

土壤环境质量数据库设计与实现^①

姚 凌^{1,2,3}, 章莉萍⁴, 张淑杰^{1,2}, 周增坡^{1,2}

(1 中国科学院资源与环境信息系统国家重点实验室, 北京 100101; 2 中国科学院研究生院, 北京 100049;

3 南京师范大学, 南京 210046; 4 常州市城市规划管理信息中心, 江苏常州 213022)

摘 要: 本文在研究和分析现有国内外土壤数据库建设方法及其应用现状的基础上, 参考国内外土壤数据库设计与管理的相关标准和行业规范, 探讨了土壤环境质量数据库的编码设计、结构设计及数据组织与存储方式, 并以江苏省土壤环境质量数据为案例建立了示范性的土壤环境质量数据库管理系统。

关键词: 土壤环境质量; SOTER; ArcGIS Engine

中图分类号: S159.2

土壤是构成生态系统的基本要素之一, 是我国重要的自然资源和农业生产资料, 也是人类赖以生存和发展的物质基础。土壤污染问题已经成为影响群众身体健康、损害群众利益的重要因素, 也是社会各界关心的热点和“两会”关注的焦点^[1]。近几年来, 随着科学技术的不断发展, 土壤环境质量方面出现的资料数量巨大、内容繁多、关系复杂。对于大量的野外采样、多要素的实验室分析与测试等方式所产生的海量土壤数据, 目前还没有一个统一的数据库标准来指导形成一致的土壤环境质量数据库。从长期发展看, 土壤普查将成为一项日常性的工作, 成为环保系统日常监测的重要内容, 数据的累积将会迅速增加, 实现统一空间框架下土壤环境质量数据的野外采样、实验室分析、模型分析计算等的综合集成就显得尤为重要。

土壤数据库是联系基础土壤研究和生产应用的重要桥梁^[2]。国外关于建立土壤资源数据库的工作已开展多年。1966 年美国的土壤保护司 (SCS) 建立了全新的、综合性的国家土壤信息系统 (NASIS), 国际上公认的最早成立建立并运行的土壤信息系统是加拿大土壤信息系统 (CanSIS), 它从 1963 年开始研制实施, 1971 年投入运行^[3], 利用地理信息技术存储、管理土壤和土地资源数据, 然后按照统一的标准实现全国范围的集成和应用^[4]。之后, 苏格兰的 Macaulay 土壤研究所所在 1975 年建立了英国第一个土壤数据库^[5], 加拿大也在 2002 年建立了相应的国家级土壤资源系统^[6]。澳大利亚也建立了一个国家尺度的土壤数据库 (ASRIS)^[3]。1987 年 9 月国际上开始实施 1:100 万

世界土壤和土地数字的数据库 (SOTER, Global and national soils and terrain digital database) 计划, 首次用 1:100 万的比例尺对全球的土壤和土地资源信息进行管理, 覆盖了其他资源环境数据如地形、植被、气候、人口密度等, 提高了数据的管理和兼容能力, 以及人们对土壤和土地资源相关资源信息的理解能力。欧亚和地中海国家也建立了欧亚和地中海国家 1:100 万土壤基础数据库, 该数据库集中体现了人们当前对该地区土壤资源的认识水平, 对土壤资源的管理和保护起到了非常重要的作用^[3]。

中国土壤数据库的研究与应用起步较晚, 但发展迅速。20 世纪 80 年代中期, 中国土壤工作者开始进行土壤数据库建设的尝试^[7]。1986 年底, 北京大学遥感中心主持的“土壤侵蚀信息系统研究”实现了区域土壤侵蚀信息数据库的建设, 这是我国较早关于土壤数据库方面的研究。中国科学院南京土壤所 1989 年建成了东北三江平原土壤信息系统^[3]。国家“七五”攻关专题“全国土壤环境背景值调查研究”的子课题“中国土壤元素背景值数据库”, 收集了全国除台湾以外的 29 个省、自治区、直辖市 (立题时海南省尚未成立) 和 5 个沿海开放城市共 41 个土壤类的全部信息和数据^[8]。1991 年, 中国科学院沈阳应用生态研究所进行了区域土壤信息系统 (RSIS) 的建立和应用研究^[9], 但是该数据库规范化和通用化程度不高, 且未与 GIS 制图相结合, 评价结果图也多以符号或者数字的形式进行渲染。本世纪初, 中国科学院南京土壤研究所的史学正和于东升^[10]提出了“数字土壤”的概念, 该概念

