

现代地理空间数据集成的设计与实现^{*}

赵健赞¹, 张治国²

(1. 青海大学 地质工程系, 青海 西宁 810016; 2. 西北勘测设计研究院 工程勘察研究分院测绘工程大队, 甘肃 兰州 730050)

摘要:在分析 GIS 和 CAD 数据模型的基础上, 探讨了数据集成的方法。提出利用直接转换模式进行图形数据集成, 利用集成数据字典方式解决 CAD 图形数据的转换和编码、语义表达的统一问题, 设计了数据集成结构和具体的集成数据字典, 并结合实例进行了论证。研究表明, 利用该方法不仅可以节约成本, 而且还能提高生产效率。

关键词: CAD; GIS; 数据字典; 集成

中图分类号: P 208 **文献标识码:** A **文章编号:** 1007-9394(2010)01-0006-04

Design and Realization of Modern Geographic Spatial Data Integration

ZHAO Jian-yun¹, ZHANG Zhi-guo²

(1. Dept. of Geologic Engineering, Qinghai University, Xining Qinghai 810016, China; 2. Engineering Brigade of Surveying and Mapping, Dept. of Engineering Investigation & Research, Northwest Investigation, Design & Research Institute, Lanzhou Gansu 730050, China)

Abstract: Based on the expatiation to data model of GIS and CAD, the method of data integration is analyzed. It is put forward that used the direct conversion scheme to integrate the graph data, and used the integrated data dictionary to resolve the unification problem of code and semantic expression in the course of data convert. Designed the data integrated structure and integrated data dictionary, demonstrated the data with the example, It is indicated that the method is economic, effective and valuable to practice and spread.

Key words: CAD; GIS; data dictionary; integration

0 引言

随着计算机和信息技术的发展, 地理信息系统 (Geographic Information System, GIS) 技术由于其强大的数据存储、管理和分析功能, 受到测绘、农业、林业、国土等相关决策管理部门的青睐。考虑到目前国内大多数规划成果和建筑工程空间信息均以 CAD 数据格式存储, 且 AutoCAD 在数据编辑方面更灵活易用。所以, 必须解决 CAD 数据和 GIS 数据的集成和共享问题。

CAD 与 GIS 的集成, 已经有较为广泛的研究和应用。如文献[1]利用 Geoway 实现了 CAD 数据到 GIS 数据的转换; 文献[2]利用 VBA 对 ArcGIS 和 AutoCAD 进行二次开发, 实现了 CAD 宗地数据向 Geodatabase 数据的转换等。纵观目前研究成果, 其核心是利用第三方软件或编写程序以实现集成, 虽在一定程度上解决了问题, 但在购买软件或编写程序中也付出了昂贵的代价, 并且集成之后的数据还存在要素丢失、变形、数据冗余等问题^[3]。

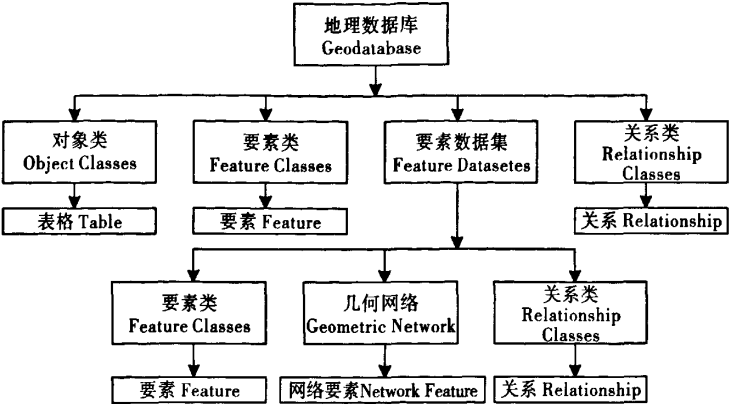
由于现代地理空间数据均为数字形式采集, 本文以基于 CAD 平台开发的数据采集、编辑系统南方 CASS 和 GIS 主流软

