

文章编号: 1007-4619 (2003)06-0451-07

城市基础地理信息集成的元数据平台开发

常原飞¹, 王伟², 孙晋岳³, 王巧年³, 庄永祥⁴, 乔彦友¹, 荐 军¹

(1. 中国科学院遥感应用研究所, 北京 100101; 2. 扬州市信息化领导小组办公室, 扬州 225009;

3. 上海炎黄在线集团, 上海 201300; 4. 扬州经济开发区管委会, 扬州 225009)

摘 要: 在分析国内外不同机构和组织拟订的元数据标准基础上, 提出了一种适合中国中小城市基础地理信息集成的元数据标准, 对元数据库的结构进行了设计, 并用 XML 文档的方法开发了元数据平台, 实现了元数据的管理和发布。该平台在扬州经济开发区基础地理信息平台建设中发挥了核心作用, 在其支持下实现了海量地理信息数据的无缝集成和应用服务。

关键词: 元数据; XML; 城市基础地理信息平台; 数字城市

中图分类号: TP311 **文献标识码:** A

1 引 言

近几年来, 随着以 3S 技术为代表的空间信息科学的飞速发展和网络互联技术的广泛应用, 城市信息化发展到了一个崭新的阶段, 即要建设城市基础地理信息平台, 实现广泛的城市空间信息共享, 促进整个城市信息应用的统一协调发展^[1]。

海量地理信息数据库是基础地理信息平台的核心理心, 而元数据(Metadata)则在这些数据的组织方面发挥着重要作用。元数据是对数据的描述, 空间元数据用于描述空间数据的内容、质量、表示方法、空间定位、管理方式及其他特征, 以此组织和管理空间信息, 挖掘空间信息资源, 帮助数据使用者查询所需的空间信息, 建立空间信息的数据目录, 提供数据转换方面的信息, 元数据发布可以极大提高数据共享和交换的效率。

元数据库则是为了存储元数据而搭建的一种特殊的关系型数据库, 其中包含着大量对空间特性信息的描述, 为了保证基础地理信息平台的持续、高效运行, 就必须有专门的元数据平台来对元数据库进行集中管理和操作, 包括元数据采集、元数据库建库、元数据更新、元数据查询、元数据管理和元数据发布等。

现有的元数据平台的模式大都为本地修改、WEB 发布的单机元数据平台, 例如国家空间信息交换中心的 NREDIS 标准空间元数据编辑器, 它只能采集和维护本地的元数据。而城市基础地理信息的元数据分布在不同部门, 元数据的采集和修改也不便集中在一台机器上, 因此, 需要建立支持网络操作的元数据管理平台。另外, 现有的各种元数据平台只支持特定的元数据标准, 例如 ArcGIS 现在只直接提供了对 FGDC 标准^[2]和 ISO 标准的支持, NREDIS 标准空间元数据编辑器也只是支持 NREDIS 标准^[3], 而这些数据标准并不适合中小城市的基础地理信息平台的搭建。因此, 需要根据实际情况, 制定出适合中小城市基础地理信息平台搭建的元数据标准, 并在此基础上, 开发出具有支持网络操作和易于网上传输发布特点的元数据平台。

2 元数据及其标准

元数据作为关于数据的数据, 主要包括如下说明性信息; 数据质量信息, 如定位和属性精度、数据完整性和一致性、信息来源及生产数据使用的方法; 空间数据组织信息, 即在数据集中表示空间数据的机制, 如表示空间位置的直接方法(栅格和矢量)、间接方法(街道地址或县编码)以及在数据集中空间对

收稿日期: 2003-06-23; 修订日期: 2003-07-29

基金项目: 国家 863 项目“重大行业 3S 应用示范工程”资助, 项目编号: 2002AA130078。All rights reserved. <http://www.cnki.net>

作者简介: 常原飞(1980—), 男, 2001 年毕业于北京大学, 获理学和经济学双学士。现在中国科学院遥感应用研究所攻读硕士学位, 在国内核心期刊上发表论文一篇。研究方向: 地理信息系统。Email: oaken@263.net

