

基于 Internet 的元数据管理模式研究^{*}

——以地理信息元数据管理为例

吴洪桥¹ 何建邦¹ 池天河¹ 黄裕霞²

(中国科学院地理科学与资源研究所 北京100101)¹ (美国中密歇根州立大学地理系 中密歇根州48859)²

Study on Metadata Management Pattern Based on Internet

——With Geographic Spatial Metadata Management As a Case

WU Hong-Qiao¹ HE Jian-Bang¹ CHI Tian-He¹ HUANG Yu-Xia²

(Institute of Geographical Sciences and Natural Resources Research, CAS, Beijing 100101)¹

(University of Central Michigan, Central Michigan 48859)²

Abstract Metadata mechanism, especially metadata management pattern, is a hotspot in the field of internet information sharing. This paper analyzes the characters and deficiency of current metadata management patterns. Subsequently, a new metadata management pattern is put forward combined with its application in geographic information sharing. Aiming at the demands of future Internet information sharing, this pattern is a network-oriented, open and extensible metadata management system. Meanwhile, this management pattern is also a management model of trinity of data users, metadata and datasets. Finally, this paper analyzes the architecture and function flow of this network-oriented integrated metadata management pattern.

Keywords Internet information sharing, Geographic information, Metadata management pattern, Network-oriented integrated management

1 引言

通讯和计算机技术的发展使信息成为国民经济和社会发展的资源,目前社会各个部门都在挖掘自己的信息资源整理和改善行业内部的数据库,在管理和应用这些海量信息的同时,一个突出的问题就是怎样有效地管理和组织海量的数据。元数据机制是解决上述问题的有效手段,随着各行业元数据标准的推出,行业内部海量数据的组织有了一定的规则。但是在基于这些规则的基础上怎样有效地实现海量数据的管理,这就为元数据的管理提出了更高的要求。

对元数据管理模式的研究逐渐成为信息共享的一个重要内容,目前一些部门已经针对数据在行业内部的应用特点来制定自己的元数据管理模式和开发管理工具。地理信息是元数据应用的重点和前沿领域,这是由地理信息的特点来决定的:首先,地理信息除了描述实体的属性信息以外,更重要的就是描述了实体空间位置的信息;地理信息是随时间的变化而变化,具有时态性;地理信息是海量的。正是由于地理信息的以上特点,增加了对其组织和管理的难度,因而利用元数据机制,采用合理的元数据管理模式是实现海量地理空间信息的组织、管理和信息共享的关键。本文就是从当前基于 Internet 信息共享的角度,来对地理信息元数据的管理模式进行研究,分析海量地理空间信息在实现 Internet 网络共享的过程中,元数据管理模式上的发展及合理的模式框架。

2 当前基于 Internet 的元数据管理模式的特点与不足

2.1 元数据独立管理模式

该模式只是一种元数据库的管理,对元数据的管理独立于数据集的管理,用户和数据生产者、元数据库、数据库可以

建立起连通关系,用户从元数据库中得不到确切的数据信息,只能通过数据生产者。这种管理模式反映在用户、数据生产者、元数据库之间的关系上如图1所示。

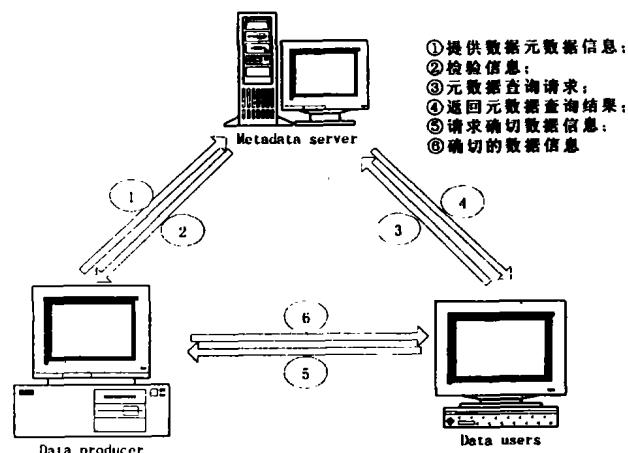


图1 元数据独立管理模型^[6]

元数据独立管理模式只能解决元数据的管理。但是由于元数据管理和数据集管理属于两种不同的管理体系,当数据集的内容发生变化时,元数据的内容缺乏相应的更新机制,影响了元数据内容的准确性;另一方面,用户对数据集的直接获取也比较困难,往往使用户获取到与元数据内容不相符合的数据集。

