

## РАЗРАБОТКА И СОЗДАНИЕ БАЗЫ ДАННЫХ ИВЭП СО РАН

Петров А.В., Ловцкая О.В., Суторихин И.А.

*Институт водных и экологических проблем СО РАН (ИВЭП СО РАН),  
656038, г. Барнаул, ул. Молодежная 1, [pav@iwep.asu.ru](mailto:pav@iwep.asu.ru)*

*В ходе многолетней работы лабораториями института накоплен большой объем разноплановой информации. С целью ее упорядочивания и структуризации было принято решение о создании базы данных института. Создание базы данных разбито на два основных этапа: На первом создается база метаданных, на втором на основе баз данных подразделений создается общепитутская база данных.*

### Введение

Институт водных и экологических проблем был основан в 1987 году на базе нескольких лабораторий ряда институтов СО РАН. В настоящий момент в состав института, расположенного в городе Барнауле входят лаборатории: биогеохимии, водной экологии, гидрологии и геоинформатики, регионального природопользования, экологии атмосферы, эколого-географического картографирования и химико-аналитический центр. Кроме этих структурных подразделений в состав института входят – Новосибирский филиал: лаборатория моделирования гидрофизических и экологических процессов, центр водно-экспедиционных исследований и группа советника РАН; Горно-Алтайский филиал; Кемеровская лаборатория геоэкологических и водных проблем; Кызыл-Озекский почвенно-биологический стационар (п. Кызыл-Озек, р. Горный Алтай); Нижнее-Обской стационар (п. Карымкары, Тюменская область). За все время существования института в структурных подразделениях накоплено большое количество разнообразной и разноплановой информации включающей данные о проводимых исследованиях в различные моменты времени.

В связи с большим объемом накопленной информации и трудностью проведения поиска среди разрозненных данных возникла необходимость структурировать для каждого подразделения имеющиеся данные с использованием международных стандартов описания, создать единую базу данных Института с возможностью доступа к ней по сети Интернет.

Прежде чем перейти к описанию разрабатываемой базы данных введем понятия и определения, которыми мы будем в дальнейшем оперировать.

Под *информацией* в данной работе мы будем понимать любые сведения о каком-либо объекте, событии, процессе, в отличии от классического определения информации по Шеннону: информация – уменьшение степени неопределенности о каком-либо объекте, процессе или явлении, или изменение неопределенности состояния самого объекта, системы, явления, процесса [1].

*Данные* – это обработанная информация, представленная в определенном виде, позволяющем автоматизировать ее сбор, хранение и дальнейшую обработку.

*Запись* – именованная структура, содержащую один или более именованных элементов данных, каждый из которых обладает своим форматом.

*Метаданные* – представлены в виде *метазаписей*, каждая из которых соответствует определенному научному результату.

*Обработка данных* – специальный класс решаемых на ЭВМ задач, связанных с видом, хранением, сортировкой, отбором по заданному условию и группировкой данных однородной структуры. При этом предусматривается генерация отчетов, как правило, табличной формы.

*База данных (БД)* – именованная совокупность взаимосвязанных данных при минимальной избыточности, отражающая состояние объектов и их отношений в рассматриваемой предметной области, и допускающей их использование оптимальным образом.

*Хранилище данных* – предметно-ориентированный, интегрированный, привязанный ко времени и неизменяемый набор данных, предназначенный для поддержки принятия решений [2].

*Сетевое хранилище данных* – распределенное хранилище данных, реализованное в сети Интернет, и не имеющее центрального репозитория данных.

*Система управления базой данных (СУБД)* – совокупность языковых и программных средств, предназначенных для создания, ведения и совместного использования базы данных многими пользователями [3].

Создание базы данных Института было решено разбить на два основных этапа: на первом этапе предполагалась создать базу метаданных – описание характеристик данных, которые будут размещены в хранилище данных. После наполнения базы метаданных предполагалось провести анализ метаданных, определить необходимость и очередность создания витрин данных, порядок размещения их в хранилище данных, определить уровень доступа к различным данным.

На втором этапе разрабатываются концептуальная, логическая и физическая модели реализации базы данных. По построенным моделям создается каркас базы данных (хранилища данных). В рамках каждого подразделения создаются витрины данных, которые потом интегрируются в общеинститутскую базу данных.

