

ГИБРИДНЫЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ ДЛЯ ПОДДЕРЖКИ МЕЖДИСЦИПЛИНАРНЫХ ИССЛЕДОВАНИЙ

Н. Н. ДОБРЕЦОВ, И. И. БОЛДЫРЕВ, Р. Д. ЮСУПОВ

Институт геологии и минералогии СО РАН, Новосибирск, Россия

e-mail: dnn@uiggm.nsc.ru, boldyrev@uiggm.nsc.ru

Several fundamental and functional system requirements for the development and support of an informational system aimed at interdisciplinary research projects have been considered. For such systems a two tier “hybrid” architecture has been proposed.

Введение

В настоящее время гетерогенность, распределенность и большие объемы накопленных геоинформационных ресурсов являются препятствием для их интеграции в единые банки данных. Об этом достаточно много говорится в научной и технической литературе. Особое место занимают научно-исследовательские и научно-технические проекты, поскольку именно в них, как правило, порождаются новейшие знания и данные, их сопровождающие. Все больше исследований проводится на стыке классических областей знания, появляются так называемые междисциплинарные проекты как форма организации научных исследований. Эти проекты объединяются по какой-либо тематике или географическому признаку и уже не несут жесткой привязки к конкретным областям знания. Как следствие, полученные результаты не попадают в существующие тематические стандарты и классификаторы, что является серьезным препятствием для создания электронных информационных ресурсов и соответствующих баз данных (БД). При этом все чаще появляются требования к обязательному созданию электронных информационных ресурсов по исследовательским проектам и полученным в их рамках результатам.

В течение 1998–2006 гг. лаборатория геоинформационных систем и дистанционного зондирования Института геологии и минералогии СО РАН принимала участие в международных проектах GASHYDAT (MAS3-CT98-0176), CRIMEA (EVK2-CT-2002-00162) и интеграционном проекте СО РАН “Алтайский экорегион” (№ 124) [1–5]. Общее для этих междисциплинарных проектов — то, что их результатами являются данные различной природы, объединенные по тематическому и/или географическому признаку, но представленные в разных форматах (и на разных носителях). В задачу группы входила разработка информационных систем, предназначенных для хранения, поиска (мета) данных, а также доступа к данным, размещенным в системе, в их оригинальном виде. В сверхзадачу группы входила разработка информационного ресурса, который бы

не выродился в простую веб-страничку проекта с текстовыми описаниями его целей и задач в начале и отчетными материалами в конце. Таким образом, необходимо было решить две основные проблемы: априорной разнородности данных и конечного вырождения информационного ресурса. Авторы настоящей работы поставили перед собой задачу обобщить практический опыт создания информационных ресурсов сопровождения и информационной поддержки междисциплинарных научно-исследовательских проектов, а также предложить концептуальные и архитектурные решения по их реализации.

1. Специфика междисциплинарных научных проектов (границы условия для информационной системы)

В процессе создания информационных систем выявились довольно существенная специфика самих проектов, которая позволила сформулировать границы условия для разработки информационной системы.

1. Короткий жизненный цикл проекта (в среднем три года). Отсутствие понимания участниками необходимой структуры информационного ресурса к началу реализации проекта. Появление большого количества первичных и/или исходных данных сразу после начала проекта.

2. Отсутствие выделенных групп для создания и сопровождения информационного ресурса. Как правило, предусматривается выполнение проекта силами профильных специалистов — участников проекта, из которых далеко не все имеют необходимую квалификацию и способны принимать непосредственное участие даже в использовании готовых технологий и решений задач в области информатики. В свою очередь, специалисты-предметники не желают тратить значительное время на работы, не связанные с их прямыми интересами.

3. Разнообразие структур и форматов первичных данных, которое определяется не только самими структурами и форматами, но и числом тематических направлений, используемых аналитических методов и т. п.

4. Число провайдеров данных, как правило, невелико и ограничивается числом участников проекта, тогда как число предполагаемых пользователей информационным ресурсом может быть (должно быть!) большим.

5. Доступ к информации иерархически регламентирован. Подразумевается, что в научно-исследовательских проектах нет ограничений на информацию о наличии тех или иных данных (нет ограничений на метаданные), тогда как сама информация представляется с авторскими ограничениями. Причем ограничения доступа де-факто не коррелируют с ограничениями на число участников (зарегистрированных пользователей). К сожалению, главным ограничением на доступ к информации является ее “первичность”. Пока результат научного исследования не опубликован, авторы не заинтересованы в предоставлении прямого доступа к первичным данным. В лучшем случае они готовы предоставлять результаты некоей предварительной обработки, например данные в виде диаграмм отношений элементов вместо исходных таблиц химических анализов.

6. Подавляющее большинство исходных и/или создаваемых в процессе реализации проекта данных не попадает под информационные стандарты содержания. Исключение составляют формальные отчетные материалы в виде научных публикаций, которые могут быть автоматически задействованы в библиографических информационных

системах. Но публикации по результатам проекта, как правило, появляются только в конце его жизненного цикла. Попытки разработки внутритематических стандартов [6], предпринимаемые в рамках целого ряда рабочих групп, на наш взгляд, не всегда эффективны. И связано это в первую очередь с бурным развитием технологий обработки данных, появлением новых аналитических методов, взаимным проникновением различных дисциплин и т. п., часто опережающим возможности формализации структурных связей и отношений.

7. Количество потенциальных “записей” в БД информационной системы проекта невелико. Речь идет не о количестве возможных измерений (анализов и т. п.), а о количестве экспериментов, натурных наблюдений, объектов исследования. Опыт создания информационных систем в рамках трех междисциплинарных проектов показал, что за время их существования в них появилось несколько сотен (первые тысячи) полноценных записей. Этот показатель важен, поскольку позволяет более свободно относиться к проблеме производительности системы.

