

МЕЛИОРАЦИЯ

УДК 631.6

ОРГАНИЗАЦИЯ МОДЕЛИ ДАННЫХ В ГИС ИЗМЕНЯЮЩИХСЯ ХАРАКТЕРИСТИК МЕЛИОРАТИВНЫХ СЕЛЬСКОХОЗЯЙСТВЕННЫХ ОБЪЕКТОВ

Н.Н.Вахонин, кандидат технических наук
РУП «Институт мелиорации»

Ключевые слова: информационное обеспечение, книга истории полей, геоинформационная система, модель данных, агрегирование

Введение

При организации хранения информации в различных предметных областях необходимо учитывать, что зачастую помимо пространственной распределенности данных имеет место их изменчивость во времени, что в общем случае обеспечивается при реализации информационной системы на основе разработанной геореляционной концепции [1, 2].

Хранение данных по процессам, имеющим большие характерные времена изменения, может быть реализована в ГИС. В частности в ГИС может быть организовано хранение информации по имеющим годовую и большую периодичность осуществления бизнес-процессам в различных подсистемах мелиорированных сельскохозяйственных объектов: книга истории полей, данные по турам обследований земель (подсистемы мелиорируемые земли и сельскохозяйственная растительность), паспортизация и кадастр технического состояния по турам инвентаризаций, обследований и т.п. (подсистема мелиоративная сеть).

Однако, так как ГИС оптимизированы на хранение пространственно-распределенной: географической, а не исторической информации, то возникает проблема организации структуры данных с учетом временной оси. Решение этой проблемы рассматривается ниже на наиболее характерном примере – автоматизированной книге истории полей, реализация которой может быть осуществлена на основе различных базисных программных средств: электронных таблиц, баз данных и т.д., но наиболее эффективна в ГИС при соблюдении во всех случаях неременного требования - атомарности используемых двумерных таблиц (первая нормальная форма).

Анализ и выбор структуры данных автоматизированной книги истории полей на основе ГИС для обеспечения эффективного их хранения и агрегирования

Характеристики определяющие содержание книги истории полей агропредприятия можно подразделить на названия сущностей (мерности осей гиперкуба, по которым необходимо вращать данные): с/х поле, вид культуры, год и однозначно определяемые ими атрибутивные данные, которые условно можно классифицировать на «данные причина» - управляющие технологические воздействия, естественно-природные характеристики и случайные погодные, гидрологические процессы в вегетационный период и «данные следствие» - данные мониторинга почвы и растений, во время вегетации и результирующая урожайность. Ведение книги истории полей в автоматизированном виде обеспечивает возможность осуществления эффективных расчетов, т.е. не просто сбора данных, а задействования их для анализа, оценки состояния и принятия оптимальных решений по различным факторам урожайности (севообороты, мелиорация, удобрения и т.д.) и формирования отчетов.

Проблемы выбора структуры данных в ГИС возникают при необходимости хранения объектов, изменяющихся в пространстве, времени и т.п., т.е. имеющих n-мерный набор системообразующих свойств, тогда как плоская реляционная таблица для хранения данных двумерна.

В частности это имеет место при организации хранения данных по урожайности сельскохозяйственных полей, являющихся типичными площадными объектами ГИС. Данные по полям необходимо хранить по годам, так как во времени их основное свойство – выращиваемая культура, а также дозы вносимых удобрений, водный режим, другие управляющие факторы и получаемая урожайность от года к году меняются. Очевидно, что строго говоря, в этих условиях объектом является не просто поле (пространственно, если отсутствует перенарезка, оно неизменно), а поле-год, что обеспечивает учет изменения его системообразующих свойств во времени.

Проблемы и альтернативные варианты организации хранения информации в ГИС связаны с тем, что атрибутивные данные книги истории полей однозначно идентифицируется тремя сущностями (сложный ключ): поля (n полей), годы наблюдений (T лет), культуры (m культур). В то же время dbf таблица двумерная: одна мерность – колонки (ограниченные в dbf числом 256) стандартно задействованы под набор атрибутивных данных, а вторая – строки задействованы под пространственные объекты. Поэтому остальные две сущности должны распределяться между другими возможными мерностями информационной системы. В частности в файле проекта ГИС это имена тем и имена видов.

Таким образом стандартно в одном проекте ГИС могут храниться данные имеющие четыре мерности. Однако учитывая то, что процедуры обработки данных организованы внутри вида, то для удобства расчетов в проекте целесообразно хранить данные по отдельной проблеме в одном виде. В результате этого в ГИС реально задействуются три мерности: набор тем, внутри которых двумерные таблицы. Такая организация и

представляет модель данных векторных тем ГИС.

