

文章编号: 1007-4619 (2002) 04-0272-07

综合省情地理信息系统空间数据元数据设计研究

梁湖清, 马荣华

(南京大学 城市与资源学系, 江苏 南京 210093)

摘要: 综合省情地理信息系统的建立与运行需要元数据的支持, 在分别对现有元数据标准和综合省情地理信息系统特点进行分析的基础上, 指出了使用元数据的优点, 提出了综合省情地理信息系统空间数据元数据的内容, 以提高系统查询检索的速度和系统分析效率为主要目的, 对元数据的存储采用了一种以数据库为基础的新的存储模式, 以关系表格来管理元数据, 建立元数据库。元数据库的管理以图形数据的管理为基础, 二者必须协调进行。

关键词: 综合省情地理信息系统; 元数据; 关系表格

中图分类号: TP311/P208 **文献标识码:** A

1 引言

综合省情地理信息系统是一个综合反映某个省范围内的基础地理信息(地形地势、河流、湖泊等)、自然资源、经济和社会发展等综合状况的地理信息系统。其目的是为政府高层领导的宏观决策分析提供客观依据并进行辅助决策, 实现灾害预警甚至为灾害预测的实施做准备。该系统涉及到矢量数据、影像数据和 DEM 数据, 数据量是巨大的甚至是海量的, 如何获取、组织和管理这些数据, 加快数据查询浏览速度、方便数据维护和更新等, 是建立综合省情地理信息系统数据库面临的挑战。另外, “数字地球”的提出, 使我们对建设中国的“数字地球”或“数字中国”有了一种紧迫感的责任感, 从而促使我们必须考虑大量数据共享问题, 让用户知道我们存在什么样的数据, 数据的质量如何, 以及怎样访问和使用这些数据成果。综合省情地理信息系统对元数据的需求是必要的、迫切的。当然, 通过元数据的有效组织, 还可以提高系统的运行效率。

2 现有元数据标准

2.1 元数据定义^[1]

元数据是关于数据集的数据, 数据集的说明或

描述。有关元数据的讨论很多, 到目前为止, 对元数据仍没有统一的定义和理解, 观点很多, 大致可分为两类, 一类认为元数据是数据集的描述与说明, 另一类认为元数据是应用系统的辅助信息, 能提高数据的利用价值, 有关详细讨论参见文献[1]。

2.2 现有的元数据标准^[1,2]

目前, 地球空间元数据已形成了一些区域性或部门性的标准(表 1), 但由于人为的和客观的原因, 仍没有一个标准成为全球统一的空间数据元数据标准。有关组织标准草案的对比如表 2。

表 1 目前已使用的主要元数据标准^[2]

Table 1 Current standard for geo-spatial metadata developed at present

元数据标准名称	建立标准的组织
地学空间数据元数据内容标准	美国联邦地球空间数据委员会 (FGDC)
GDDD 数据集描述方法	MEGRIN, 欧洲地图事务组织
CGSB 地球空间数据集描述	加拿大标准委员会
CEN 地学信息—数据描述—元数据	CEN/TC287
DIF	NASA
ISO 地理信息	ISO/TC211

其中由美国联邦地球空间数据委员会 (FGDC)

收稿日期: 2000-12-05; 修订日期: 2001-03-13

作者简介: 梁湖清(1964—), 男, 南京大学城市与资源学系博士生, 主要从事土地信息系统研究。

组织编写并发布的地理空间数据元数据内容标准于

表2 部分元数据标准草案的内容比较^[1]

Table 2 Comparison of the content of some metadata standard drafts

CEN/TC287	FGDC	ISO/TC211	建议体系
数据集标识信息	标识信息	标识信息	标识信息
数据集综述信息	数据质量信息	数据质量信息	数据质量信息
数据集质量信息	空间数据组织信息	数据集继承信息	数据集继承信息
空间参照系信息	空间参照系信息	空间数据表示信息	空间数据表示信息
范围信息	实体和属性信息	空间参照系信息	空间参照系信息
数据定义	发行信息	应用要素分类信息	实体和属性信息
分类信息	元数据参考信息	发行信息	发行信息
元数据参考		元数据参考信息	元数据参考信息
元数据语言			