8. Провайдеры данных, одновременно являющиеся участниками проекта и специалистами в предметных областях, предъявляют чрезвычайно высокие требования к пользовательским интерфейсам информационной системы. Интерфейсы должны быть просты, интуитивно понятны и максимально адаптированы к задачам проекта, а не к функциональным возможностям системы. Эти требования актуальны и для подсистемы ввода данных, поскольку последний должен осуществляться непосредственно провайдером.

Исходя из вышеперечисленных граничных условий можно сформулировать основные функциональные требования к информационной системе, ориентированной на поддержку междисциплинарных научных исследований.

1. Банк данных проекта должен предоставлять возможность сбора и хранения гетерогенной информации в различных форматах и различной семантике. Однако при этом должна предоставляться возможность гибкого унифицированного поиска элементов представленного информационного массива. Очевидным решением является использование метаданных — описательной информации, в унифицированном виде характеризующей непосредственно данные достаточно полно.

2. Структура описательной информации определяется стандартом содержания метаданных. Очевидно, что для различных предметных областей применимы различные стандарты содержания. Поэтому особенностью рассматриваемой системы должно быть отсутствие жесткой привязки к определенному стандарту содержания для ввода метаданных, обеспечиваемое возможностью настройки шаблонов ввода метаданных.

3. Система должна обеспечивать прямой доступ к открытым данным, указание которых не допускается набором дескрипторов. Для этого можно использовать вложенные ссылки как механизм “прикрепления” к метазаписи файлов любого формата и/или указателей на веб-адреса информационных ресурсов.

2. Практическая реализация

В рамках проектов Gas Hydrate Database и “Алтайский экорегион” разработана система, реализованная в виде Интернет-сайтов [<http://www.giscenter.ru/alta1>], предоставляющая пользователям следующие функциональные возможности: поиск метаданных, удаленный ввод новых и редактирование существующих метаданных (для зарегистри-

рованных пользователей с разграничением прав). Доступны следующие виды поиска: по вхождению подстроки (с использованием масок и логических операторов) в значения определенных полей метаданных, по кодам классификатора, по координатам описывавшего прямоугольника и расширенный атрибутивный поиск. Система успешно апробирована предметными пользователями в течение всего срока реализации проекта. Она удобна для пользователей с различным уровнем знаний компьютерных технологий.

Каждый элемент банка метаданных (метазапись) представляет собой описание конкретных данных (цифровой карты, электронной таблицы, книги, текстового документа, компакт-диска и т. д.). Набор представленных в метазаписи характеристик определяется стандартом содержания метаданных. Самыми распространенными стандартами в области геоинформационных технологий являются:

- CSDGM. Content Standard for Digital Geospatial Metadata [7];
- ISO 19115. Geographic information — Metadata standard [8];
- Z39.5. ISO 23950 Standard Information Retrieval (Z39.50): Application Service Definition and Protocol Specification and ANSI/NISO Z39.50 [9].

Вследствие своей универсальности эти стандарты довольно избыточны для реализации конкретного решения. Поэтому в рассматриваемом проекте в качестве стандарта содержания метаданных решено использовать подмножество дескрипторов из стандарта ISO 19115 с целью обеспечения совместимости со стандартом и возможности конвертирования в него метаданных [<http://www.giscenter.ru/altai/?page=info;metadata;metaform>] [4].

Каждая метазапись содержит набор обязательных дескрипторов: заголовок, аннотацию, коды классификатора, ключевые слова, информацию о производителях данных, описание носителя данных, язык оригинала данных, формы разрешений на данные, информацию об авторах метаданных и язык метазаписи. Кроме того, допускается заполнение таких необязательных дескрипторов, как: персоны, связанные с данными; первая публикация; дата создания данных; пределы параметров; не первая публикация; вложенные ссылки; координаты; географическая привязка и т. п. Дескрипторы делятся на уникальные и повторяемые: уникальные могут быть указаны в единственном экземпляре, а повторяемые — в нескольких. К уникальным дескрипторам относятся такие, как заголовок, аннотация, язык метазаписи, дата создания и др. Повторяемыми являются код классификатора, описание вложенной ссылки, информация о производителе данных и информация о лице, связанном с данными.

Для обеспечения возможности тематической идентификации описанных данных можно использовать тематический классификатор. Первый уровень классификатора базируется на тематическом составе проекта, представленном тематическими блоками (рис. 1). Второй и выше уровни классификатора разрабатываются ответственными исполнителями тематических блоков и представляют собой древовидные структуры для классификации информационных ресурсов [www.giscenter.ru/altai/?page=metabank;thes]. Каждая метазапись содержит список кодов классификатора, позволяющих тематически классифицировать данные, которые она описывает.

Фактически базовая структура метазаписи не отличается от подобных, реализуемых в библиографических и многих других БД, и как следствие предполагается наличие большого числа развитых технологических решений, которые могут быть применены при создании подобного компонента информационной системы.