象的数量;空间参考信息,说明数据集中的参考框架、编码和坐标的含义,如地图投影或网格坐标系参数的名称、水平和垂直基准以及坐标系统的分辨率;实体和属性信息,关于数据集内容的信息,包括实体类型、属性及其取值范围等;数据分布信息,关于获得数据集内容的信息,包括如何与分发者联系、可用格式、如何联机或在物理介质上获取数据集以及数据价格等;元数据参考信息,关于元数据现势性以及负责定位的信息。

为了准确地描述数据情况,就必须对上述元数据的内容进行细化,研究元数据的标准。

2.1 现有标准

国际上比较著名的元数据标准有美国联邦地理数据委员会(FGDC)的《地理空间元数据的内容标准》^[2]和国际标准化组织(ISO)的《ISO 19115 Geographic Information Metadata》。在参考这两个标准的基础上,我国九五攻关课题“国土资源环境和区域经济信息系统及国家空间信息基础设施关键技术研究”(97—759)项目提出了信息共享元数据内容的标准草案^[3]。

该草案按树状结构组织数据元素(elements)和复合元素。元素是最基本的信息单元,一个复合元素是由数据元素和其它复合元素组成的,它通常用来表示较高层次的概念。一个元素包括元素名、定义、类型、值域、域名缩写。元素类型有 integer、real、text、date、time 五种类型,分别用来表示整数、实数、ASCII 字符、日期和时间信息。该树状结构的根节点为复合元素“Metadata”,该复合元素由其它表示数据集不同方面特征的复合元素组成。

2.2 城市基础地理信息平台的元数据标准

上述元数据的内容非常复杂具体,对于所有的空间信息应用来说都具有很重要的指导作用。而对于每个具体应用领域来说,不可能也不需要全部采用上述标准中的所有内容,往往只需要其中的一个子集。当然,还必须根据情况补充一些特殊的内容。

在上述标准和草案的基础上,我们结合中小城市的空间数据特点,提出了适合中小城市的元数据标准。这套标准沿用了97—759项目为空间数据集定义的十大类描述复合元素,用于描述空间数据集的各方面的特征。这十类描述元素分别是“标识信息”、“数据质量信息”、“空间数据表示信息”、“空间参照系信息”、“实体和属性信息”、“发布信息”、“元

数据参考信息”、“引用信息”、“时间范围信息”和“联系信息”。其中“标识信息”和“元数据参考信息”为所有数据必须采集的元数据,其它信息根据实际情况决定是否选择其它元数据描述元素。“引用信息”、“时间范围信息”和“联系信息”三类描述元素并不单独存在而是被嵌套在其它七类描述元素中。

另一方面,由于我们针对的是中小城市地理信息的元数据,上述标准中的一些基本无内容的非必选字段,如标识信息中的“数据集 G 多边形”、数据质量信息“定量属性精度评价”等与城市基础信息平台搭建意义不大的字段,我们予以抛弃。根据城市基础地理信息平台的需要,我们定义了一些字段格式,如“引用信息”中的‘标题’字段为‘数据集@数据源’格式等。另外,由于系统的需要,我们在“标识信息”中增加了‘数据集访问参考’字段,它用来定位元数据的存储位置。此外,还进行了许多其它内容的补充,由于篇幅问题不再详述。

经过上述对《国家空间信息基础设施空间元数据内容标准草案》^[3]的添加,修改和删除,我们得到了城市基础地理信息平台元数据的标准,元数据结构如图 1。

3 元数据描述

为了易于城市基础地理信息元数据的网络传输和共享,我们采用 XML(eXtensible Markup Language)语言来描述元数据。在 Internet 环境下,XML 已经成为一种工业标准,基于 XML 的数据表示可以很好地在不同的系统和平台间进行交互。与 HTML 不同的是,XML 可以通过定义不同的标记(Tags)给数据赋予不同的语义。这些特点使得人们可以很方便地用 XML 来定义元数据标准中使用的数据结构,同时使用 XML 存储树状结构的元数据,并且可以和网上发布实现无缝连接。