说明:上表中“建议体系”系原作者提出的内容,此处未作任何修改,以便于比较

1992年7月开始起草,1994年7月正式确认为美国国家地球空间数据元数据标准,1997年4月发布修订版。该标准的内容项采用数字式章节组织结构,共分7个部分(表2),数据要素有219项。该标准的主要缺点有:(1)数据内容多,获得如此多的元数据内容,需要增加15%—20%的工作时间,编写该标准的元数据要大量的时间和资金,并要专业人员才能完成;(2)专业词汇多,没有一定专业背景的用户很难使用;(3)标准不实用,由于专业性太强,从而限制了大学科范围内统一地球空间数据的形成。国际标准化组织第三工作组组织研究的ISO地理信息元数据标准于1996年2月通过1.0版草案,1997-01-20发布2.0版标准。该标准主要以数据集的实时性、精度、数据内容和属性、数据来源、价格、图层以及适用性等为考虑对象,内容采用逐项逐行的方式表达,共分8类(表2),每一类中又包括若干子类或具体元数据项。欧洲地理信息标准化委员会(GEN/TC287)组织编写的CEN地学信息—数据描述—元数据内容标准于1996年推出。该标准以实用为主,涉及地理信息标准化框架、地理信息模型和应用、地理信息传输以及地理信息参考系统等方面的内容。

从以上分析和表2可以看出,尽管各个组织对元数据在内容划分上存在一定的差异,但其内容体系总体上反映出了元数据的下列几个特点:首先元数据是用于描述信息资源的高度结构化数据;其次

元数据可以管理和组织信息,并可以挖掘信息资源,通过它可以在Intranet或Internet上准确地识别、定位和访问信息;另外,元数据还可以帮助准确地查询所需要的信息并能维护和延续一个机构对数据的投资^[3]。

3 综合省情地理信息系统空间数据元数据设计

3.1 综合省情地理信息系统的特点

综合省情地理信息系统属于政府GIS,除具备一般GIS的基本特征外还具有以下特点。

(1)信息综合

政府领导需要的是某一地区的宏观信息,系统必须能够反映人文、社会、经济、资源与环境等各个领域的综合状况,因此系统数据必须在这些领域的基础上进行综合加工,体现出“综合”的特点以满足实际需要。

(2)数据来源广泛

信息综合决定了数据的来源各种各样。首先反映在提供数据的部门不同,例如基础地形数据由测绘部门提供,土地利用现状数据由土地部门提供,社会经济数据由统计部门提供,环境保护(污染)数据由环保部门提供,等等;其次反映在获取数据的时间不同,预测与辅助决策是综合省情地理信息系统的重要功能,实现这一功能需要时间序列数据的支持,数据库中存放数据的时间版本越多,预测越准确,越可靠;再次反映在存放数据的比例尺不同,某一方面、某一专题,领导需要的数据的详细程度不同,因此必须提供从宏观到微观的不同详细程度、不同比例尺的数据;最后反映在数据来源的格式不同,不同部门,制作数据的软件不同,提供的数据的格式也就不同。

(3)要具有按行政单元进行查询检索的功能

政府领导关心的是某一行政单元的综合状况或某些行政单元的比较状况,按行政单元进行查询检索是系统必不可少的一项查询检索功能。

(4)支持社会信息化服务

该系统的部分信息必须面向广大公众服务,通过服务窗口,民众希望了解一些有关数据的描述信息,如数据的来源部门以及数据的可靠程度等。

(5)数据具有很高的安全性

高安全性是系统的又一特点。有些数据,属于机密信息,不能向外泄露;有些数据针对不同的使用

对象,具有不同的保密等级规定。

3.2 建立使用元数据的目的

综合省情地理信息系统的特点决定了建立使用元数据的迫切性和必要性;另外,建立使用元数据的目的还有以下需要。

(1)提高系统查询检索速度的需要

通过元数据提供的数据自身的信息和数据存在的环境信息,迅速查找所需信息的存放位置,发现原始数据的存在状况,并判断需对数据采取的动作(如生成新数据、改动原有数据和个别属性数据的变动等)。

(2)进行数据分析的需要^[4]

综合省情地理信息系统数据分析的每一过程都要有空间元数据的支持。如在叠加分析中,首先系统要获取所分析的多层地球空间数据的范围、坐标体系、空间位置坐标等信息,以便将数据显示到一起;然后获取数据的空间信息,将相应的空间特征合并到一起;接下来获取属性数据的结构信息,把相应的属性合并到一起,最后通过数据的结构信息把分析结果存储到新的数据中。其它的数据分析和处理过程也要用到元数据。