### **Создание базы метаданных**

Для определения потребностей каждого подразделения в доступе к данным научных исследований и определения приоритетных направлений наполнения общеинститутской базы данных разработана база метаданных. Иногда метаданные называют данными о данных [2].

Метазаписи в базе метаданных оформляются по единому образцу и содержат:

1. Заголовок
2. Аннотация
3. Код классификатора.
4. Ключевые слова
5. Производители данных
6. Персоны, связанные с данными (держатели данных)
7. Первая публикация
8. Носитель информации
9. Дата создания данных
10. Формы разрешений на данные (открытый доступ, для служебного пользования)
11. Авторы метаданных
12. Вложенная Ссылка (Характеристика типа набора данных текст, снимок, HTML и т.п. Собственно ссылка Имя прилагаемого файла, ссылка на источник в Интернет и т.п.)
13. Координаты или географическая привязка

В случае если какой-либо из вышеперечисленных пунктов отсутствует, он опускается.

Пример реализации метазаписи приведен на рисунке 1.

Формирование базы метаданных осуществлялось по смешанной схеме, позволяющей учитывать неоднородность, различную степень заполнения и изменчивость сложных связей данных природопользования.

На начальном этапе отбор включаемой в базу метаданных информации проводился руководителем научного подразделения, далее база метаданных развивается с использованием подхода с “отложенной интеграцией”, т.е. развиваются только те подсистемы, в которых ощущается необходимость.

# БАЗА ДАННЫХ

## Лаборатория гидрологии и геоинформатики

**Карта постов гидрологических и гидрохимических наблюдений сети Роскомгидромета и других организаций на бассейне Верхней Оби 1:200 000**

Аннотация	Карта постов гидрологических и гидрохимических наблюдений сети Роскомгидромета и других организаций на бассейне Верхней Оби 1:200 000. Карта постов гидрохимических наблюдений создана на территорию в пределах Алтайского края, карта постов гидрологических наблюдений создана на весь бассейн Верхней Оби. Используется для решения задачи устойчивого подслаживания территории.
Код классификатора	1.4 - гидрология 1.7 - антропогенные образования
Ключевые слова	Верхняя Обь, ГИС, гидрология, гидрохимия, Роскомгидромет.
Производители данных	Жоров В.А. 66 64 74 Яковченко С.Г. 66 65 01 Ловская О.В. 66 65 01
Первая публикация	Жоров В.А., Яковченко С.Г., Ловская О.В., Постнова И.С., Зырянова Т.А. Геоинформационное обеспечение водозастойными расчетом для решения задач устойчивого реновирования ГИС для устойчивого развития территорий: Матер. Междунар. конф. - Владивосток-Чаянуа, 2002
Носитель информации	файлы - шой ГИС ArcView под ArcExplorer или ArcView
Дата создания данных	2002
Язык	русский
Формы разрешений на данные	открытый доступ; ссылка на авторов обязательна
Авторы метаданных	Ловская О.В., Постнова И.С., Яковченко С.Г.
Вложенная ссылка	Компонента из ГИС-проекта в формате HTML. <a href="#">Перейти</a>
Географическая привязка	Бассейн Верхней Оби, Алтайский край, Горный Алтай

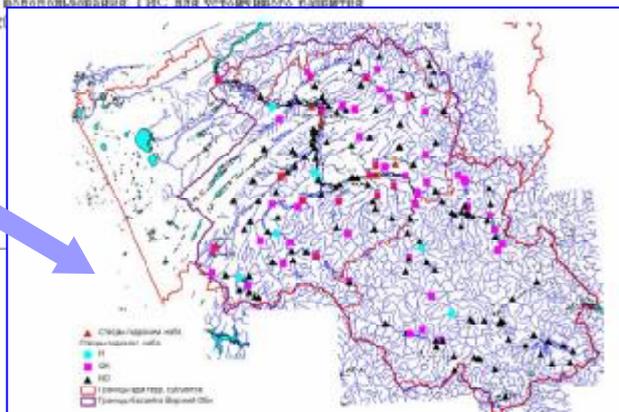


Рис. 1 Пример метазписи.

Интерфейс доступа к имеющимся метаданным реализован в виде HTML страниц со связанными ссылками и логически структурированы с учетом организационной структуры института. База метаданных располагается на сайте института <http://iwep.asu.ru> в открытом доступе.