①基金项目: 中国科学院“十一五”信息化专项 (INFO-115-C01-SDB3-02) 和国家科技基础平台项目资助。

作者简介: 姚凌 (1985—), 男, 江苏南通人, 博士研究生, 主要从事地球科学数据共享平台建设研究。E-mail: yaoling@lreis.ac.cn

的提出标志着中国土壤数据库的研究工作发展到了一个新阶段。在 SOTER 方法引入我国之后,各地区和相关部门相继进行了相关研究:中国科学院南京土壤研究所所以海南省为典型示范区,开展关于 SOTER 的理论和方法的研究,基本完成了 1:50 万海南省 SOTER 及制图工作^[7];张学雷和杨玉建^[11]进行了山东省 1:100 万 SOTER 数据库的建设。另外,中国科学院南京土壤研究所在中国科学院“十五”信息化建设项目的资助下,开展了中国土壤数据库的建库工作,并建立了中国土壤信息系统(SISChina)^[6],其内容包括土壤分类数据库、1:400 万土壤空间数据库、1:100 万土壤空间数据库、中国土种数据库、土壤质量动态监测数据库、氮磷钾养分循环数据库和区域农田生态研究数据库,为土壤数据库的进一步发展提供了经验和技術,并在 1:100 万中国土壤数据库建设中提出以“土壤类型 GIS 连接法”来融合土壤属性数据和空间数据,即利用 GIS 平台,根据土壤类型一致与相似性、土壤成土母质相同或相近、土壤剖面点位置与分布区域,将土壤属性数据库中的剖面数据连接到空间数据库中相应的土壤类型图斑单元中^[12-13]。

针对国家土壤环境质量普查的需求,使用 SOTER 方法建库主要存在两个方面的问题:

(1) 采用 SOTER 方法建立土壤数据库的重点在于 SOTER 单元的勾绘,需要进行专家认证,工作量巨大,而国家土壤环境质量普查需求的数据库设计重在多种类型土壤采样点(土壤环境质量调查点、背景点及重点区调查点等)点源数据的管理及应用,要求将剖面样点数据作为空间数据进行存储,进而对土壤环境质量及 20 年来土壤环境质量变化状况进行进一步的分析。土壤环境质量普查项目的特殊性决定了土壤环境质量数据库的建设不能使用现有的 SOTER 方法论。

(2) 国际上已有的最齐全的 SOTER 属性数据库包含的土壤和土地属性只包含 101 个指标,而本研究中所收集的土壤样点数据包括采样点的野外采样记录 50 项、理化指标检测数据 14 项、有机污染物检测数据 66 项以及无机污染物检测数据 27 项,远远超过了目前 SOTER 属性数据库所包含的 101 个指标,尤其在土壤环境质量指标方面, SOTER 仅仅涉及了包括全氮、全磷、钠、钙等 29 个指标。庞大的数据内容决定了在进行土壤环境质量数据库设计的时候必须选择合理的数据组织方式,才能提高数据库分析运算的效率。

2 土壤环境质量数据库设计

土壤环境质量数据所涉及的范围非常广泛,类型

种类繁多。根据土壤形成及其土壤质量的特点,可以将土壤环境质量数据分为基础地理环境背景数据、土壤基础数据以及土壤质量专题调查数据等 3 类。基础地理环境背景数据主要反映土壤形成的地理背景要素状况,包括地形、地貌、植被等;土壤基础数据指的是土壤本身的属性如土壤类型、可耕性等,包括土壤类型、土壤物理特征等;土壤环境质量专题调查数据包括土壤环境背景值、土壤污染要素状况调查数据等。

根据土壤数据来源和处理加工特点,可以将土壤环境质量数据分为专题调查数据和分析加工产品数据。专题调查数据主要是各类实际调查、考察等得到的原始性数据和资料,如各类土壤剖面调查数据、土壤类型考察数据等;分析加工产品数据主要是通过各种方法对原始数据进行处理和分析,形成服务于一定分析目的的衍生数据,如利用多个采样点通过空间内插所得到的相关土壤特征的面上数据。在数据库中,原始采集分析的资料最为重要,也是土壤数据库需要重点处理的对象。