3.1 XML 结构的定义与 XML Schema

目前有两种定义 XML 文件结构的主要方法,即 DTD 和 XML SCHEMA,本文采用了后者。XML Schema 的优点在于:

1) 它具有丰富的数据类型,支持的数据类型包括字符串、字符型、整数、浮点数、统一资源标识(URL)、全球惟一标识(UUID)等常见的数据类型,来自 XML 本身的类型(entity, entities, enumeration 等),而且还支持由这些简单类型生成的更加复杂的类型;

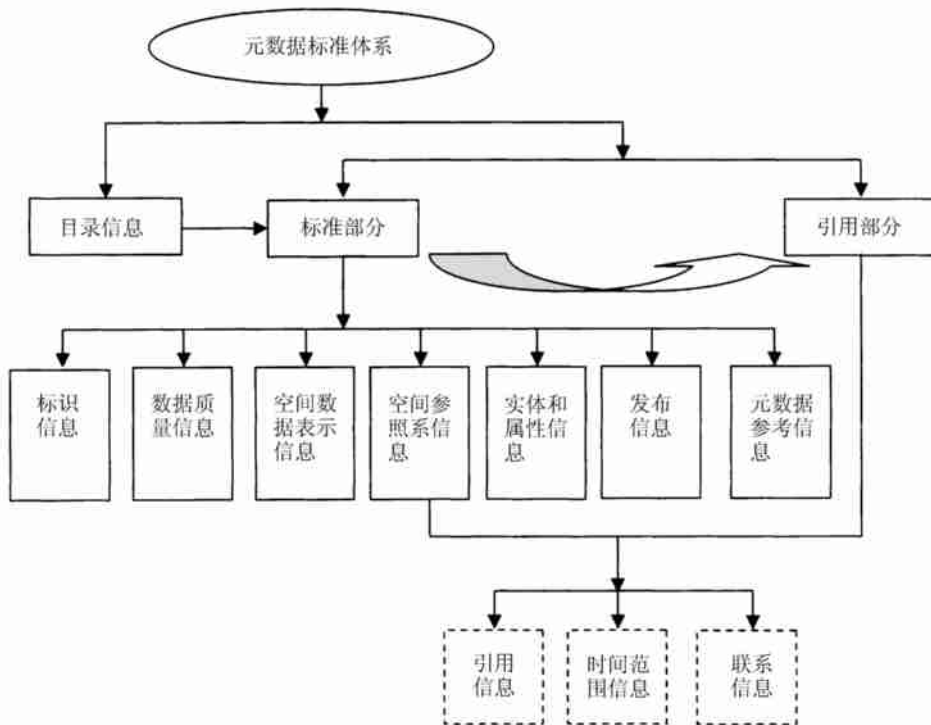


图1 元数据标准体系结构

Fig.1 The Architecture of the Metadata Standard System

2) 可以由用户自己定义数据类型;

3) 支持属性分组:属性的应用范围是多种多样的,有的是针对所有元素,有的则专门针对图形元素;

4) 支持名字空间:允许把文档中特殊的节点与模式中的类型说明联系起来。联系 XML 节点和 DTD 的惟一方法是通过 DOCTYPE 说明,即每一个文档只能使用一个 DTD,但是可以由多个 XML 模式来描述。对于城市基础地理信息来说,由于涉及到不同部门,而不同部门的可能有自己不同的数据对象定义,为了避免冲突,可以使用自己的名字空间。

3.2 元数据结构描述和数据存储

根据所提出的元数据标准,可以用 xml schema 的形式进行元数据结构的精确定义。具体做法就是根据 xml schema 规范,编写一个名为 .xsd 的架构文件,用来进行元数据结构的定义和有效性检查。定义是根据规范文本,定义元数据中各元素的字段长度,类型格式等信息;验证就是在程序实现中可以根据此架构文档,来进行验证 XML 格式正确性的检查,从而保证存储数据的 XML 文档都是符合我们定义的标准规范。

