元数据专用标准的构成

Fig.1 Constitution of metadata profile for geographic information

国家标准《地理信息元数据》给出了用于地理信息的专用标准,作为该标准的资料性附录。这一专用标准是经过在国家“九五”、“十五”科技攻关项目和基

基础性工作项目研究、试验的基础上提炼出来的,并开发了相应的工具软件,实现了元数据的标准化、规范化采集、编辑、处理和发布,具有相当好的实用性。

6 元数据实现

元数据实现应提供元数据元素结构和内容编码方法综述,用于元数据查询和检索、元数据交换和表示。以便在统一对地理数据元数据理解的同时,允许元数据局部管理的灵活性,并为在局域网或广域网上开展地理数据服务(交换网络)提供指导。

为在数据管理系统间交换元数据,实现方法应提供用多种形式和语言表示的元数据元素标记,并提供评价元数据一致性的可行方法。在元数据标准的实现过程中,应充分考虑到标准定义各个实体和要素间相互嵌套与引用关系,而且还包括大量对其它标准中定义的相关实体和数据类型的引用。目前,一个较灵活的方法是采用 XML 技术来实现。包括利用 XML 模式对标准中定义的实体和要素进行表达,并用于在元数据采集过程中对输入信息进行有效性校验,同时利用 XML XSL T 实现针对不同用户视图的转换,还可以利用 XML 数据库技术对元数据记录进行管理。当然,也可以利用关系数据库来管理和采集元数据,但相比较而言,缺乏灵活性。元数据实现的过程中,还需要考虑元数据与其描述对象—数据之间的关联性,尽可能提供通过元数据访问数据的途径。另外,由于元数据描述对象的不同层次之间具有隶属和继承关系,其元数据也存在一定的继承关系。因此,在元数据实现的过程中,尽量利用这种关系可以简化对元数据的操作。目前,许多地理信息系统软件逐渐把元数据和地理信息的管理纳入到一体化的体系中,用户可以直接通过相关地理信息系统软件在操作数据的同时采集其元数据,并通过元数据对数据进行管理。

7 元数据分层结构的应用

由于元数据主要应用于数据集,也可应用于数据集系列,或要素,或要素属性。初看起来,似乎有许多层次的元数据需要维护。但在大多数情况下,并非如此,因为只有例外元数据元素才在较低层次定义。如果元数据值不变,则该元数据聚集在较高层次中。这是最一般的情况,仅在经过一段时间原始数据的维护以后,才会增加元数据的层次。

当增加较低层次的元数据层级时,只记录更新的元数据值。因此,如果数据的发布者不变,就不需要记录下层的元数据。

例如,地理数据提供者生产 3 个行政区(A、B 和 C)的矢量地图数据。起初用传统纸质地图系列生产矢量地图,用同样方法将其转换为矢量格式。这种初始数据的元数据可以是单一的层次(数据集系列);经过一段时间以后,行政区 A 的数据进行了更新,因而它的元数据应进行扩充,说明新质量日期的值,用于代替原来数据集系列给出的值,但仅限于行政区 A,行政区 B 和 C 应保持不变。这时,行政区 A 的新元数据应在数据集层次上记录;再经过一定时间,行政区 A 的道路网数据进行了更新。这意味着需要受影响的道路网要素类型的新元数据。该元数据应在行政区 A 的要素类型层次上记录。所有与其它要素类型相关的其它元数据不受影响。仅仅行政区 A 道路的元数据改变,记录在要素类型层次上,依此类推。

8 标准研制策略的几点考虑

8.1 力求与国际标准接轨

我国标准化主管部门强调采用国际标准的重要性,指出采用国际标准和国外先进标准是国民经济发展的重要技术基础工作,推进采用国际标准和国外先进标准的工作在经济发展中已越来越重要。做好推进采用国际标准和国外先进标准的工作,对提高我国产品质量的整体水平和在国际市场上的竞争能力,扩大对外贸易和保持我国经济持续、健康、快速发展极为重要。我国加入 WTO 为采标工作提供了难得的发展机遇,以国际标准为基础制定本国家标准,已成为 WTO 对各成员的要求。采用国际标准工作是一项涉及到全社会事关经济发展的全局性系统工程,是一项重要的基础工作,是遵守 WTO 有关协议、积极履行我国入世承诺的必然要求,是有效消除贸易中的技术壁垒,促进对外贸易的重要措施。我国也需按 ISO/IEC 指南 21 的原则选择适合我国的采用国际标准的规则,以确保其结果得到国际上的承认,促进我国的对外贸易和交流的发展。《地理信息 元数据》国家标准的制订遵循了这一原则,力求与国际标准接轨。

8.2 立足通用性、基础性标准

考虑到地理信息内容的复杂性、表现形式的多样性、应用领域的广泛性和我国国情的特殊性,该标准应为通用性、基础性标准,作为各专业、各领域、各部门制定元数据专用标准的共同基础。元数据内容以共性为前提,尽可能详尽,能够包容描述各种类型地理信息数据所需的元数据实体和元素;同时,又充分重视我国的国情,以提高标准的实用性。为此,既