2.2 元数据和数据集集中管理模型

针对上述问题,目前提出了在元数据管理和数据集管理的层面上提供统一的综合管理^[5],以保证元数据管理和数据集管理的一致性和同步性。这种元数据和数据集集中管理的体系结构如图2所示。

^{*}资助项目:中科院知识创新工程项目(KZCX2-308-2)。吴洪桥 博士,主要从事网络 GIS 和图像并行处理研究。

该模型可以划分为3个层面,综合管理层、服务层和应用层。

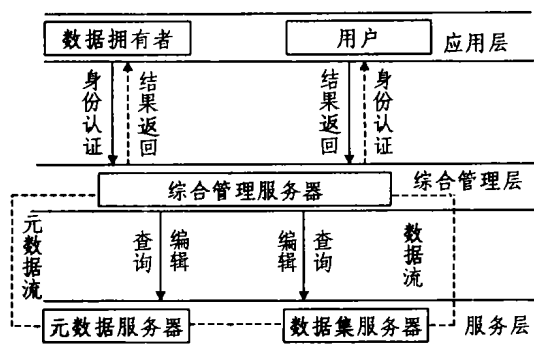


图2 元数据和数据集集中管理模型体系结构图

应用层包括数据提供者和用户,两者都可以对元数据和数据集进行查询和浏览,但在这种元数据和数据集集中管理模式中,所有查询的请求都通过一个综合管理服务器来完成。

服务层包括两种类型的服务器,分别是元数据服务器和数据集服务器,由于数据集的内容具有多样性和数据量大的特点,数据集是分布式存储的。

综合管理层是负责数据提供者和用户的安全和身份管理,更重要的就是建立用户对数据(包括元数据和数据集)操作的监听机制,当用户对元数据或数据集的某一种数据进行编辑时提示用户对相应的元数据或数据集进行对应的改动。综合管理层在逻辑上是独立于元数据服务器和数据集服务器;在物理上,综合管理服务器可以是单独的服务器,也可以与元数据服务器或数据集位于同一台服务器上。

这种管理模式的主要优点就是提供元数据和数据集的综合管理,这样就保证了元数据和数据集在内容上能够做到同步更新,但缺点是这种模式是固定的,它只为有权限的用户(一般是数据提供者)提供对数据的管理,普通用户只能浏览和查询数据。

3 基于 Internet 信息共享对地理信息元数据管理的需求

随着 Internet 的普及,基于 Internet 的数据发布成为信息共享的主要形式,将海量地理空间数据进行 Internet 共享应考虑到如下的需求与特点:

- (1) 对海量地理空间数据进行有效的组织,使用户能够了解原始数据的详尽信息;
- (2) Internet 是一个开放的信息体系,信息共享首先要考虑到信息的安全,只有在一定的安全机制下,才能保证信息的共享;
- (3) 信息共享要考虑到用户的要求,建立与用户的交互机制能够促进信息最大限度地共享和利用;
- (4) 现有的信息传输是受一定条件约束的,比如网络带宽、传输协议等;

从当前国内来看,地理信息元数据的管理往往是为了提供一个有关原始数据的准确无误的描述,元数据的管理实质上是提供元数据的不断更新,国内一些部门提供的元数据管理工具大多基于这种模式。随着国际标准化组织对行业标准的制定和完善,利用元数据来加强基于 Internet 网络的数据共享逐渐成为研究的热点,美国联邦地理数据委员会(FGDC)建立的基于 Clearinghouse 的地理信息服务系统在

这方面做出了有利的探索,在该信息服务系统中,元数据对分布在不同接点上的海量的空间数据的来源、格式、联接的位置以及使用情况提供详尽的描述信息,从而实现了将分布在若干个节点上的海量的空间数据进行管理的功能。通过这样一种元数据机制,用户可以对这些空间信息发出浏览、查询和下载的请求,服务器端通过 Z39.50 协议来响应用户请求。

但是随着信息共享发展的需求,基于 Internet 网的信息共享不再有纯粹的数据提供者 and 数据用户之分,因此元数据的管理不再是固定群体(如具有权限的数据生产者或提供者)与元数据的交互,而是一种动态和开放的管理模式。具体来讲,Internet 上的任一用户可以经过一定的机制把自己掌握的数据按照一定的格式提交到指定的服务器来共享,这样,服务器端也会通过一定的机制(主要是权限管理机制)给用户开辟一定的权限,这一权限主要包括数据的提供者对自己所提供数据的管理权限和对应的元数据的管理权限。因此未来元数据的管理要在一定的安全机制下,实现数据提供者和数据用户之间的交互,实现元数据动态和开放管理,从而促进更多信息的共享和利用。

