空间数据存储研究及其在岩石圈深部探测数据中的应用

韩 露 管 烨 高 锐

(中国地质科学院地质研究所, 北京 100037)

提要: 深部地球物理探测数据是岩石圈结构研究的重要资料, 然而由于地学数据本身的多来源、多学科、海量等特点阻碍了数据的共享, 而 GIS 技术恰好以强大的多学科交叉和空间数据处理功能为数据发布提供了新的途径。在 GIS 中信息的显示和分析方法都取决于地理信息数据模型, 因此在利用 GIS 技术建立数据共享平台时, 建立一个符合数据体查询和分析要求的数据模型尤为重要。笔者对几种空间数据存储方式和主流的空间数据模型进行了深入研究, 最终结合深部地球物理探测数据的特点, 以 Geodatabase 为基础建立了一个数据模型, 为数据发布奠定了基础。

关 键 词: 深部地球物理探测数据; 数据模型; Geodatabase

中图分类号: P313.2 文献标志码: A 文章编号: 1000- 3657(2008)01- 0172- 08

岩石圈是地球内部最接近地表、直接影响人类生存和发展的地球圈层。地球表面板块运动碰撞、山脉与盆地的形变、资源分布、地震活动、海啸发生、火山喷发等都受控于岩石圈内部结构的改变与活动, 岩石圈结构的研究是地学领域的重点[1]。而对岩石圈的研究主要以人工地震等深部地球物理探测方法最为有效[2], 因此中国系统地开展了岩石圈深部调查工作并积累了大量的宝贵资料。为了保存这些数据资料并充分发挥它的价值, 国家陆续开展了数字化工作并分别建立了专业的数据库[3]。但是由于中国特殊的地理位置, 地质构造异常复杂, 单一的数据资料无法满足研究的需要, 如何获得各种专业方法的深部探测数据已经成为当今研究中必须解决的问题, 因此针对已有的深部地球物理探测资料实现数据的共享和综合下载是推动岩石圈结构研究的必经之路。

由于地学数据的多来源、多学科、海量等特点很难将数据统一并建立数据库, 目前所建立的专业数据库普遍都是存放某一个专业方法的数据, 这既不利于数据的综合管理同时也为数据共享增加了难度。然而 GIS(Geographic Information System)的出现和发展为解决上述问题提供了新的途径。GIS 技术的显著特点是具有多学科交叉功能和空间数据处理功能, 因此被广泛地应用于地学界, 它以空间数据库为基础, 采用地理模型分析方法, 适时提供多种空间和动态的地理信息[4]。在科技部科学数据共享的发展和推动下, 岩石圈深部探测数据需要结合 GIS 技术建立共享平台, 这样才能充分发挥数据的作用并实现基于平台的数据分析功能。

GIS 是管理和存储空间和属性数据的平台, 建立 GIS 系统的目标主要是提供空间框架, 通过这个框架实现对数据的检索、查询和分析等。然而在 GIS 中信息的显示和分析方法都取决于地理信息数据模型, 因此在开发 GIS 平台时, 建立一个符合数据体查询和分析要求的数据模型尤为重要[5]。然而, 深部探测数据包含的专业很多, 专业的多样性导致了数据体的多样。从数据内容上可分为属性数据和空间数据两部分, 属性数据包括原始的探测数据、成果数据记录 、成果性的图片(如: jpeg 图片, cordrel 断面图

收稿日期: 2007- 05- 20; 改回日期: 2007- 08- 17

基金项目: 国家自然科学基金项目(40334035)、国家科学技术部国际科技合作项目(2006DFA21340)及地下信息探测技术与仪器教育部重点实验室项目(GDL0501)联合资助。

作者简介: 韩露, 1979 年生, 女, 博士, 主要从事 GIS 与地球物理研究; E- mail: hanlu311@163.com。

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

像); 空间数据包括矢量数据、遥感影像数据、高程数据等, 数据内容丰富、类型复杂。要基于 GIS 建立深部地球物理探测数据平台, 就需要将上述所有的属性数据与空间数据融合, 建立一个统一的空间数据模型, 并实现海量异构数据的存储和管理。为了实现该目的, 笔者对空间数据的管理方式和几种流行的 GIS 数据模型进行了深入研究, 最终选择了 ArcGIS 的新一代数据模型 Geodatabase, 并通过 ArcSDE 和 SQl Server 实现了一个深部探测数据库的设计。

