

海洋地质调查数据库数据录入方法

宋怀荣, 林 峰, 苏国辉, 戴勤奋, 何书锋, 孙记红

(中国地质调查局青岛海洋地质研究所, 青岛 266071;

国土资源部海洋油气资源与环境地质重点实验室, 青岛 266071)

摘 要:海洋地质调查数据库通常根据已有数据库结构标准创建各数据表和要素类,采用相关软件工具进行数据管理。基于数据的录入实践,探讨了海洋地质调查数据录入的几个关键环节:表空间和用户对应创建、表之间关联的创建、大对象数据加载等。结果表明,合理的创建用户和表空间、适当的创建数据表与公共索引表和空间站位信息表之间的关联、大对象数据采用 BLOB 格式存储,有助于实现高效、正确的数据录入。

关键词:海洋地质调查;数据库;数据录入;表关联

中图分类号:TP392

文献标识码:A

DOI:10.16028/j.1009-2722.2016.02010

加强海洋地质调查工作是我国建设海洋强国的战略所需^[1],海洋数据的管理、应用与共享已成为衡量国家海洋科技水平的重要标志。海洋地质数据涉及基础地理、地形地貌、海洋地质、构造地质、地球物理、地球化学、工程及环境等各专业领域,其应用价值和综合利用潜在价值不可小觑。

随着计算机技术、数据库技术和 GIS 技术不断引入海洋地质数据处理过程中,我国海洋地质数据的采集、存储能力不断增强。关于数据库管理系统的设计与实现,已有诸多工作者进行了详细的阐述^[2-11]。这些研究的主体思路是在一个系统下创建多个数据库,并建立权限表、角色表、角色权限表等来约束,或者利用关键字段将数据关联起来组成完整的数据库,但是,对数据的验证约束并没有特别突出,或者是缺少了核心、公有实体的创建以及表之间数据的约束验证,不能完全确保数据的唯一性、正确性、完整性。数据入库的高质量体现在高效性和库中数据的正确性,合理的入库思路和方法是保障数据库质量的前提。

1 海洋地质调查数据库的数据结构

根据标准规范,通过收集项目已有的资料,整理并分析主要研究内容,在此基础上详细定义数据结构。在海洋地质调查数据库(Marine Geology Survey Database, MGSDB)中,根据海洋地质调查数据的特点,以数据集为主体,将其划分为公共索引集、空间信息集、调查数据集、分析数据集、成果数据集等。

其中:①公共索引集用于对核心、公有实体数据集的创建,以实现数据入库时的约束验证,确保数据的无丢失性、正确性;②空间信息集是海洋地质数据的空间信息,可以实现海洋地质数据的空间定位和可视化,能直观的查看每年度海洋地质调查项目实施区域的空间位置;③调查数据集是外业调查采集的数据,包括取样(底质取样、大气取样、水体取样)、钻探、连续走航观测(单波束测深、多波束测深、侧扫声呐、浅地层剖面、船载重力、船载磁力、单道地震、多道地震、走航海流等)和定点观测(遥感、海底视像、海底热流测量、定点海流与温盐深观测等)等;④分析数据集包括底质及钻孔样品的描述、样品的测试分析(粒度分析、

收稿日期:2015-10-16

基金项目:中国地质调查局项目(GZH201100313)

作者简介:宋怀荣(1985—),女,硕士,助理工程师,主要从事海洋地质数据库管理工作。Email:songhr2010@126.com

矿物鉴定、古生物鉴定、化学分析、测年、古地磁与物理力学性质测定等);⑤成果数据集包括多道地震、浅地层剖面的成果处理数据。

图 1 为海洋地质调查数据库部分 E-R 图结构,根据实体 E-R 图来创建数据字典和逻辑表。

公共信息集主要包括元数据、质量评价、数据集、公用代码表等;空间信息集主要包括调查区、走航测线、站位信息、多波束覆盖区等;调查数据集、分析数据集、成果数据集分别有更详细的 E-R 图。