元数据本身也采用 xml 文件来存放。考虑到城市基础地理信息数据分布式存储的特点,我们对元数据提供了临时存放和系统注册存放两种方式。临时存放的元数据可以在任何地方(一般在各部门机器上),主要用于对不完整元数据的存储。这点是考虑到元数据规范中信息众多,很难一次输入成功,而且本部门可以在本地方便的修改自己未注册的数据。因此在客户端不要求用户一次性输入单条元数据记录的完整信息,用户可以将未完成或完成的元数据信息先保存下来,以备日后一起注册。系统注册存放是指将临时存放的元数据信息向中心服务器注册,从而作为一条记录存储在中心服务器的 XML 数据库中。该数据文件存放在中心服务器上,因此很容易维护元数据的完整性和一致性。

当元数据总量不大时,XML 文件数据库管理模式效率较高。如果数据量很大,可以把这些数据信息存放在 ORACLE 大型数据库管理系统的 XDB (XML 数据库)中,具体做法是把定义元数据结构的 XSD 文档添加到 ORACLE 的 XML 方案中,然后增加元数据的记录。由于我们针对的是中小城市,我们采用 XML 文件数据库管理模式,即中心服务器上的元数据以多个 XML 文件(按照 XML 记录的属性分

类,例如部门)进行存储,每个 XML 文件中存放多条元数据记录。

4 元数据平台

元数据平台是为了实现对元数据库进行管理所开发的功能平台,针对制定的中小城市基础地理信息集成的元数据标准,我们完成了支持不同部门网络环境下进行元数据采集和维护管理的功能,详细介绍如下:

4.1 平台功能用例分析

为了系统地搭建元数据平台,我们采用 RATIONAL 公司的系统建模工具 ROSE 对平台进行了设计。RATIONAL ROSE 采用由用户需求驱动的系统设计模式,可以保证对系统各种应有功能的精确设计。在开发底层开发平台环境发生更换时,上层的软件设计模式可以得到复用。

平台的用户对象有浏览用户和管理用户两种。浏览用户可以在用户权限许可的范围内查询和浏览元数据信息,包括客户端浏览用户和浏览器端浏览用户。为了安全起见,管理用户一般仅局限于客户端,它可以是不同部门不同空间位置的授权用户,主要用于维护元数据。从这两类用户角度出发,抽取了“新建元数据”,“打开元数据”,“保存元数据”,“查询浏览元数据”,“注册元数据”,“反注册元数据”以及“修改已注册元数据”七个系统用例。除了“查询浏览元数据”是浏览用户用例外,其它都是管理用户用例。平台的 UML 用例图如图 2 所示:

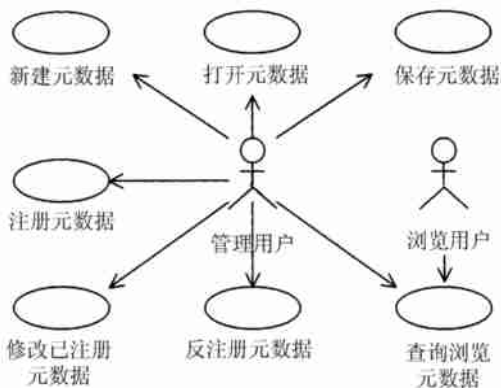


图 2 元数据管理平台的 UML 用例图

Fig.2 The Use-Case Graph of the Metadata Platform

在上述用例中,“新建元数据”,“打开元数据”是针对元数据临时存放所说的,它们都是对本地数据

进行操作:“查询浏览元数据”,“注册元数据”,“反注册元数据”以及“修改已注册元数据”则是针对中心服务器上的元数据进行操作,这也是本元数据管理平台的特色:“保存元数据”始终对当前元数据进行保存,即如果当前编辑的是临时存放的元数据,那么保存的就是本地数据;如果当前正在修改的已注册的元数据,那么更新的则为中心服务器上的元数据库文件。