Для включения в метазапись дополнительной информации о данных, указание которой не допускается набором дескрипторов, используются вложенные ссылки. Этот

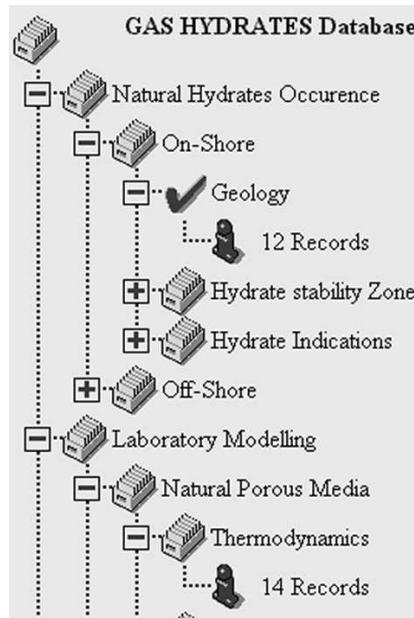


Рис. 1. Пример многоуровневого тематического классификатора БД проекта Gas Hydrate Database [3, 10]

механизм позволяет прикрепить к метазаписи файлы любого формата и указать ссылки на веб-адреса информационных ресурсов. Так, например, можно снабдить метазапись растровым изображением векторной карты, чтобы дать пользователям банка метаданных представление о содержимом карты, в том числе вместо ее подробного описания.

3. ФУНКЦИОНАЛЬНЫЕ ВОЗМОЖНОСТИ СИСТЕМЫ

Поиск метаданных. Главной функцией банка метаданных для пользователя является возможность поиска метаданных об интересующих пользователя данных. Реализованная система предоставляет следующие виды поиска: простой, тематический и географический.

Простой поиск. Этот вид поиска позволяет искать метаданные по вхождению некоторого набора строк в значения определенных дескрипторов. В рассматриваемой системе поиск осуществляется в заголовке, аннотации и ключевых словах метазаписей. Разработанный механизм позволяет задавать нетривиальные запросы для поиска с помощью конъюнкции, дизъюнкции или группировки значений.

Тематический поиск. Тематический поиск объединяет возможности простого поиска с возможностью фильтрации результатов поиска по тематическому критерию. Пользователь может явно указать, на какие узлы дерева тематического классификатора должны ссылаться метаданные, которые будут получены в результате поиска. Страница тематического поиска показана на рис. 2.

Географический поиск. Для поиска геокодированных метаданных, т. е. имеющих географическую привязку в виде координат, предусмотрена возможность ограничения области поиска прямоугольником на географической или тематической карте региона. Набор карт, доступных в качестве подложек для поиска, может быть расширен. В настоящее время на сайте проекта представлены четыре карты экорегиона: политическая,

почвенная, карта рельефа и карта антропогенных образований. Страница географического поиска с почвенной картой показана на рис. 3.

Просмотр результатов поиска. Все поисковые интерфейсы генерируют страницу с результатами поиска в виде таблицы заголовков метазаписей, удовлетворяющих критериям поиска. Пользователь имеет возможность просмотра найденных метазаписей с различной степенью детализации (только заголовок, основная информация или полная информация). Если метазапись содержит вложенные ссылки, страница ее просмотра содержит гипертекстовые ссылки на прикрепленные объекты. Также страница просмотра содержит ссылки на адреса электронной почты лиц и организаций, информация о которых представлена в соответствующих дескрипторах (производитель дан-

