

О. И. Христуло

ИНТЕГРАЦИЯ И ГЕНЕРАЛИЗАЦИЯ РАЗНОРОДНОЙ ТЕРРИТОРИАЛЬНО РАСПРЕДЕЛЕННОЙ ПРОСТРАНСТВЕННОЙ ИНФОРМАЦИИ НА ОСНОВЕ МНОГОМЕРНЫХ МОДЕЛЕЙ ДАННЫХ ДЛЯ ЗАДАЧ, СВЯЗАННЫХ С УПРАВЛЕНИЕМ ПРИРОДНЫМИ РЕСУРСАМИ

Описывается метод интеграции и генерализации разнородной распределенной пространственной информации по природным ресурсам, заключающийся в использовании многомерных информационных объектов для описания различных частей распределенных пространственных данных и их генерализации в единую модель. *Интеграция и генерализация; разнородная пространственная информация; геоинформационная система; многомерная модель данных; управление природными ресурсами*

Пространственная информация имеет большое значение при формировании исходных данных для принятия управленческих решений в задачах, связанных с контролем состояния природных ресурсов и охраной окружающей среды. Управление природными ресурсами в Республике Башкортостан может быть организовано на трех организационных уровнях: республиканском, территориальном (территориальные подразделения и комитеты) и местном (административные районы республики).

Для получения целостного представления о территориально распределенных объектах необходимо единое формальное описание пространственной информации, которая является разнородной (по типу, формату и способу описания), распределенной (по территориальному месту хранения и по принадлежности к действующим информационным системам). Особенностью этих объектов является их существенная протяженность и распределенность по территории. При этом наличие информации о точном местоположении объектов, их взаимном расположении и взаимосвязи существенно влияет на качество управленческих решений, принимаемых и реализуемых на различных уровнях.

Поэтому для создания распределенной системы обработки пространственной информации в управлении территориально распределенными объектами и системами, объединяющей (интегрирующей) подсистемы всех уровней управления, в качестве технологической основы целесообразно использовать геоинформационные системы, позволяющие с единых методических позиций описывать как пространственные (географические), так и атрибутивные (описатель-

ные) характеристики различных объектов, процессов и явлений.

В информационной системе обработки пространственной информации важно обеспечить хранение, обработку, обмен и предоставление пользователям информации, которая может существовать в виде:

- текстовых документов;
- файлов определенной структуры, баз данных;
- космических снимков;
- картографических баз данных;
- растровых данных, представленных в форме текстовых и графических документов.

На основании вышесказанного возникает самостоятельная задача интеграции и генерализации разнородной (по типу, формату, способу описания) пространственной информации о состоянии природных ресурсов, распределенной по территориальному месту хранения (в различных организациях, территориальных подразделениях) и по принадлежности к действующим информационным системам. Система может состоять из отдельных подсистем, каждая из которых решает свою собственную задачу.

Также в ее основе может лежать иерархическая структура с центральной и периферийными подсистемами, где подсистемы нижнего уровня связаны с подсистемами верхних по отношению к ним уровней путем передачи предварительно обобщенной информации, а подсистемы верхних уровней связаны с подчиненными им подсистемами через управляющие воздействия. При этом структуры хранения информации (модели данных) в каждой организации, территориальном подразделении, административно-территориальном образовании могут различаться. В качестве примера такой распределенной системы рассмотрим автоматизированную систему

поддержки принятия решений по управлению природными ресурсами и охране окружающей среды Республики Башкортостан (АСППР УПР и ООС РБ).

Одной из организаций, в которой пространственная информация территориально распределенных объектов (природные и водные объекты, места размещения отходов и другие) играет значительную роль и существует реальная необходимость в автоматизации задач, связанных с контролем состояния природных ресурсов, является Министерство природопользования и охраны окружающей среды РБ (Минэкологии РБ). Архитектура АСППР УПР и ООС РБ с компонентами комплексной ГИС приведена на рис. 1. Информационной основой управления природными ресурсами РБ является пространственная информация о водных объектах, объектах размещения отходов и др. территориях, которая представляется в виде географических (топографических) цифровых карт соответствующей территории: Республики Башкортостан, территориального подразделения, административного района РБ. При этом структуры хранения этой информации (модели данных) в каждом административно-территориальном образовании различаются.