4.2 技术实现

4.2.1 平台开发环境与 XML 相关技术

平台开发环境上采用 Microsoft 的 NET 框架,其 XML 设计具有符合 W3C 标准、可插接式结构、扩展性好、性能高并且与 ADO.NET 紧密集成等优点。开发语言选取 C#. C# 由 C 和 C++ 演变而来,是一种类型安全和面向对象的语言。因为它被编译为托管代码,所以受益于公共语言运行库的服务,如语言互操作性、增强的安全性和垃圾回收。

.NET 框架中有一个 XML 类集提供架构对象模型(SOM),使用该模型可以以编程方式生成和编译 XSD 架构。XmlSchema 类表示 XSD 架构,使用 XmlReader 和 XmlWriter 类可以加载和保持这些架构。文档对象模型(DOM)类是 XML 文档在内存中的表示形式。XmlDocument 类实现文档对象模型(DOM)的级别 1 和级别 2 建议,这是根据 .NET 框架的通用设计指南定制的。XPath 功能是在数据存储区中查询某一节点或节点集的功能。XPathNavigator 类提供实现对任何数据存储区的 XPath 查询所需的方法,并采用 XML 路径语言 XPath1.0 建议(www.w3.org/TR/xpath.html)中描述的数据模型。数据存储区定义为文件系统、数据库或对象(如 XmlDocument)中的数据。XPathNavigator 使用允许向前和向后移动的游标模型从数据存储区中读取数据,它是只读游标,不允许编辑。

4.2.2 界面设计

由于 xml 是一种树型结构,很适合用 Windows 程序中的树形视图去显示。如图 3 所示,左侧视图为 TREEVIEW,右侧为 RICHEDIT 控件,右下方的注释为当前选择元素的注释。

4.2.3 核心功能实现

实现平台 7 个用例的核心过程只有 5 个,即读写 XML 文档,验证文档有效性,将 XmlDocument 对象展开为 TREE 状结构、查询 XML 和 XML 发布。具体实现如下:

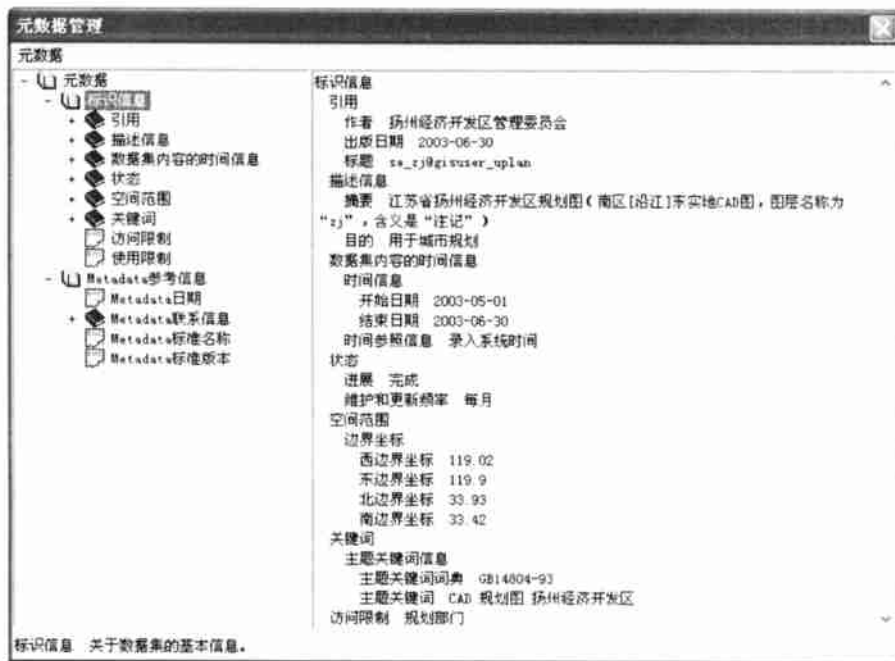


图3 元数据管理界面

Fig. 3 The Interface of the Metadata Platform

4.2.3.1 读写XML文档

虽然 XmlReader 类提供了快速高效读取 XML 文档的方法,但它是非缓存的只进、只读访问,即使用 XmlReader 类无法编辑属性值或元素内容,或无法插入和移除节点。因此我们采用以编程方式读取、操作和修改 XML 文档。在程序中新建一个 XmlDocument 类对象,通过该对象 Load 方法将指定路径的 XML 文档读到新建的 XmlDocument 内存对象中。写 XML 文档则使用 XmlWriter 类来实现。