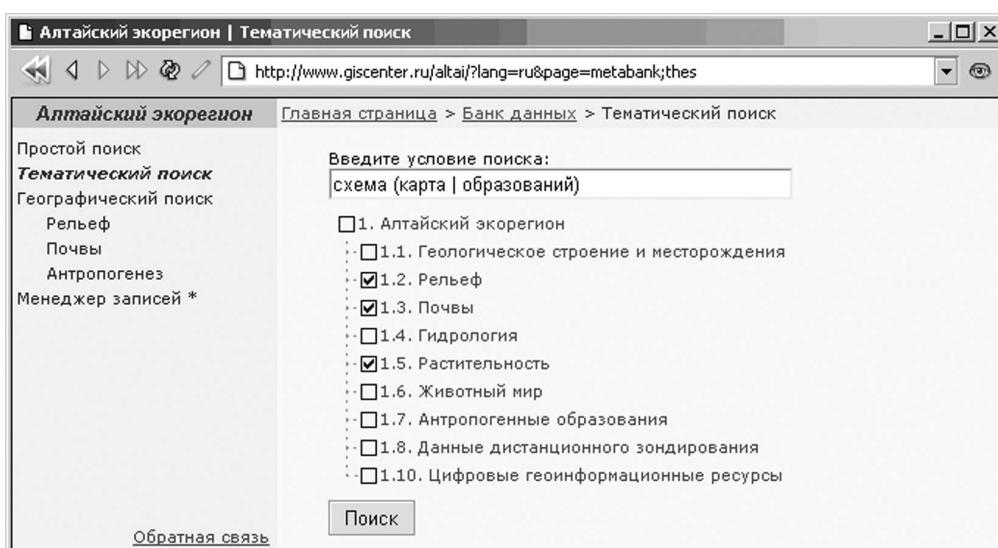


Рис. 2. Страница тематического поиска

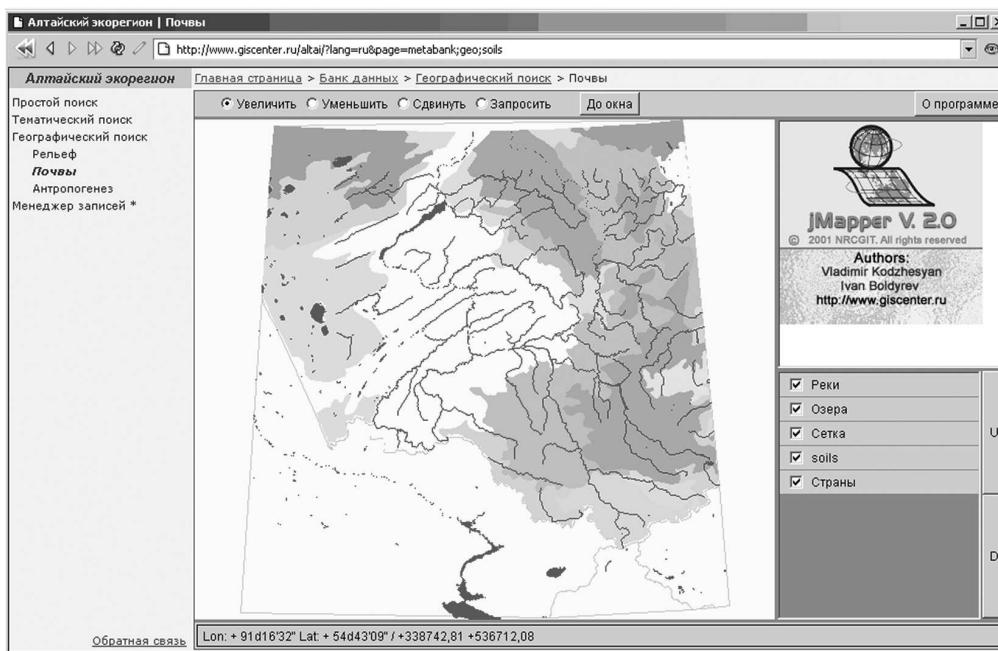


Рис. 3. Географический поиск. Почвенная карта

ных, автор метазаписи, заполнитель метазаписи и лица, связанные с данными). Страница просмотра снабжена панелью навигации по результатам поиска, которая позволяет перемещаться к первой, предыдущей, следующей и последней метазаписи в списке. Страницы просмотра результатов поиска, метазаписи и вложенных ссылок показаны на рис. 4.

Screenshot 1: Search Results Page

Запрос выбирает записи, в которых:
- индексы классификатора соответствуют ветвям "1.2 Рельеф" или "1.3 Почвы" или "1.5 Расщелинность"
- заголовок, аннотация или ключевые слова содержит ("схема") или ("образование")

Найдено 31 записи

№	Заголовок
1	Государственная геологическая карта кайнозойских образований; масштаб 1:200000; М-45-1 (Солнечное); издание второе, Алтайская серия. Министерство природных ресурсов Российской Федерации. Просмотр
2	Государственная геологическая карта кайнозойских образований; масштаб 1:200000; М-45-XV (Катанда); издание второе, Алтайская серия. Министерство природных ресурсов Российской Федерации. Просмотр
3	Государственная геологическая карта неоген-четвертичных образований; масштаб 1:200000; М-44-XI (Энгигито); издание второе, Алтайская серия. Министерство природных ресурсов Российской Федерации. Просмотр
4	Государственная геологическая карта неоген-четвертичных образований; масштаб 1:200000; М-45-VII (Усть-Каменогорск); издание второе, Алтайская серия. Министерство природных ресурсов Российской Федерации. Просмотр
5	Государственная геологическая карта СССР (новая серия) масштаб 1:1 000000. Карта четвертичных отложений, лист И - (44), 45 (Усть-Каменогорск) Просмотр
6	Геоинформационная система (ПИС) Алтай-Саянского экорегиона на CD Просмотр
7	Геонорфологическая карта СССР масштаб 1 :1 000 000, лист Н - (44), 45 (Усть-Каменогорск) охватывает территорию Алтая. Просмотр
8	Геонорфологическая схема окрестностей оз. Ая. Просмотр
9	Геонорфологическая схема пло...
10	Геонорфологическая схема дол...
11	Геонорфологическая схема дол...
12	Разрез "Чаган-Узун-2" и схема...
13	К вопросу о размерах среднег...
14	Карта четвертичных отложений...
15	Карта четвертичных отложений...
16	Карта четвертичных отложений...
17	Карто-схема комплексных форм...
18	О генезисе и возрасте пектонча...
19	О происхождении Майминского...
20	Палеогеографическая схема по...
21	Палеогеографическая схема по...
22	Палеогеографическое устремле...
23	Почвенный покров инзекций...
24	Структура глациального перио...
25	Структурно-геонорфологическая...
26	Схема геонорфологического ст...
27	Схема расположения конечных...
28	Схема катастрофических прор...
29	Схема интенсивности тектонич...
30	Схема типов берегов Чуйско-К...
31	Схема деградации позднеплей...

Screenshot 2: Detailed Record View

Запись 7 / 31 Отображать основную информацию ▾ Первая Предыдущая Следующая Последняя

Заголовок Геонорфологическая карта СССР масштаба 1 :1 000 000, лист М - (44), 45 (Усть-Каменогорск) охватывает территорию Алтая.

Аннотация Геонорфологическая карта СССР масштаба 1:1 000 000, лист М - (44), 45 охватывает территорию Алтая между 48° и 58° северной широты и 81° и 90° восточной долготы. На карте отображены рельеф и генетические типы отложений палеоген-четвертичного возраста. Карта снабжена схемами: политико-административного деления; структурно-геонорфологического районирования (масштаб 1 :1 500 000); схемой использованных материалов (масштаб 1 :1 500 000); схемой расположения листа.

Ключевые слова Геонорфология, рельеф.

Дата создания данных Кarta составлена в 1975 году.

Первая публикация "Государственная геологическая карта СССР масштаба 1 :1 000 000 (новая серия), Геонорфологическая карта, лист М - (44), 45 (Усть-Каменогорск), 1978 г.

Геологическая карта СССР масштаба 1 :1 000 000 (новая серия). Объяснительная записка. лист М - (44), 45 (Усть-Каменогорск), 1980, с. 145 (Министерство геологии СССР, ВСЕГЕИ)."

