

多源异构空间数据库整合技术探讨与实践

——以1/5万区域地质图空间数据库为例

李磊,郑锦娜,孙义伟,王心华,李敏
(中国地质调查局天津地质调查中心,天津 300170)

摘要:本文在1/5万区域地质图空间数据库两种不同标准数据的基础上,开展多源异构空间数据库整合研究实例性探索,基于空间数据应用转换模型及相关标准,实现了两类空间数据库向应用模型的整合,取得了良好的整合效果,为下一步开展空间数据一体化管理与共享服务提供了理论与实践基础。

关键词:空间数据库;数字填图;数据整合;一体化管理

中图分类号: P208

文献标识码: A

文章编号: 1672-4135(2016)03-0237-04

随着地质大调查工作的开展,积累了大量的国家级基础地质数据库,为政府部门决策、矿产资源评价以及经济建设等方面提供了数据资源保障。随着互联网大数据时代的来临,社会对于地理空间信息的精细化需求日益增加,对数据信息的分析挖掘也越来越迫切,然而在实际应用中,由于技术、资金与管理方面的原因,各个机构与部门间形成了大量的“信息孤岛”^[1],并且由于数据采集方式、标准规范与管理平台的不同,导致大量空间数据在数据标准与格式方面存在很大的差异性,造成数据共享与一体化管理十分困难。因此,研究数据整合技术是实现网络环境下空间数据无缝集成与共享应用的关键技术问题^[2]。

1/5万区域地质图空间数据库建设是地质调查成果信息化及国家基础地质数据库的重要组成部分,目前数据库基于两类标准建立,两种数据在数据模型、数据格式、表达方式与命名存储上都存在很大的差异性,本文在此数据基础上开展数据整合研究,基于空间数据应用转换模型及相关标准,实现两类数据一体化管理与集成实践,为多源异构数据整合提供实例依据。

1 数据整合技术流程

由于多源异构空间数据很难直接满足空间数据网络共享与服务的要求,需要解决的问题是对数据库中存在的数据库格式差异、坐标体系差异、时空尺度

差异和属性定义等差异进行整合。为此需建立有用于两类数据库的应用数据模型,最大限度的保持两类数据的原本性,按照一定的数据标准、空间与属性数据转换模型对两类空间数据库进行提取、转换、加载与整合,使其形成一个有机整体,最大程度上实现数据源的完全转换与共享服务。

首先需要制定完善的数据整合方案与技术流程,在严格的质量监控措施保障下对两类数据库进行数据整合,保证数据成果的正确、规范、有效。本文将数据整合流程划分为数据抽取,数据清洗与数据转换^[3-5]三个主要步骤,同时数据质量控制伴随整个流程的各个周期,多源异构数据整合流程如图1所示。

2 整合实例数据分析

本文以1/5万区域地质图空间数据库作为数据整合数据源,1/5万区域地质图空间数据库建设分为两类标准的数据。一类为回溯性建库数据,是针对二十世纪六十年代以来我国传统填图开展的1/5万区域地质调查工作中所获取的海量基础地质成果资料,应用现代计算机技术、空间数据管理技术和信息共享技术,进行全面系统地数字化、汇总、建库和管理。主要是以地质图与地质报告为建库数据源,采用数字化的方法采集数据的建库工作。其建库流程包括地质图数字化、属性数据录入与挂接、图面整饰和投影变换等步骤。另一类为数字填图数据,其基于数字填图系统,采用面向对象数据库技术,全面实

收稿日期:2016-06-25

资助项目:地质大调查项目:区域地质图数据库建设(华北)(1212011220342)

作者简介:李磊(1985-),工程师,中国地质大学(武汉)地球信息与探测技术在读博士,长期从事地质数据库建设,地质信息软件开发,地质资料社会化服务工作,Email:cuglilei@126.com。



图1 多源异构数据整合流程图

Fig.1 Multi-source heterogeneous spatial data Integrating flow diagram

现野外数据采集、存储、管理、描述、分析和成果数据库的建立等一体化组织与管理^[6]。主要以 DGSInfo 采集整理的野外地质路线、实测地质剖面、实际材料图和地理底图为基础,采用点-点、线-线、面-面属性自动复制技术实现部分继承,使空间数据库部分继承实际材料图的空间数据结构及属性。

两类数据库主要差异如下:

(1) 数据模型标准不同

回溯性数据库以《数字地质图空间数据库建设工作指南 2.0》为基础,而数字填图以《数字地质图空间数据库(DD206-06)》为依据,两类标准所规定的数据库组织模型、图层数据属性结构,以及数据采集软件都不尽相同。

例如回溯性地质图空间数据库地质体按地质属性分层管理,断层按独立图层管理;数字填图成果数据库空间地质体未分层,无断层图层,各单位以不同

的子类型标识码区分。

(2) 属性结构与内容不同

由于采用的标准不一致,使两类数据库的属性数据项的数目、名称和字段长度基本不相同。在属性内容方面,回溯性数据库使用了 GB/T 9649 系列地质矿产术语分类代码汇编作为统一代码表示,空间数据库的属性与空间实体一一对应,并直接关联;数字填图成果数据库中多数数据项内容采用汉字填写,部分采用代码,其中图元的属性只包括基本的属性项,其它属性需通过子类码及图元编号等主键与外接数据表进行关联。