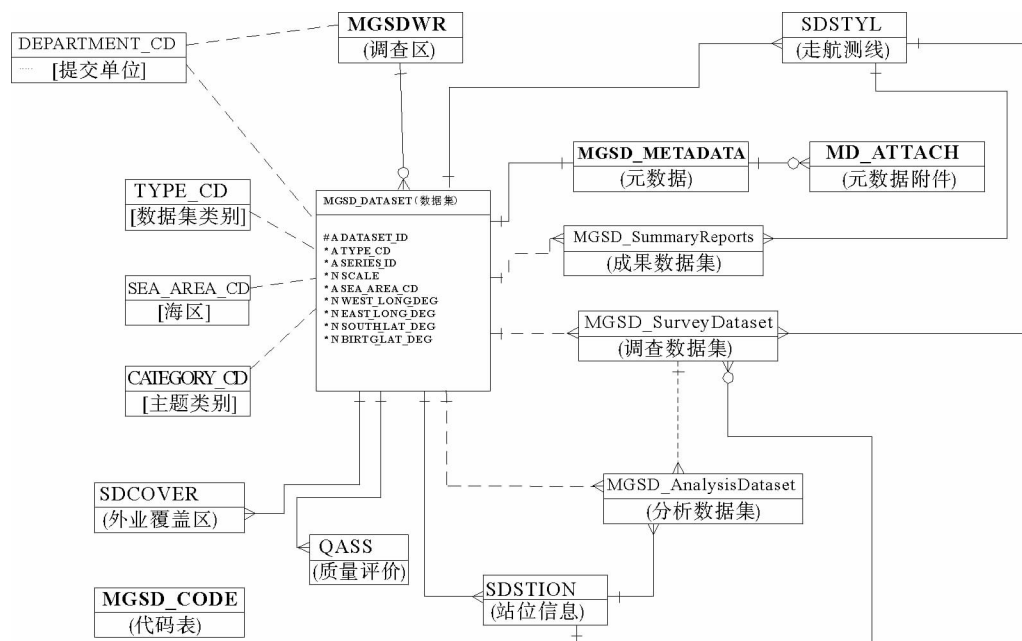


图 1 数据集实体 E-R 图

Fig. 1 The E-R diagram of dataset data entity

2 软件平台

MGSDB 系统通常由地理信息系统软件 (ArcGIS) 和关系数据库管理系统软件 (ORACLE) 共同管理。本文讨论的海洋地质调查数据库系统以 ArcGIS 10.1 和 ORACLE 11g 数据库平台为基础,利用 ArcCatalog 和 PL/SQL Developer 等组件和工具,实现空间数据和属性数据的录入工作。

ArcCatalog 工具是 ArcGIS 组件 ArcGIS-Desktop 下的功能模块,可以实现空间数据表的创建和数据导入,它使空间数据访问和管理更为容易,可以同时浏览、查找、修改地图和数据;查看和创建元数据;在功能模块 ArcMap 和 ArcToolbox 中创建和使用数据;将地图和表同步等;选择表视图可以查看地理数据源的属性或数据库中任

何数据表的内容^[12]。ArcGIS 10.1 及以上版本, ArcCatalog 能够直接连接 ORACLE 数据库,不再需要另外安装 ArcSDE 来创建表,但需要安装 ORACLE Client。

PL/SQL Developer 是一个集成开发环境,专门面向 ORACLE 数据库存储程序单元的开发,可以方便的为其他用户重新创建对象,也可以保存文件作为备份;可以重新编译全部不合法对象、查找数据库源中文本、导入或导出表格、生成测试数据、导出文本文件、监控 dbms_alert 和 dbms_pipe 事件、浏览会话信息等。

3 海洋地质调查数据库的数据录入

3.1 基本流程

海洋地质调查数据库录入工作的基本流程如

图 2 所示。在实现数据录入之前需要先完成数据结构的物理实现,然后再进行数据入库工作。本流程将数据的录入分为 3 个阶段:首先录入公共信息层数据,再录入空间信息层数据,最后录入调查、分析、成果等属性数据。由于数据结构已经设置了相应的约束条件,在数据导入时会自动检查数据是否重复、完整、匹配,从而保证了数据的安全性和可靠性。对于 BLOB 类型的数据,数据库一个记录只能接受≤4GB 的数据量,所以要保证每一行记录的数据量在此范围内。

