

МЕТОДИЧЕСКИЕ АСПЕКТЫ СОЗДАНИЯ ПОЧВЕННО-АТТРИБУТИВНОЙ БАЗЫ ДАННЫХ*Н. И. Белоусова¹, Ю. Л. Мешалкина²*¹ Почвенный институт им. В.В. Докучаева Россельхозакадемии² Факультет почвоведения МГУ им. М.В. Ломоносова

На примере базы данных Voreal 2,0, включающей химические и физико-химические свойства бореальных почв России, обсуждаются методические вопросы, возникающие при создании и развитии атрибутивных баз данных в почвоведении. Описаны процедуры унификации собранного материала и обсуждены вопросы неизбежной потери информации. Проведена географическая, классификационная и аналитическая оценка собранного материала. Введены понятия двух типов баз данных – «рабочий стол» и «камера хранения».

В 1997 г. авторами были опубликованы методические подходы к созданию почвенно-атрибутивной базы данных Voreal (далее БД Voreal), основанные на опыте создания и научно-исследовательской работы с атрибутивной базой данных достаточно большого объема в почвоведении (Белоусова, Мешалкина, 1997). За прошедшее десятилетие популярность почвенных баз данных различной направленности увеличилась (Алябина и др., 2008; Крыщенко и др., 2008; Рыжова, Подвезенная, 2008; Digital..., 2006). В значительной мере это обусловлено сокращением во всем мире и, в частности в России, экспедиционных исследований, а также повсеместным распространением ГИС и появлением цифровой почвенной картографии (ЦПК). В результате поле научной деятельности сместилось к обобщению, систематизации и осознанию большого объема уже накопленного материала (Возможности..., 2000). Появилась и получила распространение новая классификация и диагностика почв России (2004). Произшедшие изменения, а также возникающие новые научные задачи потребовали развития и расширения ранее созданной БД Voreal. В результате была создана новая версия базы данных, которую условно назвали БД Voreal 2,0. Обобщение опыта проделанной работы представлено в предлагаемой «новой версии» методических подходов расширения и развития атрибутивных баз данных в почвоведении. Таким образом, цель работы – обсуждение вопросов, возникающих при создании и развитии почвенно-атрибутивных баз данных, с особым вниманием к вопросам, возникающим при расширении ранее созданных массивов.

По характеру использования почвенные базы данных можно разделить на два типа. Назначение баз первого типа – «библиотечное»: сбор, хранение, систематизация и быстрый поиск нужных материалов. Назовем такой тип базы данных «камера хранения». Примером ее, видимо, может слу-

жить Информационная база классификации, созданная более 20 лет назад в Почвенном институте имени В.В. Докучаева (Шишов и др., 1985) или создаваемая в настоящее время Общенациональная географическая информационная почвенная базы данных в масштабе 1:2 500 000 (Алябина и др., 2008). В «камеру хранения» обычно собирается весь доступный материал. Второе направление – создание тематических рабочих баз данных, направленных на решение определенного вопроса или нескольких вопросов. Назовем базы данных такого типа «рабочим столом» или просто «рабочей базой данных». Ниже пойдет речь именно о базах данного типа, к которому относятся ранее созданная БД Voreal и ее расширенный вариант БД Voreal 2.0, а также практически любая база данных, создаваемая для решения конкретного научного или практического вопроса.

Как было показано ранее (Белоусова, Мешалкина, 1997), процесс создания любой почвенно-атрибутивной базы данных включает 4 главных этапа: 1) концептуальный, 2) сбор, унификация и структуризация информации, 3) верификационные оценки, 4) аттестация информационного содержания массива данных. Остановимся на каждом из них подробнее.

Концептуальный этап. Процесс создания любой базы данных начинается с формулирования цели и описания задач, которые предполагается решать с помощью создаваемого массива. Поставленная цель определяет объем базы (количество разрезов), ее содержание (набор параметров), пространственные границы охватываемой территории, где расположены разрезы, различного рода ограничения при отборе данных. Подчеркнем, что ограничения, обусловленные целью создания базы данных, являются отличительной особенностью рабочей базы данных¹. Они могут касаться любого аспекта: ограничиваться могут как разрезы, отбираемые в базу данных, так и набор параметров, которыми эти разрезы описаны. Целью БД Voreal было выяснение, имеют ли количественные значения химических и физико-химических свойств бореальных мезоморфных² почв классификационную или пространственную локализацию. Соответственно, собранный массив данных имел пространственные границы (бореальная зона России) был ограничен с точки зрения экологии (мезоморфные почвы бореальной зоны) и имел ограничение по набору аналитических параметров.