件 ArcGIS 为平台, 研究 CAD 数据和 GIS 数据的集成问题。

1 数据模型的对比与分析

由于 CASS 是基于 CAD 平台技术开发的, 故支持所有 CAD 的数据格式, 且具有相同的特征。常见的数据文件有线画文件 (*.DWG)、MicroStation 设计文件 (*.DGN) 和线画交换格式 (*.DXF), 其特征是利用几何要素结合颜色、线型、线宽等描述地理空间实体, 一般按图层进行数据的组织, 属性主要依靠注记来表现。ArcGIS 支持点、线、面、注记、区域、事件、路径等主要空间特征, 具有 3 种数据组织方式: Shapefile、Coverage 和 Geodatabase。其中, Geodatabase 为面向对象的数据模型, 能够表达要素的自然行为和要素之间的关系, 其数据组织结构, 如图 1 所示。

由图 1 可知, Geodatabase 用几何特征、网络、特征类的关系、平面几何拓扑和对对象组织模式扩展 Coverage 和 Shape 文件模型, 使得空间数据对象及其相互间的关系、使用 and 连接规则等都能方便的表示、存储、管理和扩展, 也使数据更具智能和面向应用领域的特性^[4]。



通过以上分析,可以得出 CASS 和 GIS 主要存在数据描述和数据组织两方面的差异。在数据描述方面,GIS 除可以用符号、颜色、注记等特征外,还可以用更加丰富的属性信息;在数据组织方面,CASS 依靠图层,GIS 根据应用目的不同以专题进行组织,每个专题下又包含点、线、面等要素。

2 数据集成方法的设计

2.1 集成方法的研究

要充分发挥 CASS 数据编辑和 GIS 数据管理、分析的优势,实现空间数据的共享和集成,就要根据各自的特点,解决图形数据转换和图形、属性链接两方面的问题。目前,图形数据转换主

要有直接转换和间接转换两种方法,直接转换又有文件转换、直接访问和数据库共享接口 3 种方式;间接转换通过在 CASS 与 GIS 间定制专用的数据读写程序实现^[5]。

对于图形和属性的链接问题,CASS 提供 3 种实现方式:一是直接应用对象特性,二是应用扩展属性,三是外挂数据库。其中,第一种方式只能利用“Thickness”、“Evaluation”等少数几个利用率很低的特性,有较大的局限性;第二种方法需要在 CASS 的“AttriBute.def”文件中,按照其数据定义结构对每个对象编辑新的属性,并在编绘图形过程中输入各个属性值(见图 2);第三种方法用数据库表的指针将图形对象与外部数据库表中的数据链接,需编写程序代码实现。



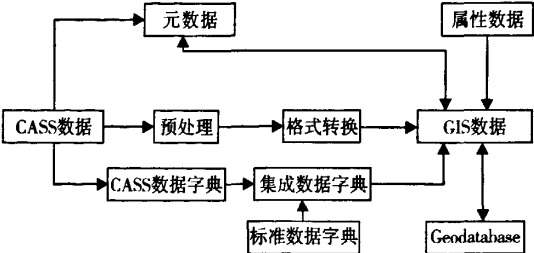
图 2 应用 CASS 的扩展属性

Fig. 2 Extension attribute of using CASS

2.2 数据集成的设计

通过以上分析,并考虑目前 GIS 转换或直接读取工具在图形数据处理过程中,要素丢失现象已经比较少;而现代属性数据的采集均以数据表的形式存在。所以,从节约成本和提高效率的角度出发,本文提出利用直接转换的方法解决图形数据的集成,属性数据则利用关键词在 GIS 中直接与图形数据进行链接,并通过建立 CASS 要素和国家基础地理信息要素间的集成数据字典,实现数据编码与语义表达的统一。在此过程中,还需要根据应用数据的目的和源数据信息建立 GIS 元数据,并对源数据进行必要的预处理,以便集成和集成后数据的维护与使用,数据集成结构,如图 3 所示。

建立数据字典的目的在于提取 CASS 数据中地理实体的表



达含义,而集成是指利用其字典项与标准数据字典项建立对应关系形成新的数据字典,以统一数据的地理编码和语义表达。具体过程分析如下。

如表 1 所示为利用 CASS 提供的地物分类标准建立的数据字典,表 2 所示为利用国家基础地理信息分类标准建立的数据字典。可以看出两者对地理实体的分类存在差异,如 CASS 中为表达围墙的信息,又将其分成边线和短线两类,一般房屋又分为砼房屋、混房屋、砖房屋等。