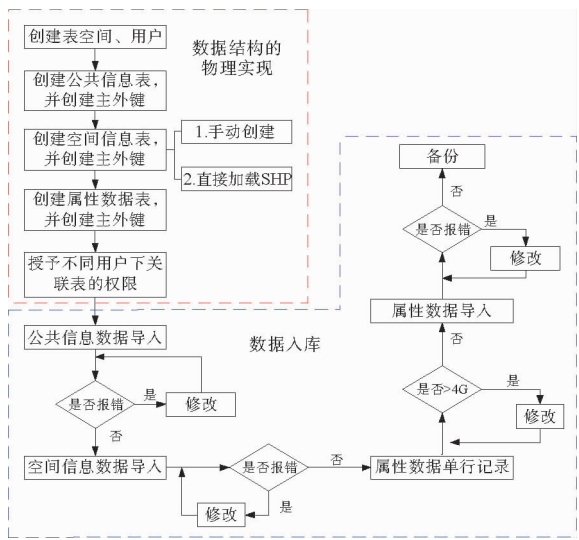


图 2 海洋地质调查数据库录入的基本流程
Fig. 2 Flow chart of data entry of MGSDB

3.2 注意事项

3.2.1 创建表空间、用户

根据不同数据集类型建立不同的表空间,这种存储方式有利于数据的集中管理,并且随着数据的增加、表空间不足时,能将数据整体转移到新的表空间中^[13]。可以给同一表空间下分配不同的用户(用户名称可参考调查类别,例如:走航调查、地质调查、钻孔调查等),并给不同的用户赋予不同的数据访问和使用权限,以确保数据的安全。

根据数据集的分类,该系统共创建 5 个表空间来存储对应的数据。MGSDB 分配的表空间、数据集和用户如表 1 所示。表空间的设置根据实际存储数据的大小设定初始值,并且设置为自动扩展方式。

3.2.2 创建空间信息表

由表 1 可知,空间信息是指点线面等要素类,其表的创建方法有 2 种:① ArcCatalog 连接数据库,进行手动创建,再加载 SHP 文件(此方法可以提前设置字段类型);② ArcCatalog 直接加载 SHP 文件(这些 SHP 文件已根据数据库结构进行属性设置)。

ArcCatalog 生成的空间表,字符的字段默认表示为 nvarchar2(),而海洋地质调查数据库结构中要求字段类型为 varchar2(),因此,为了统一表示,需要使用 SQL 语句修改 ArcCatalog 生成的空间数据表的字段类型为 varchar2()。ArcCata

表 1 海洋地质调查数据库的表空间和用户

Table 1 The table spaces and users of MGSDB

表空间名称	数据集名称	主要用户	定义
元数据及索引表空间	公共索引数据集	UPUBLIC	包括元数据(表)、数据集(表)、质量评价(表)、公共代码(表)
调查数据表空间	调查数据集	USD_TSY, USD_SAM, USD_BOR, USD_SSY	外业调查数据,主要分为走航测线调查、地质取样调查钻孔、定点观测调查等
分析数据表空间	分析数据集	UAD_ ANA, UAD_ DRA, UAD_DI	化学分析和数据解释等
成果表空间	成果数据集	USR	成果数据,例如地震+浅地层剖面解释数据等
空间数据表空间	空间信息数据集	UGEO	用于存储各空间信息,例如走航测线(线)、站位(点)、调查区(面)、多波束覆盖区(面)等

log 生成的空间数据表,没有主外键的约束,为了确保数据的完整性约束,需要 PL/SQL Developer 连接数据库对其创建主外键约束。

3.2.3 表之间的关联

为了实现数据的一致性、非重复性、完整性,各表之间需要建立关联。2 个用户之间存在表关联,需要对外键所在表用户授予对另一用户的限定关联权限。MGSDb 最主要的 2 个关联是:①各数据表与公共索引用户 UPUBLIC 中 DATASET(数据集)的关联;②各数据表与空间站位信息表(走航测线、站位、覆盖区等)之间的关联。