土壤环境质量数据库与一般土壤数据库的建设相比具有其自身的特殊性,因为土壤剖面数据在空间上具有一维多层有序的特点,在土壤剖面的每一层上都存在大量的属性信息,而对于不同类型的采样点来说情况就更加的复杂。本研究拟管理的土壤环境质量数据大致可分为以下几类:点源土壤剖面数据、空间插值数据、等值线数据、基础地理信息数据以及照片、文档、图片数据。其中点源土壤剖面数据又有 3 种类型,只采集土壤样品距地表 0~20 cm 的表层混合样数据、按土壤剖面分层采样数据以及在土壤剖面上加测地表水、地下水、农产品信息数据。这 3 类数据分别对应环境质量普查设定的质量调查点数据、背景调查点数据以及重点污染区调查点数据,空间插值数据、等值线数据对应评价成果数据。

此外,由于土壤环境质量信息评价涉及的因子众多,所需要采集的信息量大,其中土壤环境质量调查专题的必测项目有 22 个,选测项目有 16 个;土壤典型剖面背景点对比调查必测项目有 20 个,土壤主剖面背景点对比调查必测项目包括 61 个元素全量、13 种元素的有效态、4 类有机污染物和部分土壤理化性质指标;典型污染区土壤环境质量调查必测项目有 22 个,选测项目有近 70 个,如何合理地设计土壤环境质量数据库对于进行土壤环境质量的分析评价就显得尤为重要。

2.1 数据库编码设计

土壤环境质量数据库主要涉及到的数据包括背景

基础地理数据、调查与分析评价数据、质量控制数据、元数据及科学文献数据 4 大类。其中背景基础地理数据对编码设计的要求并不高,而调查与分析评价数据则对编码设计提出了比较高的要求。主要是因为不同研究部门生产的原始数据本身就存在着编码不一致的情况,此外由于各个领域对土壤信息数据的概念并不统一,导致不同行业对于相同对象的分类标准并不一致,出现了编码的不一致问题。这些问题不仅为编码的重新设计提出了高的要求,同时也为土壤环境质量数据库的最终实现带来了难题。

在土壤环境质量数据库中,基础地理数据编码和土地利用编码目前都是采用国家标准,前者主要是采用国家 1:100 万和 1:25 万地形图的标准分幅、分层和编码,后者采用国土资源部的标准土地类型编码。此外,在土壤信息领域的分类代码方面已经有的国标、行标有:1:500、1:1 000、1:2 000 的地形图要素分类与代码,专题地图信息分类与代码等。除标准编码外,调查与分析评价数据没有任何现成国家标准或行业标准,必须自行设计编码方案。土壤环境质量调查数据包括两种类型的点数据:仅采取土壤表层混合样的数据(调查点数据)以及对土壤剖面进行分层取样的数据(国家“七五”调查背景点数据、土壤污染重点调查区数据)。对于调查点数据,样点编码采用 9 位码。具体编码方法和各位编码的含义如下:

第一位至第六位码:为我国县及县以上行政区划代码。其中第一、第二两位码表示(自治区、直辖市);第三、四位码表示市(地区、自治州、盟)或直辖市所属市辖区(县、县级市);第五、六位表示县(自治县、县级市、自治旗、市辖区、林区)。编码方案参照中华人民共和国行政代码(GB/T2260-2002)和国家统计局于 2006 年 1 月发布的最新的县及县以上行政区划代码。如江苏省前两位编码为 32。

第七至九位码:为以县(自治县、县级市、自治旗、市辖区、林区)为空间计算单元的采样点总数。例如,第 50 号采样点的编码号为 185。如图 1 所示。

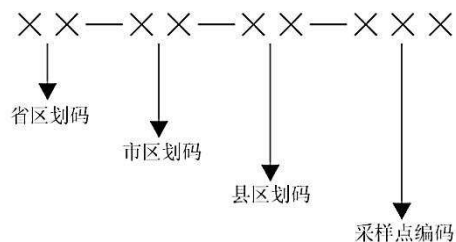


图 1 调查点数据编码设计

Fig. 1 Code design of soil survey point data

对于土壤污染重点调查区数据,由于背景点和重点区(污灌区)样点采取的是沿剖面分层采样,因此在属性项中增加一项作为外键,采用 12 位码,前 9 位与样点编码一致。经实验证明,使用数字型字段作检索的效率明显高于字符型字段,因此第十至十二位的编码设计使用土壤采样剖面的下限值(单位为厘米)代替剖面层次(图 2),例如采样深度下限值为 25 cm 时,编码为 025;取样深度下限为 185 cm 时,编码为 185。