由表 1、2 可知,要实现两种数据字典的集成,首先需要对 CASS 字典中的字段进行整理,以能体现详细地物分类为原则,并结合国家基础地理信息分类代码扩充原则,对其进行合理扩充,最终形成如表 3 所示的集成数据字典。

表 1 CASS 数据字典

Tab. 1 Data dictionary of CASS

CASS 编码	图层	类别	参数 1	参数 2	地物名称
141111	JMD	8	continuous	砼	砼房屋
141161	JMD	8	continuous	混	混房屋
141121	JMD	8	continuous	砖	砖房屋
141500	JMD	8	pf	0	棚房
141500-1	JMD	18	pf	0	棚房短线
144301	JMD	11	wall	0.5	依比例围墙
144301-1	JMD	11	wall	0.5	依比例边线
144301-2	JMD	11	wall	0.5	依比例短线

表 3 集成数据字典

Tab. 3 Integrated data dictionary

CASS 编码	分类编码	分类要素名称	地物名称	图 层	颜 色	线 型	线型比例	线 宽
141111	310300	普通房屋	砼房屋	JMD	6	continuous	1	0
141161	310300	普通房屋	混房屋	JMD	6	continuous	1	0
141121	310300	普通房屋	砖房屋	JMD	6	continuous	1	0
141500	310600	棚房	棚房	JMD	6	continuous	1	0
141500-1	310600	棚房	棚房短线	JMD	6	continuous	1	0
143120	310900	廊房	门廊	JMD	6	x5	1	0
144301	380201	围墙	依比例围墙	JMD	6	continuous	1	0
144301-1	380201	围墙	依比例边线	JMD	6	continuous	1	0
144301-2	380201	围墙	依比例短线	JMD	6	continuous	1	0

CASS 数据直接集成后常会出现要素丢失、冗余等现象,为保证数据质量并减少后期编辑工作量,需要对源数据进行预处理,其主要过程有:根据使用数据的要求删除无需转换的图元和图层;根据编码分离各个图层,检查修复图层间的混杂和错误现象;利用图形实体检查功能,确保每个要素编码、颜色、线型等特征的正确性。

表 2 标准分类数据字典

Tab. 2 Data dictionary of normal sort

分类代码	要素名称	描 述
310300	单幢房屋、普通房屋	单体及普通房屋
310301	建成房屋	已经建成并投入使用
310302	建筑中房屋	尚未建成
310500	高层房屋	高层建筑
310600	棚房	简单房屋及棚房
380201	围墙	围墙

3 数据集成的实现

考虑到 CASS 和 GIS 平台组织地理数据的差异,还应根据 GIS 由点、线、面构成专题以组织数据的特点进行转换要素对照关系的设计,如 JMD 图层中“141161”实体应设计转换为 GIS 的面要素,而“143400-1”实体应设计转换为 GIS 的线要素等。本文以某中学校舍安全评价数据为例进行集成,以实现校园安全的信息化管理。

3.1 元数据的建立与数据预处理

为了便于空间数据的有效定位、评价、比较、获取和使用,需要设计一个描述能力强、内容完善的元数据。根据本项目使用数据的目的并结合图形相关信息,由 ArcCatalog 提供的元数据编辑器直接编写所对应的元数据。

3.2 图形转换与数据集成

图形转换的主要步骤有:在 Arcmap 视图中添加源数据,利用 Export Data 将图形转换为 shapefiles 格式文件;通过 Define Projection 或 Spatial Adjustment 工具并结合控制点信息进行坐标系的定义或转换。然后,利用图形和属性表的标识字符进行属性数据表的链接;利用 CASS 编码链接已设计的集成数据字典,本例使用的属性数据表,如表 4 所示。

表 4 链接的属性数据表
Tab. 4 Attribute data table of interlinkage

房屋编号	用途	修建时间/年	结构形式	基础形式	基础埋深	屋面形式	檐口高度	层数	可靠级别	抗震级别
1	教学	1997	砖混	条形	3 000	—	16 000	5	B	B
2	教学	2008	框架	条形	4 500	—	13 000	4	A	A
3	教学	2005	三全砖木	浅	300 ~ 500	坡	3 300	1	A	B
4	教学	2005	三全砖木	浅	300 ~ 500	坡	3 300	1	A	B
5	宿舍	1997	简易砖木	浅	300 ~ 500	坡	3 300	1	C	D
6	办公	1997	简易砖木	浅	300 ~ 500	坡	3 300	1	C	D
7	教学	1997	简易砖木	浅	300 ~ 500	坡	3 300	1	C	D
8	教学	1997	简易砖木	浅	300 ~ 500	坡	3 300	1	C	D
9	教学	1997	简易砖木	浅	300 ~ 500	坡	3 300	1	C	D
10	食堂	1997	简易砖木	浅	300 ~ 500	坡	3 300	1	C	D