1　空间数据管理和存储方式研究

1.1　文件与关系数据库的混合管理

空间数据属于非结构化数据, 早期 GIS 数据库都采用文件和关系数据库混合的方式管理。文件主要用于存储空间图层及其拓扑关系, 关系数据库用于存储属性数据[6]。通过唯一的标识符来建立属性和空间之间的联系(图 1)。这种方式的文件管理系统的功能弱, 在互连网上进行数据共享时, 在数据的安全性、一致性、完整性、并发控制以及数据恢复等方面的功能比较差, 特别是多用户的并发控制比较逊色。

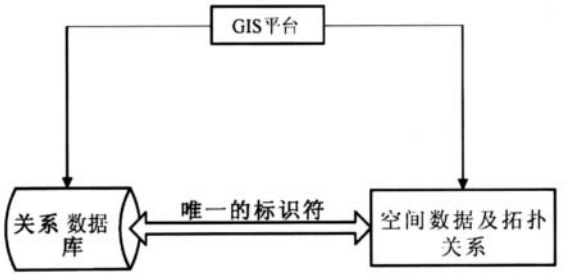

图 1　文件与数据库的混合管理方式

Fig.1　Mixed management way of files and database.

1.2　关系数据库管理

为了解决数据的安全性、并发操作等问题, 出现了用关系数据库管理空间数据的存储方式[6]。这种存储方式是将图形数据与属性数据都采用现有的关系型数据库存储, 对于空间几何数据主要采用以下两种方式: 一是对变长的几何数据进行关系范式的分解, 分解成定长记录的数据进行存储; 二是将图形数据的变长部分处理成 Binary 二进制 Block 字块, 目前通用的数据访问接口(如 ADO,ODBC)都支持二进制数据块的访问, 对于数据的操作和运算都要有相应的应用程序来支持。这种数据管理方式可将空间数据与属性数据统一存储, 避免了混合存储的缺陷, 但是每当用户请求一个空间对象的时候, 就需要大量的连接操作, 二进制的读写效率也比定长的属性字段要低, 因此操作费时, 系统开销也比较大。

1.3　对象——关系数据库的管理

由于 GIS 应用的发展, 人们对空间数据模型的要求不断提高, 于是出现了面向对象的数据模型, 这种方式能够准确地描述空间对象的属性及其行为。但是面向对象的方法缺少结构化查询语言的支持, 很难将属性信息与空间信息结合起来, 而关系数据库恰恰拥有完善的数据管理、查询等功能, 能够弥补面向对象数据库的不足。如果将两者结合起来, 则是较为理想的 GIS 数据库管理方式[6]。于是近年来出现了对象——关系数据库的管理系统, 它是一种用于空间几何对象数据库管理的专用软件模块, 能够在传统的关系数据库管理系统上进行扩展同时管理矢量图形数据和属性数据。这种扩展的方式有两种: 一是一些 GIS 软件商在原有的关系数据库上的扩展, 通过外加一个空间数据管理引擎来实现, 比如 ESRI 的 ArcSDE, MapInfo 的 Spatial Ware 等; 另外一种是数据库管理系统的软件商本身在自己的管理系统上进行的扩展, 使其能够直接存储和管理矢量空间数据, 比如 Informix 和 Oracle 等都推出了自己的空间数据管理扩展模块。但是两者的原理基本是一致的, 由于对象——关系数据库提供了非结构化的数据管理的扩展, 其效率比纯关系数据库管理方式要高的多, 同时它又具有数据的安全性、一致性、完整性、并发控制以及数据损坏后的恢复功能, 因此是目前最理想的空间数据管理方式。

2　空间数据模型研究

2.1　几种主流空间数据模型

空间数据模型是对真实世界的抽象, 它是由一系列支持地图显示、查询、编辑和分析的数据对象组成的。笔者重点研究了空间数据模型的发展过程中主流的三代数据模型: CAD, Coverage, Geodatabase。CAD 是早期的数据模型, 它的模型结构相对简单, 以点、线、面、二进制数据存储地理数据, 只能提供少量的属性数据, 难以提供 GIS 分析所要求的数据。Corverage 是第二代数据模型, 它是由 ESRI 公司在 20 世纪 80 年代推出它的第一个 GIS 软件 Arc/Info

