

文章编号: 1001-1986(2008)05-0011-04

地质元数据共享平台构架

刘明军¹, 郑建功², 姜在炳¹, 祁满意¹

(1. 煤炭科学研究总院西安研究院, 陕西 西安 710054;

2. 西安市勘察测绘院, 陕西 西安 710048)

摘要: 长期以来, 各地质勘探部门积累了大量地质数据, 为了有效管理和共享这些信息, 需建立数据共享平台。探讨了地理元数据的现状和地理元数据共享存在的问题, 以及用 XML 文件存储和关系数据库存储的优缺点等问题; 提出用 XML 与数据库技术相结合的方式构建地质元数据, 解决地质数据共享问题, 并以此为基础构建了用于管理、使用和发布地质数据的共享平台。

关键词: 地质元数据; 存储策略; XML Schema; XML 数据库

中图分类号: P628.4; TP393.17 **文献标识码:** A

Geological metadata sharing platform

LIU Ming-jun¹, ZHENG Jian-gong², JIANG Zai-bing¹, QI Man-yi¹

(1. Xi'an Branch, China Coal Research Institute, Xi'an 710054, China;

2. Xi'an Institute of Prospecting and Mapping, Xi'an 710048, China)

Abstract: For a long time, a large number of geological prospecting data is produced; therefore the sharing platform is established for managing and sharing the geological prospecting data. This paper discussed the status and the sharing of geographic metadata and the advantages and shortcomings of using XML file storage and database storage, including the adoption of XML technology of geographic metadata designed to solve the coal fields of geological data; Based on that build the management, use and release geographical data sharing platform.

Key words: geological metadata; storage strategies; XML schema; XML database

长期以来, 许多地质、勘测等部门获得了大量的地质数据, 而这些数据仍以“数据孤岛”的形式存在, 致使重复建设和浪费问题严重。那么, 如何在“海量”数据里筛选自己需要的数据, 数据的质量如何, 以及如何访问和使用这些数据成果呢。实际上, 只要解决了不同数据部门之间的资源共享问题, 上述问题就会迎刃而解。而采用 XML 与数据库技术相结合的方式构建的元数据和共享平台, 是解决资源共享问题的一个有效途径。

1 地理元数据

1.1 元数据

元数据(metadata)是“关于数据的数据(The data about data)”, 其具有以下原则:

- 元数据是有关数据属性的数据, 即相关原则;
- 元数据所描述的属性必须具有某种应用需求, 即需求原则;
- 元数据必须体现对数据的抽象或概括描述,

即概括原则;

d. 元数据必须直接或间接关联所描述的数据, 即关联原则;

e. 元数据必须有利于数据的组织管理、维护优化、查询搜索和查询交换等。

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、发现数据和使用数据的一种重要的工具与手段, 它是促进地理信息共享的关键和有效资源。

1.2 地理元数据

地理元数据是关于地理数据特征和相关信息资源的描述性信息, 是对于地理特征的概括和抽象。它通过对地理数据的内容、质量、状况、位置和其他特征进行描述与说明^[1], 用来辅助和管理地理数据。地理元数据可提供地理数据集的特征资料, 使数据用户可据此确定该数据的名称、来源、组织结构和适用范围等。例如, 地图图例就是一种地理元数据, 它提供了关于地图出版者、出版日期、地图类型、地图描述、地图比例尺及其精度等方面的信息^[2]。