Вложенные ссылки
Файлы
[изображение карты \(395.3 Kb, файл jpg\)](#)

Производители данных

- Всесоюзный аэрогеологически трестест.
- Восточно-Казахстанское территориальное геологическое управление.
- Западно-Сибирское территориальное геологическое управление.
- Всесоюзном ордена Ленина научно-исследовательский геологический институт (ВСЕГЕИ).

Screenshot 3: Attached Map

http://www.giscenter.ru/altai/metabank/data/mart/2_mart_geommap.jpg

Рис. 4. Просмотр метаданных и вложенных ссылок

Удаленное создание и изменение метаданных. В информационной системе реализована возможность удаленного добавления, создания новых и редактирования уже размещенных в банке метазаписей. Эти операции доступны зарегистрированным пользователям банка метаданных, выполнившим процедуру аутентификации. В системе предусмотрено ролевое разделение полномочий пользователей, позволяющее ограничить набор доступных операций с метаданными и определить множество метаданных, с которыми эти операции могут быть выполнены. Так, зарегистрированные пользователи, имеющие статус заполнителей метаданных, могут создавать новые метазаписи и редактировать добавленные ими метазаписи. Пользователи со статусом администратора имеют право создавать, редактировать и удалять любые метазаписи. Страница со списком метазаписей, доступных для редактирования, и страница редактирования метазаписи показаны на рис. 5.

Функциональные возможности систем могут быть расширены в случае возможности представления исходных наборов данных в виде таблиц, значения в столбцах которых (например, содержание химических соединений в пробе воды) можно рассматривать как дискретные функции от определенного параметра (например, глубины измерения).

В проекте CRIMEA [11] подавляющее большинство результатов исследований представляет собой таблицы значений различных характеристик воды на различной глубине. Для этой системы реализована возможность визуализации динамически сформи-

The screenshot displays two windows related to metadata management:

- Редактирование метаданных (Top Window):** A form titled "Алтайский экорегион" showing a classification tree under "Тематический классификатор". The tree includes categories like "Алтайский экорегион", "Геологическое строение и месторождения", "Рельеф", "Почвы", "Гидрология", "Растительность", "Животный мир", "Антропогенные образования", "Данные дистанционного зондирования", and "Цифровые геоинформационные ресурсы". On the left, a sidebar lists various metadata fields such as Title, Annotation, Classifier, Keywords, Persons, First publication, Holder of information, Date of data creation, Language of original data, Data formats, Authors of metadata, Contact information (with sub-fields for Name, Organization, Address, City, Postcode, State, Country, Phone, Email), Attached files, Coordinates, Geographical links, and Language of the metadata.
- Банк метаданных (Bottom Window):** A table titled "В банке 188 записей" showing a list of metadata entries. The columns are RID, Заголовок (Title), Действие (Action), Обновлена (Updated), and Создатель (Creator). The table lists entries such as "300 лет горно-геологической службе России: история горнорудного дела, геологическое строение и полезные ископаемые Алтая. Материалы региональной научно-практической конференции (14-15 апреля 2000 г.)", "База данных по эпицентрам землетрясений по территории Республики Алтай.", "Автодороги и пути Алтас-Саянского региона.", "Биогеохимия никрозлементов в Горном Алтасе.", "Гляциологические исследования в горном узле Белухи", "Гляциологические исследования в горном узле Белухи", "Горный Алтай", "Горный Алтай и Россия 240 лет (материалы симпозиума)", "Горные крепости Алтас-Саянского региона.", "Города (полигоны) Алтас-Саянского региона.", and "Города Монголии Алтас-Саянского региона.".

Рис. 5. Редактирование метаданных

рованных поднаборов данных (выборок) из различных дисциплин, удовлетворяющих определенным критериям и сгруппированных по выбранным параметрам. Исходные наборы данных можно рассматривать как частные случаи такой выборки. Для визуализации полученных выборок могут быть использованы несколько представлений с опциональным разделением выборки на характеристики (рис. 6):

- табличный вид (с несколькими вариантами группировки) и двумерные графики значений параметров как функции от выбранного параметра с опцией интерполяции значений;
- значения характеристик или пузырьковые диаграммы на географической карте;
- трехмерные диаграммы, отображающие значения характеристик в виде столбцов на географической карте.

Задача формирования динамических поднаборов данных, так же как и задача их предварительной обработки для последующей визуализации, была поставлена участниками проекта в качестве задачи реализации ограниченного доступа к исходным данным. Участники проекта (провайдеры) категорически отказались разрешить прямой доступ к исходным данным до публикации результатов исследования. Вместе с тем