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

时推出的。Coverage 提出了具有拓扑关系的数据组织,这种模型有两个关键之处,一是空间数据与属性数据相结合。空间数据存储在二进制索引文件中,使得显示和访问最优化。属性数据存储在表格中,用等于二进制文件中要素数目的行来存储,并且属性和要素使用同一 ID 连接,可通过 ID 进行双向查询;二是矢量要素之间的拓扑关系存储在数据表中,通过这个关系可以确定某一要素和其他要素之间的空间关系。Coverage 数据模型的优势是用户可以自定义要素表格,不仅可以添加字段并且还可以建立与外部数据表格的关联。但是 Coverage 的属性数据和空间数据是分开存储的,不利于数据的统一管理。Geodatabase 是继前两个数据模型之后的第三代数据模型,它采用面向对象技术,将现实世界抽象为由若干对象类,每个对象类有各自的属性、行为、规则,这些对象类同样具有继承性、封装性、多态性等特点,用户可以在已有地理数据模型的基础上,建立符合自己需求的扩展模型。Geodatabase 类似于 Corverage 的矢量数据模型,提供对要素类及其拓扑关系建模的支持,在此基础上它还扩展到对复合网络、要素间关系以及其他面向对象要素的支持。最重要的是 Geodatabase 模型可以建立在标准的关系数据库之上。Geodatabase 有 3 种形式:personal Geodatabase ,ArcSDE Geodatabase,还有 ESRI 在 ArcGIS 9.2 以后推出的 File Geodatabase。利用 ArcSDE 空间数据引擎,可以将空间数据和属性数据存储统一在关系数据库中,建立一个对象——关系管理模式的空间数据库。

2.2 Geodatabase 的数据组织方式

在 Geodatabase 中将数据分为四类描述,这 4 类分别为用于表现要素(Feature)的矢量数据,表现影像(Image)、专题格网数据和表面的栅格数据,描述表面的不规则三角网络(TIN),地理寻址的地址定位数据[5]。Geodatabase 按层次将地理数据组织成数据对象,并以标准的关系数据库为基础,将空间数据按照一定的空间关系和几何形状逻辑分组(图 2)。处于层次顶端的是空间数据库,往下是根据不同的分类原则组织的要素集和要素类、对象类、关系类以



