

С. В. Павлов, О. И. Христодуло

РАЗРАБОТКА МЕТОДА ОБЪЕДИНЕНИЯ ДАННЫХ ИЗ РАЗЛИЧНЫХ ИНФОРМАЦИОННЫХ СИСТЕМ В ЕДИНУЮ ИНФОРМАЦИОННУЮ СИСТЕМУ МИНЭКОЛОГИИ РБ

Описывается метод формирования структуры базы общих данных при объединении информационных систем в единую информационную систему организации на примере Министерства природопользования и экологии Республики Башкортостан, основанный на формальном анализе баз атрибутивных и пространственных данных, входящих в состав действующих в министерстве информационных систем, и позволяющий выявить общую для нескольких баз данных атрибутивную и пространственную информацию и организовать многопользовательский доступ к распределенно-хранящимся разнородным данным Минэкологии РБ. *Единая база данных; пространственные и атрибутивные данные; информационное обеспечение; управление природными ресурсами*

Важнейшей функцией Министерства природопользования и экологии Республики Башкортостан (далее Минэкологии РБ, министерство) является эффективное управление природными ресурсами Республики. Для его достижения необходима информация об объектах управления: природных ресурсах, их окружении, природопользователях, экологически опасных объектах, потенциально затопляемой территории в период весенних паводков и пр.

Министерство создает и ведет республиканский фонд информации по природным ресурсам, охране окружающей среды и экологической безопасности, осуществляет обмен, передачу и распространение этой информации, участвует в создании информационных систем для нужд природопользования, охраны окружающей среды, экологической безопасности и сохранения биологического разнообразия.

В процессе развития информационного обеспечения министерства число задач, для решения которых используется пространственная информация о территориально распределенных природных и техногенных объектах, например, водных или природных ресурсах, увеличивается. Особенностью этих объектов является их существенная протяженность и распределенность по территории Республики Башкортостан. Наличие информации о точном местоположении объектов, их взаимном расположении и взаимосвязи существенно влияет на качество управленческих решений, поэтому при создании интегрированной информационной системы Минэкологии РБ необходимо применение геоинформационных технологий. При этом все блоки информации должны содержать

как описательную (атрибутивную) информацию (наименование, количественные характеристики, пространственную информацию, принадлежность и др.), так и характеризующую местоположение объекта на поверхности Земли.

На сегодняшний день в министерстве функционируют следующие информационные системы (ИС) [1–3] (рис. 1):

- подсистема информационного обеспечения контроля развития паводковой ситуации (ГИС «Паводок»);

- автоматизированная система ведения реестра территорий, загрязненных нефтепродуктами (автоматизированная информационная система «Реестр загрязненных территорий»);

- информационная система по ведению и формированию регионального кадастра отходов на территории РБ (ИС «Кадастр отходов»);

- информационная система «Инспекционная деятельность»;

несколько локальных геоинформационных систем:

- ГИС «Особо охраняемые природные территории»;

- ГИС «Расположение ГТС РБ»;

- ГИС «Расположение рыбопромысловых участков»;

- ГИС «Регистрационная карта кадастра расположения подземных вод по РБ»;

- ГИС «Месторождения и участки подземных вод с утвержденными запасами»;

а также локальные информационные системы территориальных управлений Минэкологии РБ. Практически во всех вышеперечисленных ИС ведется обработка разнородной информации (пространственной и атрибутивной), кроме того,

периодически возникает необходимость в обмене данными между различными системами, каждая отдельная ИС Минэкологии РБ содержит свою базу данных, при этом некоторые из ИС должны использовать информацию из баз данных других информационных систем. Например, для специалистов министерства очень актуально создание и ведение единого и централизованного фонда информации о природопользователях на территории РБ – «Справочника природопользователей РБ», при этом одни и те же атрибутивные данные о природопользователях хранятся в различных информационных системах: «Кадастр отходов» и «Инспекционная деятельность».

Аналогично общегеографические данные, такие как гидрография, растительность, населенные пункты, дороги и др., которые являются общими для большинства информационных систем, обрабатывающих пространственную информацию, хранятся одновременно в базах данных ГИС «Паводок» и АИС «Реестр загрязненных территорий».

Для формирования 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 (территориальные подразделения министерства) и по принадлежности к действующим информационным системам. Для возможности объединения и представления разнородных пространственных данных, получаемых (создаваемых) в министерстве, очевидно, они должны быть унифицированы, то есть иметь единую модель данных и базу данных. В силу этого актуальной задачей является интеграция разнородных пространственных данных существующих информационных систем в единую базу данных и последующая разработка интегрированной информационной системы Минэкологии РБ.

Для организации хранения информации необходимо выявить общие данные для всех информационных систем Минэкологии РБ, провести анализ требований к этим данным, чтобы минимизировать дублирование данных (с точки зрения хранения) и время получения пользователем информации.

Обозначим i -ю информационную систему министерства – ИС $_i$, а ее базу данных БД $_i$, где $i = \overline{1, n}$, n – общее количество ИС.

Выделим в каждой из БД $_i$ две части: БД $_i^1$ – а часть БД $_i$, данные из которой находятся как в БД $_i$, так и в других базах данных, БД $_i^2$ – часть

БД $_i$, использующаяся только в ИС $_i$ и не содержащаяся в других базах данных, при этом:

$$\text{БД}_i = \text{БД}_i^1 \cup \text{БД}_i^2, i = \overline{1, n}. \quad (1)$$

Для исключения (или минимизации) дублирования данных необходимо, чтобы общие для всех (или нескольких) БД $_i$ данные хранились в единой базе данных.

Обозначим БД $_0$ – единая база данных Минэкологии РБ, тогда

$$\text{БД}_i \subset \text{БД}_0, \text{ для всех } i = \overline{1, n}. \quad (2)$$

Поскольку базы данных информационных систем министерства содержат как пространственные, так и атрибутивные данные, то представим БД $_i$ как совокупность двух частей, описывающих атрибутивную – БД $_i^A$ и пространственную – БД $_i^П$ информацию [4]:

$$\text{БД}_i = \text{БД}_i^A \cup \text{БД}_i^П, i = \overline{1, n}. \quad (3)$$

Тогда БД $_0$ будет также состоять из двух частей:

$$\text{БД}_0 = \text{БД}_0^A \cup \text{БД}_0^П. \quad (4)$$

В соответствии с (1), в каждой i -й БД $_i^П$ выделим две части: БД $_i^{П1}$ – та часть БД $_i^П$, которая входит в состав соответствующей БД $_i^1$, БД $_i^{П2}$ – часть БД $_i^П$, входящая в состав БД $_i^2$. Тогда эти части могут быть представлены:

$$\text{БД}_i^П = \text{БД}_i^{П1} \cup \text{БД}_i^{П2}, i = \overline{1, n}. \quad (5)$$

Аналогично, для каждой i -й БД $_i^A$: БД $_i^{A1}$ – та часть БД $_i^A$ которая входит в состав БД $_i^1$, БД $_i^{A2}$ – часть БД $_i^A$ входящая в состав БД $_i^2$. Тогда:

$$\text{БД}_i^A = \text{БД}_i^{A1} \cup \text{БД}_i^{A2}, i = \overline{1, n}. \quad (6)$$

Следовательно,

$$\text{БД}_i = (\text{БД}_i^{A1} \cup \text{БД}_i^{A2}) \cup (\text{БД}_i^{П1} \cup \text{БД}_i^{П2}), i = \overline{1, n}. \quad (7)$$

БД $_i^П$ содержит пространственную информацию, хранящуюся, как правило, в БД $_i$ в виде цифровых топографических карт определенного масштаба, состоящих из тематических слоев или классов пространственных объектов S_i^j , то есть:

$$\text{БД}_i^П = \{S_i^j\}, i = \overline{1, n}, j = \overline{1, m_i}, \quad (8)$$

где m_i – количество слоев, хранящихся в БД $_i^П$. В каждой БД $_i$ каждый из слоев S_i^j может содержаться либо в части БД $_i^{П1}$, либо в БД $_i^{П2}$, но не в обеих одновременно. Обозначим M_i^1 – мно-

жество номеров j слоев, содержащихся в $БД_i^{1П}$, т. е. для которых справедливо $S_i^j \in БД_i^{1П}$ при $j \in M_i^1$ и $i = \overline{1, n}$, тогда:

$$БД_i^{1П} = \{S_i^j\}, j \in M_i^1, \quad (9)$$

и M_i^2 – множество номеров j слоев, для которых справедливо $S_i^j \in БД_i^{2П}$, при $j \in M_i^2$ и $i = \overline{1, n}$, тогда:

$$БД_i^{2П} = \{S_i^j\}, j \in M_i^2, \quad (10)$$

С учетом (9) и (10), часть $БД_0^П$ из (4) включает:

$$БД_0^П = \{S_0^1, S_0^2, \dots, S_0^M\}, \quad (11)$$

где M – общее число слоев в $БД_0^П$, и $S_0^i \neq S_0^g, j \neq g, \bigcup_{r=1, M} S_0^r = \bigcup_{\substack{i=1, n \\ j=1, m_i}} S_i^j$.

Приведем аналогичное описание атрибутивных данных. $БД_i^A$ содержит атрибутивную информацию, хранящуюся в $БД_i$, в виде реляционных таблиц, в которых атрибутивная информация для пространственного объекта представлена совокупностью значений атрибутов. Поэтому пусть $БД_i^A$ представлена в виде совокупности отношений:

$$БД_i^A = \{R_0^1, R_0^2, \dots, R_0^{k_i}\}, i = \overline{1, n}, \quad (12)$$

где k_i – количество отношений в i -й $БД_i^A$, содержащих атрибутивную информацию.

По аналогии с $БД_i^П$ в каждой i -й $БД_i^A$ выделим две части: $БД_i^{1A}$ – та часть $БД_i^A$, которая входит в состав $БД_i^1$, $БД_i^{2A}$ – часть $БД_i^A$, входящая в состав $БД_i^2$. Каждое из отношений R_i^j может содержаться либо в части $БД_i^{1A}$, либо в $БД_i^{2A}$, но не в обеих одновременно.

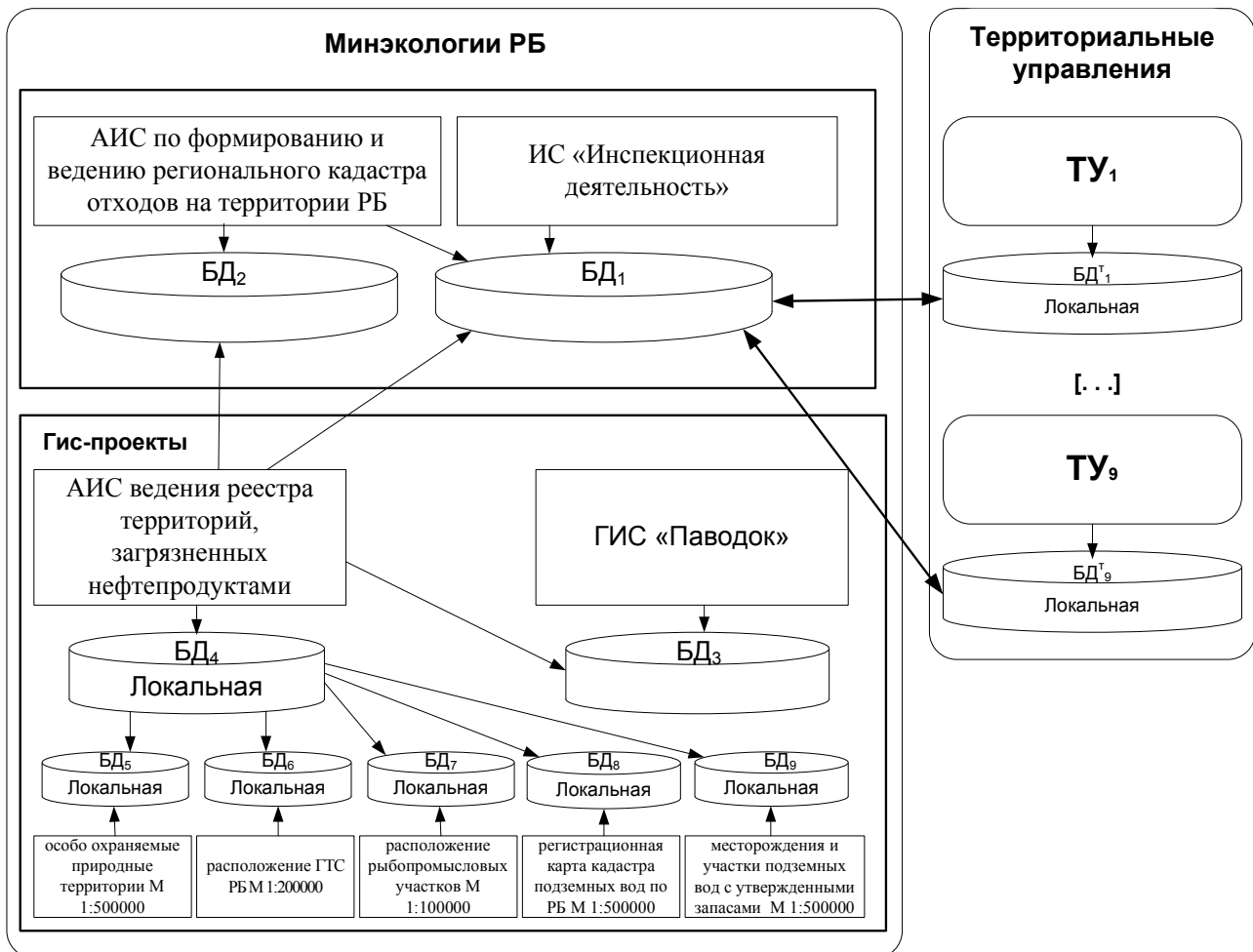


Рис. 1. Схема информационного взаимодействия информационных систем территориальных управлений и центрального аппарата Минэкологии РБ

Обозначим $K_i^1, i = \overline{1, n}$ – множество номеров отношений R_i^1 , таких, что $R_i^1 \in \text{БД}_i^{1A}$ при $j \in K_i^1$, тогда:

$$\text{БД}_i^{1A} = \{R_i^j, j \in K_i^1\}, \quad (13)$$

и K_i^2 – множество номеров отношений из БД_i^{2A} таких, что $R_i^j \in \text{БД}_i^{2A}$, при $j \in K_i^2$, тогда:

$$\text{БД}_i^{2A} = \{R_i^j, j \in K_i^2\}, \quad (14)$$

С учетом (13) и (14), часть БД_0^A из (4) включает:

$$\text{БД}_0^A = \{R_0^1, R_0^2, \dots, R_0^K\}, \quad (15)$$

где K – общее число отношений в БД_0^A и $R_0^i \neq R_0^g, j \neq g, \bigcup_{k=1, \overline{K}} R_0^k = \bigcup_{\substack{i=1, \overline{n} \\ j=1, \overline{k_i}}} R_i^j$.

По (4), имеем:

$$\text{БД}_0 = (\text{БД}_0^{1A} \cup \text{БД}_0^{2A}) \cup (\text{БД}_0^{1\Pi} \cup \text{БД}_0^{2\Pi}). \quad (16)$$

И из (6), (11), (15) получаем:

$$\text{БД}_0 = \{R_0^1, R_0^2, \dots, R_0^K, S_0^1, S_0^2, \dots, S_0^M\}. \quad (17)$$

Поскольку после создания базы общих пространственных данных производится ряд преобразований БД_i (например, перенос слоя гидрографии из БД_i^{1A} в БД_0^A), а необходимость в пространственной информации у пользователей информационных систем остается, необходимо установить соответствие между местом хранения данных в БД_i^{1A} до преобразований и нынешним местом хранения данных в единой БД_0^A . Таким образом, для создания единой базы данных необходимо установить соответствие между элементами множеств БД_i и БД_0 , а для этого задать функцию, такую что:

$$F_i : \text{БД}_i \rightarrow \text{БД}_0, \text{ где } i = \overline{1, n}. \quad (18)$$

Каждая F_i является множеством отображений f_i^{Aj} и $f_i^{i\Pi j}$,

$$F_i = \{f_i^{A1}, f_i^{A2}, \dots, f_i^{Ak_i}, f_i^{i\Pi 1}, f_i^{i\Pi 2}, \dots, f_i^{i\Pi m_i}\}, \quad (19)$$

где f_i^{Aj} – отображение, устанавливающее соответствие между каждым отношением из множества БД_i^{1A} и определенным отношением из множества БД_0 , то есть:

$$R_0^j = f_i^{Aj}(R_i^j), i = \overline{1, n}, j = \overline{1, k_i}, \quad (20)$$

а $f_i^{i\Pi j}$ – отображение, устанавливающее соответствие между каждым слоем из множества БД_i^{2A} и слоем из БД_0 :

$$S_0^j = f_i^{i\Pi j}(S_i^j), i = \overline{1, n}, j = \overline{1, m_i}. \quad (21)$$

На рис. 2 показан пример установления соответствия между слоями БД_0^A и слоями в БД_i^{1A} , необходимого для нахождения соответствующего слоя в БД_0 в процессе обновления информации, производящегося из БД_i^1 .

В результате построения всех отображений F_i появится возможность создать непротиворечивую структуру данных, необходимую для реализации единой базы данных министерства, содержащую все отношения и слои, входящие ранее в $\text{БД}_i, i = \overline{1, n}$, а соотношения (1)–(21) можно рассматривать как формализацию процесса объединения данных из различных информационных систем при их объединении в единую информационную систему Минэкологии РБ (рис. 3).

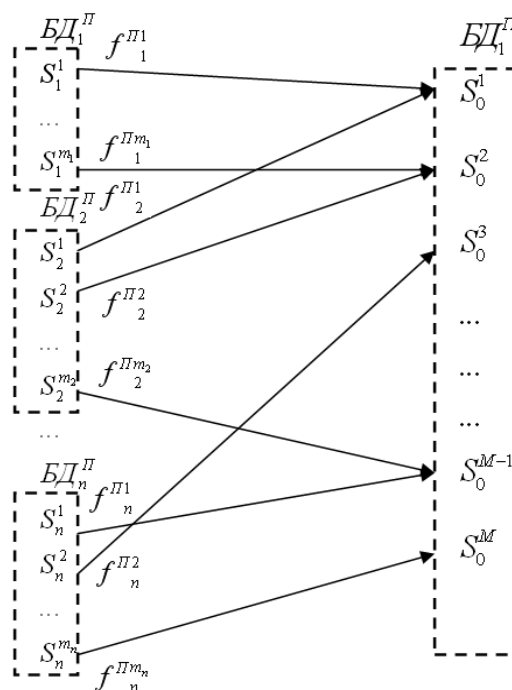


Рис. 2. Схема построения отображений для установления соответствия между слоями БД_0 и слоями в существующих БД_i

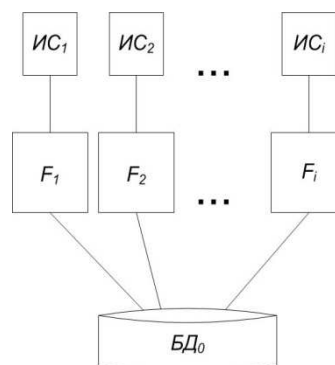


Рис. 3. Схема получения единой базы данных Минэкологии РБ

ЗАКЛЮЧЕНИЕ

Предложен метод объединения разнородных данных из различных информационных систем в единую информационную систему, основанный на формальном анализе баз атрибутивных и пространственных данных, входящих в состав действующих информационных систем, и позволяющий выявить общую для нескольких баз данных атрибутивную и пространственную информацию и установить соответствие множеств $БД_i$ и $БД_0$ для создания единой базы данных Минэкологии РБ.

СПИСОК ЛИТЕРАТУРЫ

1. Павлов С. В., Христовуло О. И., Шарфутдинов Р. Р. Разработка геоинформационной модели речной сети с учетом картографической, гидрологической и морфометрической информации для определения границ зон затопления при изменении уровня воды в водных объектах // Вестник УГАТУ: Науч. ж-л Уфимск. гос. авиац. техн. ун-та. Сер. «Управление, вычислительная техника и информатика». 2008. Т. 11, № 1 (28). С. 18–27.

2. Интеграция информационных систем по формированию и ведению регионального кадастра отходов и «Инспекционная деятельность» для Минэкологии РБ / И. Н. Заитов [и др.] // Геоинформацион-

ные технологии в проектировании и создании корпоративных информационных систем: Межвузовск. науч. сб. Уфа: УГАТУ, 2010. С. 87–92.

3. Абдуллин А. Х., Мальцева Е. А., Христовуло О. И. Разработка автоматизированной системы ведения реестра территорий, загрязненных нефтепродуктами // Геоинформационные технологии в проектировании и создании корпоративных информационных систем: Межвузовск. науч. сб. Уфа: УГАТУ, 2010. С. 118–126.

4. Бадамшин Р. А., Павлов А. С. Многопользовательская обработка распределенно хранящейся пространственной информации в научно-образовательной ГИС РБ // Вестник УГАТУ: науч. ж-л Уфимск. гос. авиац. техн. ун-та. 2009. Т. 12, № 1(30). С. 3–8.

ОБ АВТОРАХ

Павлов Сергей Владимирович, проф., зав. каф. геоинф. систем. Дипл. математик по спец. «Вычислительная математика» (БГУ, 1977). Д-р техн. наук (УГАТУ, 1998). Иссл. в обл. обработки пространств. данных.

Христовуло Ольга Игоревна, доц. той же каф. 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.