收稿日期: 2008-02-02

作者简介: 刘明军(1976—), 男, 陕西神木人, 工程师, 从事地理信息系统设计与软件开发研究。

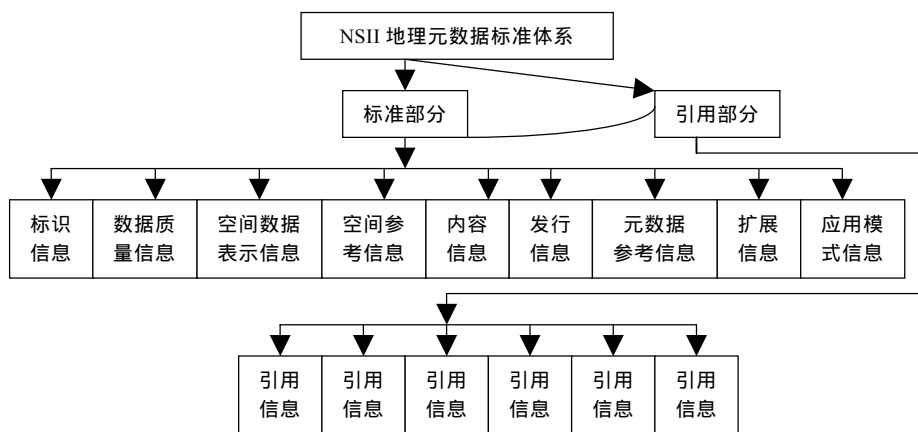


图 1 NSII 地理元数据标准体系

Fig.1 Standard system of NSII geographic metadata

当前,人们对地理元数据的研究已广泛开展,并形成了多个标准,其中主要有美国联邦数据委员会(FGDC)的地理元数据标准和国际标准化组织 ISO/TC211 的地理元数据标准,我国也建立了 NSII (National Spatial Information Infrastructure, 国家空间信息基础设施)^[3]地理元数据标准体系(图 1)。

1.3 地理元数据共享存在的问题

1.3.1 各自为营 缺乏标准

由于缺乏地理数据统一标准,地理数据的生产者生产出了大量不同数据格式和信息标准的数据,这些格式和标准都是由各科研、生产部门根据各自的需要制订的。各个不同机构组织根据各自生产的需要对空间元数据信息进行了不同的描述,于是地理元数据也处于同样的境地,从而很难将不同来源的空间元数据集集成到一起^[4]。

1.3.2 地理元数据不统一

空间数据的特征涉及到空间数据集的组织结构、采集手段、质量等级、参照系统及时间范围等诸多方面。在实际生产过程中,不同机构组织为满足其特殊需要增加了一些其他信息,很不规范,严重影响了空间元数据的共享。

1.3.3 地理元数据管理过于分散

地理元数据隶属于不同的部门或不同的项目,缺乏必要的集中统一管理,而且多为异地分布,致使元数据的共享难以实现。

1.3.4 落后的元数据描述

目前,地理元数据的表达大都采用文本性的描述,如用“陕西省、西安市、碑林区、友谊西路”来描述地理位置,用“10 km×10 km”来描述范围信息,用“煤炭科学研究总院西安研究院成立于1956年”来描述建设信息等,它有通俗易懂、易于理解的优点,但却难于描述大数据量、复杂的信息,而且基于文本描述过于松散、数据之间联系不紧密、

表述冗余,不利于数据的管理、共享和发布。但是,XML 技术的出现为解决这一问题带来了契机。

2 实现共享的关键技术

2.1 XML 语言

XML (eXtensible Markup Language)是全球信息网协会(W3C) 推荐在 Web 上使用的数据表示和交换标准,是标准通用置标语言 SGML 的一个有限子集。由于 HTML 的缺陷,严重制约了 Web 技术的发展,使其无法实现 Internet 海量信息的资源共享和互操作以及数据库功能,由此,XML 应运而生。在 Internet 环境下,XML 已经成为一种工业标准,基于 XML 的数据表示,可以很好地在不同系统和平台间进行交互。与 HTML(仅给出了处理对象的显示信息,未给出描述对象的属性信息)不同的是,XML 可通过定义不同的标记(Tags)给数据赋予不同的语义。另外,信息的内容和信息的表示是分开的。这些特点使人们可以很方便地用 XML 来定义地理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,还可以和网上发布实现无缝连接。另外,它还具有下列优点:a. 系统可利用 XML 方便地表达地理元数据;b. 系统可利用 XML 的样式语言实现地理元数据间的转换和显示;c. 可以方便地实现系统地理元数据的查询。