图 2 Geodatabase 的组织结构

Fig.2 Organization structure of geodatabase

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

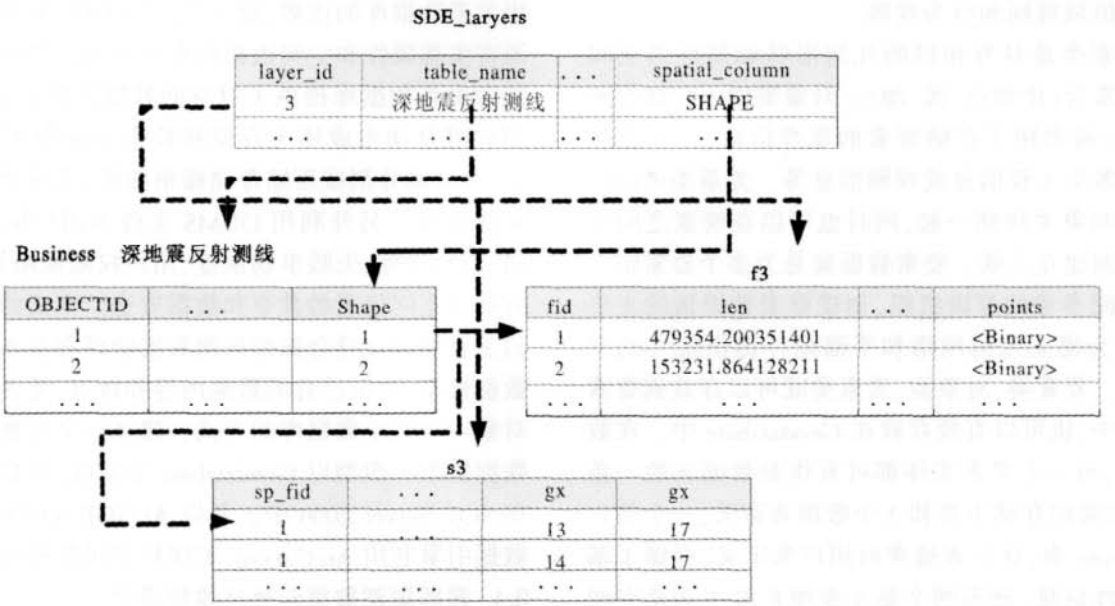

图 3　要素类在 Geodatabase 中的存储结构

Fig.3　Storage structure of elements in geodatabase

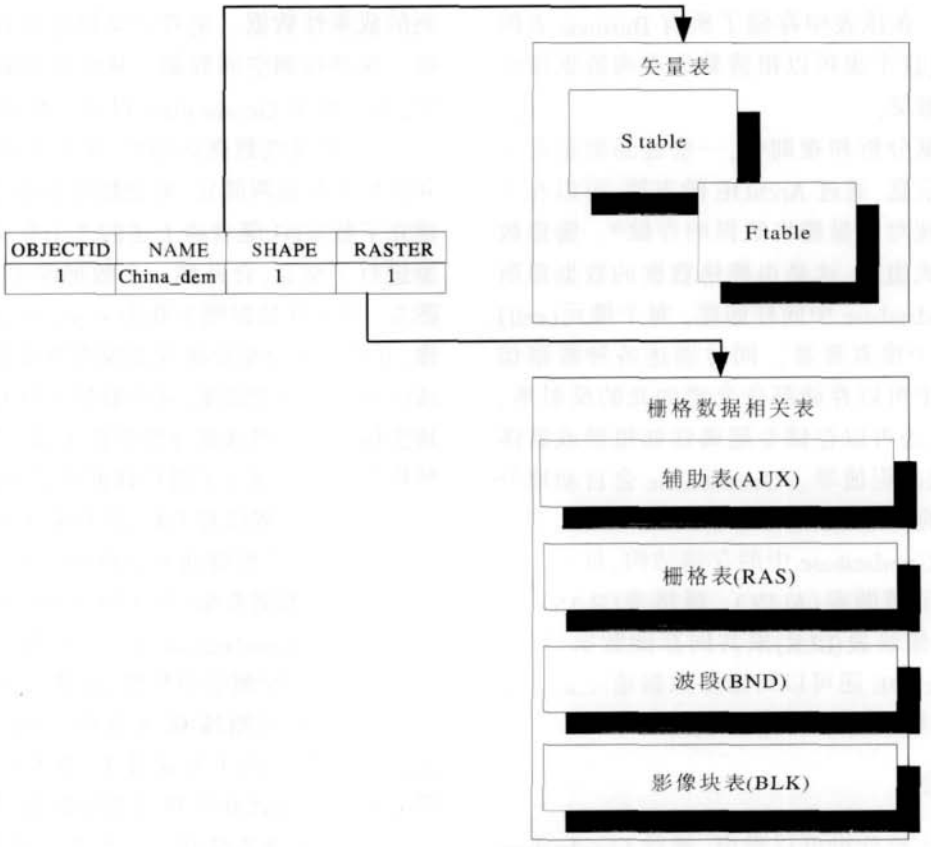

图 4　栅格数据在 Geodatabase 中的存储结构

Fig.4　Storage structure of grid data in geodatabase

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

及各种值域规则和行为规则。

要素类是具有相同的几何形状和属性的空间实体的集合(比如点、线、面)。对象类则不存储空间信息,它通常用于存储要素的属性信息,比如各种专业的数据工程信息或观测信息等。关系类可以将要素和对象关联到一起,同时也可以在要素之间或对象之间建立关联。要素数据集是为多个要素定义相同空间参考的逻辑组织,创建要素数据集的主要目的是为建立几何网络和平面拓扑提供统一的空间参考。要素类、对象类、关系类既可以存放到要素数据集中,也可以直接存放在 Geodatabase 中。在数据库中,每一个要素实体都可看作是数据表的一条记录。要素的存储主要和 3 个数据表有关,一个是它的 Business 表,这个表通常由用户来定义,存储了基本的属性信息。还有两个是 S 表和 F 表,F 表是空间数据表,S 表为空间索引表,通过 Business 中的 shape 字段对应 S 表和 F 表的 FID 字段即可查询到一个要素实体,其基本的结构关系见图 3。另外在空间数据库中数据都是按层划分的,ArcSDE 自动生成了 LARYERS 表,在该表中存储了所有 Business 表的基本信息,通过这个表可以很清楚地查询数据库中所有的空间数据层。