Для формирования единого описания пространственной информации о природных ресурсах необходимо определить виды структур хранения этой информации на территориях и описать методы их преобразования в единую многомерную модель пространственных данных по природным ресурсам. Геоинформационную систему (ГИС) можно определить как:

$$\text{ГИС} = \{\text{БД}, P\},$$

где БД – база данных, P – множество процедур обработки данных. При этом база данных является совокупностью базы географических данных (БГД), которая содержит пространственные характеристики (карты, космические снимки, координаты) и базы атрибутивных данных (БАД), которая содержит все остальные (описательные) характеристики объектов:

$$\text{БД} = \{\text{БГД}, \text{БАД}\}.$$

Таким образом, геоинформационную систему ГИС можно определить как:

$$\text{ГИС} = \{\text{БГД}, \text{БАД}, P\}. \quad (1)$$

Определение (1) можно использовать для описания иерархической взаимосвязи элементов распределенной геоинформационной системы природных ресурсов Минэкологии РБ по организационным уровням управления. Иерархия распределенной информации по природным ресурсам в РБ представлена на рис. 2.

Здесь ГИС_{РБ} – геоинформационная система, содержащая в себе информацию по всей территории Республики Башкортостан.

{ГИС _{i} } – множество геоинформационных систем территориальных подразделений и комитетов, где $i = \overline{1, N}$, где N – количество территориальных подразделений и комитетов.

{БАД _{i,j} } – множество баз атрибутивных данных местного уровня для каждого административного района РБ, где $i = \overline{1, N}$ и $j = \overline{1, M}$, N – количество ТУ и К, M_i – количество административных районов i -го ТУ и К.

Согласно (1):

$$\text{ГИС}_{\text{РБ}} = \{\text{БГД}_{\text{РБ}}, \text{БАД}_{\text{РБ}}, P_{\text{РБ}}\}, \quad (2)$$

где БГД_{РБ} – интегрированная база географических данных республиканского уровня, БАД_{РБ} – интегрированная база атрибутивных данных республиканского уровня, $P_{\text{РБ}}$ – множество процедур обработки данных на республиканском уровне.

При этом каждый элемент множества {ГИС _{i} } (геоинформационная система территориального подразделения) содержит в себе набор информации, который можно обозначить:

$$\text{ГИС}_i = \{\text{БГД}_i, \text{БАД}_i, P_i\}, \quad (3)$$

где БГД _{i} – база географических данных, БАД _{i} – база атрибутивных данных, а P_i – множество процедур обработки данных в i -м территориальном подразделении.

Аналогично определяется геоинформационная система местного уровня (для каждого административного района РБ) – ГИС _{i,j} :

$$\text{ГИС}_{i,j} = \{\text{БГД}_{i,j}, \text{БАД}_{i,j}, P_{i,j}\}, \quad (4)$$

где БГД _{i,j} – база географических данных, БАД _{i,j} – база атрибутивных данных, а $P_{i,j}$ – множество процедур обработки данных в j -м административном районе i -го территориального подразделения.

Основная проблема, возникающая при проектировании и создании баз данных, заключается в необходимости обеспечения одновременной работы с разнородными источниками данных, несогласованности их структур и форматов с масштабами и количеством исходных данных. Наиболее важной и сложной в интеллектуальном плане проблемой является разработка структуры хранимых данных, так как она существенно влияет на эффективность работы БД по таким критериям, как требуемая для хранения всей информации память, время, необходимое на ввод, корректировку и поиск информации, адаптация к возможным изменениям запросов.