Развитие базы данных может проводиться по разным направлениям: увеличение числа разрезов в рамках прежних требований; увеличение чис-

ла характеристик; частичное или полное снятие ограничений. БД Voreal 2.0 по сравнению с первым вариантом дополнена разрезами почв: а) расположенными в труднодоступных и малоизученных районах России с низким насыщением материалами; б) разрезами в мерзлотных резкоконтинентальных таежных, но не гумидных, а семигумидных и даже семиаридных районах; в) разрезами пахотных почв с аналитическими характеристиками, принятыми для естественных почв; г) почвами легкого гранулометрического состава (песчаного, супесчаного и легкосуглинистого). В дальнейшем планируется расширить экологические границы, включив в БД Voreal также почвы разного геохимического положения (автономного, транзитного, гетерономного) и водного режима (мезоморфного, полугидроморфного, гидроморфного).

В БД Voreal 2.0 относительно первого варианта были введены дополнительные аналитические характеристики: содержание общего азота, плотность сложения, пористость, групповой и фракционный состав гумуса.

Достаточно часто встречаются ситуации, когда собираемые параметры занимают в базе данных разный объем и носят разный характер. В таких случаях в рамках общей базы данных целесообразно создавать тематическую подбазу. Так, групповой и фракционный состав гумуса в почвах анализируется гораздо реже, чем общие показатели, а в работах, освещающих состав гумуса, часто отсутствуют или присутствуют в ограниченном наборе параметры, принятые в БД Voreal; в определениях группового состава гумуса наблюдается большое разнообразие методов и модификаций. Все это делает неудобным включение группового состава гумуса в общую почвенно-атрибутивную базу данных.

Таким образом, выделяются этапы формирования БД Voreal, что позволяет проводить анализ данных на каждом этапе развития до окончательного ее формирования (Белоусова, Васенёв, 1984; Белоусова, Мешалкина, 2000; Тонконогов, Белоусова, 2002; Белоусова, 2006).

Расширение и развитие базы данных может привести к такому неоперативному объему и структуре, что исследователь будет непосредственно работать только с частью базы. Иными словами первоначальная база типа «рабочий стол» превратится в «камеру хранения», и только какая-то ее часть будет рабочей – «рабочим столом».

В рамках одной базы данных нецелесообразно объединять материалы генетического и, скажем, специфического агрохимического, почвенно-биологического или какого-нибудь другого направления почвоведения (на «всякий случай»), если это не обусловлено задачами самой базы. Немотивированное расширение базы данных, как и любых других информационных источников, с некоторого «момента насыщения» затрудняет работу, не увеличивая точность и информативность получаемых ответов. Напротив, разъединение специфических данных обеспечивает рациональный

¹ База данных типа «Камера хранения» в общем случае не должна иметь ограничений.

² Мезоморфные почвы – почвы свободного внутрипрофильного дренажа независимо от положения в рельефе, развитые в гумидных климатических областях (Соколов, 1987).

объем и упрощение структуры базы данных, а значит, повышение ее операбельности. Этот пример не означает исключение из базы данных генетической направленности антропогенных почв как объекта генетического направления и запрет создания межотраслевых баз данных (почвенно-агрохимических, почвенно-климатических, почвенно-растительных и пр.). Напротив, такие смешанные базы данных открывают возможности для новых научных поисков.

Формулирование критериев отбора исходного материала является следующим важным моментом при создании или развитии почвенно-атрибутивной базы данных. Первым и обязательным условием включения разреза в базу данных является наличие его морфологического описания с указанием экологического и географического адресов. Эта информация, с одной стороны, служит объектом научного анализа, а с другой, выполняет вспомогательную функцию – она важна для контроля другой, в том числе аналитической, информации из базы данных, а также для оценки получаемых результатов и др.