(3)数据质量评价和控制的需要

建立综合省情地理信息系统的主要目的是为省

政府领导进行宏观决策提供科学依据,数据质量如何,直接影响系统的效能以及系统支持决策的可信度,影响到系统的生命力。因此,准确、客观地说明数据的质量是决定系统能否真正投入实际运行的关键之一。可见,该系统的空间数据应满足以下基本条件:有准确定义的数据字典以说明数据的组成,表征的内容等;数据的一致性;有足够的说明数据来源、数据的加工处理过程、数据释译的信息等。这些要求可通过元数据来实现。

(4)数据存储和管理的需要

通过元数据组织数据库,可以有效地降低数据存储的空间,减少数据用户查询数据库以及获取数据的时间,从而降低数据库的费用。

3.3 空间数据集元数据内容

综合省情地理信息系统空间数据是对某个省范围内的地学过程、现象、特征等认知表达的结果。所以元数据首先应包括对数据表达特征的描述,还应包括数据集整体特征的信息。考虑到国际现行元数据内容标准的有关内容,综合省情地理信息系统空间数据集元数据至少应包含质量评价、空间特征、属性特征、时间特征、元数据、数据标识、数据管理、数据形式等内容,组成结构如图 1 所示。

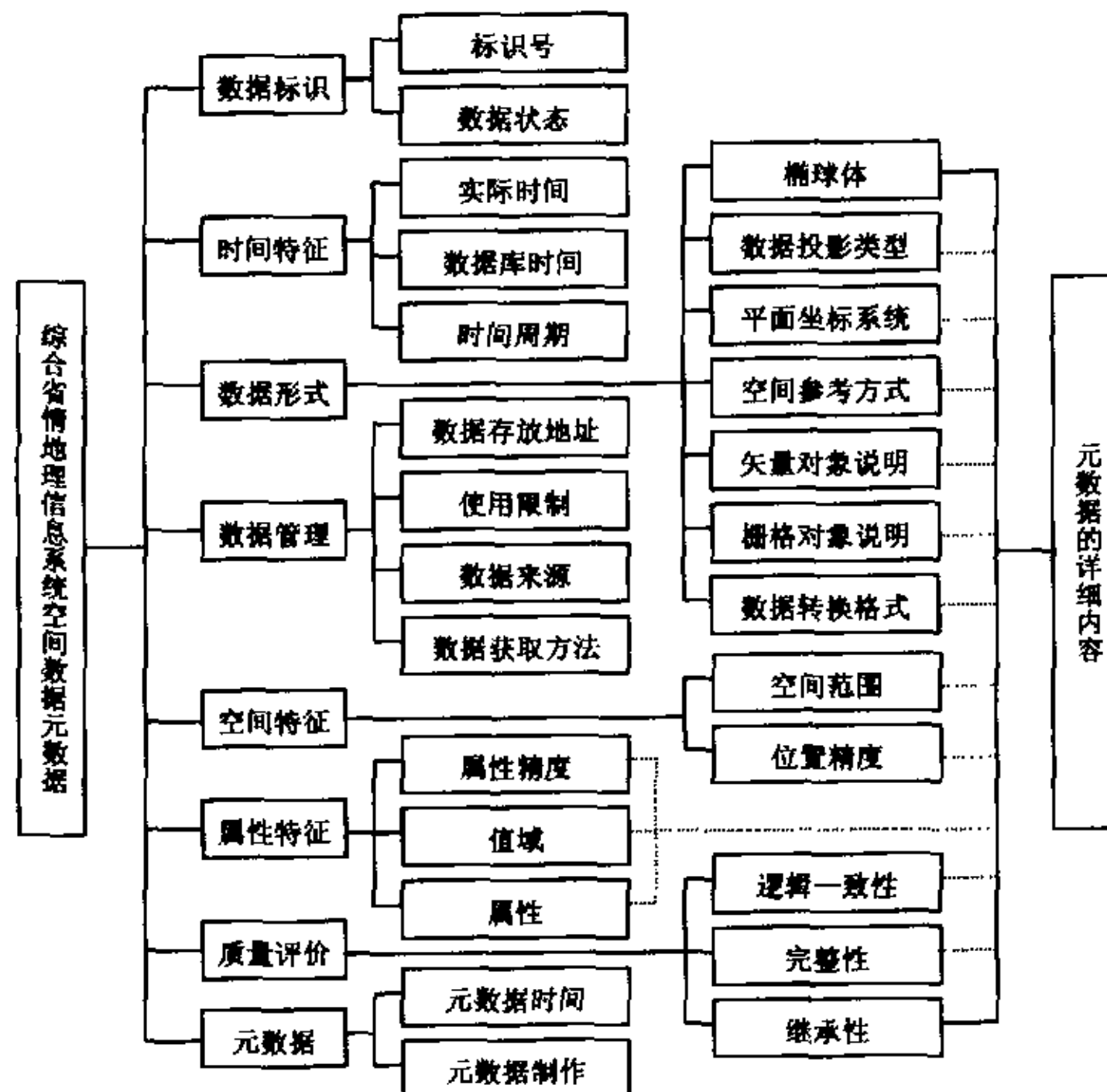


图 1 综合省情地理信息系统空间数据集元数据内容及结构

Fig.1 Structure and contents of metadata for spatial dataset in general provincial situation GIS

从图 1 可以看出,系统元数据的内容存在明显地层次关系,各层次的元数据内容存在着密切的联系,上一层次的元数据可以从下一层次的元数据中通过抽象或概括等操作获得。对综合省情地理信息系统来说,为了提高系统查询检索的速度,对海量数据的有效管理是非常重要的,必须把数据的管理放在突出的位置上,另外,图 1 中不同的元数据内容在系统中所起的作用是不同的,为此把上述元数据分为系统层、数据集层和数据特征层,其中系统层处在最高的层次上,包括数据管理等内容。