

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ И НАУКЕ

Государственное образовательное учреждение
высшего профессионального образования
«Оренбургский государственный университет»

И.А. НИКИФОРОВ

КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ ГЕОЛОГИЧЕСКИХ ЗАДАЧ

Рекомендовано Учёным советом государственного образовательного учреждения высшего профессионального образования «Оренбургский государственный университет» в качестве учебного пособия для студентов, обучающихся по программам высшего профессионального образования по специальности «Геологическая съёмка, поиски и разведка месторождений полезных ископаемых»

Оренбург 2005

УДК

ББК

Рецензент

доктор геолого-минералогических наук, профессор П.В. Панкратьев

Никифоров И.А.

Н 30

**Компьютерное моделирование геологических задач:
учебное пособие/ И.А. Никифоров. -Оренбург: ГОУ ОГУ,
2005. - 99 с.**

ISBN

В пособии изложены теоретические предпосылки и принципы компьютерного моделирования недр. Рассмотрены основные этапы освоения эмпирического материала с использованием базовых компьютерных программ.

Учебное пособие предназначено для студентов, обучающихся по программам высшего профессионального образования по специальности 130300, со специализацией "Геоинформатика".

Н

ББК

ISBN

© Никифоров И.А., 2005

©ГОУ ОГУ, 2005

Содержание

ВВЕДЕНИЕ.....	4
1 ЗАДАЧИ КОМПЬЮТЕРНОГО МОДЕЛИРОВАНИЯ В ПРИКЛАДНОЙ ГЕОЛОГИИ	6
1.1 Предмет дисциплины.....	6
1.1.1 Прикладная геология.....	7
1.1.2 Методология и техника компьютерного моделирования.....	9
1.2 Этапы создания компьютерных моделей.....	11
1.2.1 Оформительский подход.....	11
1.2.2 Кибернетический подход.....	12
2 ЦИФРОВЫЕ МОДЕЛИ ГЕОЛОГИЧЕСКИХ ДАННЫХ	15
2.1 Реляционная модель цифровой геологической карты (DGM).....	15
2.1.1 Определение модели данных цифровых геологических карт.....	15
2.1.2 Предпосылки создания модели данных ЦГК.....	17
2.1.3 Содержание ЦГК.....	18
2.2 Основы моделирования.....	21
2.2.1 Геометризация картографических элементов.....	22
2.2.2 Сингулярные и составные геологические объекты.....	23
2.3 Детальное описание реляционной модели данных	27
2.3.1 Метаданные.....	28
2.3.2 Легенда.....	30
2.3.3 Архив сингулярных объектов (Singular Object Archive).....	36
2.3.4 Архив составных объектов (Compound Object Archive).....	38
3 ОСНОВЫ ПРОСТРАНСТВЕННЫХ БАЗ ДАННЫХ.....	49
3.1 Введение в пространственные базы данных.....	49
3.1.1 Обзор вопроса.....	49
3.1.2 Кто заинтересован в управлении пространственными данными.....	50
3.1.3 ГИС и СУПБД.....	51
3.1.4 Таксономия пространства и модели данных.....	53
3.1.5 Пространственные запросы.....	54
3.1.6 Элементарное введение в SQL.....	56
3.1.7 Расширение SQL для пространственных данных.....	59
3.1.8 Обработка и оптимизация запросов.....	63
3.1.9 Пространственные операции.....	64
3.1.10 Двухэтапная обработка запроса с пространственными операциями.....	65
3.2 Введение в добычу пространственных данных.....	67
3.2.1 Поиск информационных шаблонов.....	68
3.2.2 Процесс добычи данных.....	69
3.2.3 Основной закон географии.....	70
3.2.4 Показатели пространственной формы и автокорреляции.....	72
3.2.5 Приёмы добычи данных.....	74
4 КОМПЛЕКСЫ ГЕОЛОГИЧЕСКОГО МОДЕЛИРОВАНИЯ.....	79
4.1 Функциональность и входные данные	80

4.2 Приложение Borehole Manager.....	84
4.2.1 Ввод данных.....	86
4.2.2 Обмен данными с внешними хранилищами.....	89
4.2.3 Функциональность Borehole Manager.....	91

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ.....	102
--	------------

Введение

Моделирование природных явлений относится к числу наиболее прогрессивных направлений современного естествознания. В приложении к задачам геологического плана их изучение особенно актуально. Это вызвано растущими потребностями общества в оперативном и грамотном управлении огромными материальными ресурсами, задействованными в геологической и добывающей отраслях.

Повышенный спрос на исчезающее минеральное сырьё привёл к ряду успешных программных реализаций концепции моделирования месторождений. Алгоритмическая база таких программ эксплуатирует идею косвенного подхода к изучению недр, которые понимаются, во многом, как формальная информационная структура. Применение методов математической логики, мощного статистического аппарата и развитых компьютерных технологий позволяет в реальном времени прогнозировать развитие модели геологической среды. Её состояние и реакция на внешнее воздействие, имитируемое исследователем, является основанием для принятия ключевых управленческих решений.

Данный курс состоит из описания основных этапов математического моделирования, которыми обязан владеть современный горный инженер. Материал излагается в течение одного семестра и разбит на четыре раздела:

- 1) задачи компьютерного моделирования в прикладной геологии;
- 2) цифровые модели геологических данных;
- 3) основы пространственных баз данных;
- 4) программные комплексы геологического моделирования.

Для освоения курса студент должен обладать достаточно глубокими знаниями дисциплин общегеологических направлений, понимать задачи геолого-разведочного процесса на разных его стадиях и уметь работать с компьютером. Структура изложения материала приведена в нижеследующем списке:

- 1) задачи компьютерного моделирования в прикладной геологии:
 - предмет прикладной геологии;
 - методология и техника компьютерного моделирования;
 - синхронность геологоразведочного процесса и информационного освоения недр;
- 2) цифровые модели геологических данных:
 - цифровая геологическая карта (ЦГК);
 - реляционная модель данных ЦГК;
 - архивы и легенда ЦГК;
 - реляционные таблицы модели данных ЦГК;
- 3) основы пространственных баз данных:
 - ГИС и пространственные базы данных;

- модели данных;
 - направления развития пространственных баз данных;
- 4) программные комплексы геологического моделирования:
- обзор вопроса;
 - система RockWorks. Структура комплекса и функциональность.

Изложение курса сопровождается практическими занятиями, в ходе которых студенты должны закрепить навыки компьютерной обработки данных, полученные в ходе четырёх предыдущих семестров.

Студентам предстоит исследование учебной базы данных, содержащей информацию по геохимии микроэлементов в колчеданно-полиметаллическом месторождении.

На лабораторных занятиях студентам необходимо произвести:

- оценку статистического распределения рудных компонентов в учебной базе данных;
- пространственную привязку интервалов опробования в скважинах наклонного бурения;
- загрузить трассы скважин в среду ГИС ArcView 3.2;
- сгенерировать средствами модуля Spatial Analyst серию погоризонтных планов рудных компонентов в изолиниях стандартных отклонений;
- сгенерировать средствами модуля 3D Analyst триангуляционную модель рудных тел и определить их количественные характеристики;

Результаты должны быть оформлены в виде курсового проекта.

1 Задачи компьютерного моделирования в прикладной геологии

1.1 Предмет дисциплины

Одно из самых известных определений геологии таково: *Геология — это наука о строении и истории развития Земли.* (1)

Оно включает в себя сотни научных направлений, в связи с чем нуждается в конкретизации каждого из них. Это справедливо и для нашего предмета, который ориентирован на решение сугубо практических (прикладных) задач, не претендуя на информационное описание недр в планетарном масштабе.

Надо сказать, что главные препятствия здесь состоят даже не в отсутствии вычислительных мощностей, а в недостаточности именно фундаментальных знаний о Земле. Для широкого практического использования они должны быть предельно формализованы, характеристики связей разнопорядковых геологических объектов известны и сведены до уровня математических выражений. По-видимому, это время ещё не наступило и под математическим моделированием* до сих пор подразумевается многовариантный перебор возможных геологических реализаций природных процессов. При этом роль компьютерных технологий сводится к ускорению данного перебора, а роль теории — к сокращению числа перебираемых вариантов и выбору оптимального из них.

Следует учесть, что сценой для постановки математического моделирования являются уже подготовленные в ходе геологоразведочных работ площади, т.е. наиболее изученные фрагменты земной коры. Их модельная детализация производится в предположении, что они информационно опознаны и в своём развитии следуют известным законам природы или, по крайней мере, устойчивым эмпирическим закономерностям.

Необходимо сознавать также, что речь идёт не просто о моделировании некоторых заранее заданных термодинамических условий или даже об имитации очень сложных процессов (генерации углеводородов, например). Такие задачи как раз решаемы, но не имеют пока серьёзного практического значения. Гораздо важнее, чтобы техника математического моделирования обеспечила возможность прогноза в конкретном блоке земного пространства, уже охарактеризованном некоторым фактическим материалом. Тогда любые просчёты метода сразу станут заметны и исправляемы, а моделирование из разряда экзотических новинок превратится в надёжный рабочий инструмент исследователя.

Сказанное означает, что предмет компьютерного моделирования в геологии имеет, по меньшей мере, двойственную природу. Этот дуализм объясняется, с одной стороны, уникальными особенностями геологической среды,

*Здесь и далее по тексту компьютерное и математическое моделирование понимаются как синонимы.

а с другой- методологическими подходами и технологиями, имитирующими более или менее длительное развитие данного пространства.

Здесь уместно провести некоторую параллель с формированием человеческой личности под влиянием двух главных факторов: наследственности (генетический фактор) и воспитания (приобретённый фактор). В нашем случае генетический фактор определяется материальными предпосылками, т.е. строением недр, или, что более верно, информацией о них.

Приобретённый фактор будет буквально означать приобретённые аппаратно-программные комплексы, воплощающие идеи выдающихся исследователей о процессах развития геологической материи, т.е. об её изменчивости.

Неправильная оценка этих составляющих неминуемо приведёт к провалу весьма дорогостоящих проектов или превратит сам процесс моделирования в самоцель. Чтобы избежать такого развития событий, необходимо конкретизировать само понятие предмета данной дисциплины, а это- очень непростой и во многом философский вопрос.

1.1.1 Прикладная геология

Говоря о строении недр, как о генетической компоненте нашего предмета, необходимо конкретизировать два важных для практикующего геолога момента:

- что, вообще, следует понимать под недрами и их геологическим строением?
- каковы должны быть принципы выделения геологических объектов?

Решение этих, кощунственных для фундаментальной науки вопросов, входит в прерогативу прикладных дисциплин, определяя их содержание.

Несколько упрощая можно заявить, что в любой области знаний существует некая часть реального мира, представляющая особый интерес для профильных специалистов. Вероятно, её и следует обозначить как предмет соответствующей дисциплины. Однако, чем глубже мы проникаем в профессию, тем яснее становится, что найти границы этой части невозможно. На некотором уровне познания, казалось бы, интуитивно ясная предметная специализация исчезает. Мы сталкиваемся с великим единством мироздания, диктующим необходимость глобальной интеграции наук, что хорошо понимали древние.

Как не странно, в настоящее время эта интеграция успешно идёт прямо противоположным путём- через создание дополнительных пограничных дисциплин. Именно так возникли био-, гео-, астро- химии и физики, которые уже принесли (и без сомнения ещё принесут) человечеству множество замечательных открытий. Правда, с появлением новых направлений вопрос о предмете возникших наук становится ещё более сложным, т.к. необходимо его формулировать и для них. Тем не менее, совершенно неразрешимым его назвать нельзя. Проблема сдвинулась бы с мёртвой точки, сумей мы предло-

жить некую достаточно простую метрику, по которой объекты реальности могут сравниваться друг с другом. Это создаст предпосылки их объединения по предметным направлениям, что равносильно предметной классификации.

Допустим, что в качестве такой метрики используется максимальный размер объектов материального мира, которые входят в сферу изучения конкретных наук. Распределяя некоторые из них по данному показателю, получим следующий ряд, приведённый в таблице 1:

Таблица 1 – Предметная классификация

Наука	Пороговый размер объектов	Предмет
Физика	$\leq 10^{-8}$ см	Атомы
Химия	$\leq 10^{-3}$ см	Молекулы
Биология	≤ 1000 м	Живые организмы
Геология	≤ 6371 км	Планета Земля
Астрономия	∞	Вселенная

Несмотря на очевидную условность такой классификации, она позволяет сравнивать предметы, складывающиеся из собственных объектов, которые в свою очередь состоят из предметных подсистем более низкого уровня.

Так химики, имея дело с молекулами, не могут игнорировать физический наномир, поскольку в химические соединения и реакции вовлекаются атомы вещества. Двигаясь по этой пирамиде знаний, мы обнаруживаем, что геологам для плодотворной работы требуются объекты всех вышеперечисленных наук плюс свои собственные, попадающие в воображаемое сито пороговой размерности. В нашем случае она совпадает с радиусом Земли, что просто является данью фундаментальному определению (1).

Рассуждая более здраво, следует признать, что основная работа по освоению природных богатств развёртывается в технологически доступном человечеству слое планеты. Если сравнить Землю с яблоком, то мы только царапаем его кожицу, а о том, что творится глубоко внутри плода – можно пока только догадываться.

С этих позиций специалистам по моделированию имеет смысл говорить не о геологии вообще, а о предмете прикладной геологии- информационно богатому и насыщенному, но более узкому, чем геология в классическом определении. Данный предмет включает в себя целые прикладные отрасли- такие как геология полезных ископаемых, гидрогеология, инженерная геология и др. Все они осваивают объекты общего географического пространства, причём сформированные в единое (т.н. "геологическое") время. Начало этого времени совпадает с появлением океанов и началом процессов осадконакопления [1].

В свете сказанного, справедливы два утверждения, являющиеся ответами на первый и второй вопросы параграфа, соответственно:

Для прикладной геологии недра понимаются как технологически достигаемая часть планеты.

Объективное различие вектора изменчивости по простиранию и разрезу в осадочном чехле является основополагающим принципом выделения геологических объектов любого масштаба.

В целом обе фразы можно свести в одно определение:

Предметом прикладной геологии являются технологически доступные для освоения объекты, залегающие в пределах осадочного чехла.

1.1.2 Методология и техника компьютерного моделирования

В параграфе рассматривается содержание второй (приобретённой) компоненты дисциплин, связанных с компьютерным моделированием недр.

Здесь целесообразно рассмотреть два взаимосвязанных вопроса:

- каковы цели, а значит и задачи компьютерного моделирования на разных этапах освоения недр?
- каковы основные этапы создания компьютерных моделей?

Задачи компьютерного моделирования

Планомерное промышленное освоение недр складывается в т.н. геолого-разведочный процесс (ГРП). Он характеризуется последовательным повышением детальности исследований при локализации перспективных площадей и участков. Это обстоятельство решающим образом сказывается и на выборе методик моделирования и темпах формирования информационных аналогов разнопорядковых геологических объектов.

Очевидно, полезность этой работы напрямую связана с достигаемым уровнем подобия искусственных математических моделей их реальным прототипам. Чтобы он был достаточным, необходимо следовать принципам актуализма и системного анализа недр, понимаемых в уже обсуждённом выше аспекте. Только тогда не впадая в противоречия, мы сумеем подобрать для известных природных явлений соответствующие кибернетические суррогаты. При этом необходимо учитывать как уровень абстракции, на котором они рассматриваются, так и возможности доступного исследователю методологического инструментария.

В настоящее время накоплен великий арсенал методов пригодных для решения задач моделирования. Многие из них собраны в т.н. Пакеты Прикладных Программ (ППП), в основном западного производства. Широкому распространению их в РФ мешают два обстоятельства- чрезвычайно высокие цены и, как это не странно, глубокая отраслевая стандартизация зарубежных ППП, не всегда признаваемая отечественной промышленностью.

Одновременно с этим существует множество разрозненных, но доступных компьютерных программ и даже информационных систем, применение которых для целей моделирования оправдано и полезно [2]. Их самостоятельное комплексование делает процесс моделирования более гибким, чем в ППП, работающих по принципу "чёрного ящика". В данном случае это озна-

чает, что пользователь не в состоянии изменить функциональность ППП и не имеет доступа к его алгоритмической базе. Разумеется, оба подхода могут развиваться одновременно, используя близкий или даже одинаковый методический аппарат, который будет изучаться на нашем курсе.

В целом, главные этапы процесса компьютерного моделирования отображены на рисунке 1.

Здесь схематически показаны две параллельные иерархии-стадийности геологоразведочного процесса и иерархия типов математического моделирования недр. Особенностью схемы является то, что исследуемые объекты и модели каждой стадии составляют содержание предыдущих элементов схемы более низкого уровня.

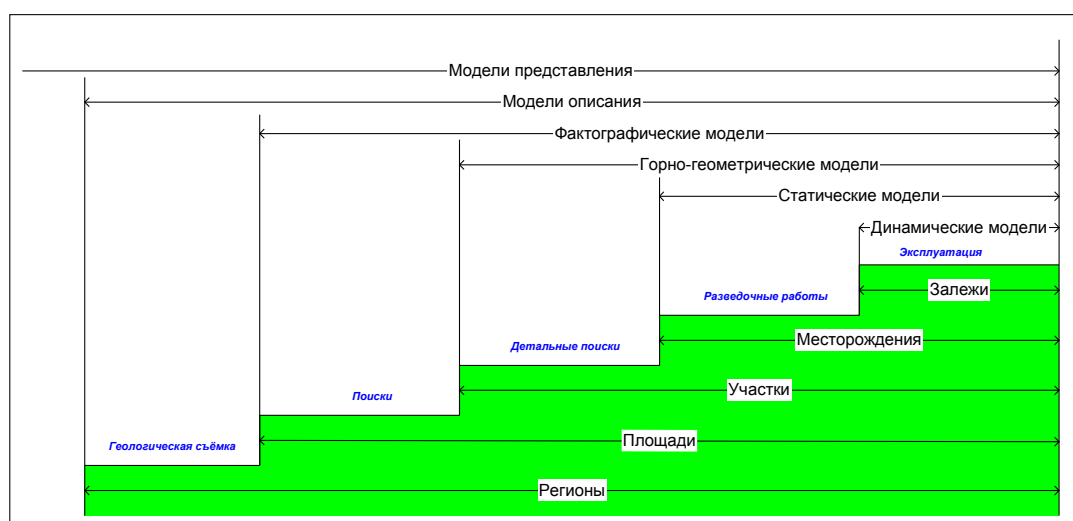


Рисунок 1- Этапы компьютерного моделирования и стадийность геолого-разведочного процесса

Таким образом, смысл геологических задач почти однозначно определяется детальностью или масштабом работ, контролирующей типизацию соответствующих математических моделей. Они аккумулируют всю информационную нагрузку, собранную в ходе предшествующих этапов геологического исследования территории. Соответственно этому задачи моделирования сведены в таблицу 2.

Таблица 2 – Задачи компьютерного моделирования по стадиям ГРП

Стадия ГРП	Задачи компьютерного моделирования
1	2
Геологическая съёмка	Составление сводных словарей (разрезов). Определение топологических взаимоотношений геологических объектов. Выделение перспективных участков. Районирование территории по перспективности на качественном уровне.

Продолжение таблицы 2

1	2
Поиски	Информационное освоение результатов поисковых работ. Формирование Баз Данных. Количественные оценки перспективности территории.
Детальные поиски	Типизация и геометризация месторождений (горно-геометрическое моделирование). Создание информационных аналогов месторождений.
Разведочные работы	Категорийный подсчёт запасов по залежам. Моделирование схем разработки месторождений (пространственных и временных)
Эксплуатация	Создание постоянно-действующих моделей залежей. Позалежный компьютерный мониторинг разработки. Автоматизированное управление разработкой путём моделирование геолого-технологических мероприятий.

1.2 Этапы создания компьютерных моделей

В общем случае сам процесс создания компьютерных моделей состоит из трёх этапов [3]:

- формализация фактического материала;
- геометризация и пространственные расчёты;
- динамическое моделирование.