Следующий критерий связан с минимальной информацией о разрезе как условии его включения в базу данных. Например, каждый разрез, включенный в базу данных, должен быть охарактеризован каким-то минимумом анализов. В БД Vogeal включались разрезы, для которых были указаны 6 из 9 собираемых аналитических признаков. Этот критерий необходим для проверки «качества» собираемых анализов и исключения неявных ошибок. Например, низкие значения рН должны сопровождаться высокими показателями гидролитической кислотности и поглощенного водорода, а высокие значения ЕКО не могут соответствовать почвам легкого гранулометрического состава и т.д. Если эти соотношения нарушены, в анализах заключена ошибка. Ошибку не всегда можно устранить. Такие разрезы или только сомнительные данные выбраковываются.

Все почвенные, агрохимические и другие анализы должны быть выполнены по общепринятым методикам. Очень важны сведения о методах анализа или модификации стандартного метода, но в публикациях они редки, и обязательно должны фиксироваться³. При их отсутствии по умолчанию принимается стандартный метод, что может не соответствовать действительности. Разнобой методов и их модификаций увеличивает статистический шум и может исказить или завуалировать характер статистических закономерностей.

Этап сбора, унификации и структуризации информации – самый сложный и самый ответственный. Он состоит из нескольких шагов нахождения

материалов, удовлетворяющих критериям отбора и унификации собранных материалов.

Нахождение материалов, удовлетворяющих критериям отбора – процесс обычно трудоемкий. В БД Vogeal включались материалы за период более 50 лет. За это время не один раз менялись классификационные подходы, которые становились раз от раза более детальными и сложными. Изменилось и усложнилось «видение» почвы и почвенного покрова от «типичной зональной почвы» до спектра зональных почв и структуры почвенного покрова. Менялась также система описания почвенного профиля от простой А–В–С до сложной, многоуровневой; были выделены новые горизонты, менялись признаки, имеющие диагностическую значимость. Например, в конце 50-х годов подбуры были выделены в самостоятельный тип и объединили почвы, однотипные по строению профиля, но прежде частично входившие в разные типы (Таргульян, 1971). Одним из важных диагностических признаков подбуров является наличие аморфных органо-минеральных образований на гранях педов и минеральных отдельностей. До конца 50-х годов этот признак практически не описывался. Подобных примеров можно привести много. Естественно, что собранный таким образом материал требует переоценки с единых позиций, представляет серьезные трудности и часто оставляет большое «поле неопределенности».

Унификация собранных материалов проводится на основе определенной классификации почв и соответствующей ей индексации горизонтов. В табл. 1 приведен пример возможной унификации для двух типов почв: дерново-подзолистых и подбуров. Она выполнена в системе Программы Почвенной карты 2,5 млн. масштаба на основе авторских описаний разрезов и для целей широкого географического обобщения. Унификация авторских разрезов в зависимости от поставленных задач может быть выполнена с разной степенью детальности.

На разных стадиях обобщения почвенных материалов происходит неизбежная *потеря информации*. Первый раз информация теряется еще до создания базы данных при переходе от описания индивидуального разреза к определению его классификационного положения, так как любая, даже самая мелкая классификационная ячейка представляет определенное обобщение индивидуальных разрезов и характеризуется диапазоном свойств. Следующая потеря информации происходит в процессе создания базы данных на этапе ее унификации. Здесь могут возникнуть две ситуации. Первая возникает, когда в публикации описание почвы, данное в прежней классификации, содержит информацию, достаточную или даже избыточную в рамках новой классификации. Эта редкая, но возможная ситуация. В этом случае кодирование почвенного описания в новой системе может сопровождаться повышением (уточнением) информации относительно ее прежнего классификационного положения. Вторая ситуация

³ Сбор и обработка таких сведений может явиться основой для изучения корреляции между аналитическими методами.

