

# 数字地球中的元数据管理模型研究

陈爱军 黄晓斌

(北京大学数字地球工作室, 北京 100871)

E-mail: Chaj@263.net

**摘要** 元数据是营造数字地球的关键技术之一,是数字地球中实现空间信息共享的主要解决途径之一,它与空间信息的共享是密不可分的。从数字地球的角度出发,对空间元数据的定义、在数字地球中的重要作用、内容、管理模型及其解决方案和应用等方面进行了较为详细的研究,并给出了基于一种解决方案的具体实现和示范系统,为数字地球中的信息共享提供了可行有效的解决途径。

**关键词** 数字地球 空间信息共享 空间元数据 空间数据交换中心

## 0 引言

在数字地球的建设过程中,海量数据的收集、组织、管理和访问的复杂性正成为数据生产者 and 用户们的最突出问题,数据生产者需要可遵循的、有效的数据管理和维护办法,用户越来越紧迫地要求提供现势性强、精度高、易管理和易访问的地理空间数据,需要更快、更加全面和有效的方法,以便从数字地球的海量数据中快速、准确地发现、访问、获取和使用所需的数据。本文从地理空间信息共享的需求出发,对数字地球的关键数据之一——地理空间元数据的定义、作用、理论体系、管理模型及其解决方案和应用前景等方面进行了研究,提出了适合我国国情的元数据标准规范草案,设计并实现了基于网络体系的元数据管理系统。

## 1 元数据在数字地球中的重要作用

戈尔在加利福尼亚科学中心的演说中指出了实现数字地球的6大关键技术,其中之一就是“元数据”<sup>[1]</sup>。那么什么是元数据呢?元数据是“关于数据的数据(The data about data)”,是有效地处理信息资源的关键因素。在地理空间数据中,元数据是说明地理空间数据的内容、质量、状况和其他有关特征的背景信息。它可以用来管理和辅助地理空间数据,满足社会对地理信息的各类需求<sup>[2]</sup>。

为什么元数据在数字地球的实现中占有如此重要的地位呢?具体说来主要体现在以下几方面:

(1) 元数据是实现数字地球中地理空间信息共享的前提条件和基本保障。可以说,没有元数据就没有地理空间信息的网络共享和全球共用。只有实现了地理空间信息的收集、存储、管理、维护、发布、销售等一系列的元数据化,才能真正实现全球信息共享,才能充分发挥数字地球的全球信息化作用<sup>[3]</sup>。

(2) 它是整理数字地球所需地理空间数据的重要指导原则。我们建立数字地球是因为迫切需要利用有关地球的各种大量信息,而大量这样的信息又散落在各处未被充分利用,那么,如何把这些散落各处的数据收集整理出来,变成数字地球可用的数据呢?这就要依靠元数据。通过元数据指定的数据标准、规范和格式,数据采集者、生产者、收集者可以整理不同种类、不同来源的各类数据、信息为数字地球所用<sup>[3]</sup>。

(3) 它是组织和管理数字地球中数据的重要手段,元数据最本质的特性之一就是它具有目录索引的作用,类似于一本书中目录的功效。通过元数据的这种目录作用,数据管理者可以利用最核心的最少的信息来有效地、清晰地管理海量的数字地球数据<sup>[3]</sup>。

(4) 它是维护数字地球中的数据的重要工具。数据生产单位通过建立所生产数据集的元数据文档,可以保证即使其主要工作人员退休或调离时,也不会失去对所生产数据集情况的了解,这样可以实

现对数据集的维护、更新,保证数据生产者对数据的投资。

(5) 它是数字地球中数据交流中心的核心内容。利用数据交流中心的元数据,数据使用者可以获得已有数据的相关信息,数据生产者可以相互交流各自已有的数据,经过再加工提供数字地球中的数据增值服务。

(6) 它是数据使用者从数字地球的海量数据中获取满足其特殊需求的数据或信息的预览器。

(7) 元数据还提供数据生产单位的数据存储、数据分类以及数据销售等方面的信息,这样便于数据使用者查询检索地理空间数据。同时,它还提供有关数据使用者处理和转换接受的外部数据的信息。

## 2 数字地球中元数据的分类

按照数字地球中元数据管理的不同类别数据来划分元数据,可以分出很多类别,如:中国生物多样性数据库元数据、中国地理空间数据库元数据等等,如图 1。这些元数据的组织方式、所包含的内容等彼此都很类似,而且与 ISO/TC211、FGDC、CEN 等标准也很接近。

