

УДК 631.41+631.95

**РЕЛЯЦИОННЫЙ ПОДХОД К СОЗДАНИЮ БАЗЫ ДАННЫХ
ШИРОКОМАСШТАБНОГО ПОЧВЕННОГО
ЭКОЛОГО-АГРОХИМИЧЕСКОГО МОНИТОРИНГА**

В. С. Крыщенко, доктор биол. наук; О. М. Голозубов

Южный Федеральный Университет, e-mail: 4Hweb@sarmat.ru

В. Н. Темников, канд. с.-х. наук

Российская инженерная академия менеджмента и агробизнеса

М. М. Овчаренко, доктор с.-х. наук

НП «Агрохимсоюз», e-mail: agrohimsouz@mail.ru

Изложен системный подход к проектированию, разработке и эксплуатации в Интернет программного комплекса и базы данных агроландшафтов, основанный на приложении теории реляционных баз данных к задачам почвоведения и земледелия. Приводится классификация информационных структур и рассматриваются технологические аспекты, вводится понятие элементарных информационных единиц (ЭИЕ) базы данных.

Ключевые слова: *реляционные базы данных, мониторинг, почвенный профиль, почвенный контур, почвенный образец, земельный участок, объект базы данных.*

Аграрное производство в настоящее время требует внедрения высокоэффективных технологий сбора, обработки и анализа информации, которые могут оказать существенную помощь в решении многочисленных задач, связанных с планированием и прогнозом развития хозяйства. В области мониторинга плодородия почв земель сельскохозяйственного назначения сформулирована задача по созданию базы данных (БД) состава и свойств почв с учетом их механического состава и типа землепользования для субъектов РФ. В это же время Обществом почвоведов им. В. В. Докучаева открыт проект создания единой государственной почвенно-географической базы данных России [1].

В БД накапливаются результаты многолетних агрохимических обследований по полям, хозяйствам, районам субъектов РФ,

а также данные по составу и свойствам выборочных образцов почв. Оба проекта предусматривают централизованный сбор большого количества разнородной и разноплановой информации, поступающей из разных источников.

Почвенные базы данных созданы и существуют на различных национальных уровнях. Россия, активно участвуя в международных проектах, серьезно отстает в создании национальной БД, но преимуществом является то, что имеющийся в Европе и США более чем 20-летний опыт (в том числе и отрицательный) позволит избежать при построении Российской почвенной БД множества излишних, трудоемких и ошибочных шагов. Информация в базах данных может быть организована по-разному. Чаще всего используется табличный способ. Базы данных с табличной формой ор-

ганизации называются реляционными БД. Реляционный подход к организации структур данных, примененный на стадии проектирования БД и последовательно контролируемый на всех стадиях цикла БД, является средством обеспечения целостности, непротиворечивости, работоспособности, модифицируемости и совместимости баз данных в дальнейшем.

Источники информации для БД: данные почвенных обследований агрохимцентров и гипроземов, включающие в себя почвенные карты и отчеты; данные дистанционного зондирования, космических снимков, аэрофотосъемки; данные кадастров, налоговых и других учетных госучреждений; данные муниципальных и федеральных научных центров почвоведения, агрохимии, агрономии, экологии, управлений сельского хозяйства; данные сельхозпроизводителей, включая схемы полей севооборота, культуры произрастания, урожайность, агрохимические мероприятия и т. д.

Частота поступления данных варьируется от нескольких дней для космического зондирования до одного раза в 5 лет для комплексных агрохимических обследований. Кроме того, имеется накопленный многими поколениями почвоведов и землеустроителей экспедиционный и опытно-экспериментальный материал, который хранится на бумажных носителях.

Этот материал вносится в базу по мере оцифровки, причем с указанием даты актуальности собранных данных, поскольку архивы даже 20-40-летней давности имеют высокую ценность для задач мониторинга. С учетом реалий допускается ввод неполной, частично полной информации и, соответственно, последующее дополнение и уточнение введенной информации.

Пользователи БД на одном уровне могут одновременно выступать и как источники пополнения и модификации данных на другом уровне. Создание и наполнение БД такого уровня немыслимо без наличия разнообразных типов пользователей с оперативным доступом к данным в сетевой среде общего пользования – Интернет. В данном контексте мы под широкомасштабностью БД понимаем возможность асинхронного ввода и совместного использования разноплановых данных от большого количества клиентов и организаций различных ведомств, но при этом относящихся к одному почвенному объекту.