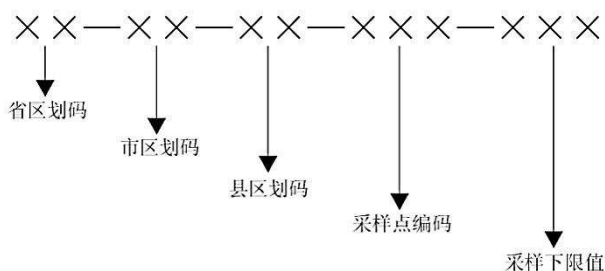


图 2 土壤剖面数据编码设计

Fig. 2 Code design of soil profile data

2.2 结构设计

土壤环境质量数据库中的数据表采用分级、分类的树状组织方式,即从子库到实体类(空间数据则采用 Feature Dataset、Raster Catalog 方式),再到具体的实体数据表;数据表的命名体现了以上的分级分类的思想,命名编码包含了其所在的子库、实体类等信息,各个分级编码间用“_”分割;数据表命名全部采用英文字符,根据实体含义以汉语拼音、英文拼写及其简写来命名。土壤环境质量数据库包括 8 个子库,5 个调查数据类。8 个子库分别为:调查(原始)数据库、工作数据库、评价数据库、(中间)成果数据库、背景地理信息数据库、质量控制数据库、元数据及数据字典数据库、文档数据库。5 个调查数据类分别为:质量调查点、背景调查点、“七五”背景调查点、背景值变化以及重点污染区调查点。

8 个子库中的具体管理内容如下:①调查数据库主要包含土壤空间点位数据和属性数据的管理;②为提高对土壤环境质量信息的分析评价效率,工作数据库管理的是对调查(原始)数据库中的某些字段进行预处理的数据;③评价数据库是在工作数据库的基础上经过评价分析得到的结果数据库;④成果数据库用于存储空间分析和专题图制作后生成的空间数据及渲染图层文件数据;⑤背景地理信息数据库是土壤污染调查数据库的必要组成,是辅助空间分析、统计分析

以及地图制图的重要数据源；⑥质量控制数据库用于存储与本次土壤污染状况调查样品的实验室分析测试的质量控制相关的数据；⑦文档数据库存储土壤污染调查项目中各省区提交的地方报告等文档，以及相关法律、法规及技术规范等；⑧元数据及数据字典数据库（简称元数据数据库）存储数据库的数据字典、编码及名称对照、评价标准以及各种工作辅助数据。

以调查数据库为例，土壤调查数据是整个数据库的核心和基础，土壤调查工作数据、评价数据等数据

都是其衍生数据，其实体-联系结构也基本上继承了土壤调查数据的关系。本节主要就土壤调查数据中的质量调查点、重点区调查点以及背景调查点的组织以实体关系视图的形式予以解释。

质量调查点属性表可以分解为野外样点记录表、理化性质测试表、无机污染物测试表和有机污染物测试表 4 个表，如图 3 所示。空间数据为主表，属性表为子表，子表和主表通过关键项（土壤样点编号）来维持。