W3C 提供了两种控制 XML 文档结构的主要方法—DTD 和 XML Schema。本文采用 XML Schema。相比 DTD,XML Schema 具有更多优点:XML Schema 文档本身也是 XML 文档,而不像 DTD 那样使用自成一体语法;支持常见的数据类型(包括字符串、字符型、整数、浮点数、统一资源标识(URL)和全球惟一标识(UUID)等);用户还可自定义数据类型,且 XML Schema 支持命名空间,这都使 XML Schema 具有较强的可扩展性。其他技术还包括可扩展

展的样式单语言(XSL)、XML 路径语言(Xpath)和 XML 链接语言等。

2.2 地理元数据的结构描述

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;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;在程序实现中根据此架构文档,验证 XML 格式的正确性,从而保证存储数据的 XML 文档都符合定义的标准规范。

2.3 地理元数据的扩展

一个完善的地理元数据库的共享平台建设,必须考虑平台的可扩展性。地理元数据包含遥感、测量、人文地理和自然地理等多学科的数据。从根本上说,地理元数据要做到扩展,可从元数据的标准化和元数据的扩展两方面入手。

建立地理元数据标准的目的是,为了提供一个描述地理数据集的过程,使用户能够定位和访问地理数据,并确定数据的适合度。另外,可借助 XML Schema 来实现地理元数据的多学科扩展。

3 地质元数据的存储策略

地质元数据是地理元数据的扩展,主要在元数据子集中增加了元数据名称元素。地质数据的来源与格式呈多样性、复杂性和海量性等特点,如地质勘查数据,包括地层数据、断层数据、钻孔数据和周围地形地貌数据等。如何有效共享这些数据,是研究的重点,而用 XML 技术建立地质元数据,是解决这一难题的关键技术。

那么,如何管理这些地质数据,存储方式非常关键。地质元数据的存储主要采用两种方式:基于 XML 文件的存储策略和基于数据库的策略。

3.1 XML 文件存储策略

下面是以地质勘查数据为例编写的一级元数据:

```
<?xml version="1.0" encoding="gb2312">
<metadata>
  <drillingInfo>
    <attributes>
      <ID>14272701</ID>
      <ExBlock>勘查区块</ExBlock>
      <longitude>107.51</longitude>
      <latitude>35.11</latitude>
      <Elevation>1100 m</Elevation>
      ...
    </attributes>
  </drillingInfo>
</metadata>
```

但是,XML 技术在数据管理方面也存在如下缺陷:首先,XML 技术是基于文件的管理机制,而文件管理存在着容量大、管理困难的缺点;第二,目前 XML 的检索是基于节点的检索,存放大量数据的 XML 文件致使检索速度极低;第三,XML 采用的解析机制包括 SAX 和 DOM 两种方式:SAX 方式是基于文件的解析,速度较慢,而 DOM 方式是基于内存的方式,资源消耗极大;第四,XML 的安全性及并发操作机制尚需解决。所以,从严格意义上说,XML 不适合大型数据管理。与 XML 相比,数据库在数据管理方面具有管理方便、存储容量小、检索速度快、修改效率高、安全性好以及并发控制良好等优点,但它需要用一种应用方式将其丰富的数据有效地发布出来,以消除平台差异,增强语义描述功能,降低环境要求。因此,应将 XML 与数据库技术相结合,消除 XML 技术在数据管理方面的缺陷。

3.2 数据库存储策略

目前,很多大型数据库都支持 XML,如 SQL SERVER 数据库。SQL SERVER 数据库对 XML 的输出支持主要体现在:

- 在 select 语句中,通过增加 FOR XML 从句来支持从数据库中以默认或自定义的 XML 形式提取数据,并可以与 IIS(Internet Information Services)结合通过 HTTP 协议直接访问数据库,在浏览器中得到 XML 数据;
- 提供了一种基于 XML 的语法(Update Grams),使数据库的更新可以 XML 的方式进行;
- 通过 XPath 进行信息定位;
- 在存储过程中,使用 OpenXML 函数写 XML 数据;
- 使用 Microsoft SQL Server 2000 OLEDB 供应器(SQL OLEDB),将 XML 文档设置为命令文本并将结果集返回为 XML 格式数据或字符串或流等。

SQL SERVER 数据库对 XML 的存储支持,是通过 OPENXML 将 XML 数据写入数据库中实现的,其语法如下:

OPENXML(<XML document handler>,<path pattern>,<flags>) with (Schema|Table)。

OPENXML 处理 XML 文档有以下 3 步^[5]:

- 用存储过程 sp_xml_preparedocument 将 XML 文档转换为数据库内部的 DOM 表示,此时得到的即为 XML document handler。
- 用与 XML 的基本元素相联系的模式来创建一个模式(XML Schema),也可以用已存在的表来替

代一个模式,此时表的字段名和 XML 名相同。

c. 用存储过程 `sp_xml_preparedocument`, 将一个转换好 XML 的文档从内存中除去, 以释放存储空间。

尽管许多大型数据库都增加了支持 XML 的功能, 但目前尚处于研究阶段, 还存在用户检索效率低、存储容量大和安全性控制不完善等缺点。为了更好地建立和管理地质元数据库, 尚需借助编程语言来实现功能的灵活扩展。

4 地质元数据共享平台的构建

4.1 应注意的问题

为了实现地质信息的网络共享, 在构建元数据系统时应注意: a. 首先应考虑该系统对元数据标准支持的可扩展性, 并能让用户方便地定义自己需要的元数据格式; b. 使用户可以通过 WWW 来访问元数据服务器; c. 要有效地组织系统的元数据记录, 以使用户在系统的海量数据记录中, 快捷地查到所需元数据, 实现不同系统之间的交互操作; d. 使系统能更好地实现异地元数据的整合。

4.2 元数据共享平台

要实现元数据真正意义上的网络共享, 必须解决网络环境下分布式异构系统互操作, 要具备平台独立、用户透明、轻松穿过防火墙等特点, 而 Web Service 技术的出现和发展为其提供了契机^[6]。在这种体系中, 网络中的每个结点都可能是服务消费者和提供者, 对外以 Web 服务的形式暴露所提供的功能和数据, 同时通过标准的 SOAP/XML 通信协议访问其他节点提供的服务, 具备很强的互操作能力^[7]。

元数据共享平台主要包括客户端、应用服务器和企业信息层 3 个部分, 其结构关系如图 2 所示。

a. 客户端: 元数据管理 UI 属于客户端层, 主要根据用户的不同权限实现对地理元数据库的浏览、查询、维护、定义、发布。通过 HTTP 协议向服务器发送查询参数, 将服务器处理后接受返回的查询结果, 进行显示和表达。

b. 应用服务器层: 提供所有的业务逻辑处理功能, 完成系统中对数据库的所有操作功能, 尤其是完成处理 Web 浏览器请求、客户端 XML 编辑和文档导入工具等请求。元数据服务器, 专门负责元数据在 Internet 上的发布任务。

c. 企业信息层: 主要包括关于地质勘探数据的元数据组织, 包括基于 XML 的存储文件和数据库两种数据载体。

5 结语

XML 缺乏关系数据库所具有的数据管理手段, 但在异构数据交换方面, XML 却具有很大优势。在数据应用和数据管理方面, XML 技术和数据库技术各有优势, 二者可以相辅相成、扬长避短。所以, “以数据库为存储手段, 以 XML 为交换载体”的数据管理模式, 利用 WebService 技术构建地理元数据共享平台构架^[8], 是地质元数据共享平台系统发展的方向。

参考文献

- [1] 中国 21 世纪议程管理中心. 中国地理信息元数据标准研究 [M]. 北京: 科学出版社, 1999.