Для ускорения поиска необходимой информации в базе данных реализован поиск. Поиск реализован на языке PHP и проводится в индексном файле [4] (рис. 2).

ИНСТИТУТ ВОДНЫХ И ЭКОЛОГИЧЕСКИХ ПРОБЛЕМ  
СИБИРСКОЕ ОТДЕЛЕНИЕ РОССИЙСКОЙ АКАДЕМИИ НАУК

- [Лаборатория биогеохимии](#)
- [Лаборатория водной экологии](#)
- [Лаборатория гидрологии и геоинформатики](#)
- [Лаборатория регионального природопользования](#)
- [Лаборатория экологии атмосферы](#)
- [Лаборатория эколого-географического картографирования](#)
- [Химико-аналитический центр](#)
- [Горноалтайский филиал](#)
- [Группа советника РАН](#)
- [Лаборатория моделирования гидрофизических процессов](#)
- [Центр водно-экологических исследований](#)
- [Лаборатория геологии и мониторинга горных систем \(Горно-Алтайск\)](#)
- [Лаборатория геоэкологических и водных проблем \(Кемьское\)](#)

### БАЗА ДАННЫХ

поиск

Результаты поиска обь

[doclab05](#)  
[doclab10](#)

Всего найдено страниц: 2

Webmaster: [Сергей](#) © 2005 IWEP SB RAS

Рис. 2 Поисковая страница базы метаданных

Индексный файл создается следующим образом: скрипт (выполняемый модуль) открывает директорию, в которой расположены метазписи, считывает каждый файл

метазаписи, удаляет повторяющиеся слова и слова короче, чем из трех букв (за исключением географических названий), создает строку в индексном файле, содержащую ключевые слова и ссылку на файл.

Поиск осуществляется по вводимому слову, скрипт проверяет каждую строку индексного файла и выдает страницу со списком метазаписей, в которых присутствуют запрашиваемые слова. Если запрашиваемые слова не найдены выдается сообщение об отсутствии информации и предлагается повторить запрос.

### База данных

На основе анализа имеющихся метаданных и информации из структурных подразделений разработана концептуальная модель базы данных института, которая не зависит от способа физической реализации базы данных.

Создание базы данных должно решить следующие задачи:

1. Инвентаризация имеющихся данных
2. Интеграция источников данных
3. Надежность хранения исходных данных
4. Поддержка проектных работ картографическими данными, данными измерений, математического моделирования, и результатами пространственного анализа
5. Обеспечение доступа к информационным ресурсам

База данных института представляет собой хранилище данных с иерархической структурой, в которое входят отдельные базы данных подразделений института, которые в свою очередь включают в себя отдельные базы данных выполненных проектов и научных результатов (рисунок 3).