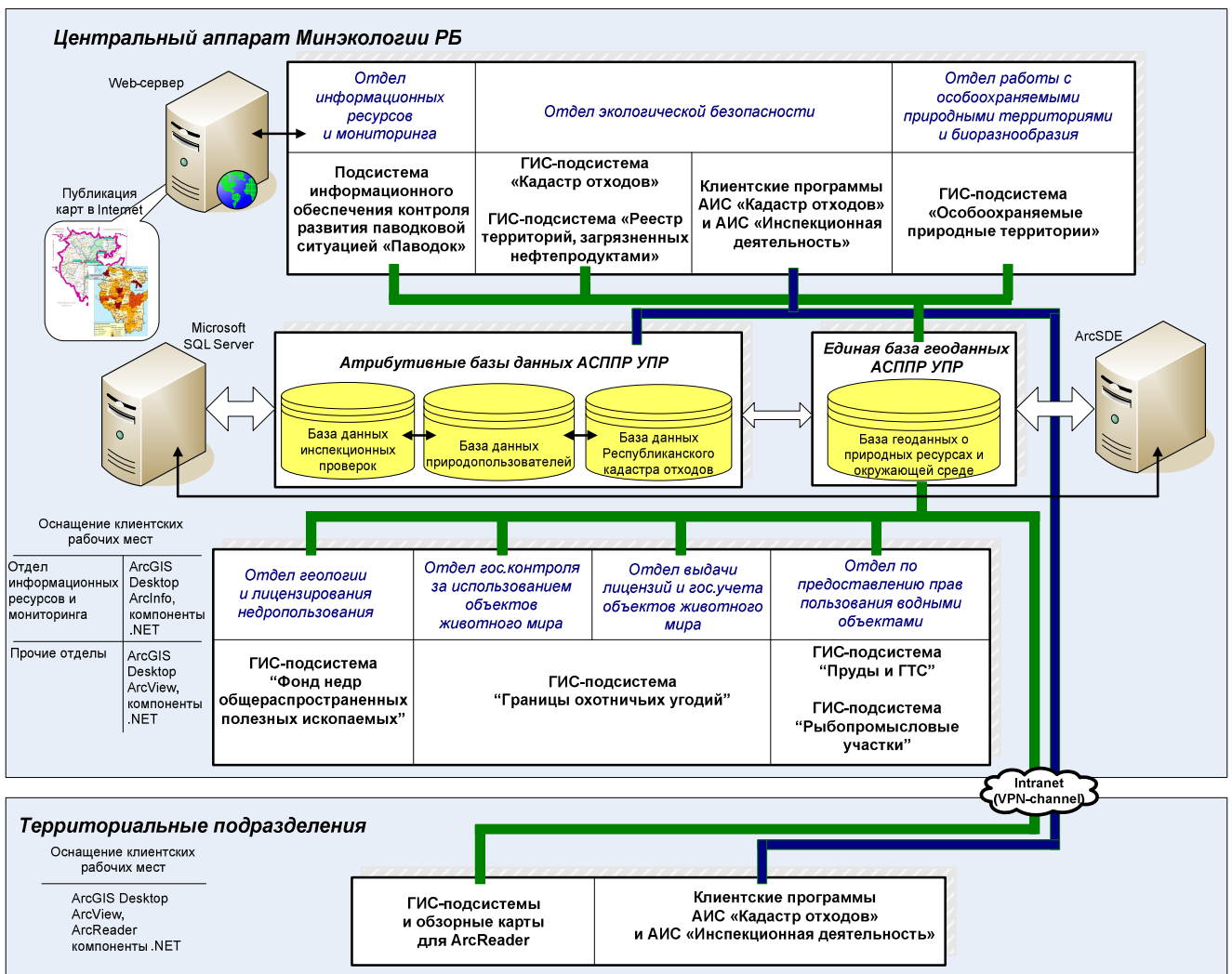


Рис. 1. Архитектура АСППР УПР и ООС РБ с компонентами комплексной ГИС

И поскольку модель данных – это структура данных вместе с множеством операций над элементами этих структур, то задача разработки модели данных является одной из самых важных при разработке БД.

База данных (БГД, БАД) геоинформационной системы на каждом уровне управления природными ресурсами строится на основе соответствующей модели данных (МД): $МД_{РБ}$ – модель данных $\{БГД_{РБ}, БАД_{РБ}\}$, $МД_i$ – модель данных $\{БГД_i, БАД_i\}$, $МД_{i,j}$ – модель данных $\{БГД_{i,j}, БАД_{i,j}\}$.