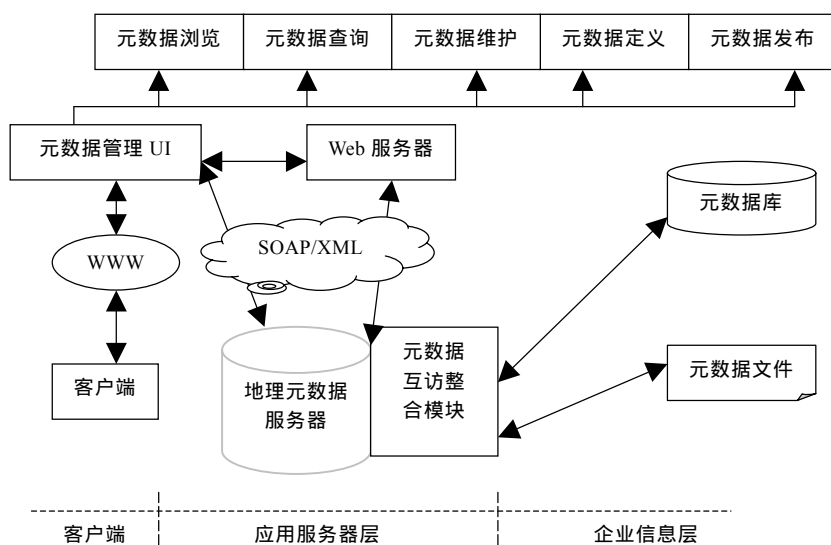


图 2 元数据共享平台构架
Fig.2 Metadata sharing platform

(下转第 19 页)

