

资源与环境元数据网络一体化 管理模式与应用研究

吴洪桥 何建邦 池天河

黄裕霞

(中国科学院地理科学与资源研究所, 北京 100101)

(美国中密歇根州立大学, 芒特普莱森特 48859)

摘 要: 资源与环境数据由于涉及的范围广、种类多、多语义、数据量大等特点, 而给资源与环境信息的共享带来了很大的困难。元数据机制是实现资源与环境数据管理和共享的重要方法和手段。本文分析了现有的资源与环境元数据的管理模式, 探讨建立一个基于网络的资源与环境数据的开放、可扩展的元数据库, 在管理上实现数据用户、元数据、数据集三位一体的管理模式, 并采用当前一些先进的开发技术实现了该模式, 设计、开发和完成了国家资源与环境数据库元数据管理工具软件 MMT-1, 得到了成功的应用。

关键词: 资源与环境元数据; 元数据管理模式; 网络一体化管理

中图分类号: P208; X82;

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重要资源, 资源与环境信息在社会中的应用和共享对充分挖掘信息资源, 推动国民经济建设和社会进步, 意义重大。

资源与环境信息是经过抽象、压缩和数字化的产品, 增加了用户对其理解的难度; 资源与环境数据涉及的范围广泛、种类多, 数据量非常庞大, 给数据的组织和发布造成了一定的困难; 对同一个资源环境数据在不同的行业和部门会有不同的理解, 造成语义上的错误和偏差。因此实现资源环境领域中的信息共享, 在加强标准研究和制定的同时, 必须在元数据实现方式和管理机制上加强研究。

当前资源与环境领域各行业和部门在加强制定本行业的元数据标准, 及资源与环境领域统一的元数据标准, 但是真正从元数据的实现方式和管理角度来加强数据的交流和信息共享的研究还不够深入。在资源与环境领域制定不同的元数据, 其实现是按部门分散完成的, 表现为多个元数据文件或多个元数据库, 由于元数据和其描述的数据集是紧密联系在一起的, 在管理上有效地实现元数据和相应

数据集的集成是实现资源环境信息共享的重要方面。目前, 无论在元数据实现方式(单个文件、多个文件或是数据库)还是从元数据与数据集的集成角度来讲, 都没有一个统一的模式, 因此应用在资源环境数据的共享上, 也存在效率的差别。

1 资源与环境元数据管理的模式与特点

元数据描述了数据的内容、质量、情况等一系列特征, 是资源与环境信息共享的关键因素。资源与环境元数据的应用有:

- (1) 组织和管理数据, 主要是建立资源环境数据文档, 并保证即使其主要工作人员变动时, 也不会影响对原始数据情况的了解;
- (2) 为用户查询和检索数据提供相关信息, 包括数据存储、分类、内容、质量以及销售等方面的信息;
- (3) 提供用户数据交换和传输的有关信息, 不同的部门和用户可以很容易地发现其所需要的数

收稿日期: 2002-04-29; 修订日期: 2002-6-12。

资助项目: 中科院知识创新工程项目——KZCX2-308-2。

作者简介: 吴洪桥 (1973-), 男, 博士, 主要从事网络 GIS 和图像并行处理研究。E-mail: wuhq@igs.nrr.ac.cn; wuhongqiao1@sina.com.

据，比如美国建立的国家地理空间数据交换中心(NGDC)便是一个很好的示例;

(4) 提供数据转换方面的信息，使用户明确数据的来源及其处理。

元数据应用在资源与环境数据的分布式存储和共享方面，其组织和管理模式可以概括为以下两种^[4, 5]:

1.1 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，但用户从元

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:

元数据独立管理模式只能解决对元数据的管理。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，往往使用户获取与元数据内容不相符合的数据集。

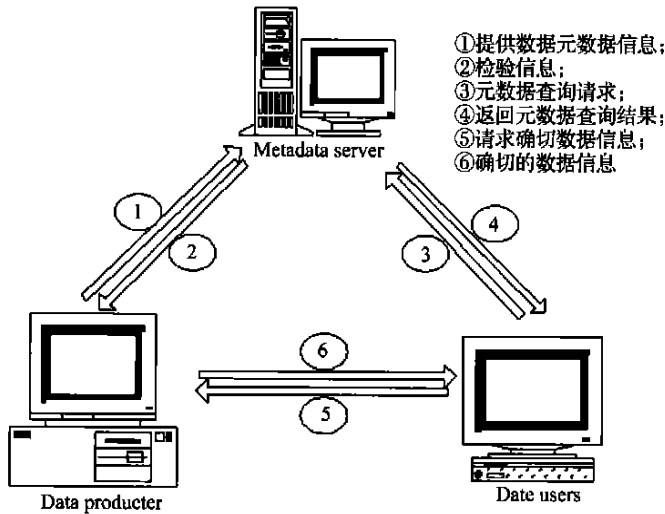


图 1 元数据独立管理模型 (据张立、龚健雅, 2001)