系统层元数据随系统存在,由(网络)数据库管理系统统一管理,系统在每次启动时先读取这部分内容,进行系统初始化,为即将进行的数据操作做准备;数据集层元数据随数据库存在;数据特征层元数据位于数据集层的内部,只能随数据集存在。

3.4 空间数据集元数据管理

元数据的管理是全面的,不但包括对描述数据的空间特征及相关的属性特征静态元数据管理,还包括对数据进行的各种操作(如数据处理、数据存储、数据更新等)过程中数据变化的动态元数据。管理是在存储的基础上进行的,有序合理的存储利于管理,否则,会给管理带来很大的麻烦。

3.4.1 系统元数据存储模式的选择

元数据的存储有两种模式^[2](图 2):其一是以数据集为基础,即每一个数据集有一个对应的元数据文档,每一个数据文件中包含对相应数据集的元数据内容;另一种是以数据库为基础,每一个地球空间数据库有一个元数据文件,该文件为一表格数据,它由若干项组成,每一项表示元数据的一个要素,其记录为每一个数据集的元数据内容。两种方式的比较如表 3。

表 3 元数据存储的两种模式比较

Table 3 Comparison of two modes of metadata storage

	优点	缺点
第一种模式	相对数据库有较强的独立性,对元数据进行检索时可以利用系统空间数据库的功能实现,也可以将元数据文件调到其它数据库系统中进行操作。	每一个数据集都有一个元数据文件,在规模巨大的数据库中则会有大量的数据文件,管理上极为不便。
第二种模式	每个数据库只有一个元数据文件,管理极为方便,添加或删除数据集只将该文件添加或删除相应的记录项即可。	获取数据集的元数据时,实际得到的只是关系表格数据的一个记录,因此要求数据用户使用的系统可以接受这种特定形式的数据。