Следует отметить, что основные положения реляционного подхода (разнообразные типы пользователей; сети удаленных баз данных; оперативный доступ мно-

гих терминальных пользователей; до предела упрощенные типы базовых структур данных) были разработаны Эдгаром Коддом еще 40 лет назад [2] и не потеряли своей актуальности и по сей день. В развиваемой на протяжении более 30 лет теории реляционных баз данных Кодд определяет три составляющих модели данных: 1. Объекты. 2. Ограничения целостности. 3. Операции.

До предела упростить типы структур данных, используемых в принципиальной схеме (или общем представлении), как рекомендует Кодд, означает обеспечить длительную жизнеспособность, модифицируемость, используемость системы управления БД (СУБД). В качестве объектов исследования при формировании БД почв агроландшафтов и мониторинга их состояния выступают элементарные (базовые индивидуумы) информационные единицы (ЭИЕ), с которых «считывается» изначально субстантивная (т. е. находящаяся в первичной, измеренной форме) информация.

Одним из базовых объектов принято считать почвенный контур. Его структура зависит от детальности. Обычно разделяют масштаб национальный – более 1:1000000, региональный – от 1:100000 и локальный (муниципальный) – от 1:10000 до 1:50000. На масштабах большей подробности (планы внутрихозяйственного устройства) почвенные контуры обычно не выделяют. Для решения задач мониторинга агроландшафтов используется локальный масштаб.

Под почвенным контуром понимают:

1. Почвенные таксоны (тип, подтип, род, вид, разновидность) по классификации 1977 г. в почвенных (контурных) картах Гипроземов масштаба 1:10000 или 1:25000 в Российских БД. Причем границы почвенных контуров увязаны с рельефом и гидрологией, и указаны в пределах хозяйства;

2. Картографическую единицу (map unit) БД почв Канады для масштабов от 1:10000 до 1:250000. В один контур может входить основной тип почвы (в Канадской классификации) и до трех дополнительных типов почв (с указанием процента занимаемой площади). Ограничение тремя дополнительными типами обусловлено исторически сложившимися недостатками файловой структуры БД;

3. Картографическую (SMU) и типологическую (STU) почвенные единицы в Европейской БД. В процессе генерализации к базовому типу почвы контура может быть указано произвольное число дополнительных типов (с процентом занимаемой в контуре площади).

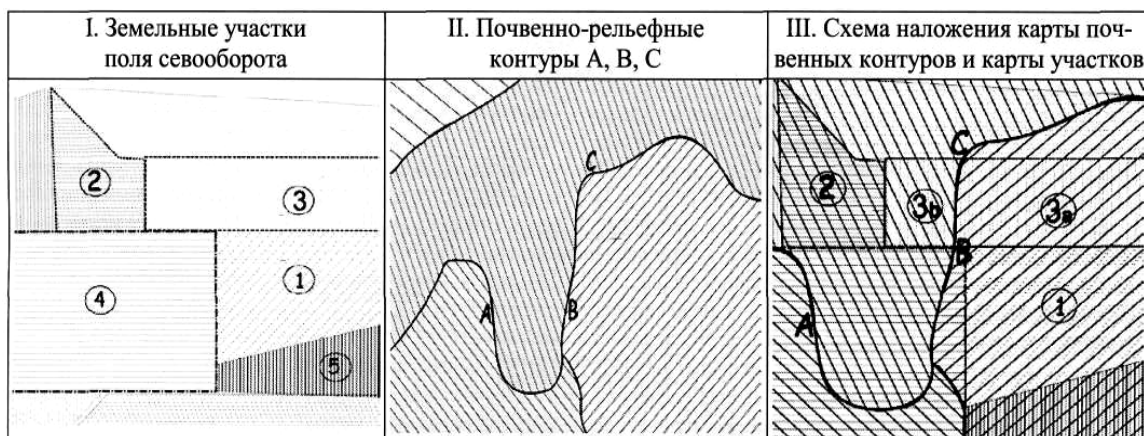


Рис. 1. Схема наложения векторной карты почвенных контуров и растровой карты космической съёмки:

А, В, С – границы почвенных контуров; □, □, □ – земельные участки

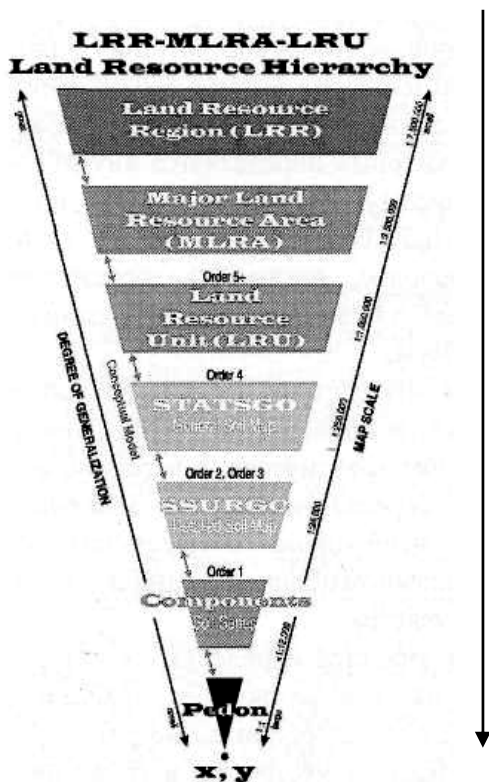
В перечисленных БД элементарной информационной единицей фактически является не просто почвенный контур, а его сочетание с однородными условиями рельефа, гидрологии, гражданских объектов, растительности и т. п.

В задачах мониторинга земель сельскохозяйственного назначения и агроландшафтов (в соответствии с разработанной в МСХ РФ формой 5-ДДЗ) необходимо учитывать еще один важнейший фактор – характер землепользования, который определяется внутрихозяйственной схемой (рис. 1-I) вплоть до участка поля севооборота. Форма 5-ДДЗ содержит данные по группам: почвенное плодородие, параметры загрязнения почв, эксплуатационные характеристики, агрохимические мероприятия и работы.

При антропогенном воздействии агрохимические мероприятия проводят по участкам поля, а не по хозяйству в целом. Фактически, в иерархии «хозяйство → поле севооборота (угодые) → земельный участок» мониторинг необходимо осуществлять на уровне элементарной информационной единицы, которой служит земельный участок. Границы земельных участков определяются внутрихозяйственными схемами и кадастровыми планами в метрической системе (с погрешностью до метра) с привязкой к реперным точкам на местности, а также дистанционным зондированием или средствами географического позиционирования в географической системе координат с аналогичной погрешностью. Привязка земельных участков к географической системе координат позволяет осуществить наложение почвенно-рельефного контура (рис. 1-II) на внутрихозяйственный план землеустройства (рис. 1-III).

Земельный участок 3 (рис. 1-I) попадает на границу почвенных контуров В и С (рис. 1-II). Применяя метод дробления без потери информации, можно выделить элементарные почвенно-земельные участки 3а и 3б, каждый из которых характеризуется как однородными почвенно-генетическими характеристиками, так и единым характером использования и антропогенным воздействием. Таким образом, элементарной информационной единицей в БД мониторинга агроландшафтов является часть земельного участка, относящаяся к одному почвенному контуру. Ниже рассмотрим особенности объекта БД – элементарного почвенно-земельного участка.

Отметим, что при работе с масштабами порядка 1:10000-1:25000 границы почвенных контуров нельзя рассматривать как точные и смежные друг с другом (в отличие от почвенных карт масштаба 1-5 млн.). Необходимо учитывать пространственную непрерывность почвенного и рельефного покрова и множество переходных форм. Кроме того, во всех странах мониторинг проводится выборочно, без смежных границ зон мониторинга. Границы участков задаются более точно, но также имеют погрешность. Эти погрешности нужно учитывать при формировании сводных интегральных показателей мониторинга. В европейской БД для учета погрешности введен так называемый уровень доверия, который зависит в том числе и от квалификации операторов, проводящих векторизацию карт. Методика дробления участка успешно применяется в технологии точного земледелия, где погрешность составляет не метры, а десятки сантиметров, и также имеются внутрипольная гетерогенность плодородия и несовпадающие границы агрохимических контуров.



Иерархия управления земельными ресурсами

Название уровня	Масштаб
LRR – Региональный земельный ресурс (штат или несколько штатов, сельскохозяйственный регион)	1:7500000
Основная земельная зона (географически связанная)	1:2500000
Обобщенная единица земельного ресурса	1:1000000
Генерализованная карта почв США	1:250000
Детальная карта почв	1:24000
Почвенные серии (более 20 000 наименований почвенных серий по классификации США)	1:12000
Педон	1:1

Рис. 2. Место ЭИЕ в иерархической административно-территориальной структуре в БД США

Элементарный почвенно-земельный участок объединяет в себе почвенно-генетические, антропогенные и административно-хозяйственные характеристики. Некую аналогию этому можно увидеть в почвенно-географической иерархии БД почв США. На рисунке 2 приведена диаграмма иерархии от регионального земельного ресурса до элементарного объекта (педона). Для БД мониторинга агроландшафтов России место элементарного почвенно-земельного участка в иерархической структуре можно изобразить следующим образом (рис. 3).