(1) 与 DATASET 的关联

由表 1 了解到用户 UPUBLIC 中的表主要是元数据和公共索引数据表,是公共信息层数据。对于同一项目、同一调查类型、同一航次、同一调查船施工,则外业调查数据拥有相同的数据集编号,相应的信息填写在表 DATASET 表中;同时,对此批数据进行的综合处理、分析和评价,仍保留对应的数据集编号,为了防止数据重复、缺漏、错误,将各数据表的数据集编号字段作为外键与 DATASET 表中的数据集编号关联。其他用户下的表关联 DATASET 表时,需要得到关联该表的权限(references on upublic, dataset)。

(2) 与空间站位信息表之间的关联

由于这些数据都是某一次海洋地质调查活动的结果(包括位置和属性信息),为了方便数据的预览、可视化使用和正确性,必须与空间信息表建

立关联。根据这些数据的空间特性,本文所指的空间位置信息包括项目既定的“调查区”信息,以及从原始调查数据中提取的“站位”、“测线”或“覆盖区”定位数据信息。根据不同的要素类,要赋予其他用户关联对应表的权限(例如,references on geo, station)。

3.2.4 BLOB 数据的存储

对于多道地震、浅地层剖面 and 侧扫声纳等数据,其原始和加工后的数据,单个文件大小都要以 GB 为单位来计量。对于此类数据的处理通常有两种方式:①直接录入库中;②建立文件名索引将其存放在服务器的指定磁盘中^[11],此方法为常用方式。本文中海洋地质调查数据库采用第 1 种方式,以 BLOB 格式进行存储,其录入数据库的过程相对复杂。BLOB 格式存储的最大数据量为 4 GB,因此,对于小于 4 GB 的文件,则可以通过创建控制文件(.ctl)和数据文件(.txt)在 cmd 中执行控制语句:sqlldr。对于地震、钻井等数据,部分单个文件的大小达到 8 GB 以上,则需将单个文件压缩成多个<4 GB 的文件,再通过以上方式录入库中。

3.3 录入结果

数据录入工作完成后对数据进行备份。如图 3 展示了 MGSDb 的 2012 年走航测线和调查取样的空间站位信息及其对应的属性表,采用本文介绍的数据录入方法,成功录入全部数据,其中

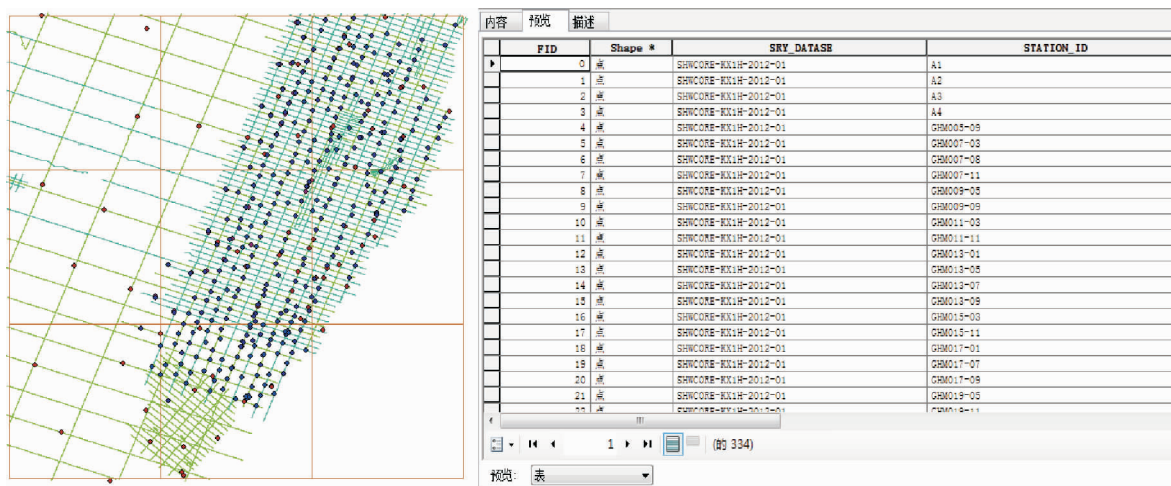


图 3 部分数据的空间位置和数据表显示

Fig. 3 The display of Spatial position and the data sheet

2012年多道地震调查共72条测线和取样调查210个站点。同时,通过数据的完整性约束,保证了调查数据与后期完成的分析数据、地震、浅剖成果处理数据的对应关系。空间站位信息有经纬度信息,附上地理底图后能很好的展示海洋地质数据的空间信息和其对应的调查、分析数据。