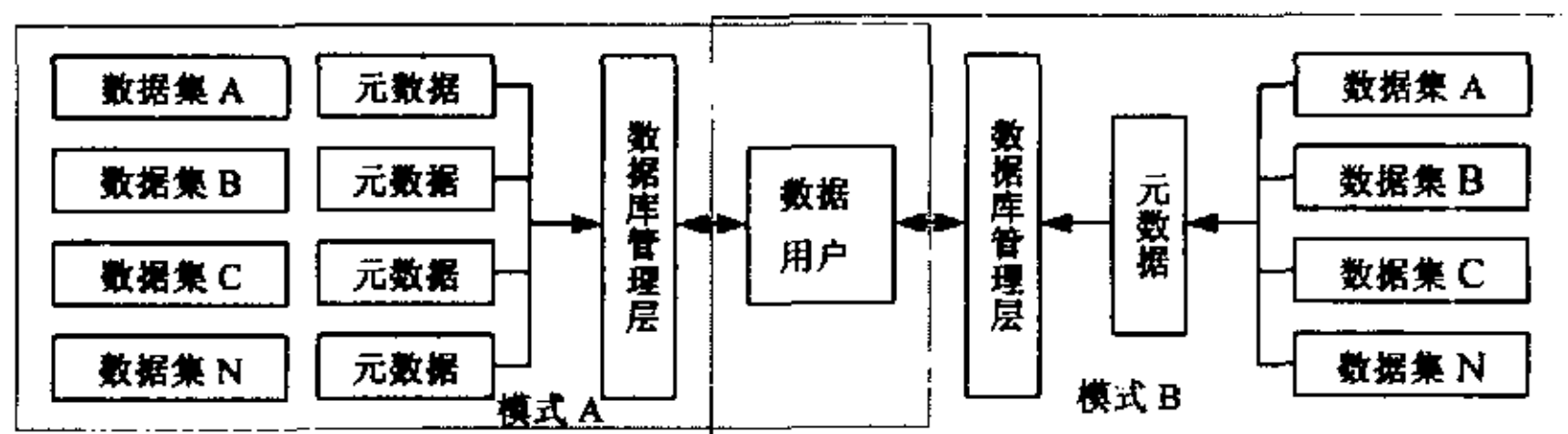


图 2 元数据存储的两种模式^[2]

Fig.2 Two ways to store the metadata

综合省情地理信息系统数据库中的空间数据按行政单元(最大为省级,最小为县级),分别存放,以适应系统经常按行政单元查询检索的需要,数据库的详细数据组织方案,另文探讨研究。根据上述关于元数据存储模式的比较,结合综合省情地理信息系统本身的特点、要求等内容,决定采用关系表格组织管理元数据,并对第二种方式进行改进。

3.4.2 系统元数据的存储

综合省情地理信息系统元数据的存储分两个层次:一层位于系统整个数据库的顶层;另一层位于和每个数据库平行的位置上,内含数据特征层元数据。

(1)在系统整个数据库(矢量数据库、影像数据库和 DEM 数据库)的顶层,建立一张系统数据库管理索引元数据表,叫做顶层元数据表,作为系统数据库顶层元数据存在,这是一张从系统打开到系统关闭始终打开的表,是系统数据库的总管,处于元数据的最高层(图 3,图 3 中 DEM 和影像数据的分层是指同一行政单元不同比例尺的数据)。该数据表的部分元数据项如表 4。另外,在系统数据库的上层再建立一张数据库元数据公共项查找表(图 3),叫做次顶层元数据表,所包含的主要元数据项如表 5。表 4 和表 5 之间通过关键字段——数据库标识号进行连接,它们都位于系统数据库的顶层,每个记录对

应系统的一个独立数据库。表面上看,似乎可以把这两张表合并为一张表;实际上,它们在整个系统中所起的作用是不同的:其一,目的不同。表 4 的主要目的是为了产生、建立整个系统的树状索引目录,查

找数据库,表 5 的主要目的是为了描述每个数据库的公共属性。其二,生存周期不同。表 4 中的内容伴随于系统从打开到关闭的全过程,表 5 的内容是随系统数据库的打开而打开,关闭而关闭。