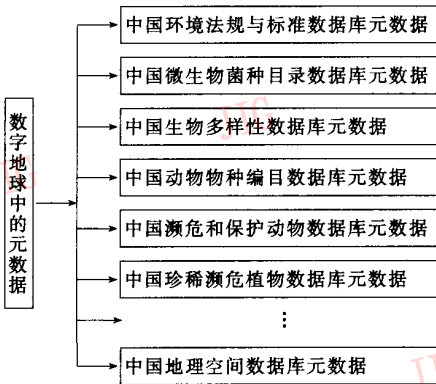


图 1 不同类别数据的元数据分类结构图

按照元数据所描述的同类别数据的不同内容,元数据可分为数据集系列元数据、数据集元数据、要素类型和要素实例元数据、属性类型和属性实例元数据<sup>[4]</sup>,如图 2。

### (1) 数据集系列元数据

数据集系列元数据是指一系列拥有共同主题、日期、分辨率以及方法等特征的空间数据系列或集合,它是用户用于概括性查询数据集的主要内容。

在软件实现上,如果拥有数据集系列元数据模块,则既可以使数据集生产者方便地描述宏观数据

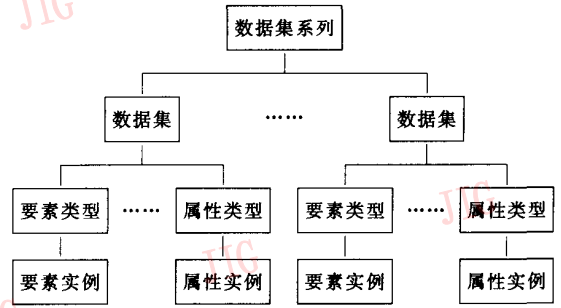


图 2 同一类别数据的元数据分类结构图

集,也可以使用户很容易地查询到数据集的相关内容,实现空间信息资源的共享。但是,对数据集详细信息的获取,还需通过数据集元数据来实现。

### (2) 数据集元数据

数据集元数据是整个元数据标准体系的核心,它既可以作为数据集系列元数据的组成部分,也可以作为后面数据集属性以及要素等内容的父元数据数据集系列。在元数据软件标准设计的初级阶段,通过该模块便可以全面反映数据集的内容。

### (3) 要素类型和要素实例元数据

要素类型在数据集内容中相对比较容易理解,它指由一系列几何对象组成的具有相似特征的集合,比如数据集中的道路层、植被层等便是具体的要素类型。

要素实例或具体要素是具体的要素实体,它用于描述数据集中的典型要素,通过它可以直接获取有关具体地理对象的信息。该模块是元数据体系中详细描述现实世界的组成部分,也是未来数字地球中走向多级分辨率查询的依据,比如武汉长江大桥是一个具体要素。

### (4) 属性类型和属性实例元数据

属性类型是用于描述空间要素某一相似特征的特征参数,比如桥梁的跨度便是一个属性类型;属性实例则是要素实例的属性,比如某一桥梁穿越某一道路的跨度。属性类型和属性实例是与要素类型和要素实例对应的模块,它们是地理数据集软件层次结构或继承关系的组成部分,是元数据标准体系的高级阶段内容。

## 3 数字地球中地理空间元数据的内容

描述数字地球中空间信息的空间元数据内容是按照部分、复合元素和数据元素来组织的<sup>[5-7]</sup>,如图 3 所示。

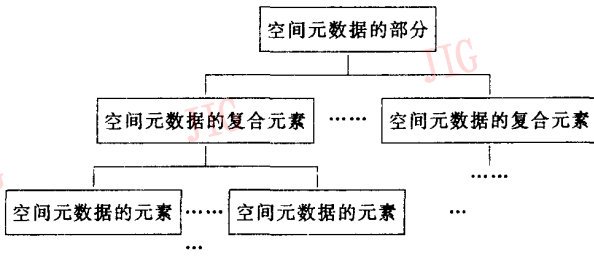


图3 空间元数据内容组织示意图

地理空间元数据标准体系的内容具体分为8个基本内容部分和4个引用部分,共12个部分组成,其中标准化内容包括标识信息、数据质量信息、数据集继承信息、空间数据表示信息、空间参照系信息、实体和属性信息、发行信息以及空间元数据参考信