Таблица 1. Пример унификации авторских названий почв

Авторское название почв	Название в БД Boreal	
Дерново-подзолистая	Дерново-подзолистые	
Дерново-палевоподзолистая		
Дерново-подзолистая со вторым гумусовым горизонтом		
Горная дерново-подзолистая		
Дерново-подзолистая глееватая		
Горная лесная бурая псевдоподзолистая		
Бурая лесная оподзоленная (при A2 > 10 см)		
Бурая оподзоленная слабogleеватая (при A2 > 10 см)		
Охристый подбур		Подбуры
Горно-лесная кислая неоподзоленная		
Горно-лесная неоподзоленная		
Примитивно-аккумулятивная		
Горно-тундровая перегнойная		
Горно-тундровая дерново-перегнойная		
Горно-тундровая торфяно-перегнойная		
Мерзлотно-таежная		

возникает, если описание разреза и индексация его горизонтов с позиции современной классификации и диагностики почв выполнены ограниченно. Тогда, привлекая косвенную информацию, разрез можно адаптировать в современную классификационную систему, при этом неизбежно происходит принятие субъективного решения. Такая введенная информация, по существу, является смещенной или может оказаться ложной.

Аналогичным образом происходит потеря информации на этапе унификации горизонтов и упрощения строения профиля. В настоящее время диагностику почвы принято сопровождать формулой морфологического строения профиля, которая включает все обязательные для данного типа диагностические горизонты. В реальных описаниях почвенных разрезов всегда было распространено выделение подгоризонтов (например, A', A'', A'''; A21, A22; BT1, BT2 и т.п.) или переходных горизонтов (AOA1, AOBhf, A1A2, A1/ВН и т.п.). Такое разделение выполнено обычно в произвольной системе. Поэтому в собранной базе данных образуется громадное разнообразие записей горизонтов, индексация которых часто не соответствует современному содержанию. Понятно, что такой материал также требует стандартизации. В табл. 2 приведен пример унификации органо-генных горизонтов, сделанный на основе описаний разрезов и химических анализов.

Таблица 2. Унификация записи органо-генных горизонтов

Авторские обозначения	Обозначения в БД Boreal
О, Т, ОА, ОА2, Дерн, А дерн, АО, АОТ, АТ, А1*, АОА2А1, А1А2*, А1АОА2*, А1АО*, АОА1*, АОА2*, А1, А1h	О
А1, АОА1**, А1АО**, АОА1В**, А1-АО**, А1А2, А1А2А1, А1А2L, А1А2ф, А1бурый, А1В, А1коричнев, А дер**, Аf, АОА2А1**, А1А2	А1

* При содержании валового гумуса > 30%.

** При содержании валового гумуса < 30%.

В заключение этого раздела подчеркнем, что нет, и не может быть стандартной процедуры перевода разреза из одной классификационной системы в другую. В процессе развития классификаций одни почвенные типы дробились и расходились в разные ячейки, другие – объединялись; в слабо изученных районах описывались новые типы и даже направления почвообразования (например, грануземы, криоземы). Все это приводит к тому, что работать приходится практически с каждым разрезом. Работа опирается, прежде всего, на морфологическое описание разрезов, при необходимости привлекаются сведения об экологическом положении разреза и аналитические показатели (см. требование наличия морфологического описания для включения разреза в базу данных).

Таким образом, нахождение материалов и включение их в базу данных требует *одновременной* оценки, редакции и кодирования по заранее принятым унифицированным нормам. Эти процессы стратегически неразрывны во времени и касаются как классификационного, так и горизонтного уровней организации почв.

Собранный материал требует определенной организации – *структуризации, подчиненной цели создаваемой базы данных*. В БД Boreal была принята многоуровневая пространственная организация: 1) уровень крупных регионов (их выделено 8), 2) для ЕТР материал дополнительно организован по административным районам и 3) в базу включались материалы ключевых участков. Такая организация материалов позволила характеризовать территории разных масштабов (Meshalkina et al., 1997; Белоусова, 2006; Белоусова, Мешалкина, 2007).

Этап верификационных оценок. После того, как данные введены в базу, требуются разнообразные проверки созданной записи. Сомнительные величины и свойства данных могут быть вызваны разными причинами и поэтому требуют разного к ним отношения. Так, количественные технические ошибки, как правило, выявляются с помощью простейших тестов.

Например, сумма поглощенных катионов не может быть равна 0, а в гранулометрическом составе сумма фракций должна равняться 100%.