例如在地质体代号表示方面差异明显,在回溯性建库中,上标用↑表示,下标用↓表示,还原用→表示;如 J3Z[↑]则表示为 J↓3→Z[↑],而数字填图则规定“\$”表示上标,“@”表示下标,每个标识只对紧邻后的一个字母或标识有效,同样 J3Z[↑]在数字填图中表示为 J@3Z[↑]。

(3) 系统库不一致

回溯性 1/5 万区域地质图数据库建设统一采用全国 1/5 万区域地质图空间数据库系统库。数字填图成果数据库在 2008 年后全国使用了统一的系统库。

(4) 其他

两类数据库在拓扑关系要求,投影方式,文件组命名方面都不尽相同,在此不再赘述。

3 整合实例技术方法

在数据整合整体流程的基础上,结合中国地质调查局发展中心提供的 GeoModel 数据整合软件^①实现了两类数据的全自动转换,结合相关空间数据质量检查软件与人工抽检,最终完成了两类数据库向应用模型数据库的无损转换。主要技术方法如下:

(1) 数据抽取

根据数据转换模型映射关系表,对数据库所需的数据项目设计抽取规则,实现数据项全自动数据抽取,完成对应用模型数据的映射,实现图层与数据项数据结构的规范化,应用模型中基本涵盖了两类数据库的所有数据项,保证两类数据向应用模型转换后图层与属性结构的一致性。

关键技术包括:1)应用模型的建立,以两类标准数据为基础,各取所长,地质内容数据项命名、长度以数字填图标准为基础,扩充回溯性数据库图层及数据项。地理部分属性数据项参照国家测绘地理信

^①刘荣梅,邓勇.传统填图和数字填图数据库成果综合集成转换技术方案,2015.

息中心相关标准的数据结构定义。2)空间与属性关系综合处理,确定主要地质内容图层数据的合并或分离,并对处理后的数据重新建立拓扑关系,补充相应的属性内容。

(2)数据清洗

针对数据项用代码表示的数据内容,需要对抽取得到的数据进行清洗,定义清洗规则,按照标准代码字典库、数字填图要素分类及相关标准等对数据代码进行翻译^①,以汉字形式统一表达,实现两类数据转换为应用模型数据后的属性内容表达的统一。

关键技术如下:属性内容统一与补充工作,按应用数据模型重新整理或关联相应属性,确定要素及对象的标识号规则、地质年代代号与地质体代号编码规则、上下标的规定、数据项下属词规定、数据项及数据项长度规定、属性数据项代码或汉字表达规定,按照统一要求汉化表达,同时对重要的属性内容进行补充完善。

(3)数据转换

定义两类数据符号库与应用模型符号库的映射关系,建立识别关系模型,完成符号库的统一,同时利用软件对数据进行格式类型转换,组织形式转换,对清洗好的数据实现投影变换,整理归并,然后按照应用模型规范要求命名,实现两类数据库整合转换后整饰表达与组织结构的统一。

关键技术包括:1)应用模型系统库与转换关系表建立,基于回溯性空间数据库 MapGIS 系统库,并进行补充扩展,建立起统一的应用模型系统库,同时建立回溯性数据库和数字填图数据库两类数据系统库与模型数据库的对应关系表,包括符号形状、

编号、大小、角度等参数,基于此实现对空间数据库符号库转换。2)投影变换,空间数据库投影统一为高斯(北京、西安)投影与经纬度投影。3)数据存储和组织、图层命名、元数据库命名规则的统一和转换,以数字填图数据组织形式为基础,补充回溯性数据库图层。

(4)数据检查

贯穿整个数据整合流程,数据预处理阶段和转换完成阶段分别要对数据进行质量检查,结合 Geo-Map 数据生产软件、GeoCheck 数据检查软件以及 DGSInfo 数字填图软件对空间数据与属性数据的准确性进行计算机自动检查,同时与人工抽检相结合,保证转换前后数据的精确性。

4 整合实例效果

本文研究选取工作区相邻的1/5万回溯性数据库建设和数字填图数据作为应用示范,经过对比,转换后的数据与源数据在文件组织、空间数据和属性信息方面都保持了良好的一致性(图2),取得了良好