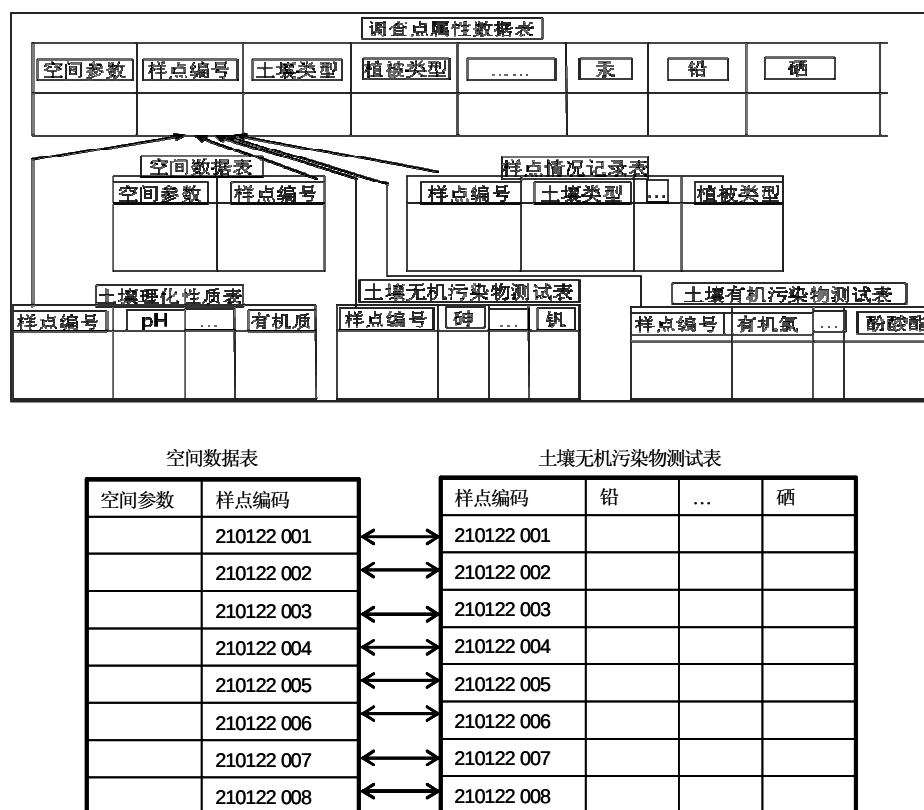


图 3 质量调查点数据表组织

Fig. 3 Data table organization of quality survey point

在土壤调查点中，主表和子表为一对一关系。图 4 为描述质量调查点如何组织的实体关系视图。

2.3 数据组织与存储设计

地理信息系统空间数据库中数据存储经历的 3 个阶段是：拓扑关系数据存储模式、Oracle Spatial 模式和 ArcSDE 模式。拓扑关系数据存储模式将空间数据存在文件中，而将属性数据存在数据库系统中，二者以一个关键字相连。这样分离存储的方式由于存在数据的管理和维护困难、数据访问速度慢、多用户数据并发共享冲突等问题而不适用本系统。而 Oracle Spatial 实际上只是在原来的数据库模型上进

行了空间数据模型的扩展，实现的是“点、线、面”等简单要素的存储和检索，所以它并不能存储数据之间复杂的拓扑关系，也不能建立一个空间几何网络^[14]。ESRI 公司所开发的 ArcSDE 解决了这些问题，并利用空间索引机制来提高查询速度，利用长事务和版本机制来实现多用户同时操纵同一类型数据，利用特殊的表结构来实现空间数据和属性数据的无缝集成等，是目前空间数据与属性数据组织得最好的产品之一，采用的是全新的 Geodatabase 数据模型^[15]，本研究中土壤环境质量数据库在底层采用 ArcSDE 存储模型。

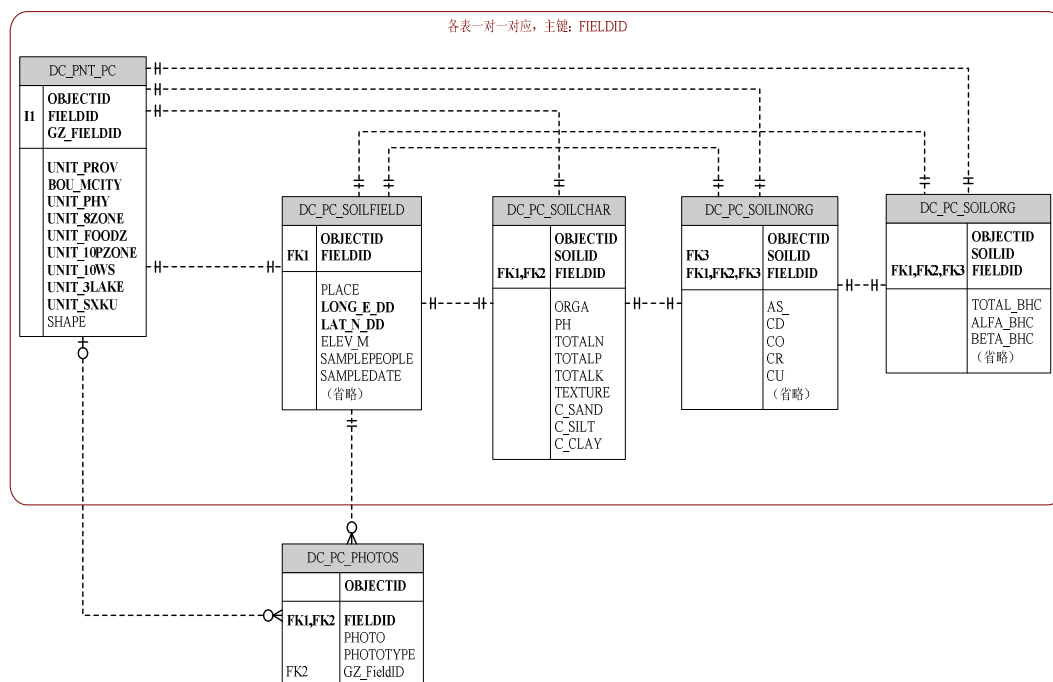


图 4 质量调查点相关数据表关联关系

Fig. 4 Correlation between data table of quality survey point

传统的数据模型主要有 3 种：网状、层状和关系型模型，本研究采用“数据库管理系统 (Oracle) + 空间数据库引擎 (SDE)”的模式来设计数据库、组织系统数据，即采用 SDE 技术提供的 MultiUser Geo Database 模型 (图 5) 中地理数据库→要素集→要素类 3 层结构组织复杂的空间数据 (土壤样点、居民地、道路、水系等)，用 Oracle 的二维关系表格组织土壤环境质量调查表格数据，统一存储在本研究选定的