В случае когда размерность подлежащей хранению информации больше, выходом является составное (неатомарное) именование любой мерности: колонок, строк, тем, проектов. При этом любая из мерностей посредством введения неатомарного наименования имени проекта, темы, столбца, строки, может быть задействована под хранение двух, трех, четырех и т.п. мерностей. Однако в результате этого усложняется возможность реализации запросов по выбору данных из таблиц.

Попутно отметим, что в файле (книге) Excel мерности реализованы:

- строки и колонки плоской таблицы – 2 мерности;

- страницы Excel – 1 мерность;

В реляционной базе данных мерности реализованы:

- каждая плоская таблица базы – 2 мерности, связанные между собой по ключевым полям.

Учитывая вышеприведенные проблемы, связанные с многомерностью данных, возможны различные варианты организации их хранения (распределения мерностей) в ГИС, в частности по урожайности культур на полях за многолетие, различающиеся удобством анализа хранимой информации. Очевидно, что выбор варианта определяется планируемыми расчетами (вычислительными моделями) для которых необходимо будет формировать различные структуры данных. Для расчета экономики текущего года (или планируемого) необходимы данные по всем технологическим операциям – определяют затраты и полученная (планируемая) урожайность – определяет доходы. Для балансовых расчетов по химическим элементам необходимы данные по содержанию каждого из них в почве, вносимым дозам и урожаю для расчета выноса.

Для решения задачи оптимизации сельхозиспользования необходимы многофакторные модели урожайности для определения которых требуются данные по всем урожаеобразующим факторам как по управляющим воздействиям, так и случайным погодноклиматическим воздействиям, а также по урожайности культур, незаменимым источником которых является книга истории полей.

При этом следует отметить, что так как расчеты могут осуществляться только используя данные собранные внутри одного вида текущего проекта, то во всех рассматриваемых далее вариантах вся информация по всем полям, культурам хозяйства за многолетие должна компоноваться в одном виде проекта.

Вариант 1 – покультурного хранения данных в проекте ГИС (табл. 1):

вид в ГИС – всё хозяйство за многолетие;

ГИС-тема – культура (число тем = m -культур);

объект – сх поле (число строк dbf-таблицы = n -полей);

колонки dbf таблиц: год-урожай, год-N, год-P, год-K, год- $T_{\text{воз}}$ (число столбцов = $P_{\text{атр}}$

$\times T_{\text{лет}}$).

Пример: тема "Пшеница" (имя dbf таблицы 1).

Таблица 1

Имя таблицы: «Пшеница»							
IDN	Имя сх поля	уро- жай 2001	уро- жай 2002	уро- жай 2003	Н 2003	Р 2003	К 2003
1	П1						
2	П2						
3	П3						
4	П7						
5	П9						

Таблица 2

Имя таблицы: "Поле 1-пшеница"

IDN	год	урожай ц/га	Н	Р	К	...
1	П1 2001					
2	П1 2003					
3	П1 2005					

Таблица 2

Имя таблицы: "Поле 2-пшеница"

IDN	год	урожай ц/га	НПК
1	П2 2002		
2	П2 2004		
3	П2 2006		

Выборка генетически однородных данных для каждого из полей за многолетие

Для расчета статистик, скользящих средних, многофакторных регрессионных зависимостей урожая от урожаяеобразующих факторов и т.п. необходимы генетически однородные данные за все годы. Для получения (запроса) генетически однородных выборок данных по конкретному полю-культуре из dbf Таблицы 1 каждой темы должно быть образовано n Таблиц 2, где строки представляют временную координату – год, а колонки – атрибуты этого поля-культуры для данного года. При наличии m культур и n полей в хозяйстве всего необходимо получить таблиц 2 в количестве $N = m \times n$ с генетически однородными данными для расчета по ним различных характеризующих поле-культуру временных характеристик, моделей урожая.

Вариант 2 – полевого хранения данных (табл.3). Сформированные вышеприведенным образом таблицы 2 не пространственные, а временные и при стандартном подходе строго говоря, не могут быть темами ГИС, но искусственным приемом введения дополнительной, учитывающей пространство-время, колонки со сложным именем – «конкретное поле – год» (где конкретное поле – это поле для которого организована данная тема, т.е. поле именем которого она названа) можно сделать таблицу 2 таблицей ГИС темы.

Очевидно, что такая организация тем с атрибутивными таблицами 2 представляет собой отдельный вариант, особенностью которого является то, что в атрибутивных таблицах собраны генетически однородные данные, в результате чего они могут сразу

без предварительных выборок использоваться в различных статистических расчетах.