Fig.1 Dependent Management Model of Metadata (According to Zhang Li, Gong Jian-Ya, 2001)

1.2 元数据和数据集综合管理模型

针对上述问题，目前在元数据管理和数据集管理的层面上提出统一的综合管理^[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综合管理模型的体系结构如图 2 所示。

该模型可以划分为 3 个层面：综合管理层、服务层和应用层。

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综合管理模式中，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。

2 网络一体化元数据管理模式与体系结构

随着资源与环境信息共享的需求，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

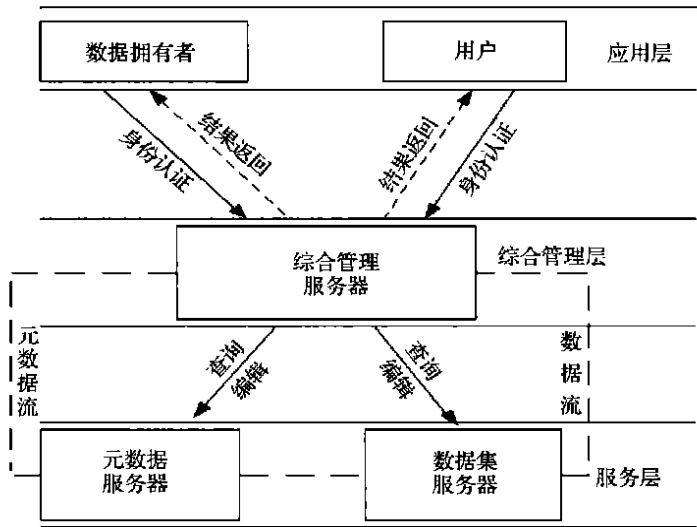


图 2 元数据和数据集综合管理模型体系结构图

Fig. 2 Framework of Integrated Management Model of Metadata and Dataset

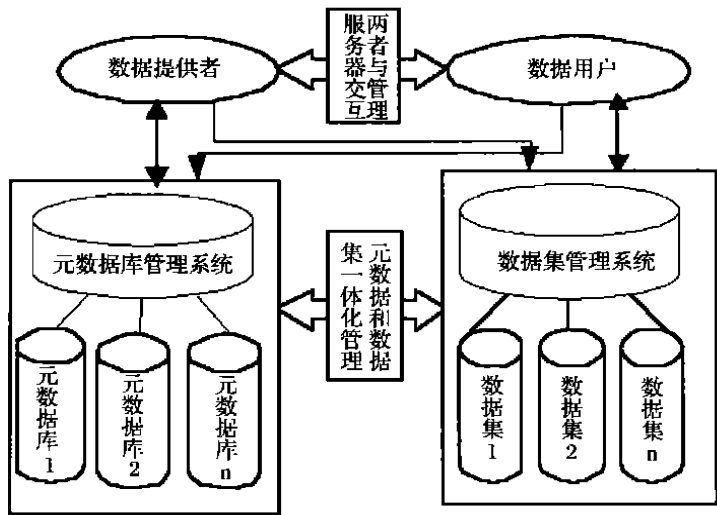


图 3 网络一体化元数据管理模式的体系结构

Fig. 3 Framework of network-oriented and integrated management model of metadata

供者) 与元数据的交互, 而是一种动态和开放的管理模式。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, 同时他也可以获得对已有部分数据的管理权限。当前的元数据模式不能很好的满足这种信息共享的需求, 针对这一问题, 我们提出了一种面向网络信息共享的需求, 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。

本文按照这样一种管理目标来组织和实现资源与环境元数据管理的应用系统, 结合当前计算机技术的发展构建了网络一体化的元数据管理模式, 其

体系结构如图 3 所示。

这种元数据管理模式有以下的特点:

(1) 动态和开放的管理: 通过服务器端与用户交互的管理层, 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, 普通用户在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集的管理权限。

(2) 网络分布式管理: 在这种模式中, 数据集管理系统管理分布在网络上不同节点的海量资源与环境数据集, 因此对数据集采用分布式管理模式; 在

对元数据的管理上，由于当前资源环境元数据的实现方式有单个文件、多个文件或者数据库等多种形式，数据库技术和分布式存储技术在海量数据的管理方面已经比较成熟，采用关系数据库（如，ORACLE，SQL Server 等）可以统一管理资源与环境的不同类型元数据，同时还可以实现元数据的分布式存储。

（3）一体化管理：在该模式中，采用了两层管理模式，元数据和数据集一体化的管理实现了元数据和数据集管理上的同步和一致性；用户与元数据交互管理解决了元数据管理动态和开放的要求，使 Internet 上的任何用户可以依据一定的步骤提供共享数据，在用户、元数据和数据集三者之间的关系上，真正实现三位一体的管理目标。