В достаточно общем виде задача создания распределенной базы данных АСППР УПР и ООС РБ заключается в построении непротиворечивой $МД$ [1], описывающей всю совокупность информации из $\{ГИС_{РБ}, ГИС_i, ГИС_{i,j}\}$, то есть:

$$\begin{aligned}
 МД_{РБ} &= \theta(M(i) = \theta(\partial_i(MM_{i,j})) = \theta(\partial_1(MM_{1,j}), \\
 &\dots, \partial_N(MM_{N,j})) = \theta(\partial_1(MM_{1,1}, \dots, МД_{1,M_1}), \dots, \\
 &\partial_N(MM_{N,1}, \dots, МД_{N,M_N})),
 \end{aligned} \quad (5)$$

для этого необходимо построить набор отображений $F = \{F_i, F_{i,j}^1, F_{i,j}^2\}$, таких что:

$$1) F_{i,j}^1: МД_{i,j} \rightarrow МД_i, \quad (6)$$

т. е. устанавливают соответствие множества моделей данных $МД_{i,j}$ и $МД_i$;

$$2) F_i: МД_i \rightarrow МД_{РБ}, \quad (7)$$

устанавливают соответствие множества моделей данных $МД_i$ и $МД_{РБ}$;

$$3) F_{i,j}^2: МД_{i,j} \rightarrow МД_{РБ}, \quad (8)$$

устанавливают соответствие некоторых территориальных моделей данных $МД_{i,j}$ и $МД_{РБ}$.