В варианте 2:

- вид в ГИС – всё хозяйство за многолетие;
- ГИС тема – конкретное поле-культура за многолетие (в виде n*m тематических слоев по числу поле-культур);

- объект – поле-год. В dbf таблице будет T_{лет} строк (хотя при отсутствии перенарезки полей по годам физически в теме будет один объект – данное неизменное по годам конкретное поле);

- колонки dbf таблицы - атрибуты данного поля-культуры (количество P_{атр}).

Формирование таблицы 2 из таблицы 1 весьма проблемно, так как для забора из неё данных в соответствующую строку таблицы 2 в связи с неатомарностью названия колонок (шапки) таблицы 1 их надо парсить. Поэтому этот вариант хранения для последующего формирования необходимых структур данных не удобен.

Вариант 3 – погодичного хранения данных (табл.3):

- вид в ГИС – всё хозяйство за многолетие;

- ГИС-тема – книга истории полей за один год (т.е. тем столько, сколько лет);

- объекты – сх поля (строки dbf-таблицы);

- колонки dbf таблиц – атрибутивные данные, включая название культур, дозы удобрений, урожаи и т.д. конкретного года.

Таблица 3. Имя таблицы: "Книга истории полей хозяйства за 1991 год "

IDN	Имя сх поля	Культура 1991	Урожай, ц/га 1991	N 1991	P 1991	K 1991
1	П1					
2	П2					
3	П3					
	...					
	П9					

Несмотря на то, что год указан в имени темы, для возможности объединения при желании ряда таблиц за разные годы в одну, целесообразно делать сложные имена колонок с добавлением года из названия таблицы в имя каждой колонки (так как год один и тот же, что и в названии таблицы, то имя становится составным номинально).

Вариант 3 очевидно лучше варианта 1, так как в нем название имен столбцов атомарны, что обеспечивает возможность осуществлять выборку используя стандартные SQL запросы без необходимости парсить имена колонок. Ещё одно преимущество этого варианта – ограниченное число необходимых столбцов, количество которых в dbf таблицах лимитировано. Сложность только в том, что для получения из таблицы 3 таблиц 2 для поля-культуры за многолетие SQL запрос должен быть не к одной ГИС теме (году), а к T_{лет} тем одновременно, но запрос стандартный, так как не надо парсить имена столб-

цов, в связи с чем для осуществления межгодовых расчетов (статистических расчетов, балансов по годам и т.п.) необходим межтабличный калькулятор таблиц разных тем.

Вариант 4 – недифференцированного по годам хранения данных по хозяйству за все годы в одной теме (табл.4):

вид в ГИС – моно-тема: все культуро-годы хозяйства за многолетие (таблица 4)

объект – сх поле (строки dbf-таблицы количество n)

колонки dbf-таблицы – атрибуты-годы (годо-культуры, годо-дозы, годо-урожаи), имеют составные неатомарные имена столбцов количеством $P_{атр} * T_{лет}$

Таблица 4. Имя таблицы: «Книга истории полей хозяйства за многолетие»

ID N	Имя сх поля	Культура 2001	Культура 2002	Культура 2003	N 20 01	N 20 02	N 20 03	P 20 01	Урожай 2001	Урожай 2002	Урожай 2003
1	П1												
2	П2												
3	П3												
4	... П9												

Преимущество варианта 4 заключается в том, что компоновка данных в одной dbf таблице позволяет в частности автоматически рассчитывать с помощью табличного калькулятора новые колонки типа баланса между годами, среднее за многолетие и т.п. Недостатки при выборе из табл. 4 генетически однородных данных за все годы для одного поля-культуры: в табл. 2 имеет место характерная проблема неатомарных имен столбцов – невозможность использования стандартных запросов и необходимость парсить имена колонок для нахождения значения года для каждого атрибутивного данного.

Ещё один недостаток варианта 4 – потребность в большом числе колонок для размещения атрибутивных данных, равно $p_{атр} * T_{лет}$, где $p_{атр}$ – число фиксируемых характеристик, тогда как число колонок в dbf таблице ограничено 256.

Формирование табл. 4 возможно либо вручную, либо автоматизированно процедурой соединения годовых таблиц 3.

Вариант 5 – хранения данных по хозяйству за все годы в одной теме с погодичными полями:

Вид в ГИС – моно-тема: все поле-годы хозяйства за многолетие (таблица 5)

- объект – годо-поле (строки dbf таблицы количество $n * T_{лет}$)

- колонки dbf таблицы – атрибутивные данные (количество $P_{атр}$)

Аналогично четвертому варианту данные за все годы собраны в одной теме (dbf таблице), но годы разворачиваются не по столбцам, а по строкам (таблица 5), что позволяет сделать имена атрибутивных колонок атомарными. Естественное условие для таблицы 5 – единый для всех лет состав колонок атрибутивных данных, очевидно с остаю-

щимися незаполненными ячейками в годы с отсутствием наблюдений данного атрибута.