针对上述分析,我们提出了一种面向 Internet 网络的,实现元数据动态和开放管理,同时也满足元数据和数据集一体化的管理模式——网络一体化管理模式。

4 网络一体化管理模式的体系结构

未来基于 Internet 的元数据管理模式需要一种在普通用户和元数据服务器之间建立一种交互管理的机制,它不仅能使普通 Internet 用户按照一定的格式和要求提交自己的数据来共享,同时能够使用户在元数据服务器建立自己的元数据库,并通过这种交互机制使用户获得对自己元数据库的管理权限,从而实现元数据的管理动态和可扩展的目标。

网络一体化元数据管理模式就是考虑到以上对元数据管理方式变化和当前技术的发展而提出的一种先进和切实可行的元数据管理模式,构建的这种一体化的地理信息元数据管理模式的体系结构如图3所示。

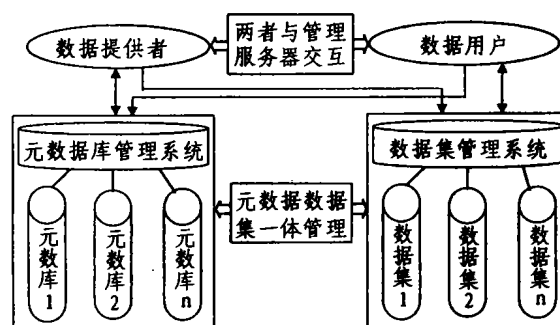


图3 网络一体化元数据管理模式的体系结构

这种模式也可以解析为三层的体系结构,带有阴影的椭圆形表示应用层;带有阴影的矩形表示管理层;虚线矩形连同其中的内容表示服务层。在这种管理模式,每一层都分为两部分。在管理层中,元数据和数据集一体化的管理层保证元数据和数据集内容上的同步和一致;用户与元数据的交互管理层来实现元数据动态和开放的管理,保证用户在提供共享数据的同时,也拥有对相应元数据的管理权限。服务层是提供信息服务的分布式元数据管理系统和分布式的数据集管理系统。应用层包括数据提供者和数据用户,他们对元数据和数据

集分别有不同的管理和访问方式。

这种元数据管理模式的特点主要体现在以下三个方面:

元数据管理是动态和开放的:这种管理模式解决元数据管理封闭和不能扩展的缺陷,通过服务器端交互的管理机制,不仅能使普通 Internet 用户按照一定的格式和要求提交自己的数据,同时还能建立自己的元数据库并提交到元数据服务器,这样元数据的管理不再是针对特定人群,普通用户通过服务器的认证以后可以获得对部分元数据和原始数据的管理权限。这样一种管理方式不仅对已有数据做到共享,而且还可以挖掘 Internet 不同接点上的地理信息,使整个地理信息的 Internet 共享是一种互动和开放的体系。

元数据分布式存储和管理:在这种模式中,首先原始数据是分布式存储和管理的。随着原始地理信息的不断增加,元数据的数据量也不断增加,当前元数据的组织完全可以采用关系数据库的表格来存储,因此完全可以采用当前比较成熟的关系数据库(如ORACLE,SQL Server等)来实现元数据的分布式存储和管理。

元数据和数据集的一体化管理:为了保证元数据管理和数据集管理的同步,在服务器端建立一种一体化管理的机制,这种一体化管理的机制可以捕捉到元数据和数据集的变动信息,当有一方变动时,能够给访问的用户一个提示信息,保证元数据和数据集的同步更新。

5 设计和应用分析

上述地理信息元数据一体化的管理模式在“国家资源环境数据库建设和数据共享研究”的项目研究中得到应用,设计和开发了元数据管理软件工具——MMT-1,该工具面向资源环境领域的各部门,基本能够满足适时提交数据和能够动态管理元数据的功能。

系统在设计中将元数据的管理功能分为三个层次来设计,即服务层、管理层和应用层。

服务层的设计 采用 Apache+Servlet+Oracle 9i 来构建元数据和数据集的 Web 服务平台。用户的请求以 XML 的形式提交到服务器端,服务器端将这些请求解析为 SQL 语句来调用响应的元数据或数据集的信息;同时,在该平台上利用 JSP 还建立一种权限认证机制,将用户的权限划分为三种,分别是:元数据浏览,元数据发布和元数据管理三种层次,不同权限的用户获得不同的服务。服务器端通过 Z39.50 协议提供服务功能。