在地学数据分析和查询中,一些遥感数据是非常重要的辅助信息,通过 ArcSDE 的支持,可以在关系数据库中实现对海量栅格数据的存储[5]。栅格数据使用网格形式组织,这是由栅格数据的数据源所决定的,在 Geodatabase 中同样如此。每个像元(cell)代表栅格的一个像素要素,同时描述各种数据信息。比如像元中可以存储部分光谱的光的反射率、图片的颜色值,还可以存储专题属性如植被或岩体类型、表面值或高程值等。Geodatabase 会自动地分割一个非常大的栅格数据,并将它们压缩存储。图 4 为栅格数据在 Geodatabase 中的存储结构,每一个栅格数据都包含了辅助表(AUX)、栅格表(RAS)、波段表(BND)、影像块表(BLK)来共同存储数据信息。另外,通过 ArcSDE 还可以对栅格数据建立金字塔索引,提高对栅格数据的访问速度。

3 应用实例

通过以上比较分析可以看出,通过 Geodatabase 可实现对多种类型数据的统一存储。它具有面向对象的优势,可以准确地描述空间实体,同时还可利用关系数据库的优势,建立广泛的关联,利用结构化语言实现属性和空间数据的双向查询。空间数据引擎为关系数据库提供了对空间数据管理的接口,利用它可自动生成用于存储和检索空间数据的数据表,并自动分割或压缩存储栅格数据,实现数据间的无缝连接。另外利用 DBMS 支持多用户的并发访问、事物管理、失败事物恢复、用户权限策略等机制,有利于空间信息的共享和数据安全。本文的研究目的主要是建立符合深部探测数据管理和发布需要的数据模型。根据已有的数据内容和特点,笔者采用了对象——关系数据库的方式,建立一个完整的空间数据模型。模型以 Geodatabase 为基础,将数据存放在 SQL Server 2000 中,并以 ArcSDE 9.0 作为空间数据引擎利用 ArcCatalog 实现对空间数据的管理。

3.1 深部探测数据分析及模型设计

岩石圈深部探测数据主要包括以下几个专业:深地震反射、深地震测深、深部重力、深部磁力、大地电磁深部探测、宽频带深地震反射、大地热流。其中大部分都包括了探测的原始数据和由原始资料所得到的成果性数据。笔者主要将这些数据按照属性数据、深部探测空间数据、基础地理数据这几大类组织,然后根据 Geodatabase 设计了数据模型。

(1)对属性数据的组织:属性数据分为原始数据和成果性数据两部分。原始数据在研究的过程中已经建立了数据库,笔者将上述的 7 个专业的原始数据模型进行了整合,合并到一个数据库中,其数据结构如图 5。成果性数据则主要为* .jpg,* .bmp,* .cdr 的图像,它们大部分是反映深部探测测线的剖面研究结果或区域性的研究结果,这些数据不具有几何投影和地理坐标,因此将这部分数据作为属性数据来管理,在数据库中专门建立了图片数据库,利用图片编码与相应的测线或工程信息关联,图片采用了二进制数直接存储在‘Image’类型的大字段中,对于这部分数据的读取和导入都需要编写专门的代码。所有的属性数据表都可作为 Geodatabase 中一个对象类的实例。

(2)深部探测空间数据:深部探测空间数据主要是根据具体的测线或测点的空间位置所绘制的 shapfile 文件。由于专业较多,基本的属性信息无法满足需要,因此在处理这部分数据时是根据专业划分的,每个专业方法建立一个要素数据集,其中存放了相应专业的测点图层、测线图层。为了便于研究和查询,在要素数据集中建立了关系类,这些关系类有