Но конкретное их количество, так же как и полнота исполнения каждого зависят от многих обстоятельств. Например, при разработке углеводородных залежей этап динамического моделирования фильтрации флюидов просто необходим, а при моделировании рудных месторождений он не требуется.

Говоря о методологической полноте компьютерного моделирования геологических задач, нельзя не упомянуть о двух полярных подходах к этому вопросу. Их несколько условно можно назвать *оформительским* и *кибернетическим* подходами.

1.2.1 Оформительский подход

Компьютерное моделирование понимается, в основном, как способ создания качественной компьютерной графики, автоматизации сложных расчётов, стандартного оформления текстовых приложений и т.п. В этом случае информационная структура модели полностью подчинена требованиям программного обеспечения, используемого для решения геологических задач. Она организована или в файлы заданного формата (.shp, .xls, .tiff и др.) или

сосредоточена в специализированных базах данных, непосредственно поддерживаемых используемые программные комплексы.

Понятно, что оформительский подход не позволяет реализовать основные преимущества компьютерного моделирования недр, состоящие в возможности многовариантного прогноза природных объектов и процессов. В нашем случае прогноз, как и много лет назад, остаётся крайне субъективным, поскольку однозначно связан с личностью геолога и заранее известными плановыми цифрами. По сути дела, самого численного моделирования здесь и не происходит, а всё дело сводится к созданию полиграфически качественной продукции.

Как не странно, такое положение вещей часто имеет место в крупных компаниях, располагающих самым дорогостоящим программным инструментарием в отрасли. Он просто не используется по прямому назначению из-за трудностей организации надлежащего контроля качества входной геологической информации, объём которой может быть крайне велик. Когда исходные данные искажены изначально, наивно ожидать практической отдачи от самых совершенных программных пакетов. Это и есть самая благодатная почва для оформительского подхода к моделированию геологических задач.

1.2.2 Кибернетический подход

В противоположность предыдущему, данный подход ориентирован на реальную автоматизацию геологических изысканий. Как оказалось, решение этой задачи требует не только мощных компьютеров, но и глубокого теоретического переосмысления традиционных представлений о строении Земли.

Чтобы созданные компьютерные модели были полезными, необходимо обеспечить выполнение двух, казалось бы, взаимно противоречивых требований:

- во-первых, неизбежные при любом моделировании упрощения не должны слишком искажать действительность, иначе результаты моделирования станут непредсказуемы;
- во-вторых, следует добиться максимально быстрой и адекватной реакции компьютерных моделей на возмущение их состояния, имитируемое вводом новых данных. В противном случае результаты моделирования потеряют актуальность и не будут востребованы. Однако, как правило, быстрая реакция информационных систем связана именно с упрощением реальности, без которой в моделировании не обойтись!

До начала восьмидесятых годов прошлого века исследователи разных стран пытались добиться удовлетворительных результатов компьютерного моделирования кустарными методами. Созданные в итоге модели месторождений оставались уникальными, несмотря даже на явное сходство строения и истории геологического развития их природных прототипов. Становилось ясно, что для обеспечения промышленной выгоды от внедрения компьюте-

ров следует провести большую подготовительную работу. Смысл её должен заключаться в унификации понятий и представлений, терминологий, единиц измерения, приборных баз всех дисциплин, чьи предметные области пересекаются при моделировании недр. В результате возникли две доминирующие методологии моделирования: *реляционная* и являющаяся её развитием *объектно-ориентированная*. Обе они имеют право на одновременное существование, обеспечивая строгую алгоритмическую базу для геологических моделей, ныне уже воплощённых в самых знаменитых программных комплексах.

Реляционная методология

Основана на известных принципах, лежащих в основе реляционных СУБД. Согласно им, формально, природные объекты могут быть описаны в виде строк реляционных таблиц. Сама таблица, состоящая из совокупности однотипных объектов, составляет т.н. *класс* объектов. Таблицы называются реляционными, поскольку имеют специальное поле связи (*relation field*), где содержится некая общая для разных классов информация (ключевое поле).

Искусно развивая реляционную идею, передовые инженеры и программисты (большинство из которых проживает в США) разработали внутренне связанные информационные структуры, имитирующие строение недр.

Примером могут служить ранние модели POSC и DGM (*Digital Geologic Map*), которая рассматривается в последующих главах.

Объектно-ориентированная методология

Используется для создания информационной модели недр последнего поколения. Она является совершенно независимой от типов прикладного программного обеспечения, используемого при моделировании.

При объектно-ориентированном подходе в качестве основного структурного блока недр выступает объект или класс объектов. В самом общем смысле объект - это сущность, хранимая и извлекаемая из словаря предметной области, а класс является описанием множества однотипных объектов. Объект может состоять из произвольного числа вложенных объектов, составляющих аналоги природных подсистем соответствующего порядка.

Обратившись к таблице 2, мы обнаружим, что словарь предметной области должен формироваться уже на стадии геологической съемки. В обычном понимании, принятом в докомпьютерную эру, самая близкая, но примитивная аналогия такому словарю - сводный геологический разрез. Каждый объект здесь обладает идентичностью (т.е. его можно отличить от прочих объектов), поведением (т.е. с ним возможно проделать некоторые манипуляции или он в состоянии что-то сделать самостоятельно) и состоянием. Налицо явная связь с принципами объектно-ориентированного программирования, которые мы изучали в шестом семестре.

Таким образом, основные усилия исследователей сводятся к формали-

зации данных, понимаемой как формирование возможно более подробного словаря системы. Если он создаётся в соответствии с логическими и общепринятыми посылками, то может использоваться любыми программными средствами, поддерживающими принятую модель данных.

Испортить объектно-ориентированную модель некачественной информацией конечно возможно, но сделать это не так просто. Концептуально верная модель должна сама указывать на противоречия в данных, ввод которых будет отвергнут, например:

- объекту типа "литологический экран" нельзя приписать свойство нефтестекобучи и т.п.;
- на скважине нельзя провести каротаж, если она уже ликвидирована на эту дату;
- сдвиг по сместителю (или вектор в объекте "грабен") всегда имеет отрицательную амплитуду, т.е. направлен вниз.

Технологически подобные декларации или как их называют программисты- "правила целостности" реализуются в схемах специализированных баз данных. Например, количество правил целостности в технологическом стандарте POSC достигает многих тысяч.

Компьютерные модели, основанные на объектно-ориентированном подходе, описываются теми же средствами, что и компьютерные программы. Это даёт возможность их унификации, тиражирования и использования повсеместно в данной отрасли. К числу самых известных объектно-ориентированных моделей следует назвать POSC (Petrotechnical Open Software Corporation), NADM (North American Geologic-Map Data Model).

2 Цифровые модели геологических данных

2.1 Реляционная модель цифровой геологической карты (DGM)

Широкое внедрение компьютерных технологий во все сферы промышленности в конце 70-х годов коснулось и геологической отрасли. Учитывая, что основным документом, отражающим строение и историю развития недр, является геологическая карта, значительные усилия сосредоточились в направлении её компьютерного конструирования. Решение этой проблемы в США было реализовано в рамках Национального Проекта Базы Данных Геологических карт (NGMDB), к которому была привлечена Ассоциация Американских Геологов (AASG). Интенсивное обсуждение вопроса привело к разработке практических рекомендаций, играющих роль стандарта *de facto* при цифровом моделировании геологических материалов. Полностью доклад о проделанной в этом направлении работе помещён по Интернет- адресу: <http://ncgmp.usgs.gov/ngmdbproject> .[4]

2.1.1 Определение модели данных цифровых геологических карт

Что же такое цифровая геологическая карта (ЦГК)? Не усложняя вопроса, можно считать, что ЦГК, представляет собой обычную геологическую карту, географические детали которой и пояснительная информация закодированы в цифровом формате, для последующей компьютерной обработки.

Тогда что же собой представляет собственно *обычная* геологическая карта? С точки зрения участников NGMDB она является отображением подмножества геологических объектов, распределённых в пространстве и времени. Эти объекты обозначены соответствующими условными знаками и имеют описания, используемые для особых целей. Геологические объекты карты могут выбираться с помощью самих атрибутов данных объектов, с помощью географического фильтра или комплексом этих методов. Формальная иллюстрация этой модели приводится на рисунке 2.



Рисунок 2- Три класса объектов, составляющих модель ЦГК

На нём представлены три главных компоненты модели данных ЦГК, отображённые в виде пересекающихся окружностей. Каждая из них обозначает единый самостоятельный класс объектов, а области пересечения демонстрируют их интерпретируемую комбинаторику. Ниже приводится описание классов данных ЦГК и их взаимоотношения:

- пространственные объекты являются оцифрованным представлением реальных геологических образований, выявленных и закартированных. Обычно они оформляются в виде комбинаций точек, линий и полигонов;
- круг «Описание данных» представляет архив характеристик или атрибутов пространственных объектов. В аспекте моделирования данных эти атрибуты относятся или к уникальным единичным объектам (размерности структуры, возраст и т.п.) или к сложным по строению образованиям, например формациям или региональным разломам;
- легенды- это инструмент классификации. На пересечении с окружностью данных возникают зоны классифицированных данных, а на пересечении с окружностью пространственных объектов возникают зоны классифицированных объектов. Кроме того, легенды включают в себя информацию о масштабе карты, используемым схемам классификации и условным обозначениям конкретных карт.

Сами карты образуют зону пересечения областей классификаций с представлением конкретного геологического объекта. Другими словами, на карте каждый пространственный объект должен быть отнесён к конкретному типу и иметь типовое описание своих характеристик.

Существует три основных вида программного обеспечения, которое может работать с с цифровыми моделями геологических карт:

- географические информационные системы (GIS);
- системы компьютерного проектирования (САД-системы);
- специализированные программы геологического картирования.

Выделяются также два главных способа или концепции использования цифровых геологических карт- *картография* и *анализ*.

Задача картографов состоит в трансформации цифрового представления картографических данных в высококачественные геологические карты, обычно на бумажной основе.

Аналитики заняты выдачей прогнозов и определением особо перспективных в практическом отношении областей пространства. Это достигается путём комбинирования цифровой картографической информации с другими полезными данными.

Среди картографов до сих пор существуют разногласия относительно скорости построения карт цифровыми методами и их полезности по сравнению с традиционными технологиями. Тем не менее, уже всем ясно, что обновление и исправление карт на цифровой основе происходит во много раз

быстрее, чем в докомпьютерную эпоху. Цифровые карты просто тиражируются, могут быть перерисованы в различных масштабах, их легко редактировать, добавляя, удаляя и модифицируя геологические объекты или меняя картографическую легенду.

Использование оцифрованных данных сближает позиции картографов и аналитиков, поскольку картографы получают возможность анализировать карты, а аналитики - отличные презентации, сопровождаемые компьютерными выкладками, пригодными для анализа и обобщения.

2.1.2 Предпосылки создания модели данных ЦГК

Создание ЦГК в широком смысле означает разработку структуры БД для организации, хранения и компьютерной обработки геологических данных. При этом цифровая модель должна определять формальную грамматику геологических карт, правила которой не зависят от объёма информационного словаря карты. Поэтому главные предпосылки создания модели ЦГК формулируются так:

- создание формальной грамматики информационного описания геологических карт;
- создание словарного запаса модели в объёме, допускающем её опытную эксплуатацию.

На момент первоначальной разработки ЦГК (это был 1997 год) основным инструментальным средством, способным реализовать идеи ЦГК, были хорошо нам знакомые ESRI продукты: ArcView 3 и Arc/Info. Как известно, ранние их версии использовали реляционные методологии, в связи, с чем и модель данных обречена была стать изначально реляционной. Тем не менее, реляционная технология смогла удовлетворить многочисленные требования к модели данных, формулируемые ниже и иллюстрируемые на рисунке 3: Ри-



сунок 3- Структура модели данных ЦГК

- модель данных должна быть лёгкой в использовании, предъявлять минимальные требования к персоналу и не усложнять общую организацию работ при создании геологических карт;
- должен быть набор минимальных требований, общих для всех геологических карт, или т.н. «ядро». С точки зрения реляционной методологии ядро рассматривается как обязательные таблицы;
- должна быть возможность учёта и отображения тех многочисленных объектов, которые не встречаются на всех без исключения картах. Их следует рассматривать как расширение ядра;
- модель данных должна допускать простые способы её расширения в виде создания дополнительных реляционных таблиц с новыми типами геологических данных. При вводе в модель новых образцов (геологических) допускается расширение легенды, добавление новых свойств и т.п. Главная цель подобных расширений заключается в возможности использования любых источников геологических данных. Со временем это сблизит модель со всеми возможными природными прототипами;
- модель должна обеспечить механизм документации первоисточников каждого геологического объекта в отдельности;
- модель данных не определяет предметный словарь отрасли, но должна иметь возможность ему следовать. Заголовки таблиц и названия полей должны выбираться из отраслевого списка терминов, что облегчает дальнейшую компьютерную эксплуатацию ЦГК и минимизирует неоднозначность геологических понятий;
- при составлении модели данных следует избегать явного кодирования понятий геологического словаря;
- атрибуты объектов геологических карт взаимосвязаны. Так, разлом может рассекать два полигона и закончится в третьем. Пространственные атрибуты таких линий должны запоминаться совместно с данными о рассекаемых полигонах. Тогда мы сумеем автоматически выбрать, например, самый верхний из рассекаемых пластов;
- модель данных должна обеспечивать возможность идентификации отдельных геологических явлений. Это могут быть обнажения горных пород, описание литологии отдельных полигонов внутри более крупных геологических образований, сегменты разломов и т.д.

2.1.3 Содержание ЦГК

Геологические карты являются чрезвычайно сложным синтетическим документом, поскольку отражают самую разнообразную информацию.

Большинство карт состоит, как бы из фона, образованного полигональными областями развития геологических образований с запутанными топологическими взаимоотношениями. Линии, разделяющие эти полигоны, также могут иметь подписи, поясняющие типы геологических границ, например, контакты. На фоне всего этого имеет место множество линейных геологических объектов, секущих и перекрывающих полигональные области. Это жилы, дайки, разломы, складчатость. Существуют и точечные объекты, к числу которых относятся условные обозначения, внесмасштабные символы и т.п.

Дополнительные сложности возникают в связи с отсутствием универсальных стандартов для условных обозначений и легенд геологических карт. Например, для обозначения одинаковых типов пород обычно пользуются одинаковыми цветами, но нет никакой гарантии, что эти же цвета будут использоваться при создании всех карт. То же самое справедливо и для образцов разнообразных типов линий и точек. Всё это порождает проблему т.н. совместимости цифровых геологических карт.

Совместимость ЦГК

Рассматривая каждую карту в отдельности, данную проблему трудно заметить. Конечно, очень хорошо, если смежные карты используют одинаковые или близкие стили для отображения сходных объектов, но такое случается редко. При картировании крупных регионов, особенно заснятых в разное время, вопросы совместимости комбинируемых карт приобретают первостепенное значение. Различают три уровня совместимости ЦГК.

Первый уровень совместимости

Он обеспечивается или зависит от исходного картографического материала, монтируемого в региональную ЦГК. При этом должны быть выяснены следующие вопросы:

- где создавались ЦГК, предназначенные для слияния?
- при их построении использовались одинаковые легенды, шкалы, единицы измерения и т.п.?

Многие случаи геологических несоответствий могут быть разрешены относительно легко или проигнорированы. Например, на одной из карт обозначены поля развития всех мезозойских вулканитов, объединённых в единую толщу. На смежной карте эта же толща полностью стратифицирована и разбита на составляющие. Такие карты могут быть успешно объединены, разумеется, с некоторой потерей детальности в сводной карте. Однако, иногда две смежные карты составляются изначально для различных целей и их комбинация становится очень трудным или, вообще, невозможным делом. Пусть на одной из карт отображено развитие пород кристаллического фундамента, а другая, смежная с ней создавалась для картирования приповерхностных месторождений. Скорее всего, на этих картах будет очень мало общих

границ и точек соприкосновения, используемых для слияния. В данном случае оно будет некорректным, в связи, с чем эти карты следует хранить в разных тематических слоях и разных информационных архивах.

Второй уровень совместимости

Совместимость ЦГК второго уровня- это совместимость атрибутивных данных комбинируемых карт. Даже если смежные карты удовлетворительно совпадают по границам друг с другом, требуется, чтобы совпадали и информационные описания геологических объектов смежных карт. Если, например, один полигон на карте охарактеризован геологическим возрастом, а на другой- нет, то карты могут быть объединены, однако возрастная целостность будет при этом нарушена, т.к. появится неопределённость.

Гораздо чаще случаются более сложные случаи. Например, на одной карте каждый полигон охарактеризован одним возрастным индексом, отражающим время, его образования. На второй карте полигоны характеризуются уже двумя атрибутами- минимального и максимального возрастов. На третьей карте также возраст определяется двумя атрибутами, но они дополнительно подтверждаются данными абсолютного возраста, полученными радиометрическими методами. Конечно, такая подробность хороша при составлении единичной карты, но как всё это совместить в сводном документе? Ответ на дилемму состоит, конечно, в глубокой переработке и унификации атрибутивных данных, что и составляют одну из целей создания совместимой модели данных.

Третий уровень совместимости

На этом уровне рассматривается совместимость кодирования. Кодирование- очень частый приём обозначения сложных, длинных выражений. Оно помогает экономить память компьютера и состоит в подмене словосочетаний другими- более короткими или, вообще, числами. Понятно, что сравниваемые карты должны быть закодированы одинаково, иначе результаты компьютерного поиска по атрибутам, содержащим эти коды, будут непредсказуемы.

Обработка данных в соответствии с тремя вышеперечисленными уровнями их совместимости не гарантирует сохранение этой совместимости в будущем. В самом деле, скорее всего на территории уже созданной ЦГК появятся новые данные, значение и характер которых, сейчас трудно предугадать. Необходимо так организовать модель, чтобы она имела открытую и гибкую архитектуру и не нуждалась в коренной переработке при встрече с новой информацией.

Кроме этого важно обеспечить совместимость модели с как можно более широким кругом пользователей. Обычно, начало работ по *цифровой геологии* иницируется в связи с какой-либо важной конкретной практической целью. Данная цель может и не требовать хранения полного набора атрибутов для ЦГК и её разработку можно произвести значительно быстрее и деше-

вле. Однако, в последующее время на базе созданной ЦГК могут возникнуть новые задачи, инициируемые другими специалистами. Поиск, подготовка и ввод новых данных в этом случае уже обойдётся существенно дороже, чем, если бы это было сделано сразу при проектировании ЦГК. Поэтому разработчикам цифровых схем необходимо тщательно учесть возможные группы пользователей ЦГК как внутри своей компании, так и в смежных с ней подразделениях.

2.2 Основы моделирования

Главной целью разработки реляционной модели геологических карт в цифровых форматах является создание их информационного эквивалента – библиотеки или хранилища. Имея их под руками, можно воссоздать для конкретной территории разнообразные карты, основанные на индексированных библиотеках скважин или информации снятой с бумажных носителей.

Эти БД позволят автоматически генерировать карты обширных территорий, собирая их как мозаику. Кроме того, появляется возможность создания разного рода тематических карт произвольных масштабов- производных этого великого цифрового архива. Сюда относятся литологические, фациальные, карты полезных ископаемых, разнообразные карты изученности и т.п. В целом процесс моделирования показан на рисунке 4. Он рассматривается с точки зрения проектировщика новых карт и/или пользователя, создающего производные карты из цифрового архива.