息等8个方面的内容,另外还有4个部分是标准化部分中必须引用的信息,即引用信息、时间范围信息、联系信息和地址信息。它们之间的相互关系如图4所示。

目前,数字地球中的地理空间元数据标准由两层组成,其中第一层是目录层,它所提供的空间元数据复合元素和数据元素是数字地球中查询空间信息的目录信息,它相对概括了第二层中的一些必选项信息,是空间元数据体系内容中比较宏观的信息;第二层是空间元数据标准的主体,由8个标准部分(1—8)和4个引用部分(9—12)组成,包括了全面描述地理空间信息的必选项、可选项和条件可选项元素。

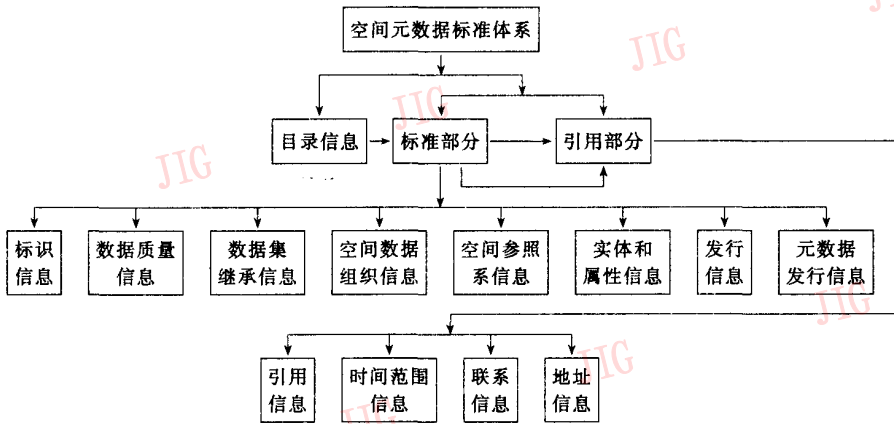


图4 空间元数据内容标准的组织框架

(0) 空间元数据: 关于数据集内容、质量、表示方式、空间参考、管理方式以及数据集的其它特征的数据,它位于整个标准体系的最上层,属于复合元素(compound),由两个层次组成。

在构成空间元数据标准内容的两个层次中,第一层目录信息主要用于对数据集信息进行宏观描述,它适合在数字地球的国家级空间信息交换中心或区域以及全球范围内管理和查询空间信息时使用。第二层则作为详细或全面描述地理空间信息的空间元数据标准内容,是数据集生产者在提供数据集时必须提供的信息。

(1) 标识信息: 是关于地理空间数据集的基本信息。通过标识信息,数据集生产者可以对有关数据集的基本信息进行详细的描述,诸如描述数据集的名称、作者信息、所采用的语言、数据集环境、专题分类、访问限制等,同时用户也可以根据这些内容对数据集有一个总体的了解。

(2) 数据质量信息: 是对空间数据集质量进行

总体评价的信息。通过这部分内容用户可以获得有关数据集的几何精度和属性精度等方面的信息,也可以知道数据集在逻辑上是否一致以及它的完备性,这是用户对数据集进行判断以及决定数据集是否满足他们需求的主要判断依据。

(3) 数据集继承信息: 是建立该数据集时所涉及的有关事件、参数、数据源等的信息,以及负责这些数据集的组织机构信息。通过这部分信息便可以对建立数据集的中间过程有一个详细的描述,比如当一副数字专题地图的建立经过了航片判读、清绘、扫描、数字地图编辑以及验收等过程时,应对每一过程有一个简要描述,使用户对数据集的建立过程比较清晰,也使数据集每一过程的责任比较清楚。

(4) 空间数据表示信息: 是数据集中表示空间信息的方式。它由空间表示类型、矢量空间表示信息、栅格空间表示信息、影像空间表示信息以及传感器波段信息等内容组成,它是决定数据转换以及数据能否在用户计算机平台上运行的必须信息。利用

空间数据表示信息,用户便可以在获取该数据集后对其进行各种处理或分析。

(5) 空间参照系信息:是有关数据集中坐标的参考框架以及编码方式的描述,它是反映现实世界与数字化地理世界之间关系的通道,诸如地理标识码参照系统、水平坐标系统、垂直坐标系统以及大地模型等。通过空间参照系中的各元素,可以知道地理实体转换成数字对象的过程以及各相关的计算参数,使数字信息成为可以度量和决策的依据。当然,它的逆过程也是成立的,即可以由数字信息反映出现实世界的特征。

(6) 实体和属性信息:是关于数据集信息内容的信息,包括实体类型、他们的属性及其属性值、域值等方面的信息。通过该部分内容,数据集生产者可以详细描述数据集中各实体的名称、标识码以及含义等内容,也可以使用户知道各地理要素属性码的名称、含义以及权威来源等。