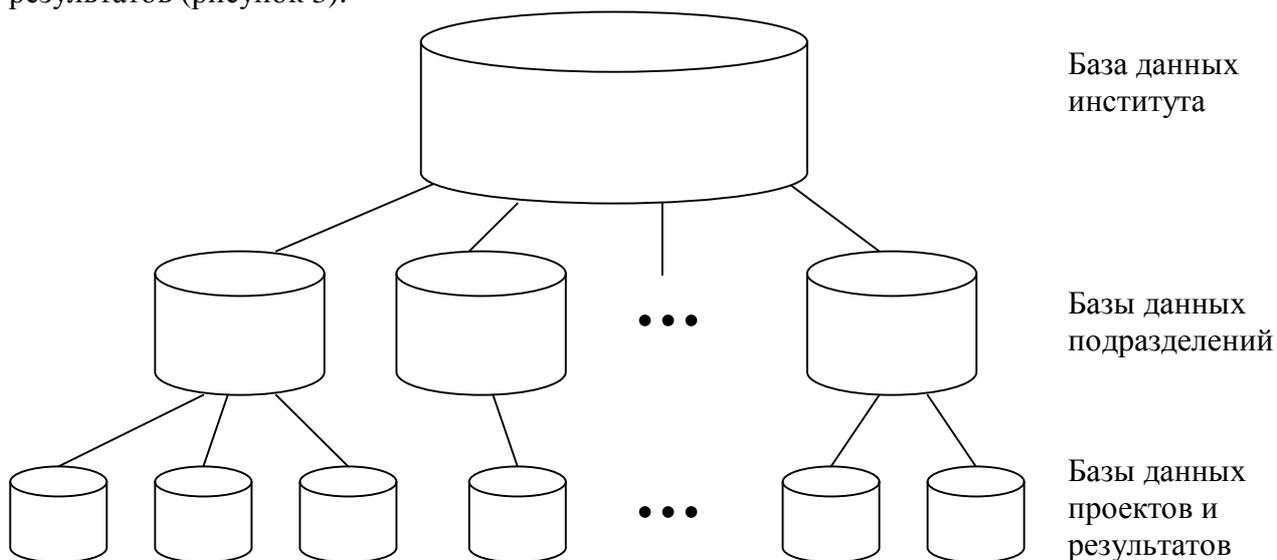


Рис. 3. Структура базы данных

Основные преимущества хранилища данных состоят в следующем [2,3]:

- Предметная ориентированность. Хранилище данных организовано вокруг основных объектов, а не вокруг прикладных задач
- Интегрированность. Для предоставления пользователю единого обобщенного представления данных необходимо создать интегрированный источник, обеспечивающий согласованность хранимой информации.
- Привязка ко времени. Данные в хранилище точны и действительны только в том случае, если они привязаны к некоторому моменту или промежутку времени. Необходимость привязки хранилища данных ко времени следует из большой длительности того периода, за который была накоплена сохраняемая в нем информация, из явной или неявной связи временных отметок со всеми сохраняемыми

данными, а также из того факта, что хранимая информация фактически представляет собой набор снимков состояния данных.

- Неизменяемость. Это означает, что данные не обновляются в оперативном режиме, а лишь регулярно пополняются за счет информации из оперативных систем обработки. При этом новые данные никогда не заменяют, а лишь дополняют прежние. Таким образом, база данных хранилища постоянно пополняется новыми данными, последовательно интегрируемыми с уже накопленной информацией.

Учитывая вышесказанное, наиболее удобной формой представления данных в базе данных института является хранилище данных со стандартной архитектурой [2] (рис. 4).