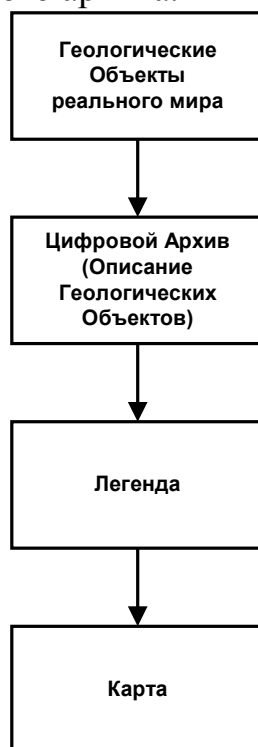


Рисунок 4- Основы Реляционной модели данных геологических карт

Как видно из рисунка, ядром реляционной модели данных геологических карт является архив оцифрованных геологических объектов. Он содержит всю информацию необходимую для создания геологических карт.

Проектирование модели производилось с учётом возможности её расширения для включения 3-х мерных объектов, однако здесь рассматривается только её двумерная реализация.

Цифровой архив содержит Базу Данных, представляющую описание любых геологических объектов. Это могут быть поля развития определённых типов пород, структуры, месторождения и т.п.

Архив может также включать географическое описание каждого объекта, его атрибутов и информационных источников этого описания. Один объект может иметь множество источников описания.

Цифровой архив связан с Легендой геологической карты. Она играет роль своеобразного фильтра для отбора специфицированных геологических объектов из архива, их обозначения и представления на карте.

Таким образом, в процессе создания новых или производных карт из архива возникают новые легенды, которые вновь будут использованы для запросов к архиву.

В целом процесс формирования ЦГК проходит следующие этапы:

- в результате изучения реальных геологических объектов, последние формализуются до географических объектов (точек, линий, полигонов и др.). Эти географические объекты снабжаются подробными атрибутивными данными;
- эта информация записывается в цифровой архив;
- для создания карты из архива (т.е. для её восстановления) пользователи должны определить Легенду;
- Легенда фильтрует Архив и восстанавливает некоторое представление первоначальной карты.

2.2.1 Геометризация картографических элементов

Рассмотрим нашу модель с ещё одной точки зрения. Она отражает интересы пользователя (обычного специалиста-геолога), которому необходимо ввести данные из существующей бумажной карты в цифровой архив. Чтобы эта информация была понятна, ему придётся вместе с уже откартированными геологическими объектами занести в архив и легенду данной карты. Помимо этого, геологу необходимо провести большую предварительную работу по структуризации элементов карты, подобрав каждому некий геометрический примитив, понятный для программной системы, поддерживающей цифровую модель данных. В принципе, таких обобщённых геометрических форм не так много и все они представлены на рисунке 5.

Массивы горных пород, интрузии и т.п. могут быть представлены объёмными телами, моделируемыми модулем 3D Analyst, например. Поверхности несогласия и складчатость имитируются цифровыми поверхностями

гид-тем*, элементы разрывной тектоники- линиями. Инженерные сооружения (скважины, колодцы), точки и результаты опробования представляются на картах соответствующими внесмасштабными условными значками. Кроме того, разнообразные подписи и заголовки, присутствующие на геологических картах в обязательном порядке обозначаются как текстовые строки.

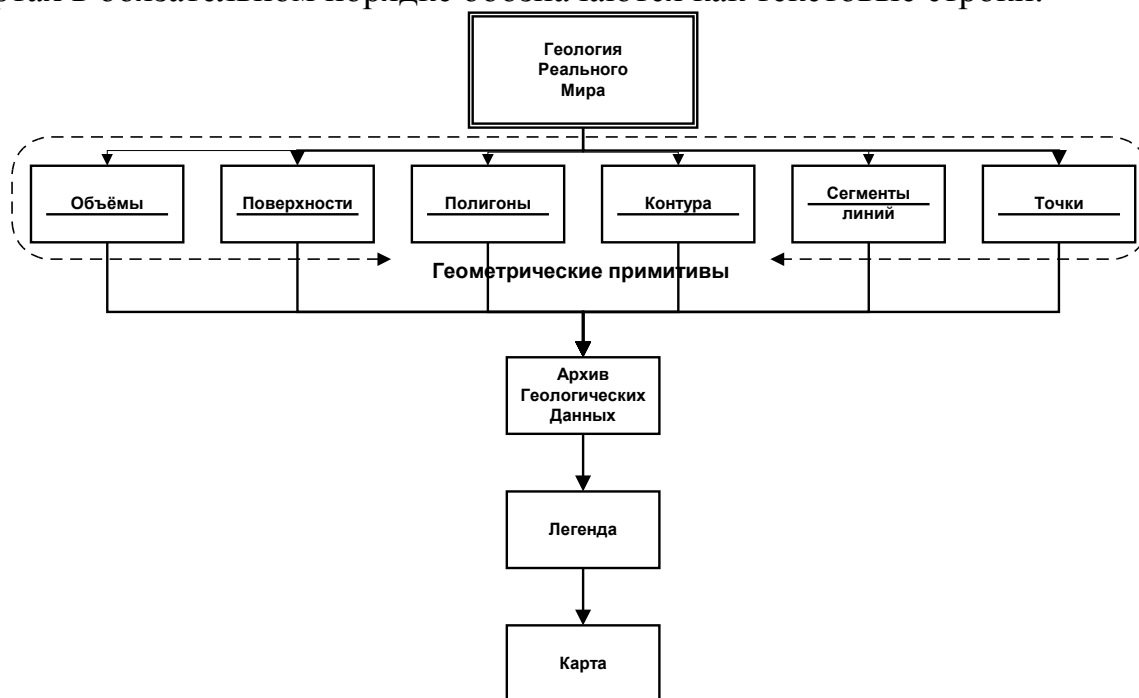


Рисунок 5- Геометрическое представление геологических объектов

2.2.2 Сингулярные и составные геологические объекты

Применительно к задачам геологического моделирования все геологические объекты могут быть условно поделены на две основные категории- сингулярные и составные.

Сингулярными объектами называются те, которые могут быть полностью рассмотрены с одной точки наблюдения. Например, это элементы залегания, описание образца, результаты химического анализа. Сингулярным объектом может быть некоторая единичная категория карты (сегмент разлома, полигон), который также может принадлежать более сложному составному объекту.

Составные геологические объекты обычно выделяются на основании множества наблюдений сингулярных и представляют собой их обобщение и интерпретацию. Геологические структуры, стратиграфически расчленённые толщи, системы разрывных нарушений- всё это примеры составных геологических объектов.

Сингулярные и составные объекты обрабатываются по разным графам и хранятся в различных разделах цифрового архива. Для непосредственного отображения на карте составные объекты непригодны в виду неоднозначно-

* Термин географических информационных систем (ГИС)

сти их природы. Картирование составных объектов возможно после разложения их на простые сингулярные компоненты, с помощью Легенды. После этой процедуры возникают атомарные геометрические примитивы геологической карты, полученные из разных источников (архива сингулярных объектов и архива составных объектов), которые уже могут быть визуализированы.

Описанная схема представлена на рисунке 6.

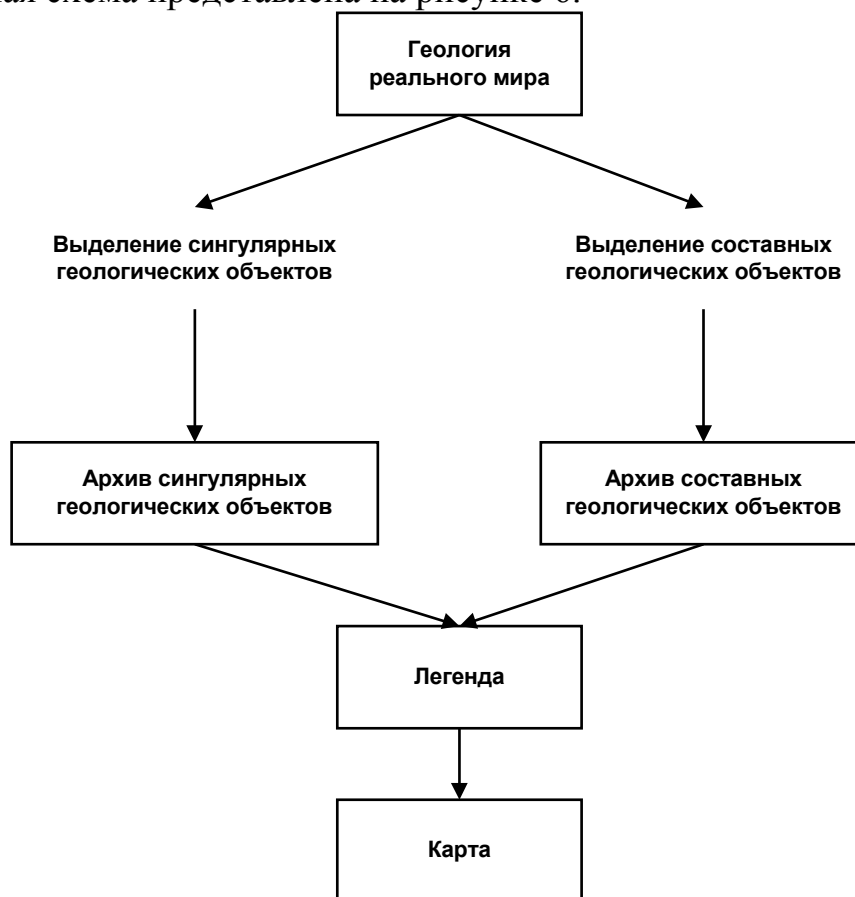


Рисунок 6- Две категории геологических объектов

Синтезируя элементы предыдущих рисунков, мы получим новую диаграмму (рисунок 7) более полно иллюстрирующую реляционную модель данных геологических карт.

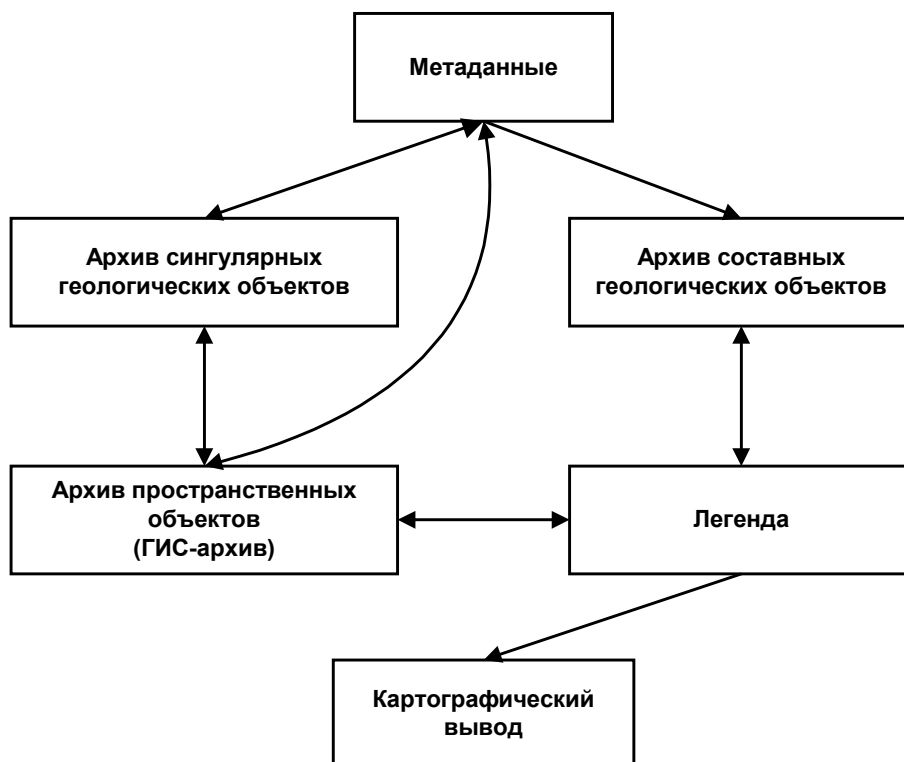


Рисунок 7- Сводная диаграмма реляционной модели ЦГК

В данный рисунок включены две новые компоненты- метаданные и ГИС-архив.

Метаданные- это данные о данных, представляющие информационное описание всей картографической информации.

Как и прежде, геологические объекты попадают на карту, фильтруясь через Легенду. Однако, пространственные данные, сопровождающие каждый объект (морфология, размеры, местоположение) теперь отделяются от атрибутивных описаний и хранятся в географических информационных системах.

ГИС-архив непосредственно связан с архивом сингулярных объектов, поскольку их пространственное положение однозначно.

Архив составных объектов может быть связан с ГИС только посредством Легенды, разбирающей его на сингулярные части. В этом процессе расчленения должна быть уже предусмотрена обратная процедура сборки составных элементов в единое картографическое целое. Данная задача решается путём составления тщательно продуманной схемы реляционной базы данных цифровой модели геологической карты (ЦМГК).

Как следует из рисунка 7, функционально ЦМГК должна обслуживаться двумя компонентами. Главные преимущества цифровых геологических карт состоят в возможности проведения пространственного анализа геологических объектов и их автоматической компоновкой в новые производные карты. Все эти возможности сосредотачиваются в ГИС, такой, например как ArcView 3.2.

Оставшаяся информационная часть модели, целиком, включающая Легенду, Сингулярный и Составной архивы, а также некоторые данные из пространственного архива (ГИС-архив) будет управляться средствами Систем Управления Реляционными Базами Данных (СУРБД). В этой связи следует

напомнить об их характеристиках, возможностях и принципах функционирования.

Данные в СУРБД хранятся в таблицах, состоящих из заголовка и собственно данных в строках или записях этих таблиц.

Отдельные таблицы связаны друг с другом с помощью содержимого специальных столбцов связи. Например, если в таких столбцах содержится наименование или код стратиграфической единицы, то записи нескольких таблиц с одинаковыми кодами могут быть объединены.

При проектировании реляционных таблиц, предназначенных для связи с ГИС, используются общие принципы разработки реляционных баз данных. Существует великое множество методик разработки баз данных, но главная цель всех этих систем состоит в снижении избыточности их структуры или её *нормализация*. Как концептуальное воплощение процесса нормализации ниже описывается схема реляционной БД в виде набора связанных таблиц. Эти таблицы созданы для обеспечения подходов, проиллюстрированных на уже приведённых выше диаграммах.

Система обозначений (нотация) отображения реляционных схем приведена на рисунке 8.

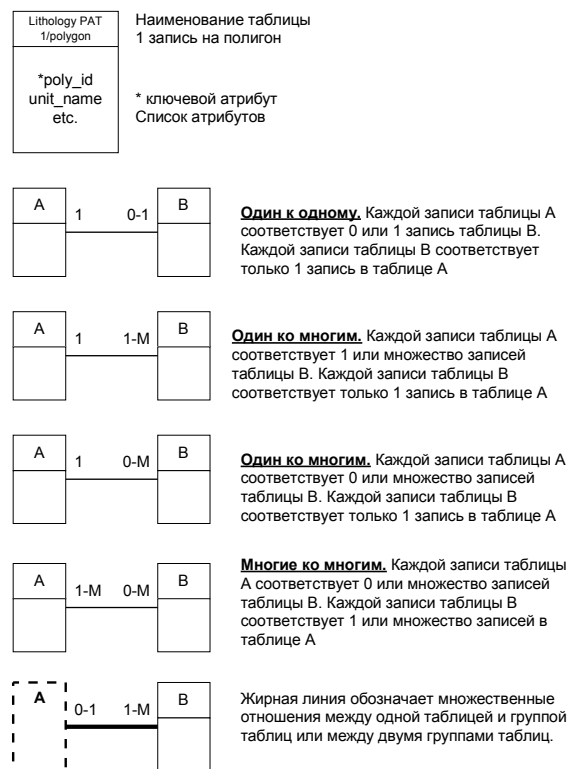


Рисунок 8- Реляционная нотация связей между таблицами

На рисунке каждый прямоугольник обозначает таблицу (или сущность) реляционной базы данных. Соединительные линии обозначают отношения, а метки на их концах- тип отношений.

Таблица разбита по горизонтали на две части. Верхняя- содержит заголовков и информацию о числе записей, на объект. Нижняя часть прямоугольника содержит список атрибутов. Ключевые атрибуты, используемые для связи, помечены звёздочкой.

Обычно, все отношения, показанные на диаграммах, являются отношениями *один-к-одному* или *один-ко-многим*. В тех случаях, когда природа связываемых сущностей и их данные определяют отношения *многие-ко-многим*, между связываемыми таблицами вставляется дополнительная таблица связи. Она преобразовывает отношения *многие-ко-многим* в два отношения *один-ко-многим*.

Данный тип преобразований диктуется самой природой реляционных баз данных и представляет собой одно из отличий реляционной и объектно-ориентированной моделей. Для объектно-ориентированных баз данных, в применении дополнительных таблиц связи нет необходимости.

2.3 Детальное описание реляционной модели данных

Ввиду особой сложности геологической реляционной модели она иллюстрируется пятью генерализованными схемами (рисунки 9-13), которые снабжены подробными комментариями. Эти схемы являются рабочими и могут служить для создания любыми заинтересованными организациями и лицами действующей реляционной модели геологических карт. Детальное рассмотрение структуры реляционных таблиц и их взаимосвязей даёт ясное представление о работе дизайнеров реляционных баз данных, моделирующих сложные природные объекты и процессы.

Соглашения по обозначениям и шрифтам

Для ясности дальнейшего изложения, наименования атрибутов (столбцов или табличных полей) даются курсивом, например *map_title*.

Названия ключевых атрибутов, используемых для связи таблиц друг с другом, дополнительно выделяются жирным шрифтом, например ***org_id***.

В пояснительном тексте наименования реляционных таблиц выделены жирным шрифтом.

Все наименования приводятся на языке первоисточника- английском (NGMDB)

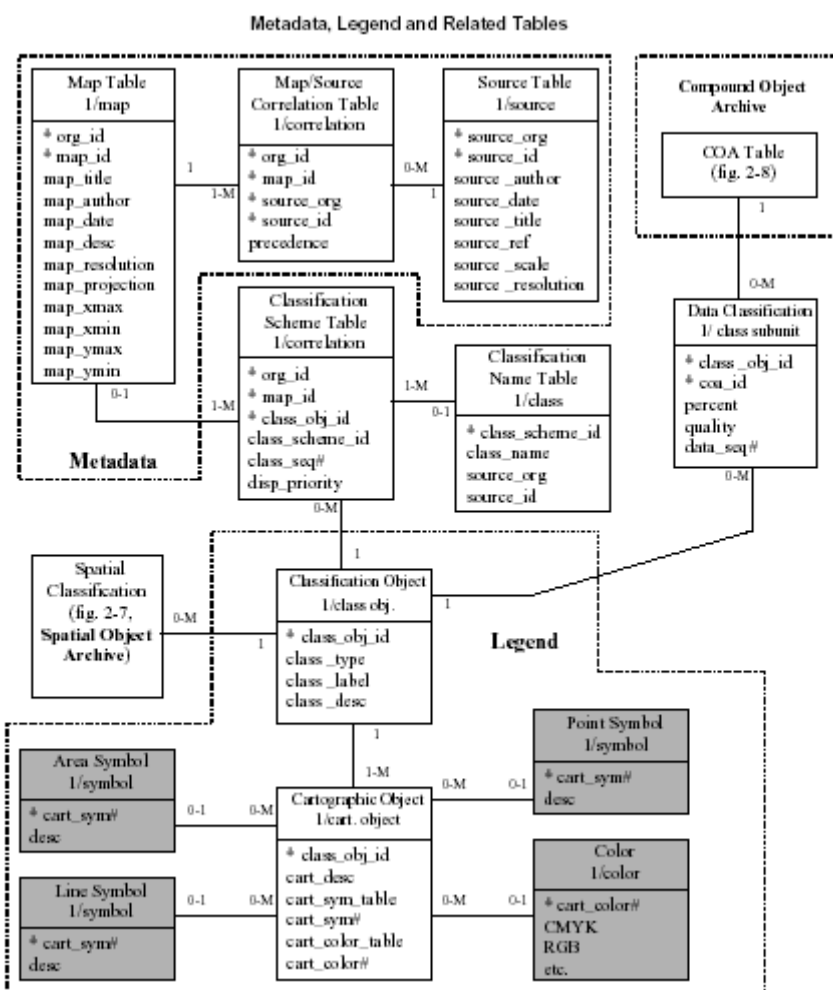


Рисунок 9- Легенда и Метаданные в реляционной модели ЦГК

2.3.1 Метаданные

Структурно модель данных геологических карт состоит из взаимосвязанных блоков, представленных на рисунке 7. Первая реляционная диаграмма, отражающая строение модели на самом высоком уровне абстракции (рисунок 9) включает такие компоненты модели как Легенду, Метаданные вместе с некоторыми связанными таблицами.