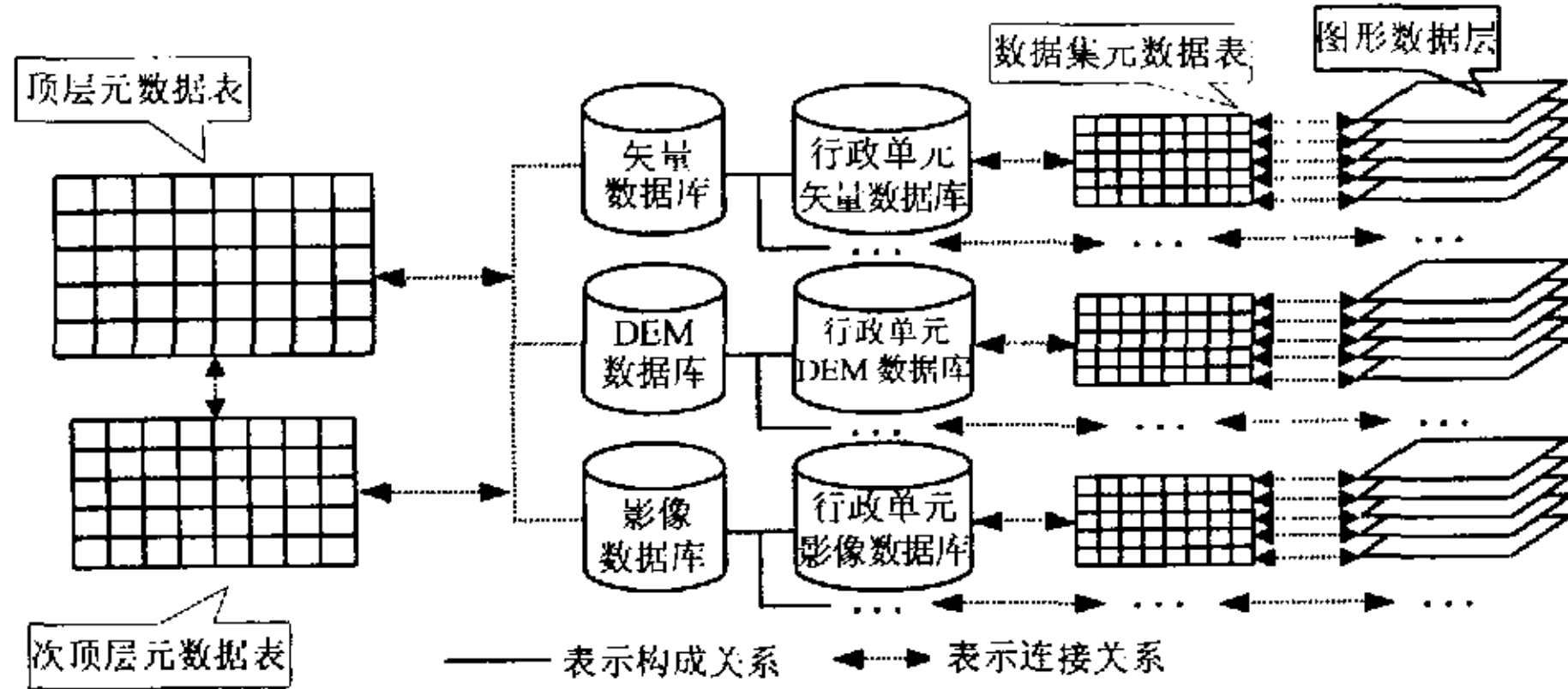


图 3 元数据存储以及元数据与数据库之间的关系描述

Fig.3 How to store metadata and description of relationship between metadata and database

表 4 顶层元数据表的部分元数据项列表

Table 4 Listing of some items in the top metadata table

元数据项	类型说明	描 述
标识 1	字符型	数据库的标识号
名称	字符型	数据库的名称描述(汉字)
文件名	字符型	数据库存放所使用的文件名
存放路径	字符型	数据库的存放路径
覆盖范围	字符型	数据库所在的行政单元名称(汉字)
实际时间	字符型	数据库中的数据所反映的实际时间
数据库时间	字符型	建立数据库的系统时间
标识 2	字符型	上一级行政单元数据库的标识号,如果为空,表示没有上一级行政单元或还没有建立上一级行政单元数据库
标识 3	字符型	下一级行政单元数据库的标识号,如果为空,表示没有下一级行政单元或还没有建立下一级行政单元数据库