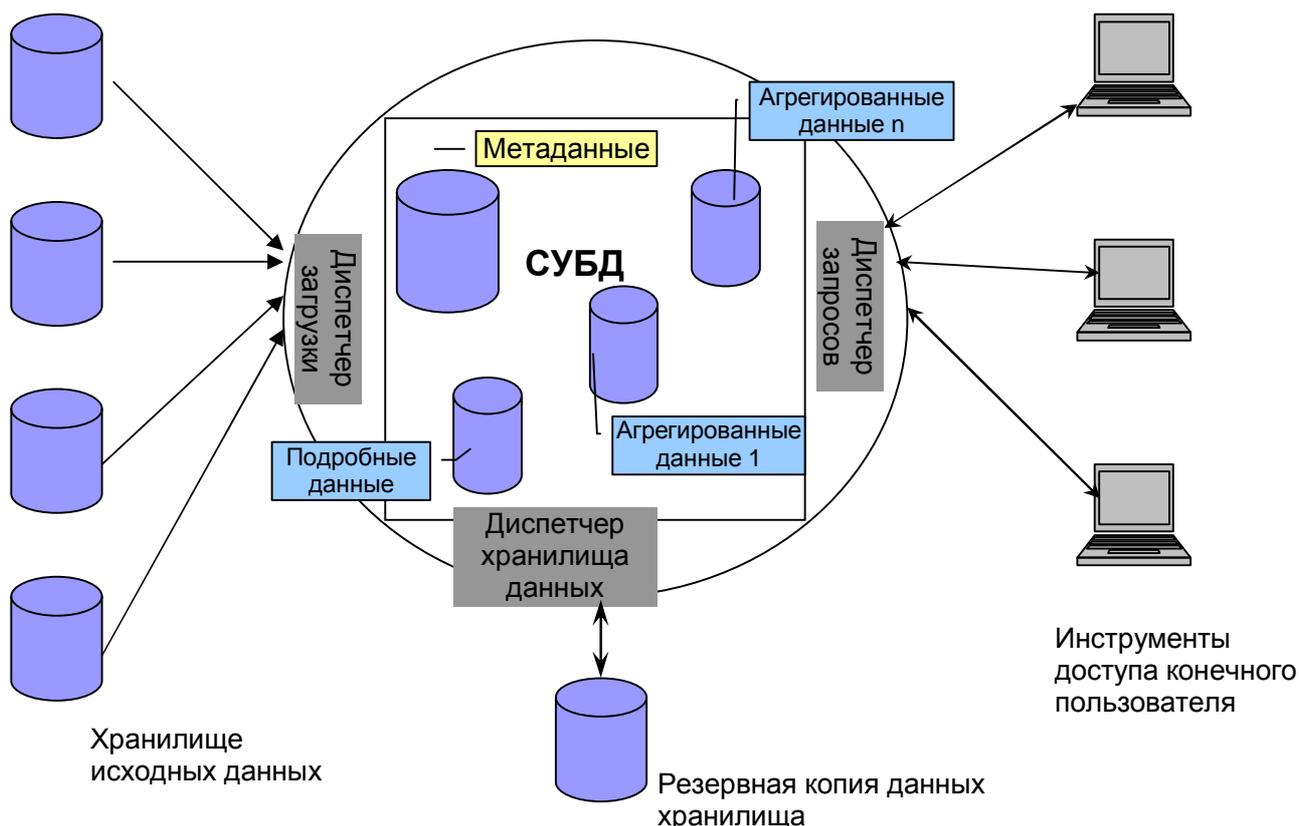


Рис. 4 Архитектура хранилища данных

Следует также отметить, что хранилищам данных присущ ряд недостатков:

- Недооценка ресурсов, необходимых для загрузки данных
- Скрытые проблемы источников данных
- Отсутствие требуемых данных в имеющихся архивах
- Повышение требований конечных пользователей
- Унификация данных
- Владение данными
- Сложное сопровождение
- Долговременный характер проектов

Однако перечисленные недостатки не снижают достоинств хранилища данных.

Проведя анализ расположения подразделений института и имеющиеся в их распоряжении данные было установлено, что наиболее приемлемой формой организации базы данных будет распределенное сетевое хранилище данных. Распределенность заключается в следующем: отдельные базы данных размещаются в компьютерной сети на различных машинах, связь между базами данных осуществляется под управлением СУБД, которое размещено на сервере. Кроме этого разрабатывается возможность работы с базой

данных при помощи СУБД, интегрированной в среду Web. В разрабатываемой системе распределение данных не будет заметно для конечного пользователя.

В организации хранилища данных в настоящее время используется файл-серверная архитектура. Суть ее в следующем [3]: база данных хранится на сервере, а копии СУБД устанавливаются на компьютерах пользователей (пример – «Реестр водных объектов Алтайского Края» см. ниже). Файл базы данных одновременно используется всеми пользователями, при помощи сетевого обеспечения. Архитектура файл-сервера позволяет добиться приемлемой производительности, т.к. в распоряжении каждой копии СУБД находятся все ресурсы компьютера пользователя. Но такая схема работы существенно загружает сеть.

Поэтому предполагается внедрение клиент-серверной и трехуровневой архитектуры [2] для институтских баз данных.

Клиент-сервер: база данных хранится на сервере, а СУБД подразделяется на две части клиентскую и серверную. Клиентская часть СУБД выполняется на стороне клиента и обеспечивает интерактивное взаимодействие с пользователем и формирование запросов к БД на языке SQL. Серверная часть работает на сервере и взаимодействует с базой данных, обеспечивая выполнение запросов клиентской части. Т.е. клиентская часть формирует и отправляет запрос, серверная часть по этому запросу производит выборку в базе данных и возвращает клиентской части только записи, которые удовлетворяют условию запроса.

Трехуровневая структура: При такой организации происходит еще большее разделение – тонкий клиент, сервер приложений, сервер базы данных. Тонкий клиент реализуется на стороне пользователя, чаще всего это интерактивное взаимодействие при помощи веб браузера с применением на HTML странице скриптов или апплетов (java, php, CGI и др.) Сервер приложений располагается на сервере и может быть либо специализированной программой (например, ArcIMS) либо же являться обычным веб сервером, вызывающим для обработки http запроса внешнюю программу при помощи скриптов. Преимущество состоит в возможности быстрой модернизации логики программы и отсутствии необходимости разрабатывать клиентскую программу.

### **Локальные базы данных**

Рассмотрим на примере локальные базы данных, которые в настоящий момент интегрируются в общеинститутскую базу данных.