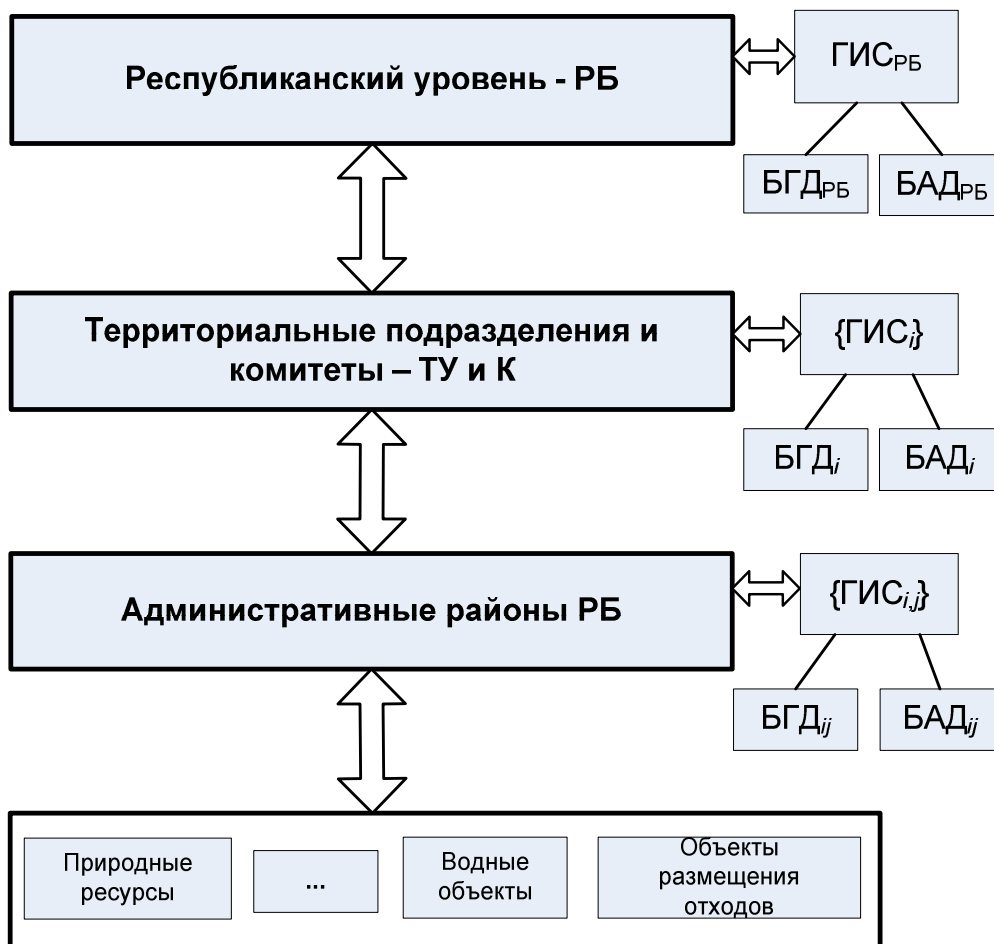


Рис. 2. Иерархия распределения информации по природным ресурсам в РБ

Географические объекты в БГД всех уровней в зависимости от степени детализации пространственных данных могут быть представлены данными разных типов: точечными, линейными и полигональными. Кроме того, географические объекты могут объединяться в тематические группы по признаку наличия между ними особых пространственных (топологии) или логических взаимосвязей.

Информационное обеспечение в иерархической системе построено таким образом, что пользователям на верхнем уровне из множества информации, возникающей на более низком уровне, нужны только отдельные или обобщенные характеристики, поэтому при построении отображений F обязательной является процедура генерализации [1]. Таким образом, задача генерализации пространственной информации о состоянии природных ресурсов заключается в описании правил обобщения информации при ее передаче из геоинформационных систем нижних уровней управления природными ресурсами (например, ГИС территориальных подразделений) на вышестоящий уровень управле-

ния. А сами эти правила могут быть описаны в виде правил преобразования МД (5)–(8).

Назовем две территориальных ГИС $ГИС_i$ и $ГИС_j$ *однотипными* в том случае, если их БД имеют одинаковую структуру, то есть $МД_i = МД_j$, в противном случае, если $МД_i \neq МД_j$, $ГИС_i$ и $ГИС_j$ назовем *разнотипными*.

На основе структуры хранения данных в БГД можно определить следующие виды разнородности моделей данных:

1) $МД_i$ и $МД_j$ являются *разнотипными по масштабам*, если количество уровней детализации представления информации в них отличается,

2) $МД_i$ и $МД_j$ являются *разнотипными по составу*, если перечень объектов, описываемых данными моделями, не совпадает.

Несмотря на различную структуру и содержание пространственных данных в различных территориальных элементах АСППР УПР и ООС РБ, необходимо совместное и непротиворечивое описание распределенных данных по природным ресурсам на единой методической основе.

Существующие многомерные модели данных, предложенные в работах [2–3], позволяют описывать пространственные данные и операции над ними на основе многомерных информационных объектов (МИО), которые обозначаются T_i^n и определяются через МИО меньшей размерности

$$T^n = \{T_i^{n-1}\}, i = \overline{1, k}, \quad (9)$$

где T – имя МИО, n – размерность МИО, i – индекс (номер) МИО при рассмотрении совокупности МИО.

Для описания внутренней структуры МИО используется понятие схемы МИО, представляющей собой множество, значения которого задают порядок вхождения МИО размерности $n-1$ в МИО размерности n , то есть

$$S^n = S\{T^n\} = \{S(T^{n-1}), S_n\} = \{S^{n-1}, S^n\} = \dots = \{S_1, S_2, \dots, S_{n-1}, S_n\}, \quad (10)$$

где S_i – i -й элемент схемы S^n .

Для реализации основных функций по обработке данных используются операции – проецирование (9) и объединение (10):

$$P_{S_{h,i}}(T^n) = T^{n-1}, \quad (11)$$

где T^n – исходный МИО размерности n , $S_{h,i}$ – значение элемента схемы, по которому выполняется проецирование, T^{n-1} – результат операции;

$$T^{n_1+n_2-1} = T^{n_1} \bigcup_{S^k} T^{n_2}, \quad (12)$$

где T^{n_1} , T^{n_2} – исходные МИО, $T^{n_1+n_2-1}$ – результат операции, S^k – элемент схемы, по которому производится объединение, $S^{n_1} \cap S^{n_2} = S^k \neq \{\emptyset\}$.