表 5 次顶层元数据表的部分元数据项列表

Table 5 Listing of some items in the sub-top metadata table

元数据项	类型说明	描 述
数据库标识	字符型	数据库的标识号
左上角 X	数值型	数据库中图形左上角最大的横坐标
左上角 Y	数值型	数据库中图形左上角最大的纵坐标
右下角 X	数值型	数据库中图形右下角最大的横坐标
右下角 Y	数值型	数据库中图形右下角最大的纵坐标
空间参考方式	字符型	表达空间对象的系统,即以何种方式表达空间特征,内容有矢量、影像和 DEM
数据投影类型	字符型	投影名称说明,如等积圆锥投影,入库时先对所有可能的投影编号
平面坐标系	字符型	用于量测空间点位置的参考体系,如经纬度
数据状态	字符型	数据库完成状态说明,如已完成
使用限制	字符型	使用该数据库的限制情况,内容有:有限制、无限制
数据来源	字符型	数据库中数据获取的简单描述,如提供该数据库数据的个人或单位
数据量	数值型	指定格式下数据需要磁盘空间的大小
垂直分辨率	数值型	对栅格数据而言,单位距离内样点垂直样点的数量
水平分辨率	数值型	对栅格数据而言,单位距离内样点水平样点的数量
数据单位	字符型	数据的度量单位
比例尺	数值型	数据库中数据的比例尺(可能多种)

(2)在和每个数据库平行的位置上,分别建立一个数据集元文件(表6),以关系表格进行元数据项的组织,每个元数据表均随相应的数据库而存在,伴随相应数据库从打开到关闭的全过程。

3.4.3 系统元数据的不同之处

综合省情地理信息系统元数据不同于其它类型的元数据,具体体现在:(1)作用不同(详见3.2);(2)内容不同,本系统元数据除包含一般类型元数据外,重点强调行政单元上下级数据的索引关系(表4);(3)组织方式不同,本系统元数据以关系型数据库存放,两层模式存储即数据库的顶层元数据表(顶层和次顶层元数据表)和与数据库处于同一层次上的元数据表(数据集元数据表),内含数据特征层元数据。

3.4.4 系统元数据的管理

前述可知,系统元数据是按照关系表格的方式组织的,所有关系表格组成一个元数据库,它与系统图形数据库或图形数据层共存。通常意义上说,元数据管理是指元数据通过各种途径形成后,对其内容的添加、删除、更新等涉及内容改变的操作和元数据内容检索、查询、放置、组织等常规性的元数据操作。从这种意义上说,元数据的管理可以通过两种

方式实现^[4],即系统管理模式和用户管理模式。系统管理模式是面向数据库的,由数据库管理系统专业人员完成,数据用户只有使用权,没有操作权,但这种方式在数据处理过程中形成的动态元数据很难及时记录下来;另一种管理方式是用户管理模式,它是面向应用项目的,即允许某些数据用户在数据应用元数据的变动信息直接反馈该元数据库,这样则能保证元数据的动态更新和新生成数据集元数据的及时捕获及写入元数据文件,但这种模式中数据用户的权限要适当控制,以避免破坏数据库。

综合省情地理信息系统的最终用户是省政府有关领导,部分信息面向大众,实行社会化信息服务;因此,该系统的用户不拥有数据的编辑修改权,只有使用权,对数据的编辑、修改以及动态更新等操作均由数据库管理员进行。可见,该系统对元数据的管理宜采用系统管理模式,开发专一的元数据管理模块,嵌入到系统总的编辑模块中,对元数据的查询、检索等操作放到系统图形的浏览模块中,能够利用系统已有的元数据对用户较为关心的问题的问题的决策作出精度分析评价。因此系统元数据的管理必须以图形数据的管理为基础,实现元数据库和图形数据库的共同协调。

表6 数据集元文件表的部分元数据项列表
Table 6 Listing of some items of metafile in the dataset

元数据项	类型说明	描 述
数据集标识	字符型	数据库中每层数据的标识
数据集名称	字符型	数据库中每层数据的中文名称
数据文件名	字符型	数据库中每层数据的存放文件名
时间域说明	字符型	数据内容表示的时间域“起始—终止日期”如1997年3月到1999年3月,表达为“199703—199903”
数据库时间	字符型	该数据集入库的系统时间
数据状态	字符型	数据完成的状态说明,如已完成
使用限制	字符型	使用该层数据的限制情况,内容有:有限制、无限制
数据量	数值型	指定格式下数据需要磁盘空间的大小
垂直分辨率	数值型	对栅格数据而言,单位距离内样点垂直样点的数量
水平分辨率	数值型	对栅格数据而言,单位距离内样点水平样点的数量
比例尺	数值型	该层数据的比例尺
数据单位	字符型	数据的度量单位
数据来源	字符型	记录该层数据来源的文件名(包括路径),具体内容查找文件名指针指向的文件(表7)