Большая часть полученных результатов имеет географическую привязку, а базы данных представляет собой проекты геоинформационных систем (ГИС). Рассмотрим более подробно ГИС «Реестр водных объектов Алтайского края» (рис. 5).

В настоящий момент в эту базу данных входят: гидрографическая сеть, озера; контура водосборных бассейнов рек и озер; створы гидрохимических наблюдений; контура административных районов; населенные пункты.

В настоящий момент продолжается наполнение данной базы, подготовлены к включению точки расположения гидрологических и гидрохимических постов, точки отбора гидробиологических проб, точки сброса и забора воды промышленными предприятиями, различные данные наблюдений.

Геоинформационная основа базы данных позволяет формировать запросы и получать результаты по этим запросам в виде картографической информации (рис. 6). Результат выполнения запроса представлен на рисунке 7.

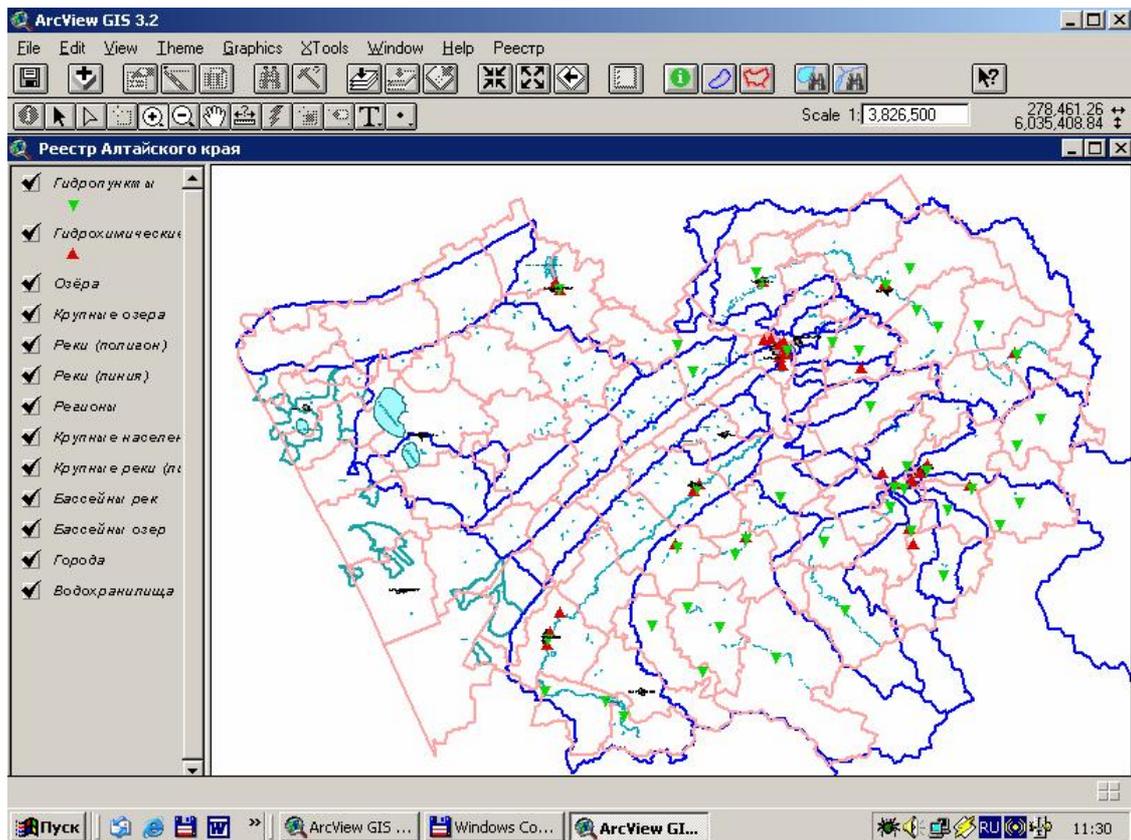
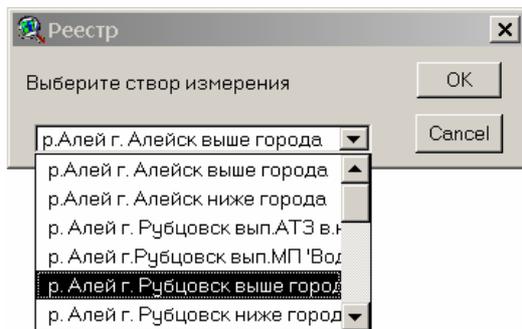
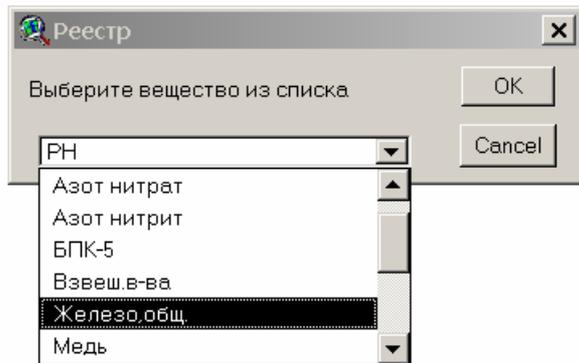