另外,读写远程服务器上的 XML 数据时,只需要把文档的网络路径作为 XmlDocument 类对象的 Load 方法的参数即可。该参数存储在系统配制文件 config.xml 文件中。由于只是读取,因此采用 XmlReader 类读取以提高效率。

4.2.3.2 验证文档有效性

XML 文档的有效性是指文档中的存储对象的值是否符合架构文件的数据类型定义。由于 XSD 架构文件惟一,因此可以采用 XmlTextReader 对象和 XmlValidatingReader 对象来验证。通过设置 template.xml 模板中 schemaLocation 属性来指定 XSD 构架文件。具体实现如下:

首先,使用系统文件流操作将模板文件 template.xml 和即将验证的 XML 文档合成临时的 XML 文档“tempfile.xml”;

其次,新建 XmlTextReader 对象实例“tr”将临时生成的文档“tempfile.xml”读入内存;

再次,“tr”作为 XmlValidatingReader 对象构造函数的参数创建新对象“vr”,并指定其 ValidationType 属性为 ValidationType.Schema;

最后,使用“vr”内嵌的 ValidationEventHandler 类完成验证,并捕获相应的异常。

此外,由于元数据和数据集是一一对应的,我们为了保证元数据的有效性,还要对元数据进行一致性检查,即检查元数据信息中的数据源和数据集是否存在。

4.2.3.3 将 XmlDocument 对象展开为 TREE 状结构

本过程是显示元数据的核心过程,它负责把在内存中的元数据对象结构解析成为 TreeView 视图中的树状结构展现出来。这要分两步来实现:

首先,取出内存中被读入的元数据的 XML 对象 (XmlDocument 对象) 的第一个子节点 (XmlNode 类型),也就是元数据的根元素节点“Metadata”;

然后,对树状 XmlNode 类型的节点进行中根遍历(递归的方法),将所有叶子节点全部添加到 TreeView 视图中。

4.2.3.4 XML 查询

按照元数据标准中定义的查询元素(标题、出版日期、作者、摘要、主题关键词、边界坐标等十几个元

素),采用 W3C 的 Xpath1.0(可以在 www.w3.org 上查看 Xpath 规范)在 C# 中的实现。

在 Xpath1.0 中,一个元素(子节点,XmlNode 类型)用它在 XML 文档中的树状结构的相对位置来表示。我们查询一个子节点使用根节点的 XmlNode,SelectNodes 方法,参数则是该节点在根节点中的

Xpath(例如“标题”元素的 Xpath 为“//ab:citation/ab:title/text()”,“ab”为“命名空间”,这里是“http://www.863digitalyz.com/2003/6”)和命名空间,返回值为在该根节点下所有相对路径为上述 Xpath 的节点集合(XmlNodeList)。界面如图 4 所示。