Oracle 数据库中，通过 SDE 的关系 (Relationships) 模型来关联这两种数据，建立一个开放的、灵活的土壤环境质量信息数据库。一方面这类数据繁杂，如仅作为空间数据的属性表保存，不利于管理和操作，加大开发工作量，另一方面地理数据库 (GeoDatabase) 模型中的关系 (Relationships) 模型很好地解决了空间数据与外部表的关联，能够提高应用系统的开发效率。

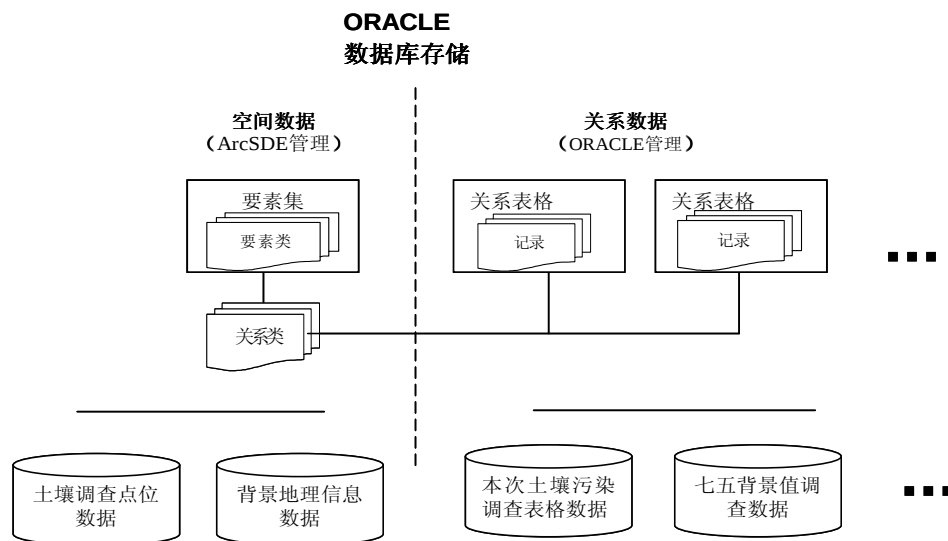


图 5 土壤环境质量数据库数据存储与组织

Fig. 5 Data storage and organization of soil environmental quality database

3 土壤环境质量数据库管理实验系统——以江苏省为例

3.1 总体架构

按照分层开发、组件式架构的设计思想，将江苏省土壤环境质量数据库管理实验系统划分为数据访问层、模块中间层、业务逻辑层和表现层，系统总体架构可用图 6 来表示，所有子系统和功能模块都采用分层设计和实现的原则。最低层为数据库层，包括各类样点（普查、背景、“七五”背景和重点）的调查数据、工作数据、评价数据以及最终成果数据，另外还包含

辅助的质量控制数据和背景地理数据。在数据层之上为基于 Oracle 10g 和 ArcSDE 9.2 的数据访问层，包括数据的存储、查询和更新等基本模块。中间层为开发平台层，在 Visual Studio.net 2005 的平台上，结合 ArcGIS Engine 9.2 进行 C/S 功能模块和应用系统的开发。再上一层为功能逻辑实现层，主要实现各应用系统功能所需的功能逻辑实现代码，并以动态链接库的方式供各应用系统的功能界面调用。逻辑实现层根据功能需求的不同，分为针对 ArcGIS Engine 的功能模块、针对属性数据的模块以及系统公共模块（用户管理、数据字典等）。