Рис. 5. Реестр водных объектов Алтайского края



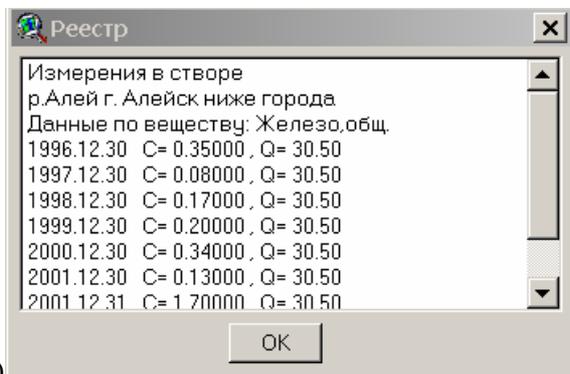
а)



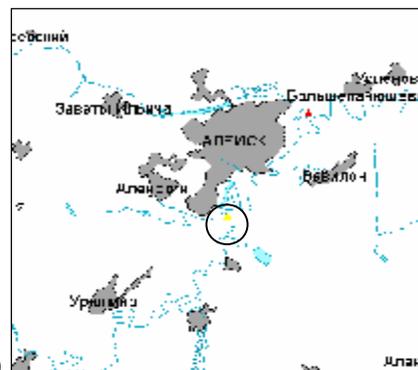
б)

Рис. 6. Пример формирования запроса по гидрохимическому пункту:

а) выбор пункта; б) выбор показателя



а)



б)

Рис. 7. Результат выполнения запроса по гидрохимическому пункту

а) информация о выбранном веществе (дата измерения, концентрация, расход);

б) фрагмент карты с выделенным пунктом.

Кроме этого существует возможность формировать запросы по бассейну, по региону (административному району), по рекам и по озерам. Также существует возможность проводить расчеты длины отрезка речной сети для решения различных задач природопользования (рис. 8).



Рис. 8. Пример выполнения запроса для расчета длины участка реки

### Заключение

Создание базы метаданных рассматривается как окончание начального этапа формирования единого общеинститутского информационного пространства с различным уровнем доступа. Интеграция уже существующих локальных баз данных в общеинститутское хранилище данных позволит увеличить надежность хранения имеющейся информации и обеспечить поддержку проектных работ информационными ресурсами.

### Литература

1. Ананьев П.И., Кайгородова М.А. Базы данных: Учебное пособие. – Барнаул: изд. АлтГТУ, 2005, - 183с.
2. Конолли Т., Брег. К. Базы данных. Проектирование, реализация и сопровождение. Теория и практика. - М.: Издательский дом "Вильямс", 2003, — 1440с.
3. Малыгина М.П. Базы данных: основы, проектирование, использование. – СПб.: БХВ-Петербург, 2004, - 512с.
4. Харрис Э. PHP/MySQL для начинающих. – М.: Кудиц-Образ, 2005, - 384с.

## ELABORATION AND CREATION OF IWEF DATA BASE

Petrov Andrey V., Lovtskaya Olga V., Sutorikhin Igor A.

*During the long-term work a great body of data including the information on the research carried out at different times has been collected in the laboratories of the Institute.*

*The creation of data base includes two stages. At the first stage, metadata base is created, i.e. the data characteristics in the data warehouse are described. The analysis of metadata is carried out, and the level of access to different data is defined. The second stage implies the creation of the unified data base including the outcomes of the Institute activity.*