

基于 GIS 技术的区域性 多源地学空间信息集成若干问题探讨

李超岭,张克信

(中国地质大学地球科学学院,湖北武汉 430074)

摘要: 地学信息分散的、非动态的管理现状,已远远不能满足市场经济发展与广泛社会需求的多元性、科学性与迫切性,也极大地制约了地学信息的社会经济效益的充分发挥。本文根据当前有关区域性多源地学空间信息集成项目研究的成果,基于 GIS 理论与技术和多源地学地理信息空间数据应用的特点,对区域性多源地学空间信息集成的有关技术方法进行了讨论。详细叙述了多源地学地理信息系统的编程与设计、数据采集技术、元数据、投影与坐标系、物理设计与约定、集成数据质量特征。

关键词: 多源地学空间信息系统;图层;数据集成;元数据。

中图分类号: TP311 **文献标识码:** A

文章编号: 1000-2383(2001)05-0545-06

作者简介: 李超岭(1957—),男,高级工程师,中国地质大学(武汉)计算机技术应用硕士,现在读博士,现主要从事 3S 与 4D 技术在地学中的应用研究工作。

地学信息分散的、非动态的管理现状,已远远不能满足市场经济发展与广泛社会需求的多元性、科学性与迫切性,也极大地制约了地学信息的社会经济效益的充分发挥。为改变这一状态,研究与建立区域性综合地学地理信息系统,不仅使多源地学资料纳入现代化计算机技术的信息系统,而且更重要的是将多源地学信息纳入科学有效的集成管理与分析应用中,为国民经济发展建设实时提供最新的、高精度的、高水平的综合地学信息依据。

区域性地学空间信息是指与地球及其表面在某一区域(如地质省)发生、发展变化及人类研究和利用其自然资源的有关信息,具有区域性、多维性和时序性,它是连接各种信息,形成一个在空间和时间上连续分布的综合信息的基础。它涉及到地理地貌、区域地质调查、地质矿产、地球化学、地球物理、大地构造、海洋、气候、环境等多源信息。基于 GIS 技术,对区域的历史与现实多源地学空间信息的空间数据进行采集、存储、检索、分析、建模和表示,为地质、地理、地球物理、地球化学等多源地学信息进行综合分析和解释提供了有力工具。多源地学空间信息的集

成不仅为公众和地学工作者提供了区域性地学空间的历史 4D 产品,而且通过 GIS 技术提供各种空间分析的方法,对多种不同的空间信息进行综合分析解释,解决空间实体的相互关系,又能提供实时性新的专题的 4D 产品。作者研究发现区域性多源地学空间信息集成技术方法,直接影响着 GIS 应用分析结果的可靠性及应用目标的真正实现。换句话说,区域性多源地学空间信息集成的数据质量在极大程度上决定了应用 GIS 进行决策的成效。本文试图基于 GIS 理论与技术,从多源地学地理信息空间数据应用的特点着手,对区域性多源地学空间信息的集成技术方法进行讨论。

1 图层划分

简单地把电子地图的数据模型代替用于空间分析的 GIS 数据模型,其结果将导致空间分析的局限性和不准确性,大大降低了数据的使用价值;对于复杂的、具有不确定性结构的环境分析、预测和过程模拟能力非常有限。从目前应用情况来看,地理信息系统的应用大致分为两大类:一类以多源信息的集成管理为主,另一类以多源信息的分析为主。数据是地

理信息系统应用的基础,因此,科学有效的图层划分不仅是以多源信息分析为主的地理信息系统应用的前提,其本身也是地理信息系统应用的重要组成部分。

图层划分是地学数据模型的基础,是为地学数据及其数据之间的关系提供定义,通过一种通用语言和一个地学信息的公共结构为地学数据理解提供一个公共的基础,以便用来描述所采集的数据,用于地学计算机系统的开发,定义数据库和定义数据转换。面向 GIS 地学数据模型是由地学要素实体和由点线面 3 个实体进行综合考虑地学空间数据模型,可以用图层和相应的属性描述,图层是用以区分空间实体的类别,即是属性一致的地理实体或特征相同的地理因子在空间分布的集合,目的是为了制图、显示、管理和分析。

目前,作者发现在许多图层划分方案中,把原来具有空间关系的地质实体根据简单制图要求进行图层划分,这一方面加大了存储量,另一个方面,把具有空间关系的信息破坏了,给今后资源分析带来困难,甚至无法建模。因此,在图层划分中,必须注意以下因素:

(1)多义性。一条断层既可以出现在断层的图层中,也可以作为地质界线出现在地质界线的图层上。在建立多源地学 GIS 数据集时,要保证不同图层的相同弧段具有相同的地理坐标,是不容易但必须要做到的。在多源地学不同图层中,为了确保有效的模拟和分析,不允许使用自动编辑工具,而应对其他图层中的复制弧段进行编辑;多义性解决方法,除了采用相同的弧段复制到派生的图层解决之外,在图层划分中比较好的方法就是采用主要图层,然后可通过空间运算产生派生图层的方法。

(2)GIS 的数据模型特点是单独空间数据与属性数据。空间数据包含几何数据及拓扑数据。几何数据即空间坐标、高度、面积;拓扑数据即空间特征的几何关系。由点线面来描述基本空间特征的三种数据模型,通常也是多源地学信息系统的基本数据模型。但这种数据对于空间分析和建模能力是相当有限的。建立多源地学信息系统的目的如果局限于自动制图的表示方法,那么是无法满足解决复杂地学问题的需要。因此,还需要用一些高级空间特征来建立新的地学数据模型,这些数据模型非常适合有关地学在网络分析和空间扩散方面的研究。它们是区划或区域(region):定义为—组相互不重复的多边

形,用于描述具有相同属性单元的不连续多边形。如同一岩石地层单位,同一构造单元等。事件(event)定义为基于基本线特征基础上离起点或终点一定距离的一点。有三种文件类型:线性文件、连续事件和点事件,如可用连续事件描述地质剖面或钻孔线上的地质体,用点事件来描述沿某一水系某一位置上发现水系沉积物某一元素异常;用线性文件来描述某一断层与有关有利成矿的地层单位及相关地质体等。路径(route)定义为基本线性特征基础上的路由,如对某一河段和控制断面的不同污染具有不同的浓度分布。

上述数据模型是基于空间特征进行描述的,因此在图层划分、属性表建立及数据采集时都必需满足这种空间特征的特点^[1]。

(3)充分考虑空间数据物理存储结构,不允许在图层及属性表的命名上出现同物异名、同名异物等问题。

(4)图层及属性表描述结构^[2]。图层描述结构:空间信息名(类 class, 图层 theme, 层文件名 coverage, 空间属性 attribute, 外挂属性库 associated look-up table, 注释 annotation coverage, 备注)。图层属性描述结构:图层名(序号, 字段名, 字段类型, 长度, 小数位, 备注)。

(5)区域性地学空间信息的集成除了用于空间分析外,还有一个目的就是产生新的专题图件,对数字地图而言,它非常注重数字化线条的颜色、分类、线型及成图质量,而不考虑空间数据之间的关系以及数据的后继处理,这一点与 GIS 的空间数据有着非常本质的区别。因此,地学空间信息的集成不可能有一个既能满足 GIS 的空间数据要求又能满足制图要求的的图层划分原则,解决的办法就是针对制图要求而增设注释图层或用于图示图例库的标准代码^[2,3]。

2 多源信息编码与设计

应该说多源信息的编码属于数据采集标准范畴,但这里强调的编码着重于面向图示多源信息的编码。除了标准化这一基本要求外,还在于不去重新编码就能适用于不同的 GIS 平台。特别是支持图示的各种库,如符号库、线型库、色标库的编码。由于不同 GIS 平台数据模型和结构不同,因此在数据模型设计时,要给予充分的考虑。多源信息的集成涉及多

种专业数据,在往常系统研究中各个专业均有相应的项目研究,如色标、图式、图例库,编码也各自成体系,这在多源信息集成和空间叠加显示中将会出现严重问题。这是因为有的GIS平台所带的库都是唯一的,而不像著名GIS平台ARC/INFO,一个图层可以带相应库,所以要解决这一问题,必需认真研究编码体系。

ARC/INFO UNIX/NT版是功能最为强大的专业GIS产品,其许多先进设计思想和概念被其他产品借鉴和采纳。在ARC/INFO的每一个coverage中,可以选用自己的符号库、色标库等。这些库的编码只能从0~999编号,也就是说把多源信息组合成一个无缝专题可以有多个相应的库来支持显示和输出,而在其他一些GIS平台,则是一种类型只用一个库来支持,而相应的编码自然超过999,因此同一数据在不同平台的应用,需要花费大量时间进行编码转化,特别是由于编码原则不一样还可能造成混乱和随意性,对标准化极为不利。另一方面许多提供的数字数据集没有足够的文档即元数据,也可能给编码的转换造成困难,使宝贵的资源数据的共享受到了很大的约束,针对这一情况,在多源信息的编码中,要充分考虑到不同平台的特点,特别是要考虑专业GIS产品设计思想。在编码过程中,应根据多源性和图示的统一性的特点来进行设计。