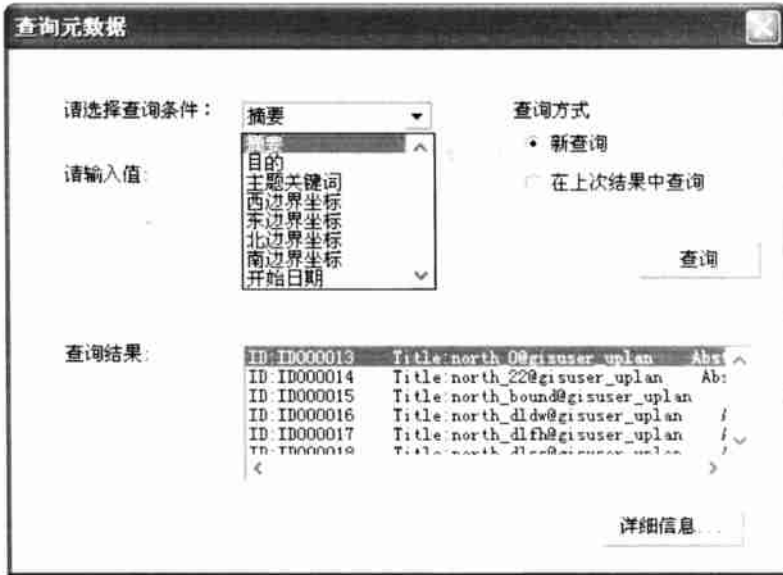


图 4 元数据查询界面

Fig.4 The Interface of the Querying of Metadata

4.2.3.5 XML 发布

采用 XML+XSL 的方式通过 Web 发布元数据。通过 Microsoft 公司的 MSXML2.DOCUMENT 组件,加载元数据文件以及为了发布元数据而编写的 xsl(xsl 文件本身也是 xml 格式的)文件,再通过一个 asp 文件来进行转换和发布,用户就可以在网络上通过 IE 浏览器来查看元数据的一些主要信息。xsl 是用来转换 xml 形式文件为其他格式文件的一种语言,在这里用它来将元数据转换为 html 格式。

5 应用与结论

扬州经济开发区基础地理信息系统是基于网络环境的,包括一个中心节点和三个分节点,即国土局节点、规划局节点和房产局节点。其中中心节点存储所有部门公用的数据,包括城市正射影像数据、数字高程模型数据、基本要素矢量数据、地名数据、数字栅格地图数据等。而国土、规划及房产节点则分别存储各部门专用的数据。在用户管理机制的支持下,所有这些节点之间,以及用户与节点之间都可以

实现数据的浏览、传输、交换和共享,以便为地理信息应用系统提供丰富统一的数据源。

基于上述的系统平台下,我们搭建了适合城市基础地理信息集成的元数据平台。它支持我们制定的城市基础地理信息平台的元数据标准,除具有良好的元数据网络传输、查询和发布功能以外,还支持网络环境下分布在不同地理位置不同部门的授权用户对中心元数据服务器中元数据库的远程维护和更新操作,大大地提高了城市基础地理信息平台数据的管理效率。

我们在每个节点上都配备了所开发的元数据平台软件,并且针对每种数据都建立了元数据库。通过元数据平台对元数据库进行管理、更新和发布,使得所有数据库都是位置透明的,用户无需知道数据的具体物理位置,就可以进行访问和调用。

扬州经济开发区基础地理信息系统的成功建立和运行,说明元数据平台的功能是稳定的,应用起来也比较方便,能够在城市基础地理信息平台的建设中发挥重要作用。

参考文献 (References)

- [1] 陈述彭主编, 城市化与城市地理信息系统, 科学出版社, 1999.
[2] Content Standard for Digital Geospatial Metadata, FGDC, (1998),

<http://www.fgdc.gov/metadata>

- [3] 《国家空间信息基础设施空间元数据内容标准草案(V1.0, V2.0)》国家空间信息交换中心 <ftp://ftp.nsi.gov.cn/pub/standard/standard02.zip>

Development of a Metadata System for Fundamental Urban GIS Framework

CHANG Yuan-fei¹, WANG Wei², SUN Jin-yue³, WANG Qiao-nian³
ZHUANG Yong-xiang⁴, QIAO Yan-you¹, JIAN Jun¹

(1. Institute of Remote Sensing Applications, Beijing 100101, China; 2. Yangzhou Information Centre, Yangzhou 225009, China;

3. Chinese Online Group, Shanghai 201300, China; 4. Yangzhou Development Administration Committee, Yangzhou 225009, China)

Abstract: Through analysis of metadata standards from home and abroad, a metadata standard system for middle and small cities is suggested, it's suitable for the building of fundamental urban GIS frameworks. The structure of the metadata database is designed, and a metadata system is developed with XML, so that the metadata can be managed and published. The metadata system has been successfully applied in building the fundamental GIS framework of Yangzhou Economic Development Region, and large volumes of spatial data has been seamlessly integrated to provide data service with the help of this system.

Key words: metadata; XML; fundamental urban GIS framework; digital city