4 结语

本文给出的海洋地质调查数据库结构模型确定了数据库的入库路线,建立了各用户的权限和表之间的关联。数据录入是数据结构的实现和检验方法,本文讨论的数据录入方法满足了海洋地质调查数据库的录入需要。文中基于数据的录入实践,总结了海洋地质数据录入的几个关键环节和入库技巧:多个用户可以方便的实现数据移植和管理;表关联能保证数据的完整性;ArcCatalog创建空间表时需要注意字符的转换和约束键的创建;使用 blob 能更好的实现大对象数据加载等。文中讨论的主要内容是海洋地质调查数据的录入方法,而对于成果图件的录入和数据的综合应用成果展示并没有进行深入探讨,这将是海洋地质数据库后期的另一个研究方向。

参考文献:

[1] 苏国辉,魏合龙,戴勤奋,等. 海洋地质调查数据库建设现

状及经验[J]. 海洋地质前沿, 2012, 28(12): 10-16.

- [2] 孙记红,魏合龙,刘展,等. 基于 ORM 的海洋地质取样应用模型设计方法[J]. 计算机应用, 2013(8): 1-6.
- [3] 严杰,梁建,李绍荣,等. 基于 WebGIS 的海洋地质调查生产信息管理系统设计[J]. 海洋地质前沿, 2013, 29(6): 67-70.
- [4] 戴勤奋,田森,蓝先洪,等. 海洋区域地质调查成果图的规范化生产流程[J]. 海洋地质与第四纪地质, 2011, 31(2): 153-159.
- [5] 毛先成. 基于.NET 的综合地质数据库管理系统的设计与实现[J]. 地质找矿论丛, 2010, 25(2): 163-166.
- [6] 黄浩. 国际海底地学信息空间数据库系统研究[D]. 杭州:杭州电子科技大学, 2011.
- [7] 李福建. 基于 GIS 的我国海岸带地质灾害基础信息管理系统的设计与实现[D]. 青岛:中国海洋大学, 2010: 4-53.
- [8] 苏天赞. 海底多维综合数据建模及可视化技术研究[D]. 青岛:中国海洋大学出版社, 2006.
- [9] 陈振. 综合地质数据模型研究与管理信息系统开发[D]. 长沙:中南大学, 2008, 4.
- [10] 李绍荣,张锡林,陈道华,等. 海洋地质调查基础数据库模式构建方法[J]. 海洋技术, 2009, 28(4): 94-97.
- [11] 李绍荣,陈宏文,邵长高,等. 海洋地质调查数据库系统的数据组织[J]. 海洋地质前沿, 2015, 31(1): 65-70.
- [12] Vienneau A, Bailey J, Harlow M, et al. ArcCatalogManual [M]. New York: Environmental Systems Research Institute Company, 2004: 2-5.
- [13] 冯斌,谭建军,李绍荣,等. 海洋地质调查数据库管理系统设计与实现[J]. 计算机工程, 2009, 35(3): 29-31.

IMPROVEMENT OF DATA ENTRY SYSTEM FOR MARINE GEOLOGICAL SURVEY DATABASE

SONG Huairong, LIN Feng, SU Guohui, DAI Qinfen, HE Shufeng, SUN Jihong

(Qingdao Institute of Marine Geology, CGS, Qingdao 266071, China; Key Laboratory of Marine Hydrocarbon Resources and Environment Geology, Ministry of Land and Resources, Qingdao 266071, China)

Abstract: The Marine geology survey database(MGSDB) usually manages data in the forms of tables and features created according to the MGSDB structure standards using related software. In this paper, a few key areas of MGSDB are selected for improvement, including the table space and corresponding system, the association between tables, and the large object data entry system. The results show that it would be helpful for efficient and correct data entry if a user and table space reasonably created, the appropriate relationship between data tables and public index table established, station location information table made available and BLOB storage used for the large object data.

Key words: marine geological survey; database; data entry; table constraints