表7 数据来源元数据项列表

Table 7 Listing of some metadata items of data sources

元数据项	类型说明	描 述
地图名称	字符型	数据来源的地图名称描述
编辑单位	字符型	原地图的编辑单位名称
出版单位	字符型	原地图的出版单位名称
出版时间	数值型	原地图的出版时间
比例尺	数值型	原地图的比例尺描述
投影方式	字符型	原地图的投影方式描述

4 结论与讨论

地理信息系统的应用越来越广泛,地球空间的数字化信息量越来越大,不同部门、不同组织为了实现某一特定的目的,可能需要同一种、同一类甚至同

·地区的数据,由于互不了解或对原数据的不信任,重复投资建设同样的数据,这实际是一种极大的浪费。因此,必须建立一种共享机制,实现数据的透明访问,解决问题的办法之一就是建立元数据。“数字地球”、“数字中国”甚至“数字城市”的建设也迫切需要元数据的支持。另外,对一个系统来说,元数据可以起到提高系统的查询检索速度、提高系统分析效率的作用,元数据的这些优点在综合省情地理信息系统中都得到了充分的体现。但我们在元数据的组织建设过程中也发现了一些问题:(1)元数据的建设加大了系统建设的投资,因而必须加强元数据自动获取软件的研究以降低元数据的生产费用;(2)目前为止,仍没有一个非常适合的数据标准可供使用,尤其中国元数据还没有完全形成一个统一的标准,客观上影响了数据共享的程度。

参 考 文 献 (References)

- [1] Cheng J C, Li Q, Yi S Z. National Spatial Data Infrastructure and Digital Earth [M]. Beijing: Tsinghua University Press, 1999, [承继成, 李琦, 易善楨. 国家空间信息基础设施与数字地球. 北京: 清华大学出版社, 1999: 146—166.]
- [2] Li J, Zhou C H. Overview on Metadata Standards of Geo-spatial Data [J]. *Progress in Geography*, 1998, 17(4): 55—63. [李军, 周成虎. 地球空间数据元数据标准初探 [J]. 地理科学进展, 1998, 17(4): 55—63.]
- [3] Li Q, Chen A J, Xia S D. Research and Design of Geo-spatial Metadata Management System on WebGIS [J]. *Journal of Image and Graphics*, 2000, 5A(10): 811—817. [李琦, 陈爱军, 夏曙东. WebGIS 中地理空间 Metadata 管理系统研究与设计. 中国图形图象学报, 2000, 5A(10): 811—817.]
- [4] Li J. Studies on Geo-spatial Metadata [J]. *Geo-information Science*, 2000(3): 8—13. [李军. 地球空间元数据的使用研究. 地球信息科学, 2000(3): 8—13.]

Research and Design of Metadata for Geo-spatial Data in General Provincial Situation GIS

LIANG Hu-qing, MA Rong-hua

(Dept. of Urban & Resources Science, Nanjing University, Nanjing, 210093, China)

Abstract: Metadata is a kind of data coming from some data-collection, and it is a description or direction of the data-collection. The advantages to utilize the metadata are as follows: (1) the speed of selecting and indexing from large volume of data in the geo-spatial database can be improved; (2) it will promote the sharing of geo-spatial information and propel the progress of Chinese standard; (3) it will be advantageous to evaluate the geo-spatial data and control their quality; and (4) taking advantage of the metadata may be helpful to store and manage the geo-spatial information. The General Provincial Situation GIS (GPSGIS) is a kind of GIS reflecting the comprehensive and general situation in some province, and the content contains: (1) the general geographic information (such as rivers, lakes, topology and so on); (2) natural resources information; (3) economic information; (4) social development information; and (5) others. So, the development and application of GPSGIS should be supported by metadata for geo-spatial data. On the basis of analyzing of the present standards for metadata, the contents of metadata in GPSGIS are put forward. The metadata is stored by a new method that is based on the relation database. Certainly the metadata is organized and managed by related tables to build the metadata database. The database is composed of some metadata tables as given below: (1) the top metadata table, located in the most top level of the all geo-spatial databases; (2) the sub-top metadata table, located in the lower level than the top one, so it is less important than the top metadata; and (3) the data-collection metadata tables, located in the equal rank to the one geo-spatial database. It is of importance that the management of the metadata databases is based on the management of the topologic data, but both of them must be coordinated.

Key words: general provincial situation GIS; metadata; related table