重视了全集元数据内容的标准化,保持与国际标准的一致性,也重视地理信息元数据专用标准内容的研究,使之成为适合我国国情、便于实施和转化为不同应用范畴元数据专用标准的资料性文件。

8.3 地理信息元数据的分级

地理信息元数据的全部内容可以用一个文件描述,即不分级。也可以将其分为详尽的和摘要的两个繁简不同的级别描述,即全集元数据和核心元数据(或称一级元数据和二级元数据)。针对 ISO 19115 的元数据内容十分复杂,在我国当前的条件下,元数据采集的工作量大,许多元数据元素的值很难获取,用户需要了解地理信息元数据的详细程度也有很大差异,为便于标准的实施,提高各种元数据专用标准的研制进程和一致性水平,该标准采用分级的方法,不仅给出全集元数据内容,且依据 ISO 19115 规定的核心元数据基本内容和扩展原则,制定了地理信息专用标准的元数据字典,作为资料性附录,为各种专用标准的研制提供方便。

8.4 “等同”还是“修改”采用相应国际标准

按照 GB/T 20000.2-2001《标准化工作指南 第 2 部分:采用国际标准的规则》的一致性程度划分类别,只有“等同”和“修改”两种类别符合一致性采标要求。“等同”采用时,国家标准和国际标准在技术内容、文本结构和措词方面必须完全相同,或者国家标准和国际标准在技术内容上相同,仅进行很少的编辑性修改。“修改”采用时,允许国家标准与国际标准存在技术性差异,但这些技术性差异应清楚地予以说明。可以进行的修改包括:国家标准的内容少于或多于相应的国际标准、更改国际标准的一部分内容、增加另一种供选择的方案等。

从对 ISO 19115 的分析,它的一些内容并不适合我国特点,例如 ISO 19115 中有关数据质量的元数据实体和元素达 39 项,非常详尽,但我国当前对数据质量的评价一般采用综述的方式,很难获取 30 多项质量元数据元素的值,需要对某些内容进行修改和扩充。起草组认为,地理信息元数据国家标准的制定不能“等同”采用,而应“修改”采用 ISO 19115。

8.5 UML 图,元数据实体、元素名称,以及 UML 包、类、属性名称的处理

ISO 19115 采用统一建模语言(UML)的静态结构图、与一组基本数据类型定义及 UML 对象约束语言(OCL)相结合,作为标准的概念模式语言。在将该标准转化为汉字描述的国家标准时,如何处理 ISO 19115 中的大量 UML 图,元数据实体、元素名

称,以及 UML 的包、类、属性名称是一个困难问题。如果完全不予翻译,在使用相应 UML 软件或开发支持该标准的软件时十分方便,但标准文本将难于阅读和理解。现方案完整地翻译了 UML 图,在元数据字典中同时列出了元数据实体、元素的中文和英文名称,并在正文中 UML 的包、类、属性名称第一次出现时用括号注明其英文名称,既保证了标准文本的可读性,又顾及了软件开发的需要。

8.6 标准的编号问题

该标准由 ISO 19100 结构化系列标准中的一项目标准转化而成,原标准属于一整套系列性标准,该标准应当与之相对应,也应属于一整套系列性国家标准。然而,迄今 ISO 19100 系列标准中仅有个别标准已转化为国家标准,大部分尚未转化,为保持该系列标准相互间的关系和结构的完整性,建议国家在审批发布该标准时,给出一个类似于 ISO 19100 的国家标准系列号。

元数据标准的研制与实施已引起各国广泛重视,许多国家不但已经完成标准的制定,开发了多种操作工具软件,而且已投入使用,并多次召开国际学术讨论会,对元数据的理论、实施、应用及标准化的有关问题进行讨论,ISO 19115 已经被多种国外商业 GIS 软件引用,随着这些商业 GIS 软件的销售和使用,ISO 19115 将得到更多用户的了解和实施。

9 结束语

近年来,我国许多单位、科研项目相继研制了不尽相同的元数据标准,缺乏一致性。鉴于这一情况,随着该标准的研制完成、批准发布,建议进行有力的宣贯,举办多种形式的培训和技术讨论,鼓励各部门、各单位与地理信息相关的信息系统、数据库及国内 GIS 软件开发商积极采用国家标准,支持开发相应的通用工具软件,支持基于国家标准制订各个领域的元数据专用标准,使该国家标准在地理信息共享和信息化进程中充分发挥作用。

参考文献:

- [1] GB/T 20000.2-2001, 标准化工作指南 第 2 部分:采用国际标准的规则[S].
- [2] ISO 19115:2003, Geographic information³/metadata[S].
- [3] GB/T XXXXX-200X, 地理信息 元数据(征求意见稿)[S].

作者简介:

蒋景瞳(1940-),男,江苏人,研究员,中国 GIS 协会标准化与质量控制专业委员会副主任,全国地理信息标准化技术委员会顾问,1981~1983 年在西德法兰克福应用测量研究所和汉诺威大学进修,现工作于国家基础地理信息中心,主要从事地理信息系统应用、信息共享和地理信息标准化研究。