管理层的设计 管理层包括元数据和数据集的一体化管理、用户与元数据服务器之间的交互管理。一体化管理的功能通过如下的过程来实现:利用现有关系数据库管理系统通过元数据和对应的数据集共同的标示码建立一种对应关系;在此基础上构建一个监听机制,能够捕捉到元数据和数据集的变动信息;当用户登录到元数据服务器或数据集服务器时,通知用户相应的数据变动信息,不同的用户可以根据自己的权限采取相应的措施。用户与元数据的交互管理就是用户(普通用户)通过客户端的应用程序向服务器发出请求,服务器在接受到请求之后在一定的时间内给用户一个反馈信息,同时在服务器端(在 MMT-1 中,这种交互程序位于元数据服务器)给用户开辟一定的权限,使他能够提交并管理自己的原始数据集和元数据。

应用层的设计 应用层面向的是 Internet 的普通用户和

数据提供者两类不同的对象。针对普通用户,客户端建立数据提交请求的应用程序,从而建立起与服务器端交互管理层的联系;同时,为方便用户浏览和查询元数据,客户端设置基于地理范围、内容、专题及关键词的查询。针对数据提供者这类有特殊权限的用户,客户端有严格的权限认证机制,通过权限认证以后,这类用户通过浏览器就可以完成元数据和数据集的编辑和更新。

系统完成的功能流程如图4所示。

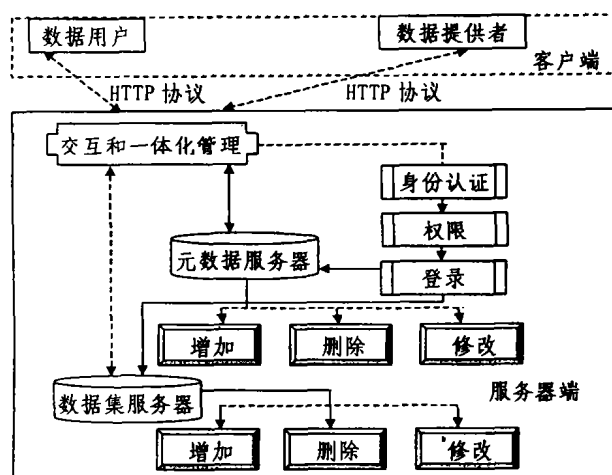


图4 MMT-1的系统结构和功能流程

该系统按照课题采用的元数据标准进行地理空间数据元数据库的建立,并将该模式初步应用在该课题元数据管理工具软件(MMT-1)的开发和应用中,说明网络一体化的元数据管理模式具有可行性,同时也为地理空间数据以外的元数据管理开辟了一个新的思路。

在当今数据量不断增大以及分布式存储的趋势下,在用户、元数据和数据集三位一体的管理模式如何解决不同数据节点上的数据冗余问题,将是一个新的研究方向。

元数据和数据集一体化的管理思想如何在实际的元数据管理中得到更好的体现,尤其是在基于网络的共享环境下采用什么样的方式来更智能、更有效地实现元数据和数据集一体化管理也将是今后元数据管理中一个重要的研究领域。

参考文献

- 1 何建邦,柯正谊,吴平生. 国家地理信息共享政策框架意见. 资源科学, 2001, 23(1):17~22
- 2 苏理宏, 黄裕震. 资源与环境信息系统元数据的组织和使用. 中国图像图形学报, 2001, 6(10):85~89
- 3 李琦, 陈爱军, 夏曙东. WebGIS 中地理空间管理系统的研究与设计. 中国图像图形学报, 2000, 10(5): 811~816
- 4 沈体雁, 程承旗. 基于空间元数据的分布式地理数据管理模型及应用研究. 测绘通报, 1999, (7): 34~36
- 5 赖志斌, 程承旗. 元数据和数据集一体化管理模型及应用研究. 测绘通报, 2001, (5): 9~10
- 6 张立, 龚健雅. 地理空间元数据管理的研究与实现. 武汉测绘科技大学学报, 2001, 25(2):127~132
- 7 乔彦友, 赵健. 分布式空间数据管理技术研究. 中国图像图形学报, 2001, 6(9):873~878
- 8 <http://www.fgdc.gov/clearinghouse/clearinghouse.html>
- 9 <http://www.esri.com>. ArcIMS Metadata Services, An ESRI White Paper, May 2002