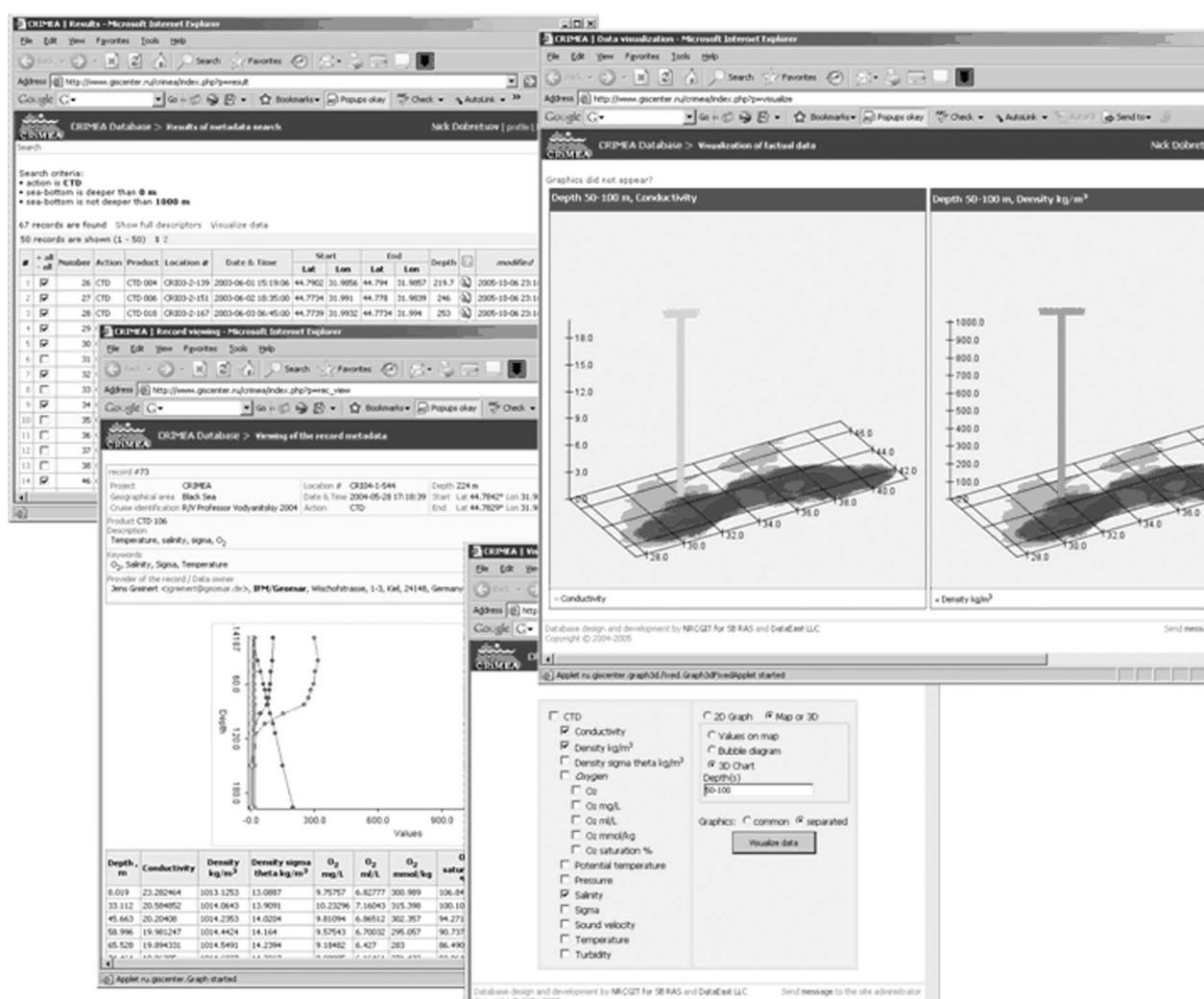


Рис. 6. Визуализация данных в виде результата предварительной обработки

была очевидна потребность в обмене информацией по мере ее получения. В качестве компромиссного решения предложено встроить в информационную систему аналитические функции, которые позволили преобразовывать данные в вид, достаточный для их восприятия и анализа, но без прямого доступа к ним. “Аналитическая” подсистема реализована в виде набора плагинов, которые динамически считывают данные из таблиц с известной структурой и формируют визуальное представление в виде графиков, диаграмм, карт и т. п. Тем самым обеспечивается возможность работать с данными без прямого доступа к ним.

4. Двухуровневая “гибридная” архитектура

“Гибридность” предлагаемой архитектуры для разработки информационной системы заключается в том, что в ее онтологии предполагается два условно независимых уровня экземпляров вместо одного.

Первый уровень — “объекты наблюдения” — реализуется в виде базы метаданных. Он представляет собой описания физических объектов (люди, географические объекты, природные явления и т. п.), для которых устанавливается иерархия отношений, в соответствии с которой их (возможные) объединения в группы и/или классы имеют четкую вложенность и, как следствие, определяют разную степень генерализации информации. В простейшем виде классы “объектов наблюдения” реализуются в виде дерева и включают в себя всю предметную (тематическую) составляющую информационной системы.

Второй уровень, который мы называем уровнем “баз данных”, в онтологии предлагаемой информационной системы также представляет собой уровень экземпляров, но “прикрепленных” к записям базы метаданных, описанных выше. В идеале этот уровень должен быть предельно абстрактным относительно предметной области, вплоть до распределения информации только по типам структур данных и/или форматов (тексты, таблицы, изображения и т. п.). В свою очередь, иерархия классов на этом уровне строится по возможностям и ограничениям обращения к конкретным типам структур.

Например, каталог космоснимков объединяет в себе набор изображений и их описания. Последние являются метаописаниями и позволяют организовать запросы к каталогу, но сами элементы каталога (снимки) этого не могут. Практически невозможно организовать запрос непосредственно к снимку (время, облачность, спектральный диапазон и т. п.). Всю эту информацию можно извлечь, только получив сам снимок “на руки” или путем наложения на него соответствующих дескрипторов. Причина — базовые свойства структуры (формата) единицы хранения информации.