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

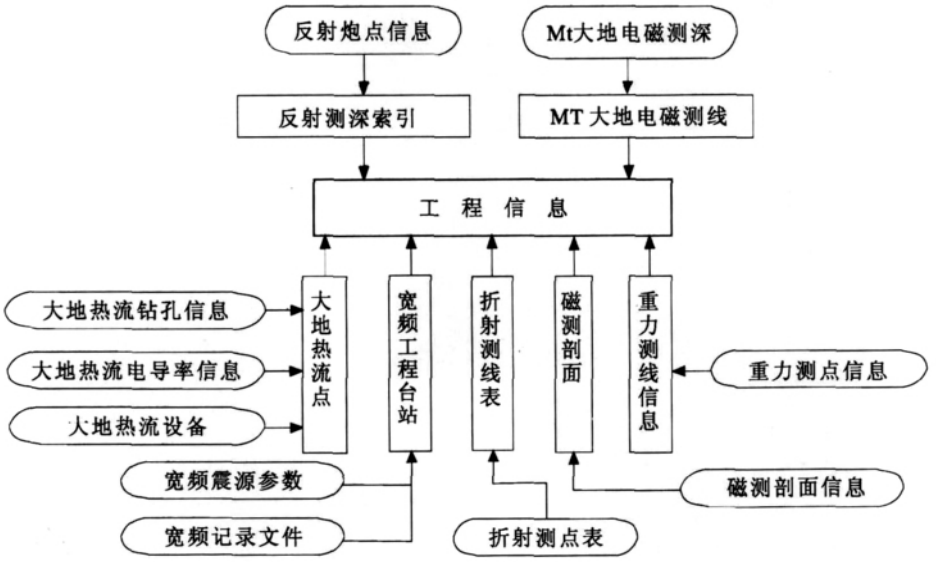

图 5　深部地球物理探测属性数据的结构

Fig.5　Structure of attribute data in deep geophysical survey

的是测线和测点之间的关联，有的是空间要素信息和相应的属性表之间的关联,因为在要素的 Business 表中只有持久的属性才有必要映射成该表中的具体字段,其他的就可以直接通过属性数据模型来实现,建立这种关系类可以扩展空间图层的属性信息,又不破坏已经建立的属性数据模型的结构，同时可以通过关系的建立实现基于空间要素和属性信息的双向查询。但是在建立要素数据集需要选择一个合适的空间域并保证同一个空间要素集的坐标投影都是一致的。选择一个合适的空间域,既要确保研究区域数据全部落在定义的空间范围之内，又能保证存储整型空间数据得到最大化的压缩，由于深部探测数据基本是全国范围的数据，因此在建立要素数据集时空间范围都是统一的全国区域，并且都采用了 Lambert 的投影方式。

(3) 基础地理数据: 基础地理数据主要是为数据检索提供背景资料和区域性地质信息,这部分数据包括矢量数据,遥感影像、高程数据。对于矢量数据依然是采用了建立要素数据集的方式将所有的矢量数据存放到一个空间系统中。深部探测的数据大多都集中在中国地学大断面附近,为了便于数据分析,实现沿地学断面的查询,建立了中国地学大断面的空间要素地理图层。在地学断面的数据中不仅有矢量图层还有用于显示断面地质信息的条带图,这部分数据是不具有空间坐标的，为了便于数据查询和背景资料的显示,对这些条带图主要采取了两种处理方式,一是对条带图栅格数据进行几何配准(配准后的图层在基础地理数据背景下显示如图 6 所示)，二是在地学断面空间要素的基本信息表中增加了‘条带图’的属性字段,并采用了和图片库相同的方法,即以‘Image’类型的方式,将 cdr 图片直接利用大字段存储,进行数据下载需编写相应的代码来访问大字段里的数据。ArcSDE 9.0 对栅格数据提供了两种管理方式,一个是栅格数据集,一个是栅格编目。在组织这部分数据时笔者采用了以下方式: 一是将多幅高程数据镶嵌后直接存放在一个栅格数据集中，因为在栅格数据集中,可在物理上实现数据的无缝连接,有利于空间分析的实现; 另外将多幅遥感影像存储在栅格编目中,因为在栅格编目中各数据物理上是独立的,而且可以具有不同的属性信息,易于更新。

3.2　深部探测数据模型实现

经过以上设计,基本确定了以 Geodatabase 为基础的深部探测数据的建模方案。Geodatabase 模型的实现方法有很多种,ArcGIS 中的 ArcCatalog 本身就提供了数据库设计的工具。但是这种方法不便于模型的修改、更新和重复利用。ArcGIS 还支持工业标准的统一建模语言和 CASE 工具。在数据库的建设过程中,由于各专业的研究人员已经建立了数据库,