根据多源信息的特点,首先参照国标将图层分类,将编码纳入图示的系统之内,然后编子类。目的是不能让某一类的编码值超过999。这些编码将作为每一图层的属性输入。不同的GIS平台可事先根据编码建立相应的库,这样就可以在不同的GIS平台中直接应用。

3 数据采集技术

我们把要素(图元)之间的连通性或相邻的关系称为拓扑关系^[4]。空间实体之间的拓扑关系,点、线、面3类实体的关系可以有6种组合,即点与点,点与线,点与面,线与线,线与面,面与面,这些组合表示了实体间拓扑关系的相邻性、连接性、闭合性、一致性等,是地理空间数据处理的依据。如果仅仅从制图的角度要求进行数据采集,不可能有健全的拓扑关系,那么GIS分析功能即使局限于对图形特征相互之间的拓扑关系分析、空间叠加、建立缓冲区的简单方法也无能为力,更不用说对于复杂的环境分析、决

策分析、过程模拟的分析能力。

在利用计算机软硬件资源进行数字地图的绘制时(如应用CAD),非常注重数字化线条的颜色、分类、线型及成图质量,而不考虑空间数据之间的关系以及数据的后继处理,这一点与GIS的空间数据有着非常本质的区别。GIS中空间实体点、线、面之间的6种组合表达了实体间拓扑关系的相邻性、连续性、闭合性、包含性、一致性等关系,是地理空间数据处理分析的依据,因此空间实体间关系的采集是GIS的基础。真实再现多源地学数据特征,同类实体和不同类实体间必须正确客观地反映地理实体间空间结构关系,且相关位置正确,包括:(1)空间实体的点线面类型定义必须正确。(2)保证多边形空间实体的封闭性。(3)保证线状空间实体的连接性。(4)保证组合实体与基础图形要素之间的正确相关性,比如组合实体“侵入岩年代单位”由岩体名称、岩石类型、岩相分带、统一编号、图元编号等基本要素组成。(5)保证编码不重复(组合实体、基础图形要素)。(6)保证符合组合实体间和基础图形要素之间的关系原则,其中包括:连接、相交、共享、落于、包含。(7)保证不出现悬挂节点和伪节点,所有线状要素相交处,都应建立节点。(8)每个线状要素必须既有从左侧相关于它的面域,又有从右侧相关于它的面域。保证空间数据之间拓扑关系的正确性是实现GIS系统数据处理和分析的基础,它的质量特性是GIS区别于CAD系统的重要内容之一。

在注重空间实体间关系的数据采集的基础上,还需考虑为满足解决复杂地学问题的需要而进行采集技术处理。

例如,采用传统的弧节点的数据模型,在数据采集中不考虑采集的方向,那么这种数据模型对模拟具有静态特征的线性特征对动态线性特征都是无能为力,这是因为河段中不同分段的污染物浓度是变化的,而且水流方向也是控制污染变化的一个因素,要解决此办法,就要求采用动态分段数据模型。该模型能够将多个属性与某一线性特征的任何部分相关联,实现对河流等线性特征及其相关属性的存贮、显示、查询和分析。要使该数据模型能够在空间分析得到正确的结果,则数据采集方法就显得格外重要,要求在数字化过程中顺着河流的方向进行采集,在每一个汇合点需有一个节点,这样才能保证空间分析的正确。在多源地学信息研究中,通过属性与把代表地质界线 and 线性特征的线,如断层、单个地质多边形

相联系以满足综合分析. 要实现这一目标, 基于制图的数据采集方法是不可能完成的. 要充分考虑地质界线不仅包括传统的推断的位置或大致的分级, 而且这些界线如果有渐变的、突变的和不整合的特征等也能表现出来, 另一方面, 在数据采集时, 还需考虑对地质界线(弧)方向进行标记以表示地质界线的哪一侧是新的或者老的单元? 对于断层来说, 通过弧段的方向表示断层哪一侧是上升的或下降的? 基于这个原因在数据采集中, 对代表断层的弧段不采用“不分割”操作, 并且代表某特殊断层的一组弧段具有能唯一识别该断层的共同属性最好, 这些要求在于面向 GIS 空间分析的数据采集是明显不同于制图数字化过程的, 应充分考虑所采集的数据应满足空间分析数据模型的要求^[2].