Другой пример, результаты геохимического опробования для точечного пространственного объекта наблюдения. Как правило, эти данные формируются в виде таблиц, причем в наиболее простом виде каждая таблица привязана к конкретному объекту наблюдения (измерения) и содержит в себе минимум две колонки: описание анализируемого элемента и его содержание в образце. Если такая “таблица” вложена в развитую структуру БД, содержащую необходимые элементы онтологии (классы, атрибуты, отношения), связывающие информацию об объекте наблюдения с содержанием конкретного химического элемента в его образце, то организация запроса по этому элементу — вопрос чисто технический. Проблема лишь в разработке и создании развитой БД. Если же таблица привязана к объекту в виде вложенной ссылки, то мы получаем логическую точку доступа к новому информационному ресурсу, который является самостоятель-

ным по отношению к первоначальной записи и представляет собой независимую БД второго уровня (рис. 7).

Доступ к этой БД второго уровня определяется ее форматом, внутренней структурой и сервисами, которые реализуют технологии поддержки обеих БД. Собственно, в описанной схеме нет ничего нового, и подобным образом организованы многочисленные решения, например по доступу к мультимедийным ресурсам через Интернет (Blogpost-сервисы и т. п.).

Организовать доступ к независимой БД, привязанной к метазаписи БД верхнего уровня, можно несколькими способами.

Первый способ, самый простой, — создать такой набор дескрипторов для записи верхнего уровня, который бы позволил организовать запрос по критериям, достаточным для получения уникальной ссылки на БД второго уровня, т. е. включить в описание объектов верхнего уровня как минимум описание свойств вложений или объектов второго уровня. Такой способ успешно реализуется в каталогах данных дистанционного зондирования. Ограничением является возможность подключать только одинаковые по структуре БД второго уровня. Иначе структура записи верхнего уровня становится бесконечно сложной.

Второй способ организации доступа, традиционно реализуемый во многих информационных системах, основан на точном знании структуры БД второго уровня и заранее определенном списке возможных запросов к ней. Очень часто реализация сопровождается созданием третьей БД, содержащей в качестве записей описание связей между БД верхнего и второго уровней. Примерами таких информационных систем могут служить БД хозяйственной деятельности предприятий (так называемые бухгалтерские системы), связывающие финансы, материальные средства, кадры и т. п. Преимущество данного способа — возможность организовать связи не только между БД верхнего и второго уровней, но и между различными БД второго уровня. Ограничением данного способа является необходимость жесткой (заранее определенной) структуры связей и знания структур самих БД, включаемых в информационную систему. Кроме того, необходима их синхронизация.

Третий способ, который получает все большее распространение в последнее время, основан на стандартизации сервисов предоставления информации. Типичным примером может служить организация потокового (stream) доступа к мультимедийной информации через Интернет. Этот способ наиболее интересен для рассматриваемых нами

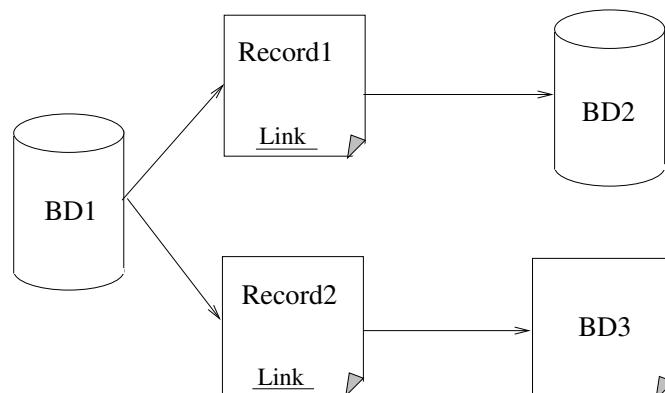


Рис. 7. Простейшая структура связанных БД через вложенные ссылки

систем, так как он не требует ни априорного знания структуры внешней БД, ни ее содержания, ни синхронизации.

Рассматриваемая нами “гибридная” архитектура предусматривает одновременную реализацию первого (для БД верхнего уровня) и третьего (для БД второго уровня) способов.

На первом этапе база данных верхнего уровня наполняется метаданными с вложенными ссылками (рис. 8). Если ссылка указывает на внешнюю базу данных второго уровня, не являющуюся документом (файлом и т. п.), запрашиваются ее дескрипторы, описывающие структуру, режим доступа, протокол взаимодействия и т. д. Эти дескрипторы сохраняются в базе метаданных.

Пользователь при обращении к системе получает доступ к релевантным метазаписям из БД верхнего уровня и описательным дескрипторам баз второго уровня, это позволяет ему сформировать запрос не только на уровне метаданных, но и включить в него подзапросы к базам второго уровня.

Получив запрос от пользователя, система последовательно обращается к соответствующим базам второго уровня, используя информацию в метазаписях. Получив ответы от баз второго уровня, система генерирует общий отчет, включающий как непосредственно данные из БД второго уровня, так и результат их обработки (диаграммы, графики, карты).