在实体和属性信息中,数据集生产者可以根据自己数据的特点在详细描述和概括描述之间选择其一,以描述数据集的属性等特征。

(7) 发行信息:是关于数据集发行及其获取方法的信息,包括发行部门、数据资源描述、发行部门责任、订购程序、用户订购过程以及使用数据集的技术要求等内容。通过发行信息,用户可以了解到数据集在何处,怎样获取、获取介质以及获取费用等信息。

(8) 空间元数据参考信息:是有关空间元数据当前现状及其负责部门的信息,包括空间元数据日期信息、联系地址、标准信息、限制条件、安全信息、以及空间元数据扩展信息等内容,它是当前数据集进行空间元数据描述的依据。通过该空间元数据描述,用户便可以了解到所使用的描述方法的实时性等信息,加深了对数据集内容的理解。

(9) 引用信息:是引用或参考该数据集时所需的简要信息,它自己从不单独使用,而是被标准内容部分的有关元素引用。它主要由标题、作者信息、参考时间、版本等信息组成。

(10) 时间范围信息:是关于有关事件的日期和时间的信息,该部分是标准内容部分的有关元素引用时要用到的信息,它自己不单独使用。

(11) 联系信息:是同与数据集有关的个人和组织联系时所需的信息,包括联系人的姓名、性别、所属单位等信息。该部分是标准内容部分的有关元素引用时要用到的信息,它自己不单独使用。

(12) 地址信息:是同组织或个人通讯的地址

信息,包括邮政地址、电子邮件地址、电话等信息。这部分是描述有关地址元素的引用信息,它自己不单独使用。

## 4 数字地球中元数据管理模型解决方案

考虑数字地球中元数据的作用、功能和可提供给用户的方式,结合当前数字地球的基石——Internet 技术的最新发展,我们提出了如下几种解决方案,并按照其中之一实现了数字地球中元数据的管理和使用。

### 4.1 基于通用网关接口

基于通用网关接口的管理模型实现逻辑结构图如图 5 所示。

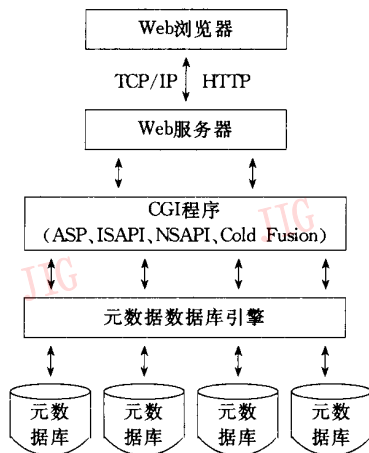


图5 基于CGI解决方案的实现逻辑结构图

可采用的实现技术包括:

- (1) 利用 C 或 Perl 编程语言;
- (2) 利用活动服务器页面(Active Server Page, ASP);
- (3) 利用 ISAPI 或 NSAPI 接口;
- (4) 利用 Cold Fusion 开发环境。

### 4.2 基于 Z39.50 协议<sup>[8]</sup>

Z39.50 是由美国国家信息标准化组织(NISO)实现的用于在线数据库摘要信息检索的美国国家标准,是符合 ISO 标准的网络通信协议。它功能强大但内容复杂,采用客户服务器计算模式面向会话实现 Internet 远程查询和检索。Z39.50 协议定义了通过 Internet 互连的不同系统间实现数据库检索和记录获取过程中必须遵循的规则。Z39.50 客户端负责

与 Z39.50 服务器端的搜索建立,发送查询,提出数据返回要求和结束搜索会话等,而 Z39.50 服务器端响应客户端请求,对客户端系统内的数据库进行操作并返回结果。基于 Z39.50 解决方案的实现逻辑结构如图 6 所示。

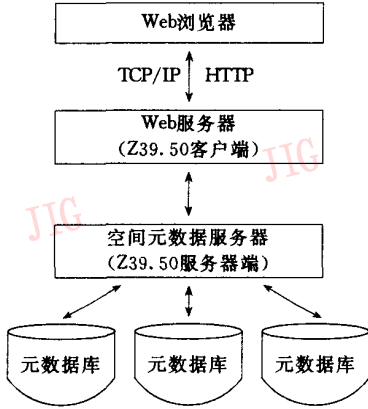


图6 基于Z39.50解决方案的实现逻辑结构图

## 5 数字地球中元数据管理模型实例与使用

我们实现的数字地球元数据网络管理系统采用第一种解决方案,后台使用 Microsoft NT 自带的 Web 服务器 IIS 4.0,前台采用 IE 4.0,编程语言利