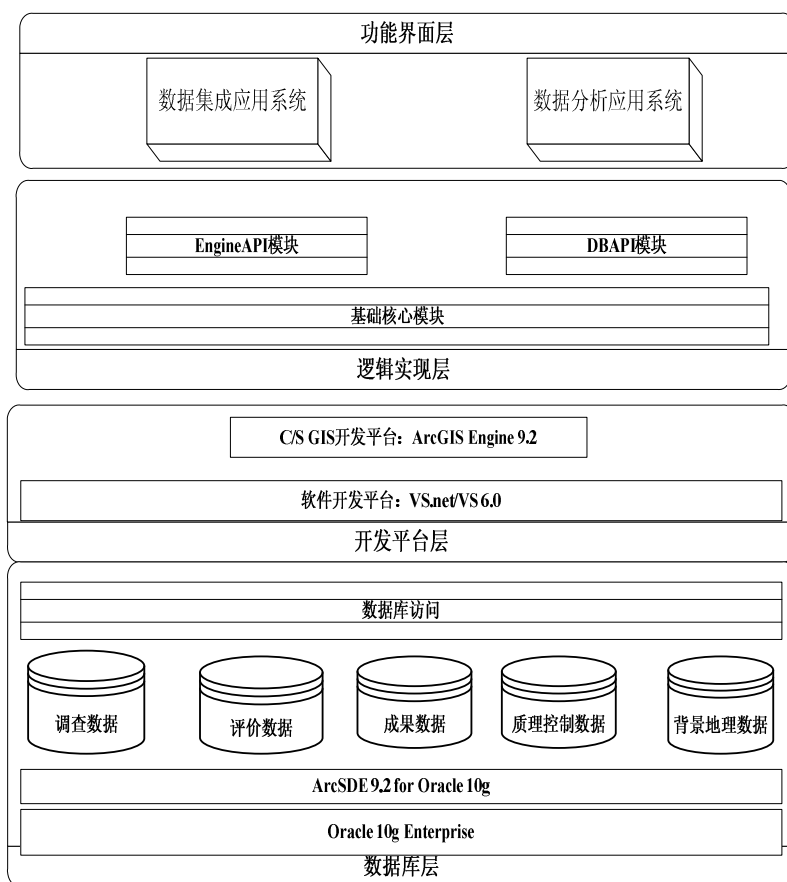


图 6 系统总体架构

Fig. 6 Overall framework of system

整个体系架构最上层为功能界面意义上的各应用系统，各应用系统通过功能界面，将功能逻辑层、数据库层统一，并输入各功能所需的参数，将结果进行展示，最终完成所有功能需求，主要包括面向不同用户的数据集成应用系统、系统管理应用系统、数据分析应用系统和信息服务应用系统。

3.2 功能模块

根据系统的应用需求，江苏省环境质量数据库管理示范系统由数据检查、导入与导出、数据预处理、数据浏览、空间选择、属性查询等功能模块构成。为提高查询浏览、空间分析的效率，通过数据预处理功能可以将经检查入库后的数据转换成工作数据存入工