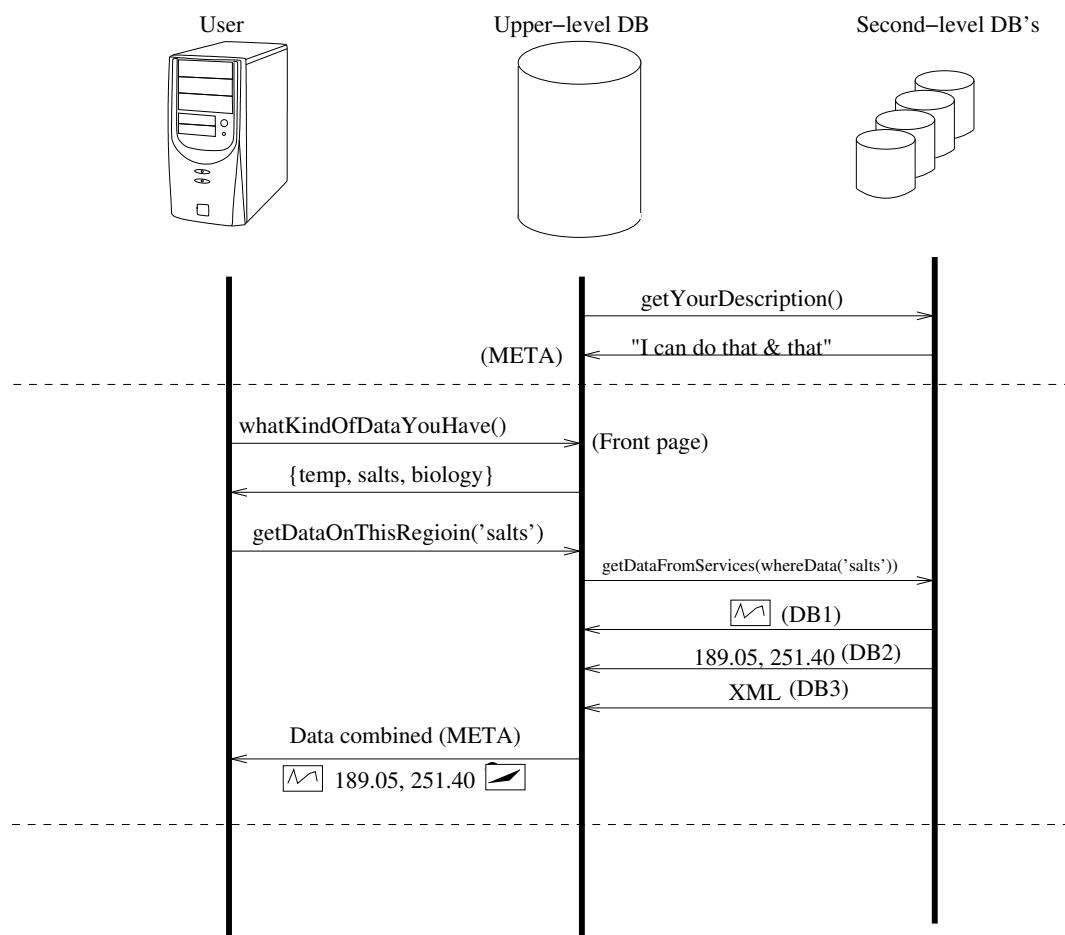


Рис. 8. Концептуальная схема гибридной базы данных

Такой подход хотя и требует разработки протоколов взаимодействия между БД верхнего и второго уровней, но позволяет избежать жесткой привязки к их структурам и не требует синхронизации.

Список литературы

- [1] Юсупов Р.Д. Опыт разработки банков метаданных о геоинформационных ресурсах на примере интеграционного проекта СО РАН “Алтайский экорегион” // Тез. докл. Междунар. конф. по измерениям, моделированию и информационным системам для изучения окружающей среды (ENVIROMIS 2004). Томск: Изд-во Том. ЦНТИ, 2004. С. 63.
- [2] YUSUPOV R., BOLDYREV I., DOBRETSOV N.N. ET AL. Functionalities for data processing and visualization within a database, as a tool for environmental purposes (monitoring) // 5th Europ. Congress on Regional Geoscientific Cartography and Information Systems, 13–16 June, 2006, Barcelona, Spain; Proc., Vol. II; Institut Geològic de Catalunya. P. 258–259.
- [3] KLERKX J., DE BATIST M., TOHIDI B., ET AL. The Gas Hydrate Database, <http://www.gashydat.org/>, 1998.
- [4] АЙРИЯНЦ А.А., ДОБРЕЦОВ Н.Н., ЗОЛЬНИКОВ И.Д. и др. Опыт систематизации и использования информации о рельефе в банке данных и метаданных проекта “Алтайский экорегион” // Рельефообразующие процессы: Теория, практика, методы исследования: Матер. XXVIII пленума Геоморфологической комиссии РАН. Новосибирск: ИГ СО РАН, 2004. С. 12–13.
- [5] ЮСУПОВ Р.Д. Геоинформационное обеспечение банка метаданных интеграционного проекта СО РАН “Алтайский экорегион” // Тез. докл. Второй Сиб. междунар. конф. молодых ученых по наукам о Земле. Новосибирск: ОИГГМ СО РАН, 2004.
- [6] SLOAN D. Development of a Gas Hydrate Markup Language, February, 2007 Final Report DE-AF26-06NT00440: [PDF-1.22MB], Center for Hydrate Research, Colorado School of Mines, http://204.154.137.14/technologies/oil-gas/publications/Hydrates/reports/NT00440_GHMLRpt.pdf
- [7] CSDGM. Content Standard for Digital Geospatial Metadata. <http://www.fgdc.gov/metadata/csdgm/>
- [8] ISO 19115. Geographic information — Metadata standard. <http://metadata.dgiwg.org/standard/index.htm>
- [9] Z39.50. ISO 23950 Standard “Information Retrieval (Z39.50): Application Service Definition and Protocol Specification” and ANSI/NISO Z39.50; <http://www.loc.gov/z3950/agency/>
- [10] SLOAN D. ET. AL. A hydrate database: vital to the technical community // Data Sci. J. 2007. Vol. 6. Gas Hydrate Issue. 16 June 2007. P. GH1–GH5.
- [11] YUSUPOV R., BOLDYREV I., DOBRETSOV N.N. ET AL. Functionalities for data management, processing and visualisation in a database system. The database of the project CRIMEA: Contribution of high-intensity gas seeps in the Black Sea to methane emission in the atmosphere // Четвертая Верещагинская байкальская конференция. Иркутск. 2005, www.crimea-info.org