4 元数据

以多源信息分析为主的地理信息系统应用涉及到多种信息的综合, 因此, 除数据采集外, 不可避免地要利用几十年以及现在仅为数字制图或电子地图而花费数不清的人力、物力和资金积累起来的数据. 而这些数据是由不同单位为了不同目的、按不同的格式采集. 由于比例尺不同, 投影方式不同, 数据采集的精度不同, 使得利用现存信息变得十分复杂, 解决这些问题已成为 GIS 应用的重要内容. 上述问题在国外引起了广泛的重视, 进而提出了元数据的概念. 元数据在 GIS 数据模型采集与集成中有着重要的作用, 是数据交换的基础, 也是数据质量的保证.

元数据(metadata)是描述数据的数据^[5], 包括数据的内容、质量、条件和其他特征. 元数据可以分为两大类: 一类是管理元数据, 它是对源数据及内容、主题、数据转换及各种操作信息的描述, 另一类是用户元数据(user metadata), 它帮助用户查询信息、理解信息、了解这些数据组织方式等. 除了图层元数据描述, 对于表中的每个属性项还应有类似的元数据描述. 元数据就相当于数据库系统中的数据字典, 但由于地理信息系统与数据库存在很大区别, 因此, 元数据在 GIS 应用中是数据交换的基础, 在 GIS 数据模型的建立中起着重要作用. 也是保证数据长期有效的基础.

在矿产资源评价、多源地学信息的管理、区域地质调查、水资源调查、环境与灾害的评价等方面, 数据资源要在 GIS 应用中发挥最大作用, 确实需要一

种元数据标准来帮助用户进行元数据信息交换, 元数据标准可以保证共享数据的一致性, 而数据共享领域的某些趋势说明 GIS 中的元数据标准需求正逐步增加. 数据模型处理的对象由传统的字符型和数值型扩展到多种媒体类型, 专用数据扩展到支持多个部门或领域共享的数据, 各个专业的数据格式向公共开放标准元数据交换格式转换, 以提高信息的流通能力, 扩大市场份额. 多源地学信息系统研究和发展需提供开放的元数据标准, 即交换标准和元数据标准. 随着地理数据的数量、复杂性和差异性的增加, 提供一种能够解释数据详细信息的元数据显得非常重要. 为了确保数据不被误用, 影响数据收集的粒度和限制, 必须彻底地进行文件化. 元数据要求全面地描述数据集, 以便为用户了解那些粒度和限制, 并评价他们准备使用数据集的适应性. 建立该标准的目的是为多源地学数据集的描述提供一个清晰的过程, 这样用户就能决定他们拥有的数据是否对他们有用和怎样获取数据, 同时也为今后上网提供基础. 区域性地理信息元数据由以图幅为单位、相对各图层、相对属性的元文件等组成.

下面是一个以图幅为单位进行描述的元数据结构例子:

元文件名称(逻辑名 C10, 定义 C20, 物理名 C8, 数据格式 C12, 图幅名 C12, 图幅代码 C8, 左下角经度 C7, 左下角纬度 C7, 右上角经度 C7, 右上角纬度 C7, 比例尺 C8, 高程系统 C10, 投影名称 C20, 投影参数 C20, 数据单位 C8, 数据精度(\pm m)C10, 相关属性可信度 C10, 应用销售限制 C10, 编制资料单位 C20, 编制出版时间 C10, 数字化数据单位 C20, 数字化方法 C10, 最新更新内容 C20, 更新时间 C10, 联系人 C12, 专家联系人 C6, 该数据存放位置及介质 C6).

5 投影和坐标系统

我国 1:2.5 万至 1:50 万地形图, 规定采用 6 度分带的高斯投影. 该投影中央经线和赤道被投影为相互垂直的直线, 且为投影的对称轴; 投影属等角性质; 中央经线投影后保持长度不变. 投影中长度变形最大为 0.138%, 面积变形最大为 0.276%. 为避免由于高斯分带投影所造成的相邻的投影带间的裂缝, 区域性多源地学空间矢量数据的集成, 应采用经纬度投影的方法, 以便建立区域性或全国范围内统

一的坐标系统,实现各图幅间矢量数据的无缝拼接,并可根据需要进行各投影系统和坐标系统间的转换.在区域性多源地学空间矢量数据的集成过程中,要特别注意 TIC 点精度,数据精度为浮点型的 GIS 应用系统,选用经纬度投影的方法,其 TIC 点精度是达不到精度要求的.这对今后的投影系统和坐标系统间的转换会带来不利的影 响.解决的办法就是选用数据精度为双精度型的 GIS 应用系统.