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

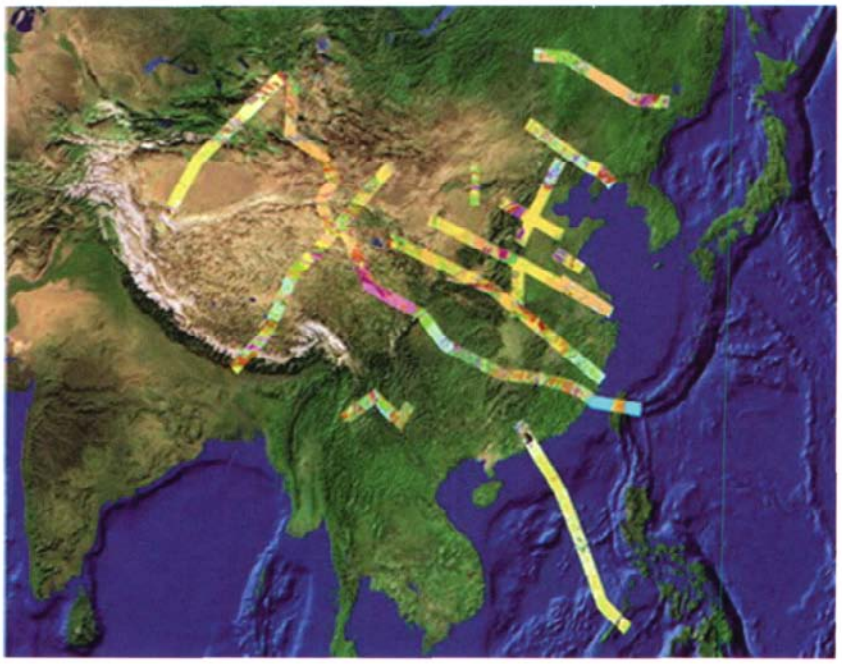

图 6 地图配准后的地学断面条带图

Fig.6 Strips of geoscience transect after map matching

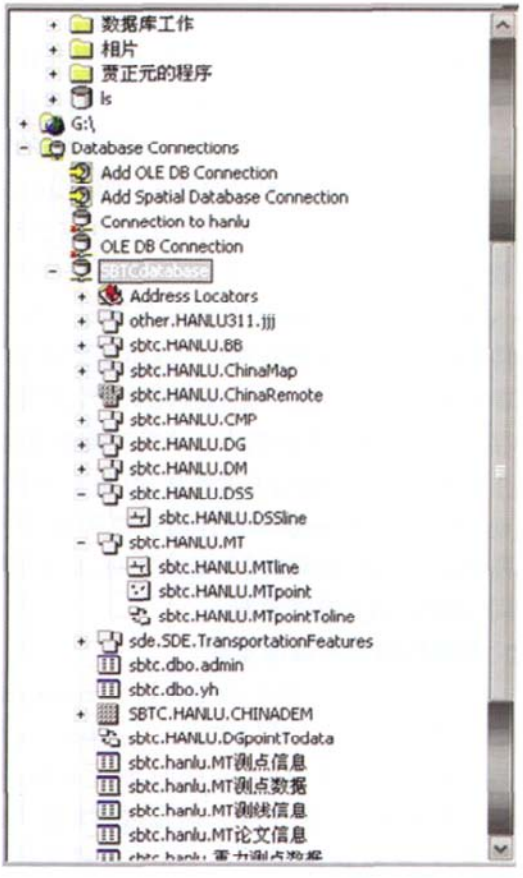

图 7 ArcCatalog 中管理深部探测数据

Fig.7 Management of deep geophysical survey data in ArcCatalog

笔者按照图 5 的结构，对已完成的数据进行了整合，将整合后的数据库直接在 ArcCatalog 中进行管理，这些属性数据在 Geodatabase 中可看成是对象类。而对于空间数据部分笔者利用了 Microsoft Visio 完成了建摸的工作[7]，它可以将建好的 Geodatabase 模型输出到一个 XMI 文件中，然后利用 ArcCatalog 提供 Case 子系统从 XMI 文件中生成 Geodatabase 模型的实例。但是对于关系类，在 ArcCatalog 中无法直接加载数据，因此，在 Microsoft Visio 中没有建立关系类，在完成数据的加载后再将关系类的模型信息重新导入到 Geodatabase 中。图 7 为在 ArcCatalog 中管理深部探测数据模型的情况。