作数据库中,平台还实现了数据浏览、属性查询、对离散点进行 IDW 或 Krigin 插值、利用插值结果编制等

值线图等功能。图 7、图 8 分别是数据浏览与查询、生成等值线结果的功能实现。

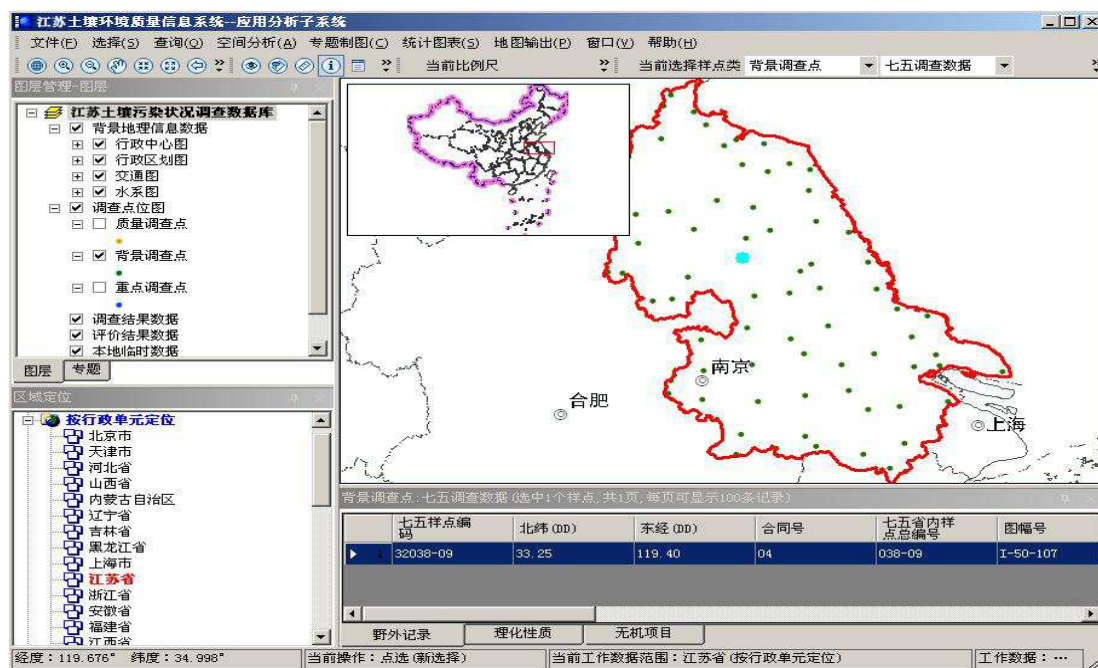


图 7 数据浏览与查询

Fig. 7 Data browsing and querying



图 8 等值线图生成结果

Fig. 8 Result of contour generation

4 结语

土壤环境质量数据的评价因子众多, 数据类型复杂多样, 土壤剖面样点的各数据类之间的关系复杂多变, 本文对土壤类型和土壤属性的编码、土壤环境质量数据库的结构及数据组织与存储方式进行了研究及设计, 提出了采用土壤区界编码加采样点编码实现土壤空间数据与土壤属性数据连接的方法, 并以江苏省土壤环境质量数据库管理示范系统加以验证。

土壤环境质量数据库目前仅限于土壤自然属性数据, 在数据库管理中进一步集成社会经济方面的数据, 并通过网络实现数据共享与信息发布是下一步的研究目标。

参考文献:

- [1] 万本太. 全国土壤污染状况调查有序推进. 环境保护, 2007, 02(A): 24-27
- [2] 吕成文, 沈德福, 陈云丰. 大比例尺土壤数据库的组织与设计研究. 土壤通报, 2004(2), 35: 122-125
- [3] 赵其国, 史学正. 土壤资源概论. 北京: 科学出版社: 2007
- [4] 李卫江, 吴永兴. 基于 WebGIS 的基本农田土壤环境质量评价系统. 农业工程学报, 2006, 22(8): 59-63
- [5] Bron. The development of soil information system. Journal of Soil Science, 1987, 38(2): 267-277
- [6] 史学正, 于东升, 高鹏, 王洪杰, 孙维侠, 赵永存, 龚子同. 中国土壤信息系统(SISChina)及其应用基础研究. 土壤, 2007, 39(3): 329-333
- [7] 周慧珍. 土壤地理信息系统. 土壤学进展, 1993, 21(6): 32-36
- [8] 蒋德珍. 中国土壤元素背景值数据库系统的研制. 中国环境监测, 1991, 7(2): 23-26
- [9] 贺红士, 侯彦林. 区域微机土壤信息系统的建立和应用. 土壤学报, 1991, 28(4): 345-354
- [10] 史学正, 于东升. “数字土壤”——21 世纪土壤学面临的机遇和挑战. 土壤通报, 2000, 31(3): 104-106
- [11] 张学雷, 杨玉建. 山东省 1:100 万 SOTER 数据库的建立与初步应用研究. 山东农业大学学报, 2001, 32(2): 136-142
- [12] 于东升, 史学正. 基于 1:100 万土壤数据库的中国土壤有机碳密度及储量研究. 应用生态学报, 2005, 16(12): 2 279-2 283
- [13] 吴克宁, 杨锋. 河南省 1:20 万土壤数据库的构建及其应用. 河南农业科学, 2007(5): 77-80
- [14] 熊丽华, 杨峰. 基于 ArcSDE 的空间数据库技术的应用研究. 计算机应用, 2004, 24(3): 96
- [15] 阚国年, 张书亮, 龚敏霞. 地理信息系统集成原理与方法. 北京: 科学出版社, 2003

Design and Implementation of Soil Environmental Quality Database

YAO Ling^{1,2,3}, ZHANG Li-ping⁴, ZHANG Shu-jie^{1,2}, ZHOU Zeng-po^{1,2}

(1 Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China;

2 Graduate University of Chinese Academy of Sciences, Beijing 100049, China; 3 Nanjing Normal University, Nanjing 210046, China;

4 Urban Planning Management Information Center of Changzhou, Changzhou, Jiangsu 213022, China)

Abstract: Through analyzing the existing soil database construction method and its application at home and abroad, and referring to the relevant standards and industrial specifications, this paper discussed the code and structural design, data organization and storage of soil environmental quality database. Finally, soil environmental quality data in Jiangsu Province was used as a case to establish the soil environmental quality database management system.

Key words: Soil environmental quality, SOTER, ArcGIS Engine