Каждая ЭИЕ имеет свои уникальные характеристики, достаточные для идентификации в БД. Иерархичность структуры нужна для интерпретируемости результатов мониторинга, получения интегральных показателей и, не в последнюю очередь, для построения «дружественного», понятного интерфейса работы пользователя СУБД.

Учитывая многолетний характер мониторинга, частую изменяемость организационно-правовых и административных границ, изменения в способе использования сельхозугодий, в СУБД должны быть определены следующие операции над ЭИЕ: расщепление ЭИЕ на несколько ЭИЕ; слияние нескольких ЭИЕ; перенос ЭИЕ по иерархическому дереву от одной родительской вершины к другой (например, от

одного хозяйства к другому) и т. п. Также, в случае уточнения границ почвенных контуров, изменения в глобальных классификаторах, может потребоваться переопределение границ ЭИЕ. Фактически ЭИЕ как базовый объект БД представляет собой часть земельного участка, принадлежащую одному почвенному контуру и определенную на момент формирования и до внесения изменений. После изменения ЭИЕ в БД записывают новый объект – измененную ЭИЕ (или несколько) с датой формирования, равной дате внесения изменений, а старый объект не удаляют из базы, а он получает дату закрытия интервала актуальности.

Вторым, не менее важным базовым объектом БД мониторинга служит почвенный профиль и индивидуальный почвенный образец. В зависимости от масштаба и точности исследований известно несколько подходов к накоплению информации по пробам почвы. Для масштаба 1:1000000 и выше в почвенных БД применяется отбор представительных профилей, характеризующих основные почвенные классификационные единицы в объеме генерализованной карты почвенных контуров [1]. Такой подход применяют в европейской БД и почвенно-географической БД России [3]. В БД США используется усредняющая техника, известная как «педон» [4]. Говорить о



Рис. 3. Иерархия структуры данных по организационно-хозяйственным (слева) и почвенно-генетическим (справа) показателям. Соединительными линиями обозначены связи (отношения) «один-к-многим» в реляционной БД

привязке «педона» или представительного разреза к земельному участку здесь не приходится. С другой стороны, в технологии точного земледелия с использованием автоматизированного отбора проб почвы (преимущественно пахотного горизонта) для каждого земельного участка создается точная сетка координат положения разрезов. В задаче мониторинга необходим синтез этих двух методик.

При обработке данных многолетних почвенных наблюдений системы НИИГипрозем на почвенных картах отмечены места заложения разрезов для получения представительных почвенных профилей обследуемого хозяйства. Эти точки заложения почвенных разрезов имеют определяемые географические координаты, но могут не быть расположены на территории конкретного элементарного земельного участка (ЭИЕ), но при этом принадлежать некоторому почвенному контуру. Кроме того, и при современных обследованиях возможно заложение серии разрезов на неиспользуемых или брошенных сельхозугодьях.

Таким образом, здесь обязательна иерархическая подчиненность разреза почвенному контуру (иерархическая подчиненность в данном тексте означает наличие отношения (реляции) типа «один-к-многим», то есть для одного контура может быть множество разрезов). С другой стороны, если известна принадлежность разреза земельному участку, то здесь также образуется

иерархическая подчиненность, но в этом случае не являющаяся обязательной. В терминах теории БД код земельного участка в описании разреза может иметь значение «не определено» (NULL).

Наличие указателя на иерархически высший уровень повышает информативность данных разреза, поскольку характеристики почвенного контура или земельного участка автоматически распространяются и на находящийся в них разрез. Данные по разрезу могут содержать также координаты точки заложения разреза. В работах [1,5] структура разреза представлена следующим образом: Почвенные характеристики → Разрез → Профиль → Горизонт → Образец почвы. Такой подход применяют в основном для крупномасштабных почвенных БД. Для задач мониторинга пять уровней иерархии являются избыточными, а все перечисленные данные (и классификация почвы, и характеристика места заложения разреза, и классификация профиля, и морфологическое описание горизонтов) разносятся по соответствующим уровням иерархической структуры профиля, схематически представленной на рис. 2. Данные по почвенному профилю в целом наследуются из объектов «контур» и «участок». Данные по каждому из обследуемых горизонтов входят в состав данных образца почвы.