Логические разночтения в данных заключаются в несовместимости, имеющейся в источниках информации. Например, наличие гор. А2 в профиле почвы, названной авторами «неоподзоленной» или дерново-подзолистая почва, в профиле которой не выделяется гумусовый горизонт. Названия органогенных горизонтов (подстилки и гумусового) могут не соответствовать содержанию гумуса. Такие разночтения требуют сопоставления всей имеющейся информации. Для них нет общих способов обнаружения и коррекции. Поэтому для критической оценки важно наличие морфологического описания разреза, которому отдается приоритет.

Однако не менее важно не выбраковать данные из массива данных, которые могут восприниматься как сомнительные из-за противоречий с установившимися стереотипами. Например, некоторое количество почв или горизонтов имеют насыщенный поглощающий комплекс при низких значениях рН (Белоусова, Васенёв, 1984). Это не так часто реализуемое в природе явление не противоречит физическому смыслу понятия ненасыщенности, т.е. относительному содержанию в поглощающем комплексе поглощенного водорода, определяемого по Гедройцу (Роде, 1937), но часто воспринимается специалистами как ошибочное аналитическое определение.

Этап аттестации информационного содержания массива данных представляет собой первый шаг научного анализа. Он включает географическую, классификационную и аналитическую оценку собранного материала. Его назначение оценить, что представляет собой созданная база с позиций поставленных задач.

Географическая оценка – это анализ пространственного размещения разрезов, включенных в базу данных. Табл. 3 показывает существенную локализацию данных БД Voreal 2,0. Это связано, во-первых, с тем, что в базу включались результаты детальных и крупномасштабных исследований. Помимо этого обнаружилось, что исследования разных лет и разных авторов расположены в одних и тех же «привлекательных местах». Ими являются, например, заповедники, стационары и опорные пункты разных научных учреждений. Вблизи крупных научных центров изученность территории гораздо выше, чем в удаленных районах. Так, более 70% разрезов, относящихся к Европейской территории России, расположено примерно поровну в пяти областях: Московской, Смоленской, Архангельской, Кировской – и республике Коми. При этом половина разрезов Московской области находится на землях Зелиноградского стационара Почвенного института им. В.В. Докучаева, а разрезы Смоленской области – совхоза «Михейковский». Можно сказать, что характер собираемого массового материала в существенной мере носит точечный характер

Таблица 3. Распределение разрезов в БД Voreal 2,0 по регионам

Регион	Число разрезов	
	шт.	%
Европейская территория России	257	33
Урал	85	11
Западная Сибирь	44	6
Средняя Сибирь	119	15
Алтайе-Саянский массив	66	9
Дальний Восток	113	14
Вулканические области	35	5
Восточная Сибирь и Забайкалье	57	7
Итого	776	100

и в ближайшее время таким и останется. Это, в свою очередь, означает, что основным методом территориальных (например, картографических) работ с использованием материалов атрибутивной базы данных остается экстраполяция.

Географический анализ базы данных Voreal показал также, что в разных регионах России работало разное количество исследователей, иногда 2–3 и даже только 1, так что влияние личности исследователя на географию и генезис почв регионов соответственно разное. Эта ситуация еще резче проявляется на классификационном уровне. Во всех российских классификациях немало выделов почв, описанных одним автором и не подтвержденных другими исследователями. Один из ярких примеров – грануземы, описанные И.А. Соколовым (Соколов, Градусов, 1981) в низовьях Нижней Тунгуски и плато Путоран. Выделен тип и подтипы. Опубликованы данные по пяти разрезам. За рамками задач БД Voreal стоит немало классификационных выделов, никак в литературе не охарактеризованных, выделенных провизорно.

Аналитическая насыщенность БД Voreal тоже очень неравномерна. Наиболее широко распространены определения гумуса, поглощенных катионов и рН солевой суспензии – ими охвачено около 90% всех разрезов базы, рН водной суспензии – менее 80%, гранулометрический состав – около 60%, а гидролитической кислотности – около 50%. Это обстоятельство имеет прямое отношение к наполнению базы, соответственно снижая ее интерполяционные возможности. Если в ходе научного анализа планируется ранжировать какой-то параметр по содержанию илистой фракции, возможности уменьшаются примерно вдвое.