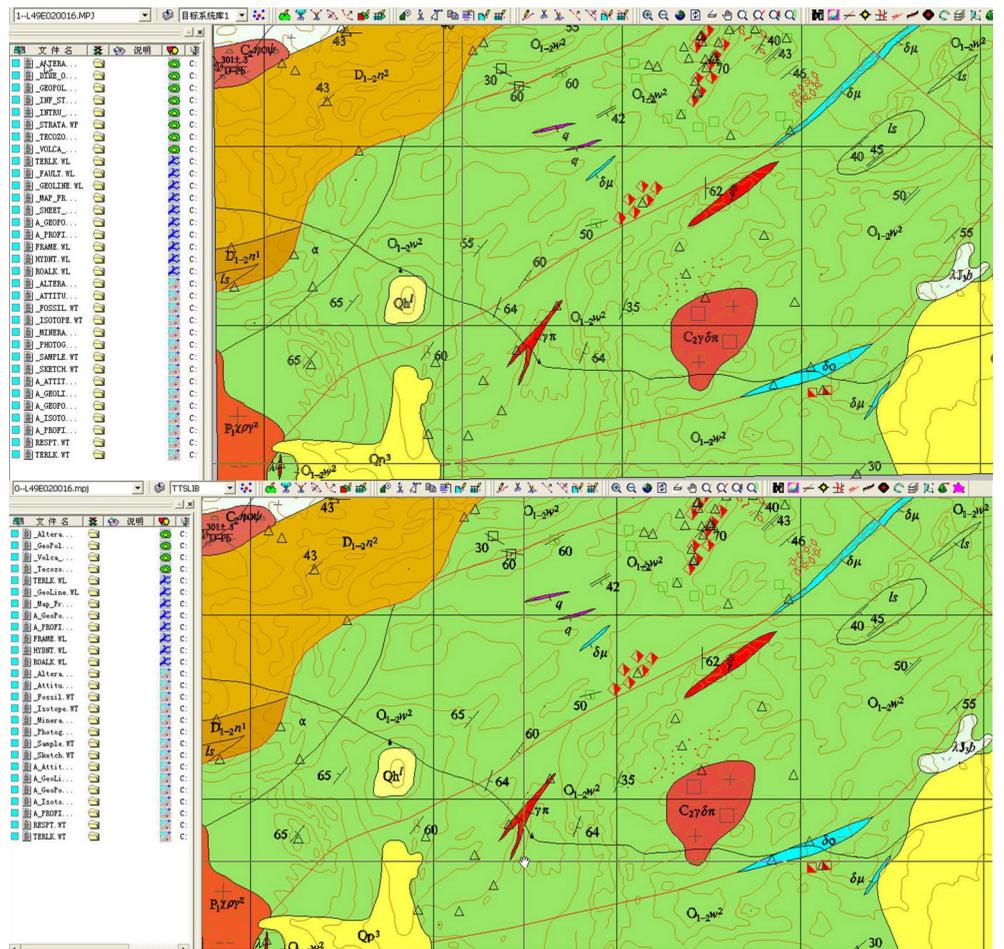


图2 数字填图转换为应用模型前后空间信息对比图

Fig.2 Spatial information of Digital Mapping before and after integration

的整合效果,为下一步开展空间数据一体化管理与共享服务提供了理论与实践依据。

5 结语

通过对多源异构空间数据开展整合研究,结合1/5万区域地质图空间数据库两类标准数据的整合探索性实践,最终形成了一套行之有效的数据整合技术方法。整合后的试验数据精度良好,分类编码正确,数据完整性好,属性数据完整,图面整饰统一,数据质量符合数据库建设要求^[7]。该数据整合技术有效的保证了两类数据库的原始性,同时又兼顾应用模型,便于下一步开展数据集成与共享研究,提高空间数据库的使用效率与服务水平,为地质找矿、灾害环境与地质科研提供基础数据支撑。

参考文献:

- [1] 李军,冯永玉,等.国土资源数据集成与应用[M].北京:科学出版社,2015,1-12.
- [2] 张梅兰,肖桂荣.区域地理空间数据整合技术研究[J].计算机与数字工程,2011,39(1):48-52.
- [3] 戴腾,张亮,等.多源空间数据整合在“数字城市”中的实现.地理空间信息[J],2013,11(3):17-19
- [4] 李广建,汪语宇,等.数字资源整合的实现机制及关键技术-对国外数字资源整合系统的实例研究[J].中国图书馆学报,2007,33(168):75-80.
- [5] 黎小红,田富鹏.异构数据库中的数据集成技术研究[J].西北民族大学学报,2006,(4):61-64.
- [6] 李丰丹,李超岭,等.大数据环境下数字填图数据集成服务技术[J].地质通报,2015,34(7):1300-1308.
- [7] 陈志强.城市基础地理数据整合与建库技术方案研究[J].城市勘测,2008,28-32.

Discussion and Practice on multi-source heterogeneous spatial data Integrating Technology, taking 1/50000 regional geologic map spatial database as an example

LI lei, ZHENG Jin-na, SUN Yi-wei, WANG Xin-hua, LI Min

(Tianjin Survey Center of China Geological Survey, TianJin 300170, China)

Abstract: Based on the 1/50 000 regional geologic map spatial database divided in two different standards, we have taken the multi-source heterogeneous spatial data Integrating Technology practice and exploration. With the database Transformation Model and other standards, we established the integration between the two spatial databases, the good result will provide useful reference for spatial database integrated management and sharing service.

Key words: spatial database; digital mapping; data integrating; integrated management

更正

本刊2016第二期95页“河北沽源一带不同时代岩浆岩地球化学特征”一文的第三作者姓名“申宗”改为“申宗义”,特此声明。