В процессе мониторинга осуществляется как точечное обследование свойств

<p style="text-align: center;">Образец почвы</p> <ul style="list-style-type: none"> - Морфометрия - Химические свойства - Физические свойства - Водно-физические свойства - Показатели плодородия - Показатели засоления - Микроагрегатный состав - Экологические показатели 	<p style="text-align: center;">Почвенный профиль</p> <ul style="list-style-type: none"> - Координаты (долгота, широта, высота над уровнем моря), другие сведения
<p style="text-align: center;">Земельный участок</p> <ul style="list-style-type: none"> - Кадастровые показатели (ОКАТО, ИНН, координаты, шифры) - Природные показатели (уклон, тип угодья, намытость) 	<p style="text-align: center;">Почвенный контур</p> <ul style="list-style-type: none"> - Координаты границ - Природная зона - Почвенная провинция - Вид - Разновидность
<p>Данные мониторинга: Культура произрастания Параметры загрязнения Агрохимические мероприятия</p>	

Рис. 4. Тематические группы данных

почвы (разрез, образец почвы), так и обследование по площади (контур, участок). Поскольку разрез закладывается разово, то данные по разрезу и являются данными наблюдения. В случае обследования площади на одном и том же объекте проводятся многократные и разноплановые наблюдения (параметры загрязнения, вид культуры произрастания и урожайность, агрохимические мероприятия и т. д.). Поэтому необходим объект нижнего уровня иерархии, находящийся в отношении «многие-к-одному» с ЭИЕ. Этот объект назовем «наблюдение». На уровне этого объекта, собственно, и происходит регистрация первичных, измеренных данных мониторинга. Количество показателей мониторинга в целом превышает сотню. Эти показатели разбиты по тематическим группам, с тем чтобы разные пользователи, возможно из разных организаций, могли независимо друг от друга вносить и модифицировать данные. В реляционной БД каждой группе соответствует таблица (объект) данных, связанная отношением «один-к-одному» с соответствующим объектом «наблюдение». Так, например, группа химических свойств некоторого образца почвы связана отношением «один-к-одному» с показателями данного образца в целом, а группа показателей по агрохимическим мероприятиям для некоторого наблюдения земельного участка связана отношением «один-к-одному» с данными наблюдения этого участка на некоторую дату. На рис. 4 перечислены основные тематические группы показате-

лей, регистрируемых в БД мониторинга агроландшафтов.

Подход к формированию структуры БД, основанный на реляционной модели данных Кодда, применен при создании БД «SoilMatrix» (www.soilmatrix.ru) – базы данных состава и свойств почв Южного Федерального округа, содержащей данные по 60000 образцам почвы для 600 хозяйств и пяти регионов ЮФО [6]. Этими же принципами разработчики руководствуются в проводимой в настоящее время адаптации СУБД «SoilMatrix» к задачам широкомаштабного мониторинга агроландшафтов.

Масштабы одной статьи не позволяют описать другие составляющие реляционной модели данных, рассмотреть другие аспекты реализации основных положений реляционного подхода в БД «SoilMatrix». Ниже приведен примерный перечень вопросов, специфичных именно для почвенно-экологических баз данных, которые могут быть рассмотрены в следующих статьях.

- Объекты «справочники», их иерархия и информативность.

- Форматы измеряемых данных, их зависимость от методик измерения, точность измерения и область допустимых значений. Контроль формата вводимых данных.

- Генерация отчетов. Принципы формирования интегральных (по площади, по типу) показателей, задачи идентификации экземпляров наблюдений, педотрансферные правила.

- Интерфейсные решения для пользователей БД в Интернет. Типы пользовате-

лей, авторизация, разграничение доступа и уровни доверия к результатам измерений.

- Специфика идентификации ЭИЕ с использованием естественных почвенных кодов и анализ субстантивных данных на математической модели – матрице дисперсности почв.

Литература

1. Почвенно-географическая база данных России. Проект Общества почвоведов им. В. В. Докучаева // www.db.soil.msu.ru/general.htm

2. Codd, E. F. The relational model for database management: version 2. Addison-Wesley Longman Publishing Co., Inc., 1990. – 567 p.

3. Rossiter, D. G. A Compendium of On-

Line Soil Survey Information Soil Geographic Databases: Europe // [www. itc.nl/~rossiter/research/rsrch_ss_digital_eu.html](http://www.itc.nl/~rossiter/research/rsrch_ss_digital_eu.html)

4. Field Book for Describing and Sampling Soils Version 2.0. National Soil Survey Center Natural Resources Conservation Service. U. S. Department of Agriculture, 2002. – 228 p.

5. Рожков, В. А. Методы оценки информативности почвенных признаков и сравнения классификаций: презентации докладов конференции «ЭКОМАТМОД-2009» // www.lem.edu.mhost.ru/doc/presentations/Rozhkov.pdf.

6. Крыщенко, В. С. Базы данных состава и свойств почв / В. С. Крыщенко, О. М. Голозубов, В. В. Колесов, Т. В. Рыбьянец. – Ростов-на-Дону: Изд-во РСЭИ, 2008. – 145 с.