Таким образом, анализ содержания БД Voreal показал крайне неравномерную изученность регионов и классификационных выделов, как и резко разную распространенность аналитических параметров.

В заключение отметим, что почвенно-атрибутивные БД позволяют ответить на очень многие научные и практические вопросы, на существенно большее их количество, чем планировалось изначально, но *не могут ответить на все вопросы*. Любая база данных – *только метод научного анализа*. Уровень получаемых результатов зависит от качества собранного материала и уровня ввода его в базу данных.

СПИСОК ЛИТЕРАТУРЫ

Алябина И.О., Шоба С.А., Урусевская И.С., Маречек М.С. Создание общенациональной географической информационной почвенной базы данных в масштабе 1:2 500 000 // Мат-лы V съезда Всерос. общ. почвоведов им. В.В. Докучаева. Ростов-на-Дону, 2008 г. С. 215

Белоусова Н.И. Оксалаторастворимые соединения Al, Fe и Si в почвах холодных гумидных областей как функция выветривания // Почвоведение. 2006. № 1. С. 7–18.

Белоусова Н.И., Мешалкина Ю.Л. Тип органопрофиля и генетический тип почвы // Лесное почвоведение: итоги, проблемы, перспективы. Тез. Докл. Междунар. науч. конф. Сыктывкар, 2007. С. 15–16.

Белоусова Н.И., Васенёв И.И. Кислотность и ненасыщенность поглощающего комплекса автономных суглинистых почв таежной зоны СССР // Почвы и почвенный покров лесной и степной зон СССР и их рациональное использование. Науч. тр. Почв. ин-та им. В.В. Докучаева. М., 1984. С. 132–142.

Белоусова Н.И., Мешалкина Ю.Л. Опыт создания унифицированной базы данных бореальных почв России (методические вопросы) // Почвоведение. 1997. № 8. С. 926–933.

Белоусова Н.И., Мешалкина Ю.Л. Некоторые параметры катионно-обменного комплекса бореальных почв России // Почвоведение. 2000. № 8. С. 951–965

Возможности современных и будущих фундаментальных исследований в почвоведении. М.: ГЕОС, 2000. 139 с.

Классификация и диагностика почв России. Смоленск: Ойкумена, 2004. 342 с.

Классификация и диагностика почв СССР. М.: Колос, 1977. 223 с.

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

Программа Почвенной карты СССР масштаба 1:2 500 000. М.: Почв. ин-т им. В.В. Докучаева ВАСХНИЛ, 1972. 158 с.

Роде А.А. Основные черты почвообразовательного процесса // Почвоведение. 1937. № 6. С. 849–862

Рыжова И.М., Подвезенная М.А. Пространственная вариабельность запасов органического углерода в почвах лесных и степных биогеоценозов // Почвоведение. 2008. №12. С. 1429–1437.

Соколов И.А. О некоторых сравнительно-генетических понятиях и темах // Почвоведение. 1967. № 10. С. 26–39.

Соколов И.А., Градусов Б.П. Об экзогенезе в области широкого распространения основных пород // История больших озер Центральной субарктики. Новосибирск: Наука, Сиб. отд-ние, 1981. С. 57–99.

Таргульян В.О. Почвообразование и выветривание в холодных гумидных областях. М.: Наука, 1971. 268 с.

Тонконозов В.Д., Белоусова Н.И. География и генезис железистых аккумуляций в почвах тундровых и таежных областей России // Почвоведение. 2002. № 6. С. 654–662

Шишов Л.Л., Рожков В.А., Столбовой В.С. Информационная база классификация почв // Почвоведение. 1985. Т. 9. С. 9–20

Digital Soil Mapping – An Introductory Perspective / Ed. Ph. Lagacherie, A. McBratney, M. Voltz. Elsevier, 2006. 600 p.

Meshalkina J.L., Belousova N.I., Vasenev I.I. Soil Absorbing Complex properties of Russian boreal soils and its dependence on the spatial scale of study area // Soil and Water Quality at Different Scales. Netherlands: Kluwer Academic Publishers, 1997. P. 35–49.