在完成图形转换和数据链接的基础上,通过叠合比较法、目视比较法和逻辑检查法^[3]等方法对形成的数据进行检查,确定无误后输入 Geodatabase 数据库文件,实现数据的集成。如图 4 所示为集成后实现数据查询的结果,达到预期的效果和目的。

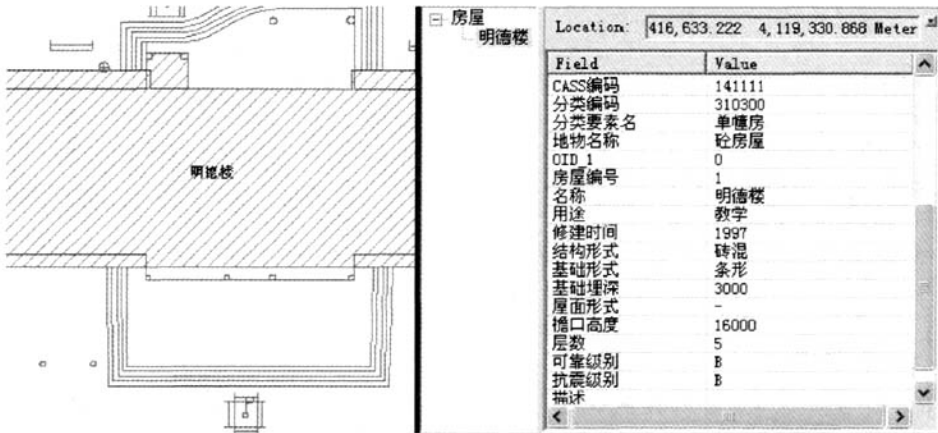


图 4 实现查询功能示意图
Fig. 4 Schematic of realizing query function

4 结论

根据现代地理数据采集的特点,本文提出基于数据字典的集成方法,在一定程度上解决了 CAD 图形数据转换和不同平台间编码、语义表达的统一问题。这种方法充分利用了现有成熟技术,不但避免了购置新软件带来的成本问题,还提高了数据生产和管理的效率,具有一定的实用和推广价值。

当然,要从根本上解决多源数据共享问题,还需要更多的研究和开发。所以,为了在信息化管理决策的趋势下,提供更加可靠、方便的基础地理信息数据,并能够更加充分利用 CAD 和 GIS 各自的技术优势,提出几点建议:在野外数据采集过程中应尽量按照现代地理信息数据的要求进行,避免过多的数据后处理;属性数据的采集应尽量完整、齐全;在图形编辑过程中应充分考虑数据共享的内在要求,为后期的数据集成做好准备。

[参 考 文 献]

[1] 张叶,孙毅中,陈年松. CAD 城市基础数据到 GIS 转换的有关问题探讨[J]. 测绘与空间地理信息, 2007, 30(1): 94 ~ 97.

[2] 韩江峰,邓敏,徐枫,等. CAD 宗地数据向 Geodatabase 自动转换方法研究[J]. 测绘通报, 2009(9): 58 ~ 61.

[3] 王强,吴孟泉,孙西兵,韩磊. dxf 与 shp 转换中的问题研究[J]. 山东国土资源, 2009(7): 46 ~ 48.

[4] 马立广,张亮. AutoCAD 环境下 Geodatabase 数据转换与可视化操作实现[J]. 测绘科学, 2008, 33(5): 196 ~ 197.

[5] 陈能,施蓓琦. AutoCAD 地形图数据转换为 GIS 空间数据的技术研究与应用[J]. 测绘通报, 2005(8): 11 ~ 14.

作者简介:赵健贤(1981 ~),男,甘肃民乐人,硕士,主要研究方向:测量数据处理及 GIS 应用。