6 物理设计及约定

区域性多源地学地理信息系统的特点就是以图幅为单位管理多源信息,包括地理、地质、物化探、水文、矿产等,根据 GIS 数据技术的特点,把多源的类型、比例尺、图幅名称等识别码冠入到图层名及属性文件名是非常不利于管理和标准化.如不同图幅的相同图层采用不同名称这是同物异名,不利于交流和共享.解决的办法就是采用物理设计的办法,通过不同的存贮路径及相应的路径名称加上统一的图层命名规则是整个系统标准化建设的重要基础.如目录物理意义可分为:一级目录:图幅代码,二级目录名:地质分类名数据,三级目录:地质分类数据及图层名,4 级目录:图层数据.

例如:TOPOGRA

|_____M50000
|_____JA006008

其意义为:地理底图类数据,比例尺是 1/50000,图幅号 006008,J 行 50 列.

7 数据质量特征

根据中国地质调查局 1:20 万地质图空间数据库建库工作质量控制要求和我们近几年工作的经验,结合其他学者提出有关质量特征^[6],将区域性多源地学空间信息集成数据质量特征概括如下几点:(1)位置精度,空间实体的位置通常以三维或二维坐标表示,而位置精度则是表示实体的坐标数据与真实数据的接近程度,因而常以坐标数据的精度来表示.位置精度包括数学基础精度、平面精度、高程精度、接边精度、形状再现精度等.(2)属性精度,是指

实体的属性值与其真实值相符的程度,它通常取决于数据的类型,且常与位置精度有关.属性精度包括要素分类与代码的正确性,要素属性值的正确性及名称的正确性.(3)逻辑一致性,指数据之间关系的可靠性.包括数据结构、数据内容、空间属性与专题属性,以及拓扑性质上的内在一致性.(4)数据完整性,包括数据范围、数据的分层、实体类型、属性数据和名称等各方面数据的完整性.(5)时间精度,主要是指数据的现势性.一般通过数据采集时间、数据更新时间及更新频度来表现.(6)元数据,指数据说明的准确性和全面性,要求对数据的来源、数据内容及数据处理过程等有准确、全面和详尽的说明.

8 结语

区域性地学空间数据集成通过一种共同的语言和结构定义地学数据及数据之间的关系,不仅为社会提供理解地学信息的公共基础,而且能为各种空间分析提供准确的的空间数据.要使区域性地学空间数据集成能够满足复杂的、具有不确定性结构的环境分析、预测和过程模拟能力的要求,区域性地学空间数据集成需要研究和解决的问题很多.上面谈到的几个问题,只是作者在全国 1/20 万地质图空间数据库工作指南及质量检验和验收工作过程中和在区域地质调查新技术新方法集成示范项目研究的部分成果的总结.这项工作还在继续进行,还有一些问题需要解决.

参考文献:

- [1] 刘学. 基于 GIS 的空间过程模拟建模方法研究[J]. 中国图象图形学报,1999, (6): 476-480
- [2] Richard B. Australian geological survey organization [J]. GIS Application, 1998, (9): 56-62
- [3] 阎正. 城市地理信息系统标准化指南[M]. 北京:科学出版社,1999.
- [4] 郭仁忠. 空间分析[M]. 武汉:武汉测绘科技大学出版社,1997. 476
- [5] 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.
- [6] 刘大杰. GIS 空间数据的精度分析与质量控制[M]. 上海:上海科学技术文献出版社,1999.

STUDY ON REGIONAL MULTI-SOURCE GEOLOGICAL SPATIAL INFORMATION SYSTEM BASED ON TECHNIQUES OF GIS

Li Chaoling, Zhang Kexin

(Faculty of Earth Sciences, China University of Geosciences, Wuhan 430074, China)

Abstract: The separate and static management status of geology information hasn't been able to meet the market economy's development and variety, scientificity and imminence of widely social demands, also extremely restrict the total exertion of social economy benefits of geology information. According to the achieved results of regional multi-source geology spatial information integration projects, based on GIS theory and technique and the characteristics of multi-source geological information spatial data application, the discussion of technical methods about regional multi-source geology spatial information integration is given, and also narrates in details the programming and design of multi-source geological geographical information system, the capturing techniques of data, metadata, the projection, reference frame, the physical design and stipulation as well as the features of the quality of integration data.

Key words: muti-source geological geographical information system; theme; integrated data; meta-data.

* * * * *

(上接 532 页)

respectively. The FCT algorithm includes two phases: one is low-order transport, in which excessive numerical diffusion may be introduced; the other is high-order flux-corrected, in which numerical dispersion and oscillation can be effectively eliminated through correcting (limiting) antidiffusion flux. Moreover, the moisture content is obtained by FUCG algorithm, which is able to conserve the mass balance. At last, the effectiveness of FCT algorithm is verified by the numerical examples.

Key words: unsaturated soil; convection-dominated; FCT algorithm.