Применительно к пространственным данным, описывающим состояние природных ресурсов в АСППР УПР и ООС РБ, многомерным информационным объектам различных размерностей могут соответствовать:

T^0 – элементарной характеристике объекта (водного, размещения отходов и др.);

T^1 – полному описанию точечного географического объекта (в зависимости от детальности информации, например, водозабор, несанкционированная свалка и т. д.);

T^2 – полному описанию одного линейного географического объекта как совокупности точек, соединенных отрезками (например, река, ручей, канал);

T^3 – полному описанию одного полигонального географического объекта как совокупности линий, ограничивающих его контур (напри-

мер, озеро, водохранилище, полигон ТБО, особо охраняемая природная территория);

T^4 – описанию группы однотипных объектов (например, речная сеть);

T^5 – описанию всех рассматриваемых объектов с некоторой степенью детализации (например, речная сеть масштаба 1:200 000);

T^6 – всей совокупности пространственных данных по природным ресурсам на территории, схема данного МИО имеет вид

$$S(T^6) = S^6 = \{S_{r1}, S_{r2}, S_{r3}, S_4, S_5, S_6\}, \quad (13)$$

где $S_{r1} = \{\text{Идентификаторы атрибутов}\}$, $S_{r2} = \{\text{Идентификаторы узловых точек}\}$, $S_{r3} = \{\text{Идентификаторы линий}\}$, $S_4 = \{\text{Идентификаторы объектов}\}$, $S_5 = \{\text{Типы объектов}\}$, $S_6 = \{\text{Уровни детализации (масштабы)}\}$.

Для описания распределенной пространственной информации по природным ресурсам расширим многомерную модель пространственных данных по природным ресурсам, описанную МИО T^6 и схемой (13), и, с учетом трехуровневой системы управления природными ресурсами на территории РБ, определим МИО T^8 . При этом имеет место следующее соответствие:

T^7 – описание всех пространственных данных по природным ресурсам для территориального подразделения или комитета РБ;

T^8 – описание пространственных данных по природным ресурсам РБ, схема данного объекта имеет вид

$$S(T^8) = S^8 = \{S_{r1}, S_{r2}, S_{r3}, S_4, S_5, S_6, S_7, S_8\}, \quad (14)$$

где S_{r1} , S_{r2} , S_{r3} , S_4 , S_5 , S_6 определены в (13), $S_7 = \{\text{Идентификатор территориальной системы}\}$, $S_8 = \{\text{Идентификатор республиканской системы}\}$.

Многомерная модель данных также содержит метаданные пространственных данных, описываемые МИО T_m^6 , со схемой:

$$S(T_m^6) = S_m^6 = \{S_{m1}, S_{m2}, S_{m3}, S_{m4}, S_{m5}, S_{m6}\}, \quad (15)$$

где $S_{m1} = \{\text{Метаатрибуты}\}$, $S_{m2} = \{\text{Код объекта}\}$, $S_{m3} = \{\text{Слой}\}$, $S_{m4} = \{\text{Масштабы}\}$, $S_{m5} = \{\text{Пользователи}\}$, $S_{m6} = \{\text{Территориальные системы}\}$.

В совокупности с метаданными пространственных данных многомерная модель пространственных данных по природным ресурсам определяется как

$$M_{\Pi} = M(T^8, T_m^6) = T^8 \bigcup_{\substack{\text{КодОбъекта} \\ \text{КодСлоя} \\ \text{КодМасштаба}}} T_m^6 = T_{\Pi}^{11}. \quad (16)$$

Для решения задач генерализации пространственной информации по природным ресурсам существующих операций над МИО недостаточно. Поэтому необходимо ввести новую опера-

цию для описания изменения отдельных характеристик МИО. Предлагается ввести операции изменения однотипных МИО.

$$C_{s_1, l}^0(T_{\text{ВХ}}^{n-3}, T_{\text{ВЫХ}}^{n-3}):$$

$$T_{\text{ВХ}}^{n-3} \xrightarrow{\Pi_{s_1, l}(T_{\text{ВХ}}^{n-3})=f(\Pi_{s_1, l}(T_{\text{ВХ}}^{n-3}))} T_{\text{ВЫХ}}^{n-3}, \quad (17)$$