Однозначная идентификация кортежа данных в табл.5 достигается введением ключевого поля, имеющего составное трехзвенное имя поле-годо-культура, чем обеспечивается его уникальность. В случае когда поле в каждом году используется только под одну культуру, уникальным является уже двухзвенное имя поле-год (табл. 5), где вторая колонка сложное ключевое имя, образуемое из имен колонок 3, 4.

В случае перенарезки полей их имена по годам в колонках 2 и 3 могут быть и неодинаковыми.

Таблица 5. Имя таблицы: «Книга истории полей хозяйства за многолетие»

ID N	Ключевое поле сх поле-год	Имя сх поля	Год выращивания	Культура	N	P	K	Урожайность
1	П1-1991	П1	1991	пшеница				
2	П1-1992	П1	1992	пшеница				
3	П1-1993	П1	1993	рожь				
4	П1-1994	П1	1994	травы				
5	П1-1995	П1	1995	ячмень				
6	П2-1991	П2	1991	картофель				
7	П2-1992	П2	1992	картофель				
8	П2-1993	П2	1993	ячмень				
9	П2-1994	П2	1994					
10	П2-1995	П2	1995					
...					

Такое составное именование сельхозполей приводит к тому, что территориально (площадно) одно и то же поле выступает в качестве $T_{лет}$ площадных объектов, в результате чего при визуальном отображении на карте они отображаются только все сразу, т.е. наложенными друг на друга за все годы или пересекающимися, если по годам происходит перенарезка полей. Отображение полей только одного конкретного года можно осуществить посредством его выбора, построителем запросов или вручную для удобства предварительно отсортировав таблицу по колонке "год выращивания". При этом визуализация распределения культур по полям может достигаться либо надписями, либо типом закрашки.

Следует отметить, что если по годам происходила перенарезка сх полей, то при визуализации конкретного года сх поле вырисовывается так, как оно реально и было нарезано в этом году. В частности при разбиении поля на два куска они и именуются как два полуполя.

Недостатком табл. 5 является невозможность автоматического расчета новых колонок по формуле среднего по годам или баланса по годам и т.п. стандартными про-

цедурами ГИС.

Из вышеизложенного следует, что каждый из проанализированных возможных вариантов имеет свои преимущества и недостатки (сложности) с точки зрения взаимоконвертации подготовки структур данных (выборок) для различных процедур (агрегирования, статистических и балансовых расчетов, получения регрессионных зависимостей и т.д.) и отображения данных на карте. То есть абсолютно предпочтительного варианта хранения данных книги истории полей хозяйства нет, но вариант 5 является наилучшим универсальным способом хранения информационного объекта наиболее общего типа – с изменяющимися во времени свойствами: перенарезка формы полей, изменение вида их использования по годам (сх поле-год-культура).

Так как исходные данные могут оказаться набранными в любом из вариантов, то желательна возможность их автоматизированной конвертации между всеми вариантами. Необходимо также выбрать наиболее удобный исходный формат хранения, который следует изначально использовать для ввода данных. Наиболее удобным в этом отношении является вариант 5, в котором за счет введения сложного имени объекта: сх поле-год-культура (сх поле-год) достигается возможность размещения всех данных в одной теме, в dbf таблице (табл.5) с атомарными полями атрибутивных данных, чем обеспечивается возможность легко реализуемого вращения данных, в частности переформирования данных в таблицы по вариантам 1-4. Так, стандартной сортировкой dbf таблицы 5 по колонке «годы», выделением и сохранением каждого года можно разбить общую тему для всех годо-полей на темы по отдельным годам, т.е. организовать их хранение по варианту 3. В свою очередь из варианта 3 стандартной процедурой соединения таблиц легко получить dbf таблицу по варианту 4. При этом следует учесть, что для однозначной идентификации данных необходимо переименовать все имена колонок ежегодно различающихся данных, добавив в их название год, либо после разбиения всех данных на отдельные годовые темы по варианту 3, либо после соединения столбцов в одну таблицу по варианту 4, но в этом случае сложно визуально отследить, к какому году относится каждая из колонок. В ArcView редактирование имен колонок невозможно и для его осуществления реализована программа ShpEdit.