4 结 论

经过对几种空间数据存储方式和地理数据模型，特别是 Geodatabase 的研究，笔者基本完成了岩石圈深部探测数据模型的建设。通过实际的应用可以看出，基于 Geodatabase 的深部探测数据模型简化了地理数据的建模工作，同时具有以下优势：

(1)将深部探测的空间数据直接存储在数据库中，采用面向对象的模型，打破了关系数据库范式的种种限制，直接支持对象的嵌套和变长记录。

(2)通过建立对象类、要素类、关系类、栅格数据

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

集等将所有的数据信息有效地融合在一起，有利于综合多种深部探测方法对复杂地质对象的研究，便于数据提取。

(3)能够定义对象和特征之间的常规和任意的关系，充分利用已有的属性数据资源，并实现属性和空间数据的有效融合。

(4)通过 Geodatabase 的版本化管理可以有效地解决网络并发操作的问题，允许多用户对同一要素的编辑、查询。

(5)通过灵活的栅格管理方式可以实现海量栅格数据存储和局部数据的快速更新。

岩石圈深部探测数据是典型的多源空间数据，它们的量纲不一、形式多样，基于 Geodatabase 的深部探测数据模型以地理对象为基础，将各种专业数据有效融合。它为建立以 GIS 为基础的数据共享平台，挖掘和利用其已有的数据资源进行深部构造研究等工作奠定了基础。

参考文献(References):

[1] 李廷栋. 中国岩石圈构造单元[J]. 中国地质, 2006, 33(4): 700- 710.
Li Tingdong. Lithospheric tectonic units of China [J]. Geology in China, 2006, 33(4): 700- 710(in Chinese with English abstract).

[2] 吴功建. 岩石圈研究的重要问题与研究方向[C]//高锐主编. 岩石圈论文集. 北京: 地质出版社, 2000.
Wu Gongjian. Import question and research aspect about lithosphere [C]// Gao Rui (ed.). Lithosphere. Beijing: Geological Publishing House,2000.

[3] 李廷栋. 中国岩石圈三维结构专项研究主要进展和成果 [J]. 中国地质, 2006, 33(4): 689- 699.
Li Tingdong. Main advances and achievements in the special - subject study of the 3D lithospheric structure of China [J]. Geology in China, 2006, 33(4): 689- 699(in Chinese with English abstract).

[4] 边馥苓. 地理信息系统原理和方法[M]. 北京: 测绘出版社, 1996.
Bian Fuling. Geographic Information System Principle and Method [M].Beijing: Mapping Publishing House, 1996.

[5] Michael Zeiler. Modeling Our World: The Esri Guideto Geodatabase Design[M]. California, USA: ESRI Press, 1999.

[6] 东凯, 方裕. 空间数据模型概念与结构研究 [J]. 地理信息世界, 2004, 2(2): 8- 16.
Dong Kai, Fang Yu. Concept of spatial database model and its arcstructure research [J]. Geomatics World, 2004, 2 (2): 8- 16(in Chinese with English abstract).

[7] ESRI, Geodatabase and Object Model Design Using CASE Tools, [EB/OL]http: //www.esri.com/devsupport/devconn/sde/resentations/uc2000/608.pdf, 2001.

Spatial data storage and its application in deep geophysical survey data

HAN Lu, GUAN Ye, GAO Rui

(Institute of Geology, Chinese Academy of Geological Sciences, Beijing 100037, China)

Abstract: Deep geophysical survey data are important information on the study of the lithospheric structure. However, data sharing was hindered by the characters of geoscience data such as multi- source, multidisciplinary and massive data. GIS can afford a new approach to data release because of its function of multiple discipline crossing and spatial data processing. The presentation and analytic methods of information in GIS depend on the data model; therefore it is important to construct a data model according with the need on data query and data analysis when the sharing platform was designed using the GIS technique. Several spatial data storage methods and main spatial data models are studied intensively. Finally, a data model is constructed for data release based on Geodatabase according to the characteristics of the deep geophysical survey data.

Key words: deep geophysical survey data; data model; geodatabase

---

About the first author: HAN Lu, emale, born in 1979, doctor, mainly engages in GIS and geophysical studies; E- mail: hanlu311@163.com.

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net