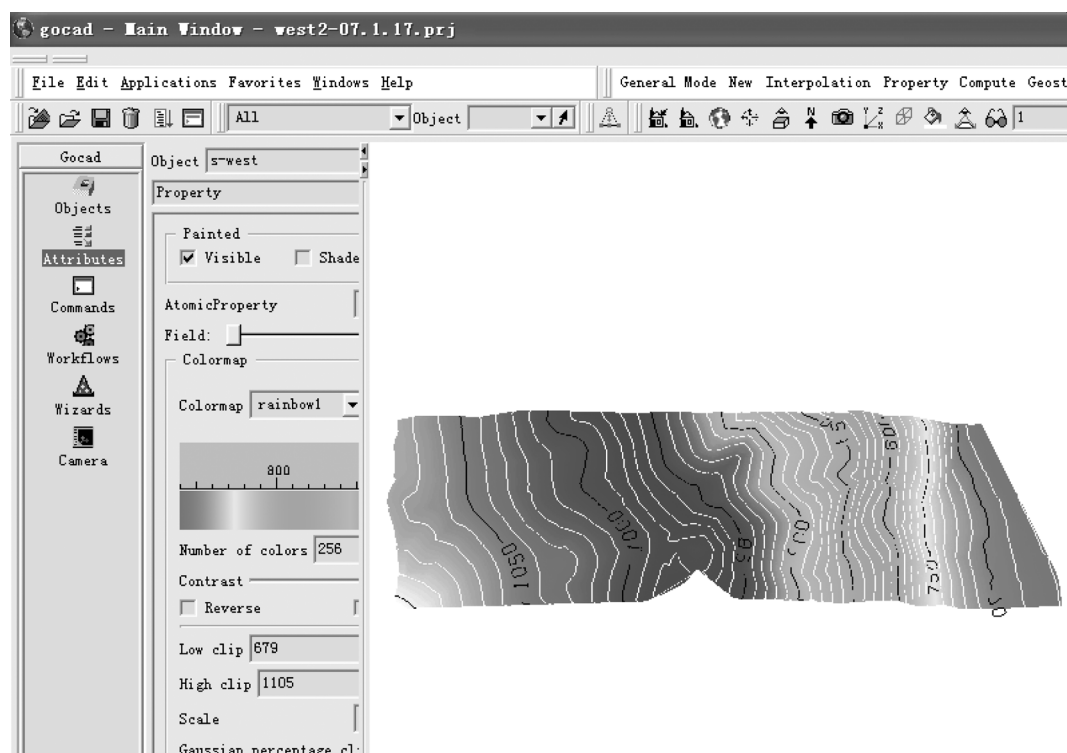


图 4 地表等高线(Z 坐标属性)显示

Fig.4 Contour map of the earth surface

参考文献

- [1] 魏群. 数字城市可视化仿真三维实体建模及虚拟现实软件系统的自主研发[C]// 宋玲, 赖明, 廖小军. 第二届中国国际数字城市建设技术研讨会论文集. 苏州: 知识产权出版社, 2006.
- [2] 魏群. CIS/2 钢结构国际标准及 CAD 系统的研发应用[C]// 第二届中国国际钢结构大会论文集. 北京: [s.n.], 2005.
- [3] 魏群. 逻辑产品模型的研究与应用[C]// 新世纪水利工程科技前沿(院士)论坛论文集. 天津: 天津大学出版社, 2005.
- [4] 张俊红, 魏群, 刘秀伟, 等. 电磁波 CT 数据在构建地质数据库中应用的研究[C]// 魏群, 张国新, 胡兴旺, 等. 三维可视化仿真与虚拟现实技术的研发应用. 郑州: [s.n.], 2006.
- [5] 张建桃. 公路边坡的工程地质三维建模与稳定性分析[D]. 北京: 中国矿业大学, 2005.
- [6] 张聚兴. 三维地质建模在工程地质环境质量评价中的应用研究[D]. 北京: 中国地质大学, 2006.
- [7] 钟登华, 李明超. 水利水电工程地质三维建模与分析理论与实践[M]. 北京: 中国水利水电出版社, 2006.
- [8] 朱焱辉, 朱培民, 金丹. 从 arcgis 平面栅格图形到 gocad 三维图像转换的开发探讨[J]. 工程地球物理学报, 2004, 10(1): 435-440.
- [9] 党丽娟. 工程地质三维建模方法及关联数据库的研究[D]. 郑州: 华北水利水电学院, 2008.
- [10] 李亦纲, 曲国胜, 陈建强. 城市钻孔数据地下三维地质建模软件的实现[J]. 地质通报, 2005, 5(24): 470-475.
- [11] 朱良峰, 吴信才, 刘修国, 等. 城市三维地层建模中虚拟孔的引入和实现[J]. 地理与地理信息科学. 2004, 20(6): 26-30, 43.

(上接第 14 页)

- [2] 李鹤元, 李波, 朱文忠, 等. 地理空间元数据和基于网络的数据分发技术[J]. 测绘学院学报. 2003(1): 58-61.
- [3] 蒋景瞳. 中国地理信息元数据标准研究[M]. 北京: 科技出版社, 1996.
- [4] 张书亮, 龚敏霞, 闫国年. 基于 XML 的地理空间元数据表达研究[J]. 遥感学报. 2003(1): 66-71.
- [5] BERTINO E, CATANIA B. Integrating XML and Databases[J]. IEEE Internet Computing. 2001(7-8): 84-88.
- [6] 王继周, 李成名, 林宗坚. 面向共享的空间元数据管理框架研究与设计[J]. 国土资源遥感. 2003(3): 67-69.
- [7] 顾蔚, 陈天滋. 基于 XML 的元数据系统在地理信息共享中的研究与设计[J]. 计算机应用研究. 2003(1): 104-107.
- [8] 吴金华. 地理空间元数据的探讨[J]. 西安工程学院学报. 2002(2): 59-61.