这种模式也可以解析为三层的体系结构：即应用层、管理层、服务层。在这种管理模式下，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来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。

3 网络一体化管理模式的设计和应用

上述元数据的管理模式在开展“元数据标准规范及其软件工具”的研究中得到应用，并开发了资源环境元数据管理工具软件 MMT-1。

系统采用瘦客户端的 Client/Server 体系结构，数据提供者和用户通过身份认证登录某一站点来使用该系统。进入该系统以后，根据系统赋予的权限，可以对元数据或者数据集进行浏览、使用和管理。客户端具备发出请求和显示结果等功能，相应用户的请求和处理如元数据添加、删除等均在服务器端实现。通过 HTTP 协议完成通讯功能，系统完成的功能流程如图 4 所示：

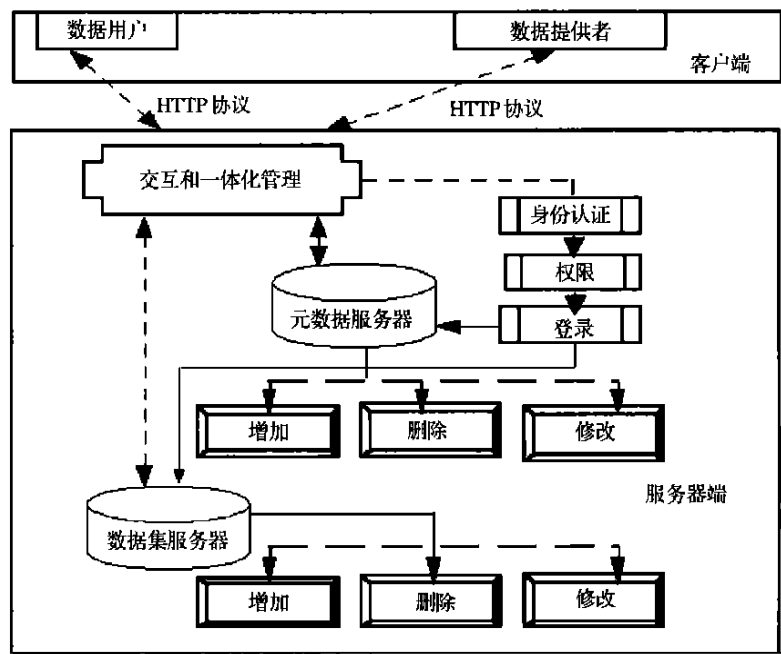


图 4 MMT-1 的系统结构和功能流程

Fig.4 Architecture and Function Flow of MMT-1

该系统在技术实现上具有以下特点：

（1）在服务器端有一种和 Internet 普通用户的交互机制，使普通用户在按照一定的步骤注册和提供共享数据以后就可以成为数据拥有者，同时获得相应的权限，可以对自己的数据和元数据进行管理

和操作。

另外，元数据和数据集的管理是一体的，这种管理机制通过以下方式来实现：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识码建立一种对应关系；在此基础上构建一个监听

机制,使它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。

同时,用 XML 来描述元数据标准,实现元数据标准规范的良好扩展性和平台之间的方便移植。XML 适合于完整表达元数据体系结构和数据项,用于灵活提取各种元数据集,使得本元数据管理工具独立于多种元数据标准。同时,实现严格的身份认证和权限管理。

(2) 在客户端,根据数据提供者和用户权限的不同,设置不同的登录界面;并为 Internet 用户提交数据,提供简单的客户端应用程序;对数据的修改通过 http 协议提交相应的服务器(元数据服务器和数据集服务器)。

上述研究表明,网络一体化的管理模式是有效可行的。在当今数据量不断增大以及分布式存储的趋势下,其在用户、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地体现,以及采用什么样的方式来更智能、更有效的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,黄裕霞. 资源与环境信息系统元数据的组织和使用. 中国图象图形学报, 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. 中国图象图形学报, 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. 中国图象图形学报, 2001, 6(9): 873~878.
- [8] <http://www.fgdc.gov/clearinghouse/clearinghouse.html>.
- [9] <http://www.esri.com/ArcDMS/Metadata/Services/> An ESRI White Paper, May 2002.

Study and Application of Network-based Integrated Management Model of Resources and Environment Metadata

WU Hongqiao¹, HE Jianbang¹, CHI Tianhe¹, HUANG Yuxia²

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

(2. University of Central Michigan, Michigan)

Abstract: There are some difficulties in sharing of resource and environmental information because of their multi-scope, multi-species, multi-significance and large data. Metadata is an important and efficient means for solving the problem. This paper has analyzed the current managing models of metadata of resources and environment. Subsequently, it discusses constructing a network-based, open and extendable metadata database management system and carrying out a management model of trinity of data users, metadata and datasets. Finally, a metadata management software MMT-1 has been completed by some advanced developing means and successfully applied.

Key words: resources and environment metadata; management model; network-based integrated management model