用 VC++。该元数据网络管理系统由权限验证功能(服务器端验证)、输入合法性校验功能(客户端校验)、查询功能(服务器端查询)与返回和显示功能(服务器端格式化查询结果并返回,客户端显示)等组成。整个空间元数据网络管理系统逻辑图及其各功能模块间的关系如图 7 所示。此系统可以作为数字地球中的子系统,许多个子系统的组合,便可以真正实现数字地球中的元数据管理体系。下面简单介绍各功能模块的功能。

(1) 权限功能:提供用户登录界面,当用户输入用户名和保密字后,该信息会被传到服务器端,保存在服务器端的合法用户名和保密字进行匹配,只有二者完全匹配,才是合法的用户,才可以查询空间元数据内容。

(2) 输入功能:输入功能由输入结构、输入空间元数据、以及输入空间数据信息 3 部分组成。

(3) 查询功能:查询功能由初始化查询条件和返回结果两部分组成。当用户在客户端提交查询条件和算法时,服务器将对规则进行解译,用来操作空间元数据数据库,并将查询结果信息返回给客户端,形成返回结果清单,由此清单进一步获取用户所需的空间信息。

(4) 返回功能:返回功能由请求空间元数据、返回空间元数据、返回结构、请求空间数据、返回空间数据等内容组成。

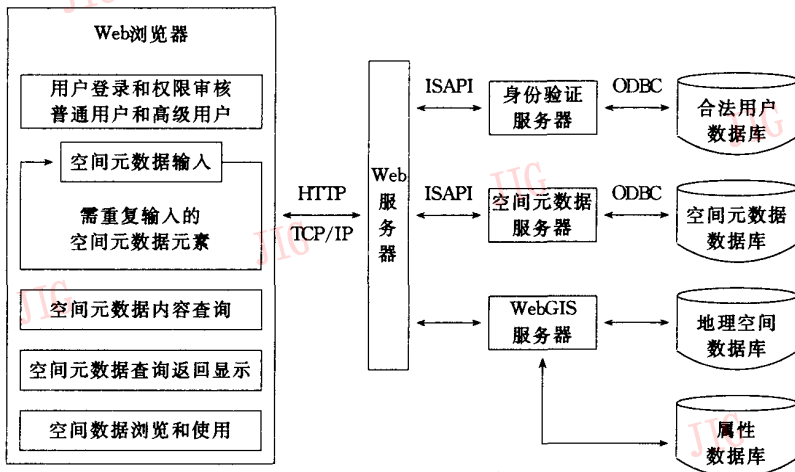


图7 空间元数据网络管理系统逻辑图

## 6 结束语

随着信息高速公路覆盖全球,知识经济蓬勃发

展,网络进入千家万户,数字地球正在迈向日新月异的未来,因此,元数据标准体系在数字地球中的重要作用日益突出。标准引导、规范制约、应用为王,服务导向,低成本扩张,将成为我们走向数字小区、数字城

市、数字区域、数字中国直到数字地球的必由之路。

### 参 考 文 献

- 1 Al Gore. Digital Earth: Understanding our planet in the 21st century, 1998.
- 2 Metadata, Clearinghouse, Standards, NSDI, <http://www.fgdc.gov/>, 1998.
- 3 Harlan, Onsrud J(eds). Metadata Issues in Sharing Geographic Data. Sharing Geographic Information the State University of New Jersey, 1995.P. 499.
- 4 <http://www.opengis.org/techno/specs.htm>
- 5 <http://www.statkart.no/isotc211/>
- 6 ISO/TC211 Metadata Working Document; Geographic Information, 1997.
- 7 赵永平,承继成. 地理信息数据描述中元数据标准化的研究. 中国标准化,1998,(3).
- 8 <http://lcweb.loc.org/z3950/>

## Research on Management Model of Metadata in Digital Earth

Chen Aijun, Huang Xiaobin

(The CyberGIS Studio, Peking University, Beijing 100871)

**Abstract** Metadata is one of key technology of constructing Digital Earth, and a main solution of implementing spatial information sharing in Digital Earth. In this paper, the definition, important role in Digital Earth, content, management model and its solutions, and application of Metadata have been researched in detail. And based on one of solutions, demonstration system of management model of Metadata have been achieved, and introduced here. It provides a feasible and practical way to spatial information sharing in Digital Earth.

**Keywords** Digital Earth, Spatial information sharing, Spatial Metadata, Spatial data clearinghouse