Для анализа данных необходимо иметь возможность осуществлять с таблицами тем ряд процедур:

агрегирование во времени – формирование набора временных рядов вида рис.2 и расчет статистик, а также скользящих средних, кривых распределения и т.п.;

агрегирование по пространству – осуществление резюмирования: расчет статистик по объектам, кластеризованным на группы по различным системообразующим признакам (атрибутивным колонкам).

Следует отметить, что разбиение при резюмировании всех записей на группы по признаку одной из колонок, например, по имени полей не обеспечивает полной генетиче-

ской однородности данных, для достижения чего необходимо многоуровневое разбиение (кластеризация). К примеру, следующий уровень вложенности кластеризации – по видам культур разбивает для каждого поля данные по годам выращивания каждой конкретной культуры. Далее, к примеру, дифференциация данных может быть осуществлена по признаку мелиорированности и т.д.

Таким образом, для получения статистик по генетически однородным выборкам необходимо резюмирование с разбиением на группы двух, трех, четырех уровней вложенности. Осуществление этого при отсутствии в ГИС процедур многоуровневой сортировки и резюмирования может быть реализовано посредством создания дополнительных колонок со сложными именами типа поле-год, поле-культура, поле-год-культура и т. п., сортировкой, по любой из которых формируется соответствующая выборка.

Из таблицы 5 могут формироваться логические группы:

- по полям
- по годам
- по культурам
- поле-культурам
- годо-культурам
- поле-годам
- поле-годо-культурам

автоматизированный расчет по которым может осуществляться используя стандартную функциональность ГИС.

Наибольший интерес представляет резюмирование данных по культурам: либо для выбранного года – все поля (подвариант – группы полей, выбираемые по мелиорированности, типам почв и т.п.); либо для выбранного поля (подвариант – группы полей, выбираемые по мелиорированности, типам почв и т.п.) – все годы (подвариант – группы лет с различными погодно-климатическими условиями).

Все варианты организации моделей данных в ГИС при числе мерностей больше трех, вышеизложенные на примере сх полей, культур, годов относятся и к любым другим предметным областям, в частности к каналам мелиоративного объекта при осуществлении многолетнего кадастра, точкам наблюдений при проведении мониторинга. Вместо годов могут приниматься более подробные временные характеристики – месяцы, дни, что также увеличивает «мерность». Также увеличивает мерность наличие третьей пространственной координаты у объектов – глубины (высоты) у точек измерений.

Выводы

1. Информационная система агропроизводства должна включать ряд составляющих, соответствующих различным подсистемам: книга истории полей, агрохимический и почвенный мониторинг, мелиоративный кадастр для информационного обеспечения принятия решений, включая этапы, оценку состояния, прогноз оптимизации, реализация

которых целесообразна на базе геоинформационных систем.

2. Контроль состояния подсистем сопряжен с периодической повторной фиксацией показателей отслеживаемых объектов: сельскохозяйственных полей при ведении книги истории полей, элементарных участков по турам агрохимических и почвенных обследований, элементов мелиоративной сети (каналы, дренаж, подпорные сооружения) при ведении инвентаризаций и кадастра мелиоративных систем. Так как при этом помимо пространственной распределенности необходимо учитывать и временную составляющую, то «мерность» данных больше чем стандартная «мерность» ГИС, в связи с чем необходима организация структуры хранения данных.

3. Проанализированные варианты организации хранения пространственно-временных данных в ГИС на примере книги истории полей показали, что наиболее эффективным для обработки, конвертации, агрегирования является вариант хранения данных за многолетие в одной теме с принятием в качестве объекта составного понятия «сх поле-год».

Литература

1. Вахонин, Н.К. Организация структуры информационного обеспечения моделей природно-технических систем на основе мониторинга./ Н.К.Вахонин // Мелиорация переувлажненных земель: Сб. научн. работ БелНИИМил, Т. XLIII. – 1996. – С. 275-282.
2. Вахонин, Н.К. Автоматизированная георегиональная система информационного обеспечения принятия решений в агропроизводстве на основе мониторинга./ Н.К.Вахонин // Стратегия формирования информационного общества в Республике Беларусь. – 2003. – С. 209-212.

Summary

Vakhonin N.K.

ORGANIZATION OF THE DATA MODEL IN GIS, TIME-VARYING CHARACTERISTICS OF THE RECLAIMING AGRICULTURAL OBJECTS

In the article on the example of book of fields history discussed some problems of organization storage of multidimensional data in a GIS. On the base of analysis for data structure possible options is obtained that the most effective treatment is an option to store all data for Many Years in a single topic, using as an object of design " agricultural field-year" /

Поступила 28 января 2011 г.