изменение значения атрибута l схемы S_1 в объекте $T_{\text{ВЫХ}}^{n-3}$ на значение некоего функционального преобразования f от значения атрибута l схемы S_1 в объекте $T_{\text{ВХ}}^{n-3}$, и

$$C_{s_1, l}^1(\{T_{\text{ВХ}, i}^{n-3}\}, T_{\text{ВЫХ}}^{n-3}):$$

$$\{T_{\text{ВХ}, i}^{n-3}\} \xrightarrow{\Pi_{s_1, l}(T_{\text{ВХ}}^{n-3})=f(\Pi_{s_1, l}(T_{\text{ВХ}, 1}^{n-3}), \dots, \Pi_{s_1, l}(T_{\text{ВХ}, m}^{n-3}))} T_{\text{ВЫХ}}^{n-3}, \quad (18)$$

изменение значения атрибута l схемы S_1 в объекте $T_{\text{ВЫХ}}^{n-3}$ на значение некоторого функционального преобразования (свертки) f от значений атрибутов l схемы S_1 в наборе объектов $\{T_{\text{ВХ}, i}^{n-3}\}$.

На основе построенной многомерной модели и операций изменения МИО можно описать отношения преобразования $F_i^{\text{одн}}$, $F_i^{\text{рм}}$, $F_i^{\text{рс}}$:

$F_i^{\text{одн}}$ – в случае, когда МД_i и $\text{МД}_{\text{РБ}, i}$ являются однородными,

$F_i^{\text{рм}}$ – в случае, когда МД_i и $\text{МД}_{\text{РБ}, i}$ являются разнотипными по масштабам,

$F_i^{\text{рс}}$ – в случае, когда МД_i и $\text{МД}_{\text{РБ}, i}$ являются разнотипными по составу.

Отношения преобразования $F_i^{\text{одн}}$ строятся на основе операции порождения МИО. Для территориального подразделения:

$$T^7 = \{T_i^6\}J, \quad (19)$$

где $i = \overline{1, N}$, N – количество территориальных подразделений, J – порядок вхождения. Аналогично формируется МИО T^8 на республиканском уровне.

Разнотипные отношения $F_i^{\text{рм}}$ и $F_i^{\text{рс}}$ строятся на основе последовательного применения операций изменения и порождения МИО.

Таким образом, метод интеграции и генерализации разнородной распределенной пространственной информации о природных ресур-

сах на основе многомерных информационных объектов и введенных операций их изменения позволяет построить модель данных, описывающую территориальные элементы распределенной автоматизированной системы поддержки принятия решений по управлению природными ресурсами и охраны окружающей среды Республики Башкортостан. Метод и модель позволяют объединить в единое информационное пространство информацию на всех уровнях управления природными ресурсами РБ при сохранении ее целостности, определяют связи распределенных пространственных данных по природным ресурсам РБ в виде совокупности МИО различной размерности на основе информации, возникающей и хранящейся в республике Башкортостан. Это дает возможность совместного представления и использования разнородной территориально распределенной информации о природных ресурсах.

СПИСОК ЛИТЕРАТУРЫ

1. Павлов С. В., Хамитов Р. З., Никитин А. Б. Структура разнородной территориально-распределенной пространственной информации при создании единой геоинформационной системы Росводресурсов // Вестник УГАТУ. 2007. Т. 9, № 4(22). С. 3–10.
2. Павлов С. В., Христодуло О. И. Методология создания многомерных баз данных на основе тензорных структур // Проблемы создания национальной академической системы баз данных и баз данных: Тезисы докл. Всероссийск. совещания. Уфа, 1995. С. 20–21.
3. Сахаров А. А. Принципы проектирования и использования многомерных баз данных (на примере Oracle Express Server) // СУБД. 1996. № 3. С. 44–59.

ОБ АВТОРАХ

Христодуло Ольга Игоревна, доц. каф. геоинф. систем. Дипл. инж. по автоматиз. и механиз. процессов обр. и выдачи информации (УАИ, 1991). Канд. техн. наук (УГАТУ, 2000). Иссл. в обл. проектир. и разработки корпорат. инф. систем.