Метаданные включают в себя три таблицы: **Map Table**, **Source Table** и **Map/Source Correlation Table**.

Source Table

Назначение данной таблицы заключается в поддержке поиска первоисточников для каждого объекта ЦГК.

Таблица содержит ссылки на опубликованные и рукописные источники информации, хранимые в цифровой библиотеке. Другими словами, эта таблица не содержит самих первоисточников (карт, отчётов, докладов, статей).

Они сосредоточены в отдельной базе данных, но доступ к ней обеспечивается организацией ссылочных связей (по одной ссылке на источник). Каждая ссылка занимает одну запись в таблице.

Таблица содержит два ключевых атрибута:

- *source_org* (идентификатор организации, где создана карта)
- *source_id* (идентификатор организации, откуда получены сведения по картируемому объекту)

Этих двух ключей достаточно, чтобы отыскать или запомнить первоисточник каждого геологического объекта, отображённого на карте.

Полный список атрибутов и их форматов приводится в таблице 3.

Таблица 3– Атрибуты **Source Table**

Атрибут	Описание	Формат
<i>source_org</i>	идентификатор организации, где создана карта	character
<i>source_id</i>	идентификатор организации, откуда получены сведения по картируемому объекту	character
<i>source_author</i>	Список авторов карты	character
<i>source_date</i>	Год публикации или создания исходной карты.	date
<i>source_title</i>	Заголовок исходной карты	character
<i>source_ref</i>	Расширение заголовка	character
<i>source_scale</i>	Масштаб исходной карты	integer
<i>source_resolution</i>	Разрешение исходной ЦГК в метрах	integer

Map Table

Таблица содержит информацию о каждой карте в архиве. Она также содержит два ключевых атрибута *org_id* (код или идентификатор организации, где составлена ЦГК) и *map_id* (идентификатор карты).

Информация в таблице используется, чтобы определить экстенд (размах) и проекцию карты и текстовое описание её местоположения.

Остальные атрибуты таблицы необходимы для описания и вывода отдельных карт.

Для новых карт, которые добавлены в систему в обеих таблицах возникает по одной записи с дублирующейся информацией.

Однако, для производных карт, восстановленных из архива, возникнет только одна запись в таблице **Map Table**, которая посредством ссылок будет связана с одной или многими записями в **Source Table**. Число этих записей зависит от того, как много карт используются в качестве источников геологических объектов создаваемой карты.

Потенциально, на основе любой исходной карты может быть построено произвольное количество производных карт. Таким образом, множество исходных карт порождает множество производных, и отношение между двумя рассматриваемыми таблицами характеризуется выражением: *многие-ко-многим*. Чтобы создать этот тип отношений в реляционной БД, необходима ещё одна таблица связи (корреляционная таблица), вставленная между ними. Перейдём к её обсуждению.

Map/Source Correlation Table

Корреляционная таблица содержит все ключевые атрибуты связываемых таблиц- по два из каждой таблицы. Каждой записи **Map Table**, имеющей ключи- *org_id*, *map_id* может соответствовать любое число записей в **Map/Source Correlation Table** (имеют одинаковые *org_id* и *map_id*, но разные *source_org* и *source_id*. Однако, для каждой записи корреляционной таблицы будет существовать только одна, соответствующая ей запись в **Map Table**. Список атрибутов, их смысловых описаний и форматов приведён в таблице 4.

Таблица 4– Атрибуты **Map/Source Correlation Table**

Атрибут	Описание	Формат
<i>org_id</i>	идентификатор организации, где составлена карта	Character
<i>map_id</i>	идентификатор карты	Character
<i>source_org</i>	идентификатор организации, где создана карта	Character
<i>source_id</i>	идентификатор организации, откуда получены сведения по объектам, составляющим карту.	Character
<i>precedence</i>	Приоритет (порядок) построения геологических объектов в областях перекрытия нескольких исходных карт. Выбирается тот, у кого выше приоритет.	Integer

Для **Source Table** взаимоотношения с корреляционной таблицей аналогичны вышеописанным.

Таким образом, корреляционная таблица преобразует отношения *многие-ко-многим* между **Map Table** и **Source Table** в два отношения *многие-к-одному*. Эта методика обычна для создания реляционных БД.

Дополнительная роль корреляционной данной таблицы состоит в поддержке механизма назначения *предшественника* (*precedence*). Для областей производных карт, создающихся из множества пространственно перекрывающихся объектов, почерпнутых из разных источников, пользователь может указать порядок, в каком они должны использоваться в областях перекрытия.

2.3.2 Легенда

Легенда является второй важнейшей частью цифровой модели показан-

ной на рисунке 9. Она состоит из двух таблиц данных и множества стандартных поисковых таблиц.

Функции легенды состоят в следующем:

- перепись объектов, включаемых в конкретную карту;
- запись описаний этих объектов, важных для составления данной карты;
- указании условных обозначений для объектов карты;
- организации связи объектов пространственного архива (Spatial Object Archive) с атрибутивным описанием их в архиве составных объектов (Compound Object Archive);

Таким образом, понятно, что легенда является "сердцем" нашей модели данных, но "сердцем" Легенды является специальная таблица 5, имеющая название **Classification Object Table**.

Classification Object Table

Таблица определяет объекты, которые появляются на карте и в легенде карты. Это могут быть названия разломов, классов сингулярных объектов со сходными условными обозначениями (обыкновенные разломы или символы сланцеватости). Таблица также используется для определения составных объектов, таких как литологические толщи или зоны метаморфических изменений. Каждому из объектов назначается:

Таблица 5– Атрибуты **Classification Object Table**

Атрибут	Описание	Формат
1	2	3
<i>class_obj_id</i>	уникальный идентификатор, играющий роль ключа таблицы. Он используется для каждого объекта или группы объектов, обозначенных условным знаком и описанных в Легенде	Integer
<i>class_type</i>	строковая подпись или метка геологического объекта. Они выбираются из классификатора модели, что исключает их неоднозначность при вводе.:	Character

Продолжение таблицы 5

1	2	3
<i>class_type</i> (<i>npod.</i>)	<i>sedimentary</i> <i>intrusive</i> <i>extrusive</i> <i>volcanic</i> <i>metamorphic</i> <i>structural symbols</i>	

	<i>Proterozoic Rocks</i> <i>Upper Plate Rocks</i>	
<i>class_desc</i>	англоязычное описание объекта или группы объектов. Для объектов не входящих в Compound Object Archive- это описание, появляющееся непосредственно в Легенде	Character

Данная таблица используется для выбора условных знаков объектов, определённых в **Classification Object Table** (таблица 6).

Каждому объекту в **Classification Object Table** может соответствовать одна или множество записей в **Cartographic Object Table**, т.е. каждый класс характеризуется множеством объектов данного класса. Это позволяет производить обозначение каждой детали результирующей карты, собранной из множества слоёв. Каждый из этих слоёв организован в соответствующие символьные таблицы и таблицу цветов.

Таблица 6– Атрибуты **Cartographic Object Table**

Атрибут	Описание	Формат
<i>class_obj_id</i>	уникальный идентификатор, играющий роль ключа таблицы. Он используется для каждого объекта или группы объектов, обозначенных условным знаком и описанных в Легенде	Integer
<i>cart_desc</i>	Описание образца символа Легенды	Character
<i>cart_sym_table</i>	Название таблицы символов	Character
<i>cart_sym#</i>	№ символа в таблице <i>cart_sym_table</i>	Integer
<i>cart_color_table</i>	Название таблицы цветов	Character
<i>cart_color#</i>	№ символа в таблице <i>cart_color_table</i>	Integer

Выбор нужной таблицы происходит по идентификатору *class_obj_id*. Могут быть следующие варианты:

Point (точечный) символ определяется таблицей **Point Symbol Tables**, Line (линейный) определяется таблицей **Line Symbol Tables** и т.д.

Введение атрибута обеспечивает возможность использования множества таблиц для каждого типа символов, определяемых ключом *class_obj_id*. Данное соглашение позволяет очень гибко и творчески организовывать картографический процесс.

В качестве примера рассмотрим случай, когда пользователю необходимо закрасить полигон или группу полигонов некоторым базовым цветом, причём последний должен перекрываться перекрещивающимися штриховками двух разных цветов.

Вместо разработки специального символа для этой комбинации пользователь просто определит три разных картографических объекта:

- первый по *cart_sym_table* выбирается из соответствующей таблицы символов (это Area Symbol), чтобы найти в ней запись с необходимым поли-

гоном (*cart_sym*), который нужно залить цветом *cart_color*;

- два других объекта определяются точно также, но выбираются из других линейных таблиц, каждому из которых соответствует свой цвет штриховки.

Legend Correlation Tables

Легенда связывается с остальными частями модели посредством трёх корреляционных таблиц. Их назначение в точности такое же как для рассмотренной выше **Map/Source Correlation Table**. Они превращают отношения *многие-ко-многим* в отношения *один-ко-многим*.

Classification Scheme Table

Эта таблица обеспечивает связь между **Map Table** и таблицей Легенды-**Classification Object Table**, объединяя наборы классифицированных объектов с определённой картой. Она также позволяет пользователю определять такие наборы как особые схемы классификации и запоминать их для последующего использования.

Более того, данная таблица включает два специальных атрибута: *class_seq#* и *disp_priority*.

Первый атрибут используется для упорядочения объектов в легенде карты, согласно требованиям пользователя.

Второй атрибут позволяет указывать порядок рисования объектов во время создания карты. Если, например, на карте имеются участки метаморфических изменений или разрывные нарушения, важно чтобы они выводились на печать, после того как будут выведены перекрываемые ими полигоны. Иначе изменённых пород просто не будет видно!

Существует ещё одна таблица **Classification Name Table**, связанная с рассматриваемой корреляционной таблицей. Эта связь позволяет именовать, разработанные пользователем классификационные схемы (атрибут: *class name*) и сохранять их в архивных первоисточниках. Впоследствии эти схемы могут быть вызваны по имени или уникальному идентификатору (**class scheme_id*)

Data Classification Table

Эта корреляционная таблица является посредником между Легендой и архивом составных объектов (Compound Object Archive-COA), осуществляя упрощение типа реляционных отношений между ними. Её структура представлена в таблице 7.

Таблица позволяет связывать единичные объекты (элементы Легенды) с множеством объектов COA. Например, может быть создан (путём выборки) новый элемент Легенды, ссылающийся более чем на одну разновидность горных пород. Среди новых атрибутов, не связанных с уже изученными табли-

цами выделяются три:

Таблица 7– Атрибуты **Data Classification Table**

Атрибут	Описание	Формат
<i>percent</i>	Используется для объектов Compound Archive. Атрибут означает, что пользователь может указать долю разновидности горных пород, в объёмных %, в составном объекте (толще, формации и т.п.). Это число появится в Легенде.	Integer
<i>quality</i>	Описание качества или доверия к подсчёту этих процентов.	Character
<i>data_seg #</i>	Определяет порядок (номер), в котором будут появляться в легенде составляющие Compound Archive.	Integer

Spatial Classification Table

Данная таблица упрощает корреляционные отношения между **Classification Object Table** и различными таблицами **Spatial Object Archive** до типа *один-ко-многим*. Она будет рассмотрена при обсуждении реляционной схемы, представленной на рисунке 10.

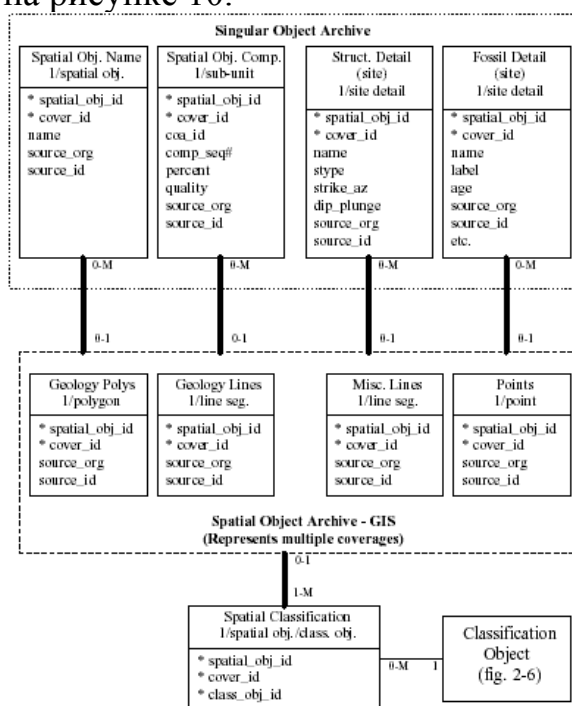


Рисунок 10- Реляционные таблицы архива сингулярных и пространственных объектов

Архив пространственных объектов (**Spatial Object Archive**) представляет собой хранилище пространственного (географического и геометрического) описания всех объектов в картографической библиотеке. Это информация касается формы, размера, местоположения каждого элемента в данном архиве, который должен обслуживать ГИС- основной изобразительный и аналитический компонент нашей модели.

В большинстве ГИС элементы различных геометрических типов хранятся отдельно. Например, точки, линии, полигоны, а также связанные с ними атрибуты обычно сохраняются в разных слоях. Кроме этого, чисто формального подразделения часто бывает удобно группировать однотипные объекты в т.н. тематические слои, объединяющие данные по некоторому полезному свойству. Это могут быть полигоны разновозрастных пород, разломы одной глубины заложения и т.п. Такие тематические слои называются «покрытиями».

Модель данных ЦГК воспринимает Архив пространственных объектов как набор таблиц, каждая из которых представляет покрытие одного геометрически определённого типа. Никаких других специальных условий по организации покрытий наша модель данных не предъявляет. Единственное требование состоит в обеспечении их уникальности. Оно выполняется введением двух атрибутов – *spatial_obj_id* и *cover_id*. Первый из них (*spatial_obj_id*) гарантирует уникальность объекта внутри покрытия, а второй (*cover_id*) гарантирует уникальность самого покрытия. Другими словами не бывает двух покрытий с одинаковыми *cover_id*.

Слои данных, представленные на рисунке 10 представляют собой обычную минимально необходимую геометрическую нагрузку карты. Если, например, карта не имеет в своём составе точечных или линейных объектов, соответствующие слои для данного конкретного случая не будут нужны. И, напротив, на карте могут появиться новые данные, заслуживающие создания специальных слоёв. Например, если на карте отображены полигоны, секущие границы породных толщ, то для них потребуется создание новых специальных слоёв.

Каждая запись в таблице покрытия соответствует одному геологическому объекту в одном слое данных одной карты. Таким образом, эта запись является самым мелким элементом или атомом, из которых геологическая карта конструируется.

Дополнительно к атрибутам *spatial_obj_id* и *cover_id* каждая запись, а следовательно и каждый объект включает информацию для его связи с первоисточником. Для этого используются уже известные атрибуты *source_org* и *source_id*. Поскольку сводные карты возникают в результате монтажа исходных карт, благодаря этим атрибутам мы всегда сумеем отыскать первоисточник каждого геологического объекта.

Корреляционные таблицы Архива пространственных объектов

Spatial Classification Table представляет собой корреляционную таблицу, соединяющую ГИС пространственного архива с **Classification Object Table** Легенды карты. В связи с тем, что между этими таблицами существуют реляционные отношения многие-ко-многим, необходима ещё одна промежуточная таблица, чтобы связать их друг с другом. Единичный вход **Classification Object Table**, такой, например, как литологическая разновидность,

обычно будет ссылаться на несколько ГИС-объектов (полигонов, линий и т.п.).

Аналогично один объект в ГИС-таблице (GIS Table) может быть включён в несколько элементов **Classification Object Table** внутри одной или множества карт. Например, единый сегмент разлома (линия или дуга) может иметь более чем одно направление развития, в связи, с чем на легенде он должен быть представлен несколькими символами.

Внутри ключевой атрибут *cover_id* означает таблицу архива, на которую ссылается каждый объект классификации (в Легенде). Остальные ключевые атрибуты *spatial_obj_id* и *class_obj_id* используются для выбора конкретного пространственного объекта в таблице покрытия и для связи с объектом классификации, к которому данный пространственный объект относится.

2.3.3 Архив сингулярных объектов (Singular Object Archive)

Singular Object Archive состоит из нескольких таблиц, используемых для запоминания описательной информации, связанной с отдельными пространственными объектами. На рисунке представлены только 4 таблицы, хотя на самом деле их может быть значительно больше! Все таблицы в архиве имеют одинаковые ключевые атрибуты *spatial_obj_id* и *cover_id*, описанные выше.

Поскольку каждая запись в таблице архива может ссылаться только на один пространственный объект отсюда и наименование архива- **Singular Object Archive**.

Кроме этого, каждая запись в любой из архивных таблиц содержит атрибуты *source_org* и *source-id*. Они позволяют найти путь до информационных первоисточников каждого объекта.

Spatial Object Name Table

Таблица используется для закрепления за каждым пространственным объектом конкретного имени. Это могут быть имена отдельных батолитов, представленные единым полигоном или название некоторой структуры, представленной единичным линейным сегментом.

Жирная линия, которая соединяет **Spatial Object Name Table** и **Spatial Object Archive** представляет реляционное отношение один-ко-многим. Для каждого пространственного объекта может быть ноль или множество имён, определённых в **Spatial Object Name**. Т.е. определение по имени допускает неоднозначность.

Вместе с тем каждому входу **Spatial Object Name** соответствует только один пространственный объект.

Данное отношение один-ко-многим характерно для отношения **Spatial Object Name Table** с любой атрибутивной таблицей **Spatial Object Archive!**

Spatial Object Composition Table

Таблица используется для хранения информации о составе конкретных пространственных объектов, главным образом полигонов развития отдельных горных пород. Через это свойство она может ссылаться на составные пространственные объекты в **Compound Object Archive** (рисунок 11).

Обычная процедура для заведения новой формации горных пород состоит в определении её и её состава в **Compound Object Archive**. Это определение позволяет впоследствии через Легенду и **Spatial Classification Table** связаться с множеством пространственных объектов.

Для большинства представленных на карте полигонов, информация о составе определённой в Легенде породной разновидности формации или толщи в полигоне или его части отсутствует. Однако, там где это известно **Spatial Object Composition Table** (таблица 8) даёт возможность запоминать информацию о заполнении конкретного полигона породными разновидностями, определёнными в **Compound Object Archive**. Для этого существуют два дополнительных атрибута *coa_id* и *comp_seq#*. Первый однозначно идентифицирует формацию, а второй её составную часть (тип пород) соответственно.

Полностью структура таблицы такова:

Таблица 8– Атрибуты **Spatial Object Composition Table**

Атрибут	Описание	Формат
1	2	3
<i>spatial_obj_id</i>	Уникальный идентификатор единичного объекта в конкретном покрытии	Integer
<i>cover_id</i>	Идентификатор покрытия	Integer
<i>coa_id</i>	Идентификатор толщи в Архиве составных объектов (COA)	Integer
<i>comp_seq#</i>	Номер разновидности горных пород, описанной в составе толщи	Integer
<i>percent</i>	Доля разновидности горных пород, в объёмных %, в составном объекте (толще, формации и т.п.).	Integer
<i>quality</i>	Описание качества или доверия к подсчёту этих процентов.	Character

Продолжение таблицы 8

1	2	3
<i>source_org</i>	Идентификатор организации, создавшей карту	Character
<i>source_id</i>	Идентификатор организации-первоисточника, где описан объект.	Integer

Structural Detail Table

Таблица предназначена для хранения преимущественно точечных

объектов, отражающих данные элементов залегания и других структурных характеристик.

В таблицу заносится минимум информации, необходимой для воссоздания соответствующих символов, показанных на исходной карте. Обычно это индикатор типа откартированного структурного элемента (простираение и падение пласта, сланцеватость, запрокинутость, трещиноватость) и т.п. Для его отображения может быть выбран подходящий символ, развёрнутый и выровнен в соответствующем направлении. После этого символа может быть размещено число, обозначающее угол простираения. Он стандартизуется в интервал 0-359°. Угол падения измеряется вправо от вектора простираения.

Fossil Detail Table

Fossil Detail Table- ещё одна таблица, предназначенная для описания тематических данных в конкретной точке карты. Её структура полностью представлена в таблице 9. Это данные по месторасположению находок определённых окаменелостей. Каждый вход (строка) таблицы может иметь следующие поля:

Таблица 9– Атрибуты **Fossil Detail Table**

Атрибут	Описание	Формат
<i>spatial_obj_id</i>	Уникальный идентификатор одиночной точки в точечном покрытии	Integer
<i>cover_id</i>	Идентификатор покрытия	Integer
<i>name</i>	Полевой номер или наименование образца	Character
<i>label</i>	Метка, связанная с картографическим символом (если она отличается от name). Сам символ определяется в Легенде (Classification Object Table)	Character
<i>age</i>	Возраст обнаруженной окаменелости	Character
<i>source_org</i>	Идентификатор организации, составившей исходную карту	Character
<i>source_id</i>	Организация- первоисточник идентификации окаменелости на исходной карте	Character

В простейшем случае такая таблица может входить в состав **Singular Object Archive**, как и показано на рисунке. При более совершенных схемах это может быть корреляционная таблица связывающая архив карты с БД палеонтологической информации.

2.3.4 Архив составных объектов (Compound Object Archive)

Архив составных объектов или **COA** состоит из множества таблиц данных и поисковых таблиц (рисунки 11-13), используемых для определения составных или комбинированных геологических объектов. Основным компонентом архива является **COA Table**. Через Легенду она связана со всеми остальными компонентами модели, где комбинированные геологические

<i>coa_id</i>	Уникальный идентификационный номер компоненты модели в СОА	Integer
<i>coa_name</i>	Наименование компоненты в СОА	Character
<i>coa_type</i>	Название таблицы с атрибутивной информацией по данной компоненте: Structural Rock Metamorphic Overlay	Character
<i>coa_desc</i>	Текстовое описание объекта (появляется в Легенде)	Character
<i>source_org</i>	Идентификатор организации, составившей исходную карту	Character
<i>source_id</i>	Организация- первоисточник идентификации выделения компоненты на исходной карте	Character

Таблица **COA Table** связана отношением один-к-одному с таблицей **Formal Unit Table** (таблица 11). Это таблица- справочник. Она необходима для хранения информации о формальном определении компонент (единиц) модели. Поскольку не каждая из них требует своего эквивалента в **Formal Unit Table**, то для каждого входа **COA Table** существует или ноль или максимум одна запись в **Formal Unit Table**.

Таблица 11– Атрибуты **Formal Unit Table**

Атрибут	Описание	Формат
<i>coa_id</i>	Уникальный идентификационный номер компоненты в СОА	Integer
<i>name</i>	Формальное наименование картографической единицы	Character
<i>type_section</i>	Местоположение описания типа разреза или площади.	Character
<i>etc</i>	Дополнительные атрибуты.	

К таблице **COA Table** присоединена с помощью отношения один-ко-многим ещё одна очень интересная таблица **COA Relation Table** (таблица 12). Она используется для хранения топологических, временных и любых других геологических взаимоотношений между объектами **COA Table**.

В качестве примера, приведём следующие геологические отношения: старше, моложе, одновозрастные, связаны с..., часть того-то, включает, протыкает, выше, ниже и т.д.

Обычно эти взаимоотношения показываются в Легенде с помощью маленьких пиктограмм, но здесь их можно хранить в рамках самой модели.

В связи с тем, что между этими двумя таблицами существует реляционная связь один-ко-многим, между любыми двумя объектами **COA Table** может существовать множество взаимоотношений, оформленных в виде записей таблицы **COA Relation Table**.

Например, соляной диапир может интродуцировать более молодые породы (протыкать, старше).

Рассмотрим структуру таблицы более подробно.

Таблица 12– Атрибуты **COA Relation Table**

Атрибут	Описание	Формат
<i>rel_id</i>	Уникальный идентификатор записи в таблице.	Integer
<i>coa_id</i>	Уникальный идентификационный номер первой компоненты в COA	Integer
<i>rel_coa_id</i>	Уникальный идентификационный номер второй компоненты в COA, с которой у первой имеются взаимоотношения.	Integer
<i>relation</i>	Широкий диапазон наименований, обозначающих временные и структурные взаимоотношения между компонентами модели (справочник,)	Character
<i>rel_desc</i>	Текстовое описание взаимоотношений <i>relation</i>	Character

COA Tree Table

Данная таблица также связана с **COA Table** отношением один-ко-многим. Она предназначена для поддержки разнообразных геологических иерархий, обычных при создании сводных карт. Это может быть иерархия стратиграфического описания, иерархия порядка линейных сегментов тектонических нарушений и т.п.

Таблица содержит только два атрибута *coa_id* и *parent_id*. Вместе они создают составной ключ данной таблицы, содержащей все списки «родителей- (*parent_id*)» для каждого геологического объекта (*coa_id*) таблицы **COA Table**.

Например, стратиграфические подразделения, к которым принадлежит конкретный горизонт, должны быть перечислены как его родители.

Если все «родственные» отношения между картографическими компонентами будут прописаны в **COA Tree Table**, её можно использовать для автоматического конструирования иерархии стратиграфического разреза или создания сводной ЦГК, путём смысловой группировки элементарных картографических единиц. Например, по пространственному запросу создана смешанная выборка разнопорядковых картографических элементов. Если иерархические взаимоотношения между ними сохранены в таблице, то автоматическое восстановление сводной карты на их основе возможно.

Другой пример из области разрывной тектоники. Крупнейший Калифорнийский разлом San Andreas состоит из великого множества разнопорядковых сегментов. Самым древним предком каждого из них будет San Andreas, т.е. его *parent_id* будет записан для каждого «сыновнего» разлома.

Rock Units

Данная таблица является основной для хранения описательной информации, по горным породам всех видов, встречающихся на моделируемых картах (рисунок 12). Сюда относятся все типы стратиграфических подразделений независимо от того слоистые они или нет, толщи, представленные неконсолидированными осадками, ледяные поля и водные пространства, даже если подстилающие их породы на карте не отображаются!

Собственно **Rock Units** включает в себя только самую необходимую информацию- идентификационный номер породной компоненты модели (ПКМ) и её ранг (атрибуты *coa_id* и *rock_rank*, соответственно).

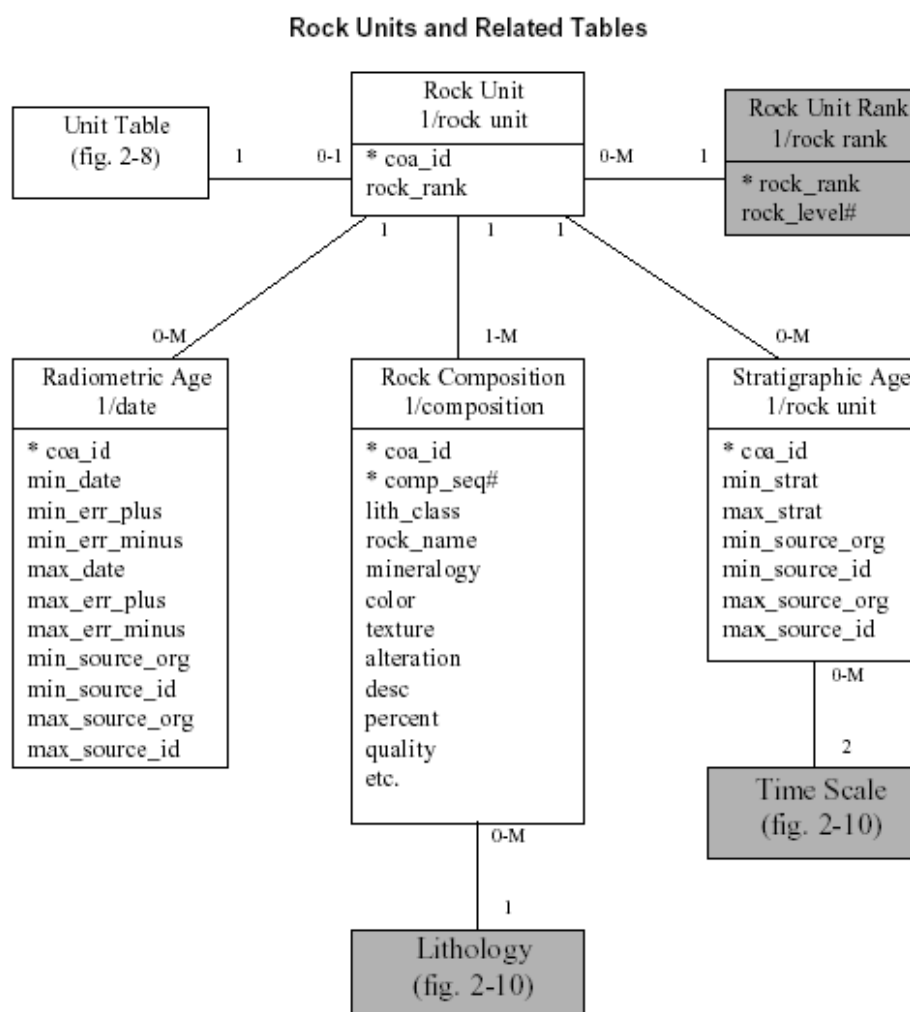


Рисунок 12- Структура модуля *Rock Units*

Все остальные сведения по ПКМ касаются или её части, т.е. состава или могут встречаться более чем однажды, например- определение возраста. Такая информация хранится во вспомогательных таблицах. Одна из них называется **Rock Unit Rank Table**.

Rock Unit Rank Table представляет собой простую справочную таблицу (один-к-одному), которая используется для связки числового атрибута *rock_level#*, с каждым наименованием ПКМ. Это даёт возможность пользо-

вателю вставлять такие слова как *формація*, например, в атрибут *rock_rank* и одновременно ссылаться на него числовым образом.

Уровни рангов организованы иерархически так, что самому высокому рангу соответствует минимальное значение (например, свита = 100, пласт = 500).

Одна ПКМ может иметь любое число определений возраста, которые запоминаются в специальной таблице **Radiometric Age Table**. По этой причине она связана с **Rock Unit Table** реляционным отношением один-ко-многим. Ключом для их связи является значения одинакового атрибута *coa_id*. Дополнительно в эту таблицу включены атрибуты для хранения информации о минимальном и максимальном возрастах, ошибках их определения и ссылки на первоисточники, где опубликованы эти данные.

Stratigraphic Age Table

Подобным же образом в модели ЦГК организовано хранение сведений о стратиграфических возрастах ПКМ. Для этого заведена специальная таблица. Она также связана с **Rock Unit Table** реляционным отношением один-ко-многим, поскольку допускается, что с одной ПКМ связано несколько стратиграфических возрастных диапазонов. В большинстве случаев для каждой ПКМ заводится одна запись стратиграфического диапазона, определяемого минимальным и максимальным значениями. Однако, встречаются случаи, когда описание возраста дискретно, т.е. ПКМ соответствуют несколько несмежных возрастных интервалов. Тогда на одну ПКМ будет приходиться несколько записей в **Stratigraphic Age Table**.

Как и **Radiometric Age Table** данная таблица содержит атрибуты для минимального и максимального возрастов, и ссылки на первоисточники возрастной диагностики. Отличие состоит в том, что в **Radiometric Age Table** для этих атрибутов используются числовые значения, в то время как в **Stratigraphic Age Table** для этих целей используются наименования стратиграфических интервалов. Они берутся из хронологической шкалы, закодированной в цифровой реляционной модели в форме трёх таблиц: **Stratigraphic Time Scale, Stratigraphic Rank, and Stratigraphic Tree** (рисунок 13).

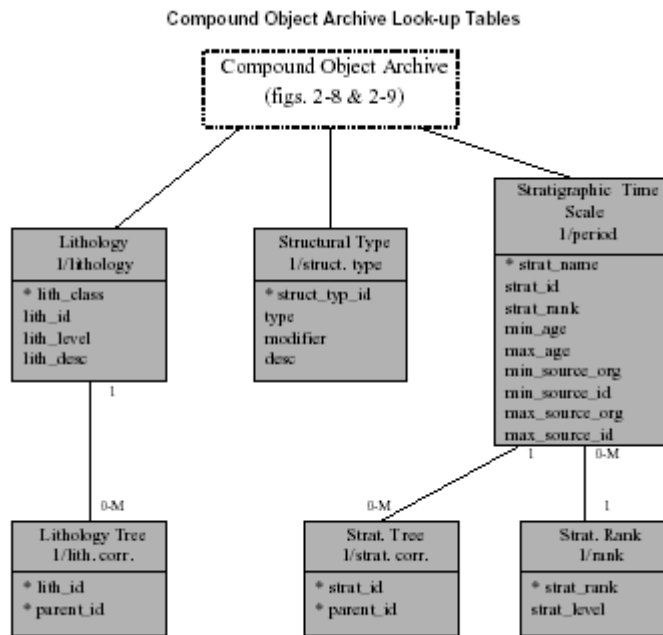


Рисунок 13- Справочные таблицы (Time Scale) COA

Справочная таблица **Stratigraphic Time Scale Table** имеет иерархическую структуру. Реляционные отношения *предок-потомок* стратиграфической шкалы хранятся в **Stratigraphic Tree Table**, а уровень внутри иерархии запоминается в **Stratigraphic Rank Table**. Такая структура позволяет решить многие проблемы хранения стратиграфических данных и облегчает создание сводных карт, базирующихся как на хронологии геологических компонент, так и генерации производных ПКМ на основе стратиграфического деления. Кроме того, существенно облегчается анализ геологических карт, требующий сопоставления возрастов природных образований.

Все три таблицы являются стандартными справочниками, входящими в состав реляционной модели ЦГК. Любая организация, использующая описываемую модель данных, может заменить их своими, более привычными таблицами, но пользователь или дизайнер карт об этом просто не должен знать! Добиться этого непросто, поскольку данные справочные таблицы используются при построении любой карты.

Stratigraphic Time Scale Table

Эта таблица содержит записи, по одной для каждого стратиграфического периода. Ключевой атрибут *strat_name* предназначен для хранения наименования периода (Мезозой, Мел и т.д.). Каждому периоду присваивается условный идентификационный номер *strat_id*. Он используется для хранения отношений *предок-потомок* между ними. Кроме того, каждому периоду назначается специальный ранг (эпоха, эра и т.д.), запоминаемый в атрибуте *strat_rank*. Дополнительно, таблица содержит атрибуты для хранения минимальных и максимальных значений радиометрического возраста и организаций, где данные определения были произведены.

Rock Composition Table

Данная таблица (таблица 13) используется для хранения сведений о составе ПКМ. Поскольку она связана отношением *один-ко-многим* с Rock Unit Table, это означает возможность существования множества записей по составу пород в **Rock Composition Table** для одной породной компоненты **Rock Unit Table**. Такой подход позволяет описывать породные компоненты смешанного состава. Причём, каждому типу породных разновидностей будет отведена своя запись в **Rock Composition Table**. Никаких ограничений по числу таких разновидностей не предусмотрено.

Таблица 13– Атрибуты **Rock Composition Table**

Атрибут	Описание	Формат
<i>coa_id</i>	Уникальный идентификационный номер породной компоненты в Compound Object Archive	Integer
<i>comp_seq#</i>	Уникальный идентификационный номер разновидности (композиции) породной компоненты. Он используется также как порядковый номер блока описательной информации по этой разновидности для данной породной компоненты. Обычно, нумерация начинается с самых значительных в процентном отношении разновидностей и заканчивается наименее представленными.	Integer
<i>lith_class</i>	Термин литологической классификации, выбранный из справочника Lithology Table (см. ниже)	Character
<i>rock_name</i>	Свободная текстовая фраза, означающая наименование разновидности, данное автором карты. её выделения.	Character
<i>mineralogy</i>	Минералогическая характеристика, связанная с <i>rock_name</i> .	Character
<i>color</i>	Цвет или цвета разновидности.	Character
<i>texture</i>	Текстура разновидности	Character
<i>alteration</i>	Описание любых изменений, которым подверглась породная разновидность в ходе геологической истории.	Character
<i>desc</i>	Авторское литологическое описание разновидности, предназначенное для публикации.	Character
<i>percent</i>	Доля разновидности в объёмных процентах от объёма породной компоненты.	Integer
<i>quality</i>	Качество определения доли разновидности (\pm nn %)	Character

Смысл большинства атрибутов понятен без объяснений. Однако значения некоторых *lith_class*, например, из них берутся из справочных таблиц, таких как **Lithology Table**.

Lithology Table

Данная таблица (таблица 14) предназначена для хранения предварительно разработанной литологической иерархии типов горных пород. В этом смысле она напоминает **Stratigraphic Time Scale Table**, но отличается от неё меньшей строгостью и стандартизацией.

Основные структурные уровни этой классификации называются:

Unconsolidated-	неконсолидированные, рыхлые
Sedimentary-	осадочные
Intrusive-	интрузивные, магматические
Extrusive-	вулканические
Metamorphic-	метаморфические.

Каждый из данных разделов делится на единицы второго уровня, такие как песчаники или гнейсы, например. Второй уровень делится на третий, внутри которого может выделяться самый специфический четвёртый (аркозы, силикатный гнейсы и т.д.)

Причина создания такой таблицы состоит в том, чтобы проводить сравнения пород по литологическому признаку внутри одной породной компоненты модели. Если диагностика проведена автором, то такое сравнение возможно в принципе. Если же такое сравнение произведено специалистами, не сведущими в геологии данного региона, то его результаты могут быть случайны и непредсказуемы.

Таблица 14– Атрибуты **Lithology Table**

Атрибут	Описание	Формат
<i>*lith_class</i>	Уникальный термин (ключ) из предварительно определённого перечня описаний литологических разновидностей. Используется в Rock Composition Table	Character
<i>lith_id</i>	Уникальный идентификационный номер литологического термина. Используется в Lithology Tree Table для хранения отношений предок-потомок.	Integer
<i>lith_level</i>	Числовое значение уровня иерархии литологического термина.	Integer
<i>lith_desc</i>	Текстовое описание литологического термина.	Character

Structural Correlation Table

Данная таблица архива составных объектов является корреляционной. Она служит для связи и преобразования реляционного отношения *многие-ко-многим* между справочной таблицей **Structural Type** и **COA Table** в два простых отношения *один-ко-многим*.

Множественные отношения между таблицами возникают оттого, что тектонические элементы в природе и на карте должны быть представлены в различных аспектах. Например, многие разломы характеризуются смещениями как по простиранию, так и по глубине. Каждому такому аспекту или характеристике структуры в таблице **Structural Type** соответствует отдельная запись.

Structural Correlation Table содержит два ключевых атрибута для связи соединяемых таблиц: *coa_id* и *struct_typ_id*. Кроме этого, существует атрибут *accuracy*, используемый для указания уровня точности картирования структуры. Это могут быть слова: "приблизительно", "условно", "скрытый" и т.п.

Справочная таблица **Structure Type**, связываемая с помощью описываемой корреляционной таблицы с **COA Table**, помимо ключевого атрибута содержит атрибуты:

type- для указания типа структуры (складка, разлом, сдвиг) и т.п.

modifier- для конкретизации *type*. Например, разлом может быть вертикальный, а складка- антиклинальная и т.п.

desc- используется для текстового описания структуры, хранимой в конкретной записи данной таблицы.

Metamorphic Overlay Units

В архиве составных объектов СОА нам осталось ещё одна таблица, структура которой несколько отличается от рассмотренных ранее.

Метаморфические образования, оформленные в стиле покрытий, включены в модель для демонстрации её (модели) способности объединять в единое целое различные типы картографических компонент. Любое полигональное образование, может быть наложено на базовый полигон, осложняя его в точности так же, как зоны метаморфизации изменяют исходные породы. В дальнейшем будем для ясности называть такое покрытие метаморфической накладкой.

На геологических картах области метаморфизации обычно представлены условными знаками или крапом, накладывающимися на цветные полигоны развития разнообразных горных пород. При этом каждой метаморфической накладке соответствует одна уникальная запись в **COA Table** со своим *coa_type*. Это означает, что между ними существует реляционное отношение *один-к-одному*. Другими словами метаморфическая накладка обслуживается моделью как полноценный составной объект. Она не хранится в **COA Table** в явном виде только потому, что имеет специфические черты, отражённые в

специальных атрибутах- нетипичных для большинства составных объектов. Это может быть атрибут *grade* обозначающий степень метаморфизма и, вообще, переработки, взятый из специального справочника. Кроме того, в данной таблице может храниться её развёрнутое текстовое описание.

Заканчивая данную главу можно сделать следующие выводы, перечисленные в порядке убывания важности.:

- эксплуатация, а тем более создание развитых программных комплексов подразумевает понимание существенных свойств моделируемых природных явлений;
- формализация этих свойств в большинстве случаев означает их ясную и формулировку, допускающую построение непротиворечивых логических схем;
- построение логических схем невозможно без создания тематических словарей и стандартов обеспечивающих однозначность информационных описаний;
- для перевода логических схем из понятийного состояния в кибернетические структуры необходимы знания по теории БД и навыки программирования;
- реализация крупных программных проектов в области геологии возможна при создании крупных коллективов квалифицированных единомышленников.

При выполнении перечисленных предпосылок даже самые сложные задачи естествознания, могут решаться методами компьютерного моделирования. Пример тому- рассмотренный нами реляционная модель ЦГК реализованная в США, в рамках Национального Проекта Базы Данных Геологических карт (NGMDB)

3 Основы пространственных баз данных

3.1 Введение в пространственные базы данных

3.1.1 Обзор вопроса

Мы переживаем разгар информационной революции. Сырьё (данные), уже не покоятся в земной коре, где они формировалось на протяжении миллионов лет, а непрерывно накапливаются многочисленными датчиками и устройствами сбора информации. Так, например, по опубликованным сведениям НАСА система Earth Observing System (EOS) ежедневно генерирует один терабайт (2^{10} Гигабайта) данных.

Снимки с искусственных спутников- яркий пример пространственных данных. Чтобы считать информацию с такого снимка, нужно обработать данные с учётом пространственной системы координат, которой, возможно, является поверхность Земли. Однако, искусственные спутники- не единственный источник пространственных данных, а земная поверхность- не единственная система координат. Системой координат может быть, и часто оказывается кремниевая микросхема. В медицинской практике в качестве системы координат используется тело человека.

Запросы или команды, составляемые в отношении пространственных данных, называются пространственными запросами. К примеру, запрос "Как называются все книжные магазины, где продают ежедневно более тысячи книг?" является примером непространственного запроса. С другой стороны, запрос "Как называются все книжные магазины в радиусе пяти километров от центра города?"- это пример пространственного запроса.

На сегодняшний день для хранения данных и управления ими используются системы управления базами данных (СУБД). Базы данных и программы, построенные на их основе- это безмолвная "история успеха" информационной эпохи. Мало-помалу они стали пронизывать все стороны повседневной жизни, и современное общество потерпит крах, если их вдруг не станет.

Несмотря на эти впечатляющие достижения большинство существующих СУБД, как полагают многие специалисты, либо вообще не способны управлять пространственными данными, либо делают это в крайне недружественной по отношению к пользователю манере. Почему так происходит?

Традиционная роль СУБД- быть простым, но эффективным хранилищем деловой и бухгалтерской информации. Используя СУБД, можно надёжно сохранять и эффективно извлекать из базы данных сведения о сотрудниках, поставщиках, клиентах и товарах. Набор вероятных запросов ограничен и база данных устроена так, чтобы эффективно на них отвечать. Из мира бизнеса СУБД безболезненно перекочевали в правительственные учреждения и промышленность.

Информация, постоянно хранящаяся в этих громадных базах данных достаточно проста.- это числа, имена, адреса, описания объектов и т.п. Подобные СУБД очень эффективны для решения тех задач, ради которых

они были спроектированы. Например, запрос "Перечислить десять ведущих компаний мира по продажам нефти за 2000 год" будет очень эффективно обработан СУБД даже в том случае, если для этого необходим просмотр чрезвычайно большой базы данных с информацией о заказчиках.

Команды такого рода принято называть "запросами", хотя вопросы в них не входят. База данных сумеет выполнить такой запрос, даже не просматривая информацию по всем заказчикам. Она воспользуется индексом, так же как если бы Вы обратились к предметному указателю книги, чтобы сократить область поиска.

С другой стороны такой сравнительно простой запрос- как "Составить список сотрудников, проживающих не далее чем в пяти километрах от главного офиса компании" поставит базу данных в тупик. Для его обработки она должна:

- преобразовать сведения о главном офисе компании и адреса клиентов в удобную систему координат, возможно на основе широты и долготы, в которой модно вычислять и сравнивать расстояния.
- просмотреть полный список сотрудников и сравнить расстояние от офиса до места проживания каждого из них.
- запомнить имя сотрудника, если это расстояние окажется менее пяти километров.

База данных не способна использовать индекс для сужения области поиска, поскольку традиционные индексы не справляются с упорядочением многомерных координат. Это означает, что относительно простой с точки зрения эффективности запрос может привести традиционную СУБД в состояние "зависания". По этой причине в настоящее время информационными продуктами первой необходимости становятся базы данных, специально предназначенные для обработки пространственных данных и пространственных запросов.

3.1.2 Кто заинтересован в управлении пространственными данными

Профессионалы во всех сферах деятельности вынуждены иметь дело с управлением пространственными данными и их анализом. В таблице 15 приводится краткий перечень специалистов в разных областях и по одному примеру запроса, который будет соответствовать сфере интересов каждого из них.

Таблица 15– Актуальные географические запросы

Профессия, служба	Актуальный запрос
1	2
Владелец мобильного телефона	Где поворот к ближайшей автозаправочной станции? Где по дороге купить корм для своей кошки?
Командующий армией	Насколько значительно перемещение войск противника по сравнению со вчерашним днём?
Управляющий страховой компанией	Какие дачи на берегу Урала пострадают от крупного наводнения в первую очередь? Имеет ли смысл браться за их страховку?
Врач	Как на основании клинической практики лечения больных зубом определить источники водоснабжения с недостатком йода?
Климатолог	Как мне проверить новую модель глобального потепления и подтвердить её правильность?
Специалист по транспорту	Как следует расширить сеть автодорог, чтобы до минимума сократить заторы.
Служба спасения	Где находится человек, зовущий на помощь? Как лучше до него добраться?

3.1.3 ГИС и СУПБД

Географические информационные системы являются основной технологией, привлекающей внимание к системам управления пространственными базами данных (СУПБД). Основные функции ГИС- это пространственный анализ и визуализация географических данных [5].

Географические данные- это пространственные данные, для которых базовой системой координат является земная поверхность. Великое множество технологий, встроенных географами в ГИС,- это именно та причина, которая вызвала феноменальный рост этих систем и послужила основанием для появления многочисленных междисциплинарных приложений. В таблице 16 приведён краткий перечень самых распространённых операций, реализованных в ГИС

Таблица 16– Перечень операций анализа, реализованных в ГИС

Операция	Задача
1	2
Поиск	Тематический поиск, поиск по областям
Анализ местоположения	Буфер, коридор, наложение
Топографический анализ	Расчёт уклона/экспозиции, дренаж, гидрографическая сеть
Анализ потоков	Связность, кратчайший путь
Распределение	Диагностика изменений, близость, ближайший сосед
Пространственный анализ/статистика	Шаблоны, центры, автокорреляция, индексы сходства, топология
Измерения	Расстояние, периметр, очертания, смежность, направление.

ГИС поддерживают большой набор операций над несколькими объектами и слоями, в то время как СУПБД обеспечивают более простые операции над совокупностями объектов и множествами слоёв. Например, если ГИС располагает сведениями о политических границах каждой страны, то она может составить список территорий, граничащих с данным государством (например, Францией). Однако, отыскание ответа на запросы множеств, скажем "перечислить страны с наибольшим количеством соседей" или "найти страны, территория которых полностью окружена одной страной" для ГИС может оказаться очень трудным делом. Но оно как раз по силам СУПБД.

СУПБД созданы для работы с очень большими объёмами пространственной информации. Она может храниться на вторичных носителях – магнитных дисках, компакт дисках, дисковых библиотеках и т.п. При этом для работы с ними используются специальные индексы и особые приёмы обработки запросов. Кроме того, СУПБД унаследовали от традиционных СУБД способность поддерживать механизмы управления параллельными заданиями, что позволяет множеству пользователей одновременно получать доступ к совместно используемым пространственным данным, сохраняя при этом их целостность.

ГИС может быть построена как интерфейсная часть СУПБД. Прежде чем ГИС сможет выполнить анализ пространственных данных, она обращается к соответствующим данным, хранящимся в СУПБД. По этой причине эффективная СУПБД может существенно повысить и производительность ГИС.

Слияние ГИС и СУПБД возникает во многом как следствие появления новых сфер информационных услуг, в частности Интернет-услуг и развития мобильной телефонной сети. Появление новых программных приложений, где решаются подобные задачи, требует очень быстрой реакции на запросы

пользователя, которые могут быть самыми неожиданными. Например, федеральное правительство США постановило, что начиная с октября 2001 года инфраструктура сотовой телефонии должна позволять отслеживать положение всех аппаратов с точностью 125 м в течении 67% времени их активного состояния. Это означает, что запрос клиента, движущегося в автомобиле со скоростью 70 км в час к ГИС-службе провайдера может иметь такой вид : "Куда повернуть, чтобы попасть на ближайшую заправочную станцию?" при этом сообщать о своём местоположении уже нет необходимости.

Парадокс состоит в том, что хотя Интернет-революция, как часто отмечают, и положила конец понятию географии, спрос на пространственные технологии постоянно растёт.

В этой связи интересна эволюция концепции аббревиатуры GIS на протяжении двух последних десятилетий:

1980-ые	Geographic Information System
1990-ые	Geographic Information Science (Геоинформатика)
2000-ые	Geographic Information Services (Географические информационные услуги).

3.1.4 Таксономия пространства и модели данных

Что такое пространство- понятие действительно очень сложно. Тем не менее, мы пытаемся в нём ориентироваться, и делаем это не всегда лучшим образом. Хуже всего у человека дело обстоит с запоминанием направления и ориентации. Другие существа (птицы, кошки, собаки) безошибочно их определяют с первого раза и запоминают (или вычисляют) навсегда. Однако, когда речь идёт о припоминании *топологических* отношений, таких как: смежный, связанный, внутри и т.д. мы, люди мыслим довольно неплохо.

Топология как раздел математики посвящена исключительно изучению отношений, остающихся неизменными при эластичных деформациях базисного пространства.

Например, если мы на куске резины нарисуем два прямоугольника так, чтобы один находился внутри другого или бы они имели общую сторону, то как бы мы не скручивали резинку, названные отношения между прямоугольниками не изменятся! По-видимому, топологическая структура пространства и есть то таинственное свойство мироздания, формирующее нашу личность и отличающую её от остальных существ.

Попытки объяснить структуру географического пространства привели к созданию т.н. пространственной таксономии (классификации). Она имеет дело с различными формами описания, пригодными для упорядочения пространства- это пространства:

- топологическое;
- сетевое;
- евклидово;

– направлений.

В зависимости от того, что в первую очередь нас интересует мы выбираем подходящее пространственное описание. В нижеследующей таблице 17 приведены различные типы пространств и связанные с ними географические понятия:

Таблица 17– Связь пространств и географических понятий

Пространство	Географическое понятие
Топологическое	Смежный
Сетевое	Кратчайший маршрут
Пространство направлений	К северу от
Евклидово	Расстояние

Однако, важно понять, что ни одно описание пространство не может быть достаточным для произвольного запроса к СУПБД, т.е не может подходить для всех случаев в жизни.

Вникая глубже в данный вопрос нельзя обойти стороной такую важную тему как модели данных.

Модель данных- это правило или набор правил идентификации и отображения объектов, представленных в некотором пространстве.

Карелия- "земля тысячи озёр". Но как отобразить эти озёра? На интуитивном уровне непосредственным представлением каждого озера служит двумерная область. Аналогично поток можно представить одномерной кривой, а месторасположение скважины- нуль-мерной точкой. *Это объектная модель.* Она идеально подходит для представления таких обладающих формой сущностей как озёра, сети дорог и города.

Объектная модель является концептуальной; она преобразуется в компьютерное представление посредством векторной структуры данных.

Полевая модель часто применяется для представления непрерывных или аморфных явлений, например температуры или земного магнетизма.

Поле называется функция, отображающая базисную систему отсчёта на область определения атрибута.

Для температуры, например, распространённой областью определения атрибута является шкала Цельсия или Фаренгейта.

Компьютерной реализацией полевой модели обычно служат растровые структуры данных. В силу пространственной автокорреляции значений поля (они непрерывны) значение в каждой ячейке растра является средним для всех точек поля внутри ячейки.

3.1.5 Пространственные запросы

Самой замечательной особенностью развитых СУПБД является возможность общаться с ними на международном стандартном языке для определения и доступа к реляционным базам данных. Он называется Structured

Query Language или язык структурированных запросов SQL.

SQL-декларативный язык, иначе говоря пользователь определяет только желаемый результат, но не способ его достижения. Так, например, может выглядеть SQL-запрос "Выбрать все земельные участки, граничащие с участком МОЙ ДОМ":

```
SELECT      M.address
FROM        land_parcel L, M
WHERE       Adjustment(L,M) AND L.address='МОЙ ДОМ'
```

Вся ответственность за реализацию операций, указанных в тексте запроса ложится на СУБД. В частности, требуется, чтобы в запросе на языке SQL можно было вызывать функцию проверки смежности **Adjustment(L,M)**. Таким образом, не обращаясь явно к ГИС, с помощью подобного запроса решается типичная аналитическая задача.

Обработка запросов подразумевает выполнения ряда шагов, которые иницируются СУБД.

Все запросы можно разделить на две категории: те что требуют *однопроходного* просмотра БД и те, что требуют *многопроходного* просмотра.

При обработке запроса первой категории обращение к запрашиваемой записи (кортежу) таблицы (отношения) должно производиться не более одного раза. Запрос "Как называются все книжные магазины в радиусе 5 км от центра Москвы?"- это пример запроса, требующего однопроходного просмотра.

Данный запрос также называется *запросом пространственного диапазона*, в котором под диапазоном понимается круг радиусом в 5 км.

Если область запроса имеет прямоугольную форму, то запрос пространственного диапазона часто называют запросом окна (window query).

Запрос соединения (join query)- это пример запроса, требующего многопроходного просмотра. Чтобы выдать ответ на запрос соединения, СУБД должна произвести выборку и комбинирование двух таблиц базы данных. Если же таблиц больше чем две, то они могут обрабатываться попарно. В контексте пространственных БД, когда атрибуты соединения являются пространственными по своей природе, такой запрос носит название "*пространственного соединения*" или (spatial-join query). Рассмотрим отличия и особенности непространственного и пространственного соединения на примере двух SQL-запросов.

Проанализируем две реляционных таблицы следующей структуры:

Senator (S)

Имя (name)	№ страховки (soc)	Пол (gender)	Округ (district-полигон)
------------	-------------------	--------------	--------------------------

Business (B)

Название (b-name)	Владелец (owner)	№ страховки (soc)	Месторасположение (point-точка)
-------------------	------------------	-------------------	---------------------------------



Люди, представленные в этих таблицах однозначно идентифицируются по № страховки. Составим следующий непространственный запрос: "Составить список имён всех сенаторов-женщин, имеющих собственные компании".

На языке SQL данный запрос может быть записан так:

```
SELECT          S.name  
FROM           Senator S, Business B  
WHERE          S.soc=B.soc AND S.gender ='Женщина'
```

Теперь рассмотрим запрос, который можно задать после выполнения первого, т.е применим его к уже готовой выборке: "Найти всех сенаторов, избранных от округов площадью более 300 квадратных миль и являющихся владельцами собственных компаний в границах своего округа"

```
SELECT    S.name  
FROM      Senator S, Business B  
WHERE     S.district.Area()>300 AND  
           Within(B.location, S.district)
```

Обратите внимание, что в случае с непространственным соединением ключевые атрибуты обязаны быть однотипными, а при пространственном соединении они могут иметь разные типы (точка и полигон).

3.1.6 Элементарное введение в SQL

SQL-коммерческий язык запросов, изначально разработанный корпорацией IBM. С момента возникновения он приобрёл статус стандартного языка запросов к реляционным базам данных. SQL является декларативным языком, другими словами, пользователь должен только описать ответ на запрос, не указывая процедуру получения этого ответа.

Язык SQL включает в себя, по меньшей мере два отдельных компонента: язык описания данных -ЯОД или DDL (data description language) и язык модификации данных- ЯМД или DML (data modification language).

ЯОД используется для создания, удаления или изменения структур описания таблиц базы данных.

На ЯМД формулируются запросы, иницирующие вставку и удаление строк из таблиц, заданных средствами ЯОД.

Кроме этого SQL содержит и другие операторы языка управления данными. В данном параграфе будут изложены начальные сведения об этом средстве, достаточные для того, чтобы понять пространственные расширения SQL, которые будут обсуждаться ниже.

Язык описания данных (ЯОД)

Создание реляционной схемы, создание и удаление таблиц выполняются средствами единственного SQL-компонента, названного нами ЯОД.

Представим, что нам необходима реляционная таблица с условным названием CITY. Она должна состоять из пяти атрибутов (столбцов), которые

называются соответственно:

Name	Название
Country	Страна
Pop	Население
Capital	Столица
Shape	Очертания

SQL –оператор для её создания выглядит так:

```
CREATE TABLE CITY {
    Name VARCHAR(35),
    Country VARCHAR(35),
    Pop INT,
    Capital CHAR(1),
    Shape CHAR(13),
    Primary KEY (Name)}
```

Разберём структуру оператора **Create Table**.

Именем таблицы является **CITY**. Таблица состоит из пяти столбцов, при этом должно быть указано название каждого столбца и соответствующий тип данных.

Атрибуты **Name** и **Country** должны быть строками ASCII-символов длиной не более 35 знаков.

Атрибут **Pop** имеет целочисленный тип.

Атрибут **Capital** является однобуквенным. Если город является столицей, то **Capital** принимает значение “Y” (да), в противном случае его значение равно “N”. И, наконец, первичным ключом таблицы служит атрибут **Name**. Это означает, что каждая строка таблицы должна иметь уникальное значение для имени города. Другая строка с таким же именем будет отвергнута!

Создадим ещё две реляционные SQL таблицы, которые понадобятся нам для дальнейшего освоения материала:

Схема Country	Схема River
<pre>CREATE TABLE Country { Name VARCHAR(35), Cont VARCHAR(35), Pop INT, GDP INT, LIFE-EXP NUMBER Shape CHAR(15), PRIMARY KEY (Name)}</pre>	<pre>CREATE TABLE River { Name VARCHAR(35), Origin VARCHAR(35), Length INT, Shape CHAR(15), PRIMARY KEY (Name)}</pre>

Язык модификации данных (ЯМД)

После того, как таблица создана по правилам ЯОД, она готова принимать данные. Эта задача часто называется «заселением таблицы» и выполняется средствами другого компонента SQL-языка модификации данных. К

примеру, следующий оператор добавляет к условной таблице River одну строку:

INSERT INTO	River (Name, Origin, Length) Values('Миссисипи', 'США', '6000')
--------------------	--

Если заданы не все атрибуты отношения (таблицы), выполняется автоматическая подстановка по умолчанию. Наиболее часто в этом качестве используется «пустое» значение NULL. Попытка добавить в таблицу **River** другую строку с Name 'Миссисипи' будет отвергнута СУБД в силу ограничений по первичному ключу.

Основная форма оператора удаления строк таблицы имеет такой вид:
DELETE FROM <ТАБЛИЦА> WHERE <УСЛОВИЯ>

Например, следующий оператор удаляет из таблицы **River** строку, которую мы только что вставили:

```
DELETE FROM River  
WHERE Name='Миссисипи'
```

Основная форма запроса на языке SQL

Когда схема базы данных описана средствами ЯОД, а таблицы заполнены данными, можно формулировать SQL-запросы на извлечение из базы нужных подмножеств данных. Основная синтаксическая конструкция запроса на SQL чрезвычайно проста:

```
SELECT      Имена-столбцов  
FROM       Отношения (таблицы, участвующие в запросе)  
WHERE      Ограничения на кортежи (строки)
```

Примеры запросов на языке SQL

- 1 Составить список всех городов в таблице City и тех стран, где эти города находятся:

```
SELECT      Ci.Name, Ci.Country  
FROM       City Ci
```

В данном запросе предложение WHERE отсутствует, поскольку на строки не накладывается ни каких ограничений. Мы просто выбираем все строки из двух заказанных столбцов.

- 2 Составить список столиц, перечисленных в таблице CITY:

```
SELECT      *  
FROM       City  
WHERE      CAPITAL='Y'
```

В данном запросе знак * означает, что на выводе должны присутствовать значения всех столбцов таблицы **City**.

- 3 Составить список атрибутов стран по таблице Country, в которых средняя продолжительность жизни меньше 70 лет:

```
SELECT      Co.Name, Co.Life-Exp
FROM        Country Co
WHERE       Co.Life-Exp <70
```

- 4 Как называется столица и какова численность населения страны, где начинается река Св. Лаврентия?

```
SELECT      Ci.Name, Ci.Pop
FROM        City Ci, Country Co, River R
WHERE       R.Origin = Co.Name AND
           Co.Name = Ci.Country AND
           R.Name= 'Св. Лаврентия' AND
           Ci.Capital = 'Y'
```

- 5 Составить список стран, ВВП которых превосходит ВВП Канады:

```
SELECT      Co.Name
FROM        Country Co
WHERE       Co.GDP > ANY (SELECT Co1.GDP
                        FROM Country Co1
                        WHERE Co1.Name = 'КАНАДА')
```

3.1.7 Расширение SQL для пространственных данных

Даже будучи самым мощным языком обработки запросов SQL имеет свои недостатки. Главный из них состоит в том, что поддерживаются лишь простые типы данных, например целые числа, даты, строки. Программы, созданные как расширения пространственных БД должны работать со сложными типами данных, таких как точки, линии, многоугольники.

Производители СУБД ответили на это двояко: для хранения пространственной информации предлагаются либо большеобъемные двоичные объекты (binary large objects- BLOB), либо гибридные системы, в которых пространственные атрибуты сохраняются в файлах операционной системы средствами ГИС. Оба этих предложения крайне неэффективны.

- в первом случае все возможности SQL блокируются, поскольку он не предназначен для обработки двоичных BLOB-файлов;
- в гибридных системах пространственные атрибуты хранятся в виде отдельных файлов операционной системы и потому не могут пользоваться преимуществами традиционных служб БД (управление запросами, параллельной работой, поддержкой индексирования и т.д.).

Основное влияние на расширение возможностей СУБД в части поддержки сложных пространственных объектов оказали объектно-ориентированные системы. Они работают с версией языка SQL, которая носит название SQL3/SQL99 и поддерживает *пользовательские типы данных* (как в языках

Java и C++). Это означает, что типы данных могут не только выбираться из регламентированного перечня, но и создаваться новые, ещё не существующие типы по усмотрению пользователя системы. Такие типы обозначаются аббревиатурой АТД от Абстрактные Типы Данных.

АТД- это пользовательский тип и связанные с ним функции. Например, если мы располагаем информацией о лицензионных участках, хранящихся в БД в виде многоугольников, то полезным АТД могла бы стать комбинация типа "многоугольник" и нескольких, связанных с ним функций(методов). Это может быть функции проверки смежности *adjacent*. Функция *adjacent* применяется по отношению к *лицензионным участкам* для того, чтобы определить, имеют ли они общие границы. Понятие "абстрактный" используется в АТД потому, что конечному пользователю нет необходимости знать детали реализации функций, связанных с этим типом. Всё, что должны знать конечные пользователи- это средства взаимодействия, то есть доступные функции и типы данных для входных параметров и конечного результата.

Стандарт OGIS как расширение языка SQL

Консорциум OGIS был создан ведущими производителями программного обеспечения с целью выработки промышленного стандарта, регламентирующего взаимодействие с ГИС. Модель пространственных данных OGIS может быть встроена в большое количество различных языков программирования, но мы остановимся на встраивании этой модели в SQL. [6]

Стандарт OGIS основан на модели геометрических данных, состоящей из базового класса GEOMETRY. Этот абстрактный класс задаёт пространственную систему координат, применимую для всех производных данного класса. Существует четыре основных класса, порождённых от предка GEOMETRY,- это Point, Curve, Surface и GeometryCollection.

С каждым из этих классов связан специфический набор операций (их всего 3 типа):

- 1) базовые операции, применимые ко всем геометрическим типам данных. Например, *SpatialReference* возвращает базовую систему координат, в которой описана геометрия объекта. К числу распространённых систем координат относится широко известная система широт и долгот, а также *Universal Transfer Mercator (UTM)*;
- 2) операции, выявляющие топологические отношения между пространственными объектами. Например, операция *Overlap* проверяет, имеют ли внутренние области двух объектов непустое множество пересечений;
- 3) общие операции пространственного анализа. Например, операция *Distance* возвращает кратчайшее расстояние между двумя пространственными объектами.

Наиболее важные операции, а также их назначение приведены в таблице 18.

Таблица 18– Операции стандарта OGIS [OGIS, 1999]

Опера- ции	Наименования	Результат
1	2	3
Базовые функции	SpatialReference()	Возвращает базовую систему координат.
	Envelope()	Возвращает минимальный описывающий прямоугольник геометрии.
	Export()	Возвращает альтернативное представление геометрии.
	IsEmpty()	Возвращает True, если геометрическое множество является пустым.
Базовые функции	IsSimple()	Возвращает True, если геометрия является простой (без самопересечений)
	Boundary()	Возвращает границы геометрии.
Топологические операции и операции над множествами	Equal	Возвращает True, если внутренние области и границы обеих геометрий равны.
	Disjoint	Возвращает True, если границы и внутренняя область не пересекаются.
	Intersect	Возвращает True, если геометрии имеют общие элементы.
	Touch	Возвращает True, если границы двух геометрий пересекаются, а внутренние области- нет.
	Cross	Возвращает True, если внутренняя область пересекается.
	Within	Возвращает True, если внутренняя область одной геометрии не пересекается с внешней областью другой геометрии.
	Contains	Проверяет, содержит ли одна геометрия другую.
	Overlap	Возвращает True, если внутренние области двух геометрий имеют непустое пересечение.

Пространственный анализ	Distance	Возвращает кратчайшее расстояние между двумя геометриями.
	Buffer	Возвращает геометрию, содержащую все точки, лежащие на расстоянии не меньшем указанного от данной геометрии.
	ConvexHull	Возвращает наименьшее выпуклое геометрическое множество, заключающую в себе данную геометрию.
	Intersection	Возвращает геометрическое пересечение двух геометрий.
	Union	Возвращает геометрическое объединение двух геометрий.
	Difference	Возвращает фрагмент геометрии, который не пересекается с другой геометрией.
	SymmDiff	Возвращает фрагменты двух геометрий, которые не пересекаются друг с другом.

Используя данные из трёх, ранее заведённых таблиц: **City**, **Country**, **River** и операции стандарта OGIS сформируем несколько SQL-запросов, иллюстрирующим отношения между пространственными сущностями в них.

1 Отыскать в таблице **Country** названия всех стран, которые являются соседями США.

```
SELECT      C1.Name AS "Соседи США"
FROM        Country C1, Country C2
WHERE       Touch(C1.Shape, C2.Shape)=1 AND C2.Name= 'США'
```

Здесь предикат **Touch** проверяет, являются ли два любых геометрических объекта смежными без наложения.

2 Для всех рек, перечисленных в таблице **River**, определить страны, по которым они протекают.

```
SELECT      R.Name, C.Name
FROM        River R, Country C
WHERE       Cross(R.Shape, C.Shape)=1
```

3 Какой город из таблицы **City** ближе других расположен ко всем рекам, перечисленным в таблице **River**?

```
SELECT      C1.Name, R1.Name
FROM        City C1, River R1
WHERE       Distance(C1.Shape, R1.Shape) <
            ALL (SELECT      Distance(C2.Shape, R1.Shape) <
                FROM        City C2
                WHERE       C1.Name <> C2.Name
            )
```

В подзапросе вычисляется расстояние между реками и городами, кроме города, с которым непосредственно производится сравнение. Оператор ALL-

групповой оператор множеств. Другими словами третья строка данного запроса на русском языке звучит так: «Где расстояние между городом и рекой меньше всех расстояний между всеми городами и реками»

4 Река Св. Лаврентия может снабжать водой города, удалённые от неё не далее чем на 300 км. Составить список городов, которые могут получать воду из реки Св. Лаврентия.

```
SELECT      C1.Name
FROM        City C1, River R
WHERE       Overlap(C1.Shape, Buffer(R.Shape, 300))=1 AND
           R.Name= 'Св. Лаврентия'
```

5 Составить список длин рек в пределах каждой страны, по которой они протекают.

```
SELECT      R.Name, C.Name, Length(Intersection(R.Shape, C.Shape))
           AS "ДЛИНА"
FROM        River R, Country C
WHERE       Cross(R.Shape, C.Shape)=1
```

Во фразе **WHERE** находятся случаи пересечения стран реками. Для каждого такого случая вычисляется длина (**Length**) каждого такого пересечения предикатом **Intersection**.

3.1.8 Обработка и оптимизация запросов

Запросы в базах данных формулируются на таком декларативном языке как SQL. Ответственность за преобразование запроса в последовательность операторов, поддерживаемых пространственными индексами и структурами хранения, возлагается на программные компоненты баз данных. Основная задача состоит в правильной обработке запроса за минимально возможное время. В этой связи интересны приёмы, используемые для обработки и оптимизации запросов в пространственных базах данных.

Для составления большинства запросов в реляционных системах применяют фиксированный перечень основных операций. Такие базовые реляционные операции образуют строительные блоки, пригодные для создания всех сложных запросов. Обработка и оптимизация запросов делится, таким образом, на два этапа:

- проектирование и тонкая настройка алгоритмов выполнения каждого базового реляционного оператора;
- преобразование запросов высокого уровня в набор базовых реляционных операторов и оптимизация их с использованием информации, полученной на первом этапе.

Эта же технология, но с несколькими существенными отличиями может применяться в контексте пространственных баз данных. [8]

Во-первых, сфера приложения СУПБД очень широка и до сих пор не

достигнуто соглашения относительно состава "строительных блоков", т.е. пространственных операций, которые могли бы покрывать все возможные случаи.

Во-вторых, алгоритмы пространственных операций предъявляют крайне высокие требования, как к ресурсам центрального процессора (ЦП), так и к ресурсам системы ввода-вывода.

Всё это усложняет процесс проектирования запросов по сравнению с традиционными СУБД, когда общепризнанными являются предположения о том, что издержки ввода-вывода преобладают над издержками ЦП, а хороший алгоритм- такой, который минимизирует число обращений к диску.

С точки зрения обработки запросов существуют, по меньшей мере три позиции, по которым различия между пространственными и реляционными базами очевидны:

- 1) в отличие от реляционных баз данных, пространственные базы не имеют фиксированного набора операторов, которые служат в качестве строительных блоков при выполнении запросов;
- 2) пространственные базы данных имеют дело с чрезвычайно большим количеством сложных объектов. Они являются неточечными пространственными объектами и не могут сохраняться в одномерном массиве;
- 3) для проверки истинности пространственных условий требуются вычислительно сложные алгоритмы, полностью загружающие ЦП. Таким образом, предположение о преобладании издержек ввода-вывода над издержками процессора в данном случае несправедливо.

3.1.9 Пространственные операции

Пространственные операции можно разделить на четыре группы.

- 1 Операции обновления. Это стандартные операции СУБД такие как изменение, создание и т.д.
- 2 1.Операции выборки. Могут быть двух типов
 - а) точечные запросы. Дана точка, найти все пространственные объекты, в которых она находится;
 - б) запросы диапазонов или запросы областей. Для данного многоугольника найти все пространственные объекты его пересекающие. Когда многоугольник является прямоугольником, подобный запрос носит название запроса окна.
- 3 Пространственное соединение. Когда две таблицы соединяются по пространственному предикату, такое соединение называется пространственным. Одним из важных вариантов пространственного соединения и важным оператором в составе ГИС является наложение или слияние карт. Путём сочетания двух множеств пространственных объектов эта операция позволяет получить новое множество. При этом границы множества новых объектов определяются непространственными атрибутами, значения которых устанавливаются операцией наложения. Например, если опера-

ция присваивает одинаковое значение непространственного атрибута двум соседним объектам, то эти объекты сливаются. Приведу некоторые примеры пространственного предиката:

- intersect(пересекаться);
- is_enclosed_by(окружён);
- distance(расстояние);
- northwest(к северо-западу);
- adjacent(смежный);
- meets(касается);
- overlap(перекрывать).

Пример пространственного соединения: "Найти все лесонасаждения (forest) и речные поймы (flood-plains), которые перекрывают друг друга". На языке SQL этот запрос будет представлен так:

```
SELECT      FS.Name, FP.Name  
FROM        Forest FS, Flood-Plain FP  
WHERE       Overlap(FS.G, FP.G)
```

4 Пространственные агрегаты. Примером агрегата является такой: "Найти реку, ближайшую к палаточному лагерю". Это один из вариантов поиска ближайших соседей.

3.1.10 Двухэтапная обработка запроса с пространственными операциями

В обработке пространственных запросов задействованы сложные типы данных; например, точное представление границы озера может потребовать указания тысяч вершин.

Эффективная обработка таких больших объектов обычно требует выполнения пространственных операций в соответствии с двухэтапным алгоритмом, включающим этап фильтрации и этап очистки (рисунок 14).

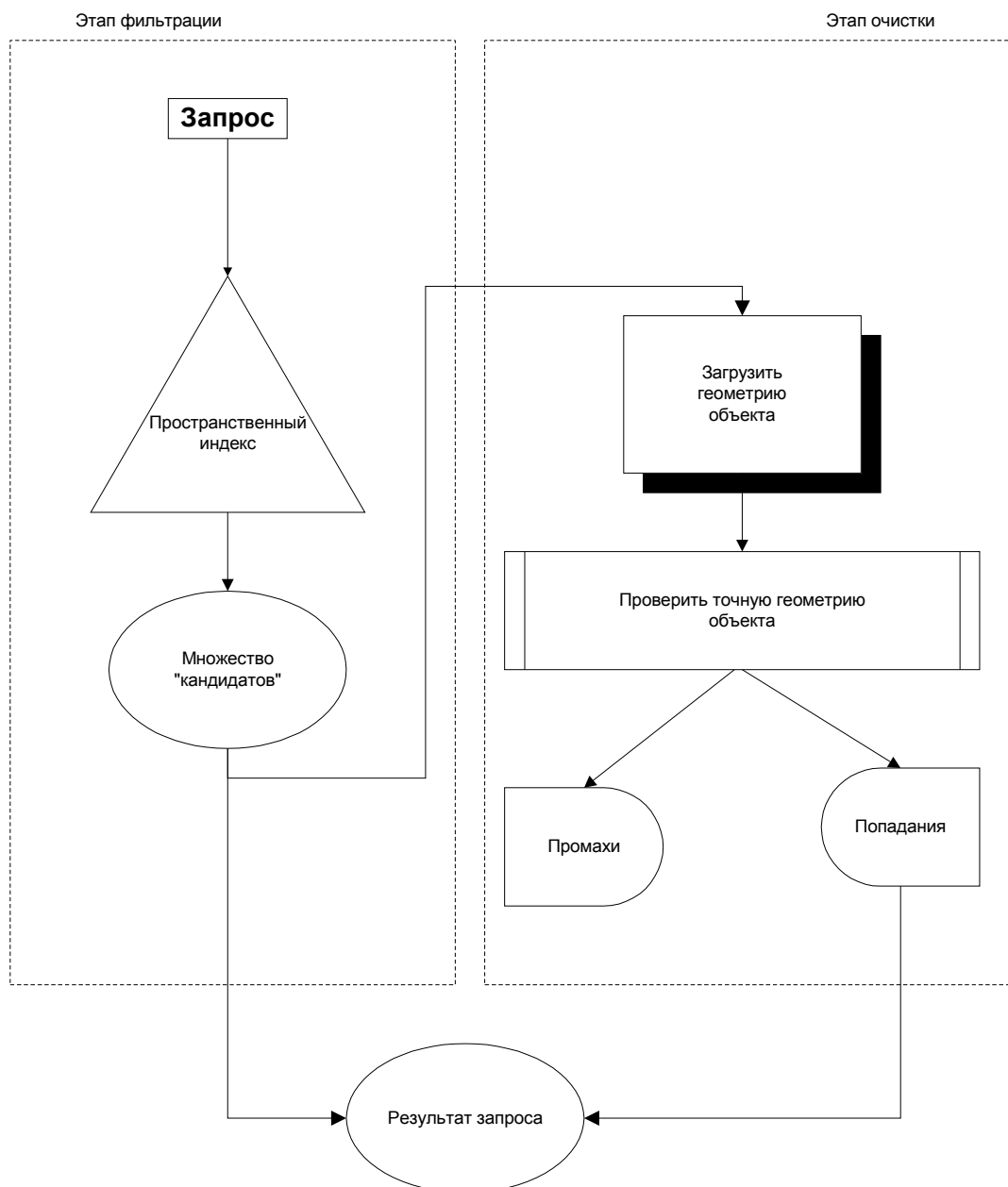


Рисунок 14– Двухэтапная обработка пространственных запросов

Этап фильтрации.

На этом этапе для представления пространственных объектов используются более простые приближённые представления. Чаще всего это минимальные ограничивающие прямоугольники. Рассмотрим такой SQL-запрос: "Найти все реки, в пойме которых стоит ЦЕРКОВЬ"

```

SELECT      River.Name
FROM        River
WHERE       Overlap(River.Flood-plain, :ЦЕРКОВЬ)

```

Теперь, приближённо представим каждую речную пойму минимальным ограничивающим прямоугольником. Очевидно, что на выяснение принадлежности конкретной точки очередному прямоугольнику потребуется меньше ресурсов, чем на выяснение того, лежит ли точка в пределах внутренней области неправильного многоугольника, т.е. точного образа поймы.

Результат проверки этого приводит к созданию множества т.н. "кандидатов".

Для упрощения системы оптимизации запроса приближённым вариантом можно заменить даже пространственный предикат. Например, часто вместо сложного Touch(касаться) используется более простой Overlap. Самое главное при таком упрощении не потерять нужные строки из окончательного набора.

Этап очистки.

Здесь изучаются точные геометрические очертания каждого из множества "кандидатов" и вычисляются точные значения неупрощённых предикатов. Обычно это требует использования алгоритма, требовательного к ресурсам процессора. Этап очистки иногда может происходить за пределами СУПБД и выполняться средствами прикладной программы, например ГИС.

3.2 Введение в добычу пространственных данных

В последнее время в сфере геоинформационных технологий возникла новая быстроразвивающаяся отрасль- добыча данных (*data mining*).

Добыча данных- это процесс поиска нетривиальных и, возможно, полезных информационных шаблонов в больших базах данных. Само наименование этого процесса- метафора *добычи* призвана создать впечатление, что искомые шаблоны- это самородки драгоценных сведений, сокрытые в огромных пластах информационного шума и ожидающие своего звёздного часа.

Деловой мир быстро взял методы добычи данных на вооружение, усмотрев в ней способ, поднять свои доходы.

Надо сказать, что методология добычи пространственных данных (ПД) и требования к ним существенно отличаются от таковых для обычных непространственных материалов. Целостный процесс добычи данных- это сочетание множества более мелких процессов, которые достойны изучения сами по себе. К ним относятся:

- 1) извлечение и очистка данных;
- 2) выделение информационных признаков;
- 3) проектирование и настройка алгоритмов;
- 4) собственно обработка данных;
- 5) анализ выходной информации.

Для пространственных данных также очень важен вопрос масштаба, то есть уровня агрегирования, на котором этот анализ данных производится. В частности, хорошо известен тот факт, что при пространственном анализе

даже идентичные эксперименты могут привести к противоречивым результатам, если они проводятся на различных уровнях детализации.

Существует ещё одно более тонкое отличие- между добычей пространственных данных и пространственным анализом. Как следует из названия, анализ ПД покрывает широкий спектр методов, имеющих дело, как с пространственными, так и с прочими характеристиками пространственных объектов. Методы добычи ПД базируются буквально на всех научных направлениях и представлениях, способствующих выделению скрытого пространственного смысла из произвольных информационных массивов. Это приёмы пространственной статистики, геостатистики, того же пространственного анализа, обучения машин и технологий СУПБД.

3.2.1 Поиск информационных шаблонов

В основе добычи ПД лежит поиск т.н. информационных шаблонов- своего рода гипотез или предположений, нетривиальных по своей природе, но практически полезных.

Шаблон может представлять собой любую отчётную статистику, как среднее, медиана, среднеквадратическое отклонение по выборкам. Это может быть и простое правило, например, "Недвижимость у моря в среднем на 40% дороже, чем на территориях удалённых от побережья". Важно, чтобы подобные утверждения были истинными и легко проверяемыми.

Пример нетривиального шаблона, который на западе вошёл в учебники по маркетингу был обнаружен при изучении транзакций БД национальной розничной торговли США. Он гласит: *"Люди, покупающие в послеобеденное время детские пелёнки, также склонны приобретать и пиво!"*. Это стало неожиданной и необычной находкой для предпринимателей, которые немедленно произвели перестановки на полках магазинов и извлекают выгоду от этого до сих пор.

Этот случай подтверждает мысль о том, что добыча данных связана с выдвижением множества любых самых невероятных гипотез, с тем, чтобы последние могли быть верифицированы стандартными статистическими средствами. Например, если магазин располагает тысячей наименований товаров, то для отыскания двух изделий, между которыми имеется значимая корреляционная взаимосвязь, потребует полумиллиона корреляционных тестов. Их прямое выполнение потребует существенных вычислительных ресурсов и расходов.

Нет нужды говорить, что разработка приёмов добычи данных в геологической отрасли имеет первостепенное значение. Они, связаны именно с возможностью быстрого выполнения автоматического поиска локальных и потенциально очень полезных шаблонов при помощи определённых компьютерных алгоритмов.

3.2.2 Процесс добычи данных

Типичный ход процесса добычи данных можно описать следующим образом (рисунок 15). Предположим, сотрудник городского полицейского управления хочет найти объяснение необычно высокого уровня преступности, наблюдаемого в этом году в городе. В данном случае он выступает в качестве эксперта. Он имеет доступ к базе данных, но отыскать нужную закономерность самостоятельно он не в состоянии. Для решения вопроса приглашается аналитик, специализирующийся именно на добыче данных.

Обсуждение вопроса подводит эксперта и аналитика к соглашению относительно постановки задачи, решение которой может дать удовлетворительный ответ на проблему.

Главная задача аналитика состоит в том, чтобы определить какой способ или комбинацию способов надо применить для решения задачи. Это может быть кластерный или факторный анализ или множественная регрессия предварительно выделенных показателей.

В идеале алгоритм добычи данных должен быть спроектирован так, чтобы напрямую обращаться к базе данных при помощи SQL (OGIS SQL для пространственных баз данных). Однако, обычно преобразование БД в формат, совместимый с алгоритмом задачи требует множества этапов и времени.

Выбор способа решения задачи и соответствующего алгоритма также является недетерминированным, итерационным процессом. Например, большинство алгоритмов требуют настройки пользовательских параметров, и во многих случаях нельзя заранее судить о том, каковы те правильные параметры, которые следует исследовать в той или иной БД. Выходом алгоритма добычи данных обычно является *гипотеза*, которая может иметь форму параметров модели (в случае регрессии), правил и т.д. Иначе говоря, выход представляет собой потенциальный шаблон.

Следующий этап- верификация, очистка и визуализация шаблона. При работе с пространственными данными эта часть процесса обычно решается с помощью ГИС.

Заключительный этап процесса добычи данных состоит в интерпретации шаблона и, когда это возможно, выдаче рекомендаций для адекватного проблеме действия.

Например, вывод может гласить, что высокий уровень преступности отражает спад в экономическом положении города и тогда представитель полиции может напрямую адресовать этот вывод чиновникам городской администрации. Или же результаты добычи данных могут указывать на то, что высокий уровень нарушений закона является следствием исключительно высокой активности в нескольких горячих точках, связанных с конкретными транспортными магистралями. В этом случае есть надежда на выявление преступной группировки.

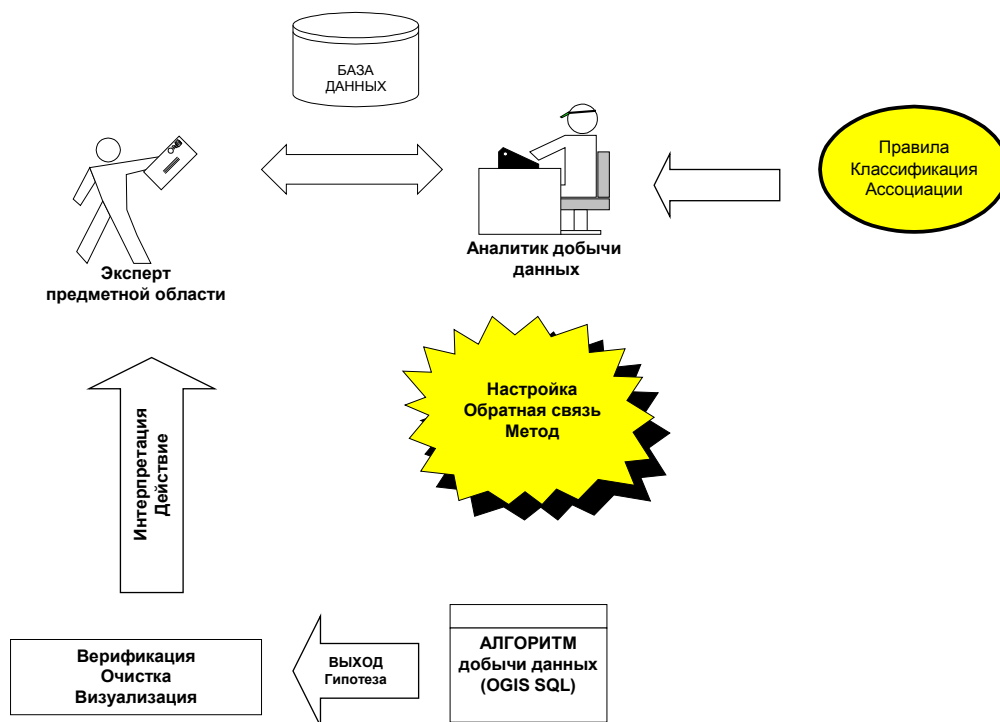


Рисунок 15– Процесс добычи данных

В целом, процесс добычи данных очень напоминает сбор статистики. Однако, более точно её следует охарактеризовать как шаг *фильтрации*, предшествующий применению точных статистических инструментов. Роль этого шага заключается в том, чтобы буквально перепахать большое количество данных и получить ряд нетривиальных гипотез, которые впоследствии можно проверить средствами обычной статистики. Такая стратегия аналогична фильтрации данных с помощью минимальных ограничивающих прямоугольников на предварительном этапе конструирования SQL запроса. При этом также возникает множество пространственных "кандидатов", число которых сократится после верификации шаблонов.

3.2.3 Основной закон географии

Огромную роль в разработке алгоритмов добычи данных играют сами свойства этих данных, связанные с правильными предположениями об их организации. Рассмотрим, к примеру, задачу поиска скрытых контуров на растровых изображениях (космоснимках и спектрональных изображениях). Допустим, мы имеем дело с изображением размером 4×4 пикселя, каждый элемент которого мы намерены отнести к одному из двух классов: *чёрное* или *белое* (рисунок 16). Самый грубый "силовой" алгоритм просмотра предполагает считывание и обработку 2^{16} различных их комбинаций. Это ровно 65536 вариантов.

Если же мы будем утверждать, т.е. определим шаблон, что большинство блоков снимка размерностью 2×2 могут принимать значения

одного из двух классов, т.е. быть полностью чёрными или белыми, то количество комбинаций сокращается до 2^4 .



Рисунок 16– Результаты поиска с использованием алгоритма добычи данных

Разумеется, это довольно сильное ограничение, которое, однако, нельзя отвергать полностью. Как правило, большинство соседних элементов растровых изображений, особенно при высоком разрешении принадлежат к одному и тому же классу благодаря пространственной автокорреляции, а также в силу крупных размеров многих деталей снимков.

Сказанное подчёркивает различие между добычей пространственной и классической информации. Можно утверждать, что оно соответствует разнице между классической и пространственной статистикой.

Одно из основополагающих предположений, оказывающих решающее влияние на статистический анализ в целом, состоит в том, что элементы выборки данных получены в ходе независимых экспериментов. Например, как это имеет место при подбрасывании монеты или метанием игральной кости.

Когда же дело доходит до анализа пространственных данных, предположении о независимости элементов выборки, вообще говоря, становится ложным. На самом деле, пространственные данные проявляют тенденцию к высокой степени автокорреляции.

Например, люди с похожими чертами, родом деятельности и происхождением часто собираются вместе в одних и тех же местах. Сходство проявляют и экономические показатели регионов. Смена природных ресурсов, дикой природы и температуры происходит в пространстве постепенно.

Фактически свойство похожих вещей и явлений объединяться в пространстве настолько фундаментально, что географы предали ему статус первого закона географии: ***Всё связано со всем, однако близкие вещи и явления связаны друг с другом сильнее, чем далёкие*** [7].

В пространственной статистике- области статистики, посвящённой анализу пространственных данных это их свойство *называется пространственной автокорреляцией*.

Итак, добыча данных- это процесс автоматизации поиска потенциально полезных шаблонов. Однако, отдельные хорошо известные примеры того, что мы теперь называем добычей пространственных данных, имели место задолго до изобретения компьютеров:

- 1) в 1855 году, когда по Лондону распространялась азиатская холера, некий эпидемиолог нанёс на карту все места вспышек заболевания и обнару-

жил, что они образуют группу, центр которой приходится на водокачку. Когда водокачка была отключена властями, холера пошла на спад. Позднее учёные подтвердили, что данное заболевание передаётся с водой.

- 2) теория Гондваны, согласно которой все материки образовывали один участок Земли, была выдвинута неким Р.Ленцом. Он обнаружил (по картам), что все континенты соответствуют друг другу по очертаниям и могли бы соединиться воедино подобно одной гигантской мозаике. В настоящее время теория дрейфа континентов является общепризнанной.

Основная задача добычи пространственных данных состоит в автоматизации обнаружения именно таких зависимостей, которые могут быть впоследствии изучены и проверены специалистами.

3.2.4 Показатели пространственной формы и автокорреляции

Как нам уже известно, пространство можно рассматривать как непрерывное и дискретное. Непрерывность пространства отличает большинство наборов данных, собираемых науками о земле.

- непрерывную форму данных часто трудно представить, поскольку непрерывное пространство характеризуется бесконечно большим множеством значений;
- в дискретном пространстве, количество значений перечисляемо, т.е. конечно;
- в непрерывном пространстве местоположение задаётся координатами, а в дискретном пространстве оно определяется как объект. Налицо явная параллель с толкованием векторных и растровых данных;

Существуют два методологических термина обозначающие основные статистические подходы к исследованию этих пространств: геостатистика и пространственная статистика. Так для изучения географической информации применяется пространственная статистика, т.е оно обычно рассматривается как дискретное.

Термин геостатистика чаще ассоциируется с непрерывным полевым пространством.

Пространственную форму набора (облака) точек характеризуют центрированность, дисперсия и очертания.

Центральным средним называется усреднённое значение, рассчитываемое как среднее по координате X и среднее по координате Y . Центральное среднее также известно как центр тяжести пространственного распределения.

Часто используется взвешенное центральное среднее, которое служит более надёжной мерой в некоторых пространственных приложениях. Например, для вычисления центра рудного поля, популяции животных и т.п.

Взвешенное центральное среднее вычисляется как отношение суммы

координат всех точек, умноженных на вес, к сумме всех весов.

Использование центрированных мер применяется главным образом в двух случаях- свёртки информации и поиска оптимального месторасположения:

Свёртка информации.

Она позволяет представить пространство не в многообразии форм объектов, а в плане размещения в нём их центроидов. Разумеется такой подход требует значительно меньших объёмов памяти и затрат на её обслуживание.

Поиск оптимального месторасположения

Иногда центрирование имеет практическое значение для отыскания самого эффективного месторасположения планируемой деятельности. Например, оптовую базу следует размещать в центре района, чтобы минимизировать транспортные расходы.

Дисперсия— это мера рассеяния геометрического множества относительно своего центра. Часто используемыми показателями дисперсии и изменчивости являются:

- диапазон распределения;
- среднеквадратическое отклонение;
- вариация и коэффициент вариации.

Все меры дисперсии географических распределений обычно вычисляются как некая сумма отношений весов географических объектов, к значениям показателей близости между ними.

Очертания многомерны, и единого критерия, способного охватить все измерения очертаний не существует. Многие описания очертаний основаны на сравнении периметра фигуры с длиной окружности той же площади.

Коэффициент Морана

Как говорилось выше, в задачах пространственного анализа существует необычное явление, которое необходимо учитывать и измерять. Это автокорреляция [9]. Она проявляется в том, что изменения в одной точке влияют на события в соседних точках.

Такая пространственная зависимость может быть определена как *"склонность переменной принимать близкие значения, функционально зависящие от расстояния между областями пространства, где производятся измерения"*.

Мерой пространственной зависимости служит пространственная автокорреляция. Например:

- значения температуры в двух точках, расположенных близко друг от

друга, будут схожи;

- народы соседних стран говорят на близких языках;
- плодородность почв, обычно изменяется постепенно. Так же изменяются уровни осадков и атмосферное давление.

В статистике существуют показатели, предназначенные для количественной оценки этой взаимозависимости. Один из таких показателей носит название I-коэффициента Морана, предназначенный для определения т.н. "матриц смежности" (рисунок 17).

Для данной переменной $x = \{x_1, \dots, x_n\}$, значение которой измерено в n точках I-коэффициента Морана есть отношение

$$I = \frac{zWz^t}{zz^t},$$

где

$$z = \{x_1 - x_{cp}, \dots, x_n - x_{cp}\},$$

x_{cp} - среднее значение x

W- нормализованная по строкам матрица смежности размерности $n \times n$

z^t - транспонированная матрица z (рисунок 17)

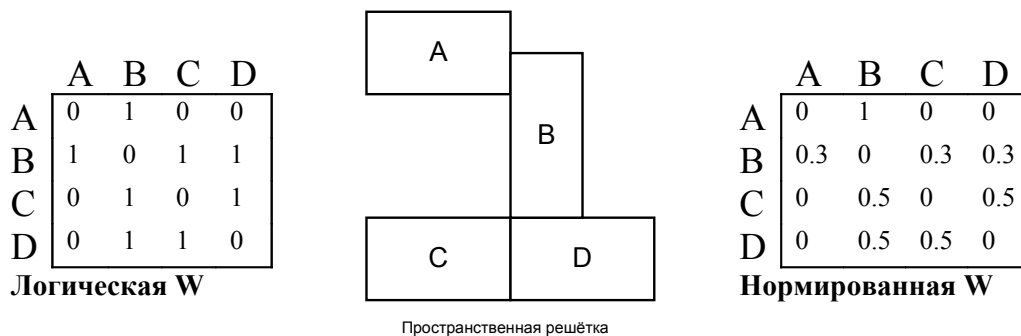


Рисунок 17 – Матрицы смежности коэффициента Морана

3.2.5 Приёмы добычи данных

Уже существуют и постоянно возникают новые методы добычи данных. Однако все их можно сгруппировать по трём непротиворечивым категориям: **классификация, кластеризация и правила ассоциации.**

Предсказание местоположения и тематическая классификация

Целью классификации в широком смысле этого слова является оценка значений одного атрибута отношения (таблицы) на основе значений других его атрибутов. Очень многие проблемы можно сформулировать как задачи классификации. Например, выбор оптимального по тем или иным критериям места для реализации любого проекта или явления полностью ей соответствует. По этой причине классификация и предсказание местоположения в контексте ГИС рассматриваются почти как синонимы. Вот некоторые типичные примеры классификации:

Проблемой предсказания местоположения можно считать прогноз того, в каких "горячих точках" следует ожидать роста преступности.

Розничные торговцы, выбирая место будущего магазина по сути дела также прогнозируют свой успех или неудачи.

Целью тематической классификации является распределение по классам пикселей спутниковых изображений. Эти классы могут быть обозначены как вода, городские застройки, леса и т.д. Значения пикселей возникают на основе т.н. "спектральных сигнатур", регистрируемых приёмником на борту спутников. Здесь задача тематической классификации тесно связана с анализом пространства, поскольку в большинстве случаев соседние элементы принадлежат к одному и тому же классу.

Существует большое количество доступных методов решения задачи классификации. Это методы максимального правдоподобия, особенно часто применяющиеся при дистанционном зондировании. Классическим приёмом добычи данных является метод классификации основанный на линейной регрессии. С помощью матрицы смежности он легко адаптируется под пространственные данные с учётом их автокорреляции.

Следует иметь в виду, что задачи классификации работают, если нам заранее известно число классов и их содержание. Другими словами классификация возможна при наличии некоторых предварительных гипотез о пространственных данных и заключается в их распределении по объектам. По этой причине процесс добычи данных на её основе называется "предсказание по эталонам"

Кластеризация

С точки зрения нашего предмета кластеризацией можно назвать процесс обнаружения "групп" или кластеров в больших базах данных.

В отличие от классификации при кластеризации не используется какая-либо априорная информация о количестве групп и их характеристики. Таким образом, в кластеризации нет понятия тренировочных или обучающих данных, поэтому её называют также "обучением без учителя".

Кластеры формируются на основе критерия "сходства", используемого для определения связей между каждой парой кортежей (строк) в базе данных. Строки, которые оказываются сходными, обычно объединяются в группу, и такая группа помечается.

Например, элементы спутникового изображения часто подвергаются кластеризации на основе спектральной сигнатуры. Благодаря этому снимок, полученный путём удалённого зондирования, может быть быстро сегментирован при минимальном вмешательстве человека. Разумеется, эксперт в предметной области должен проверить, верифицировать и, возможно подвергнуть кластеры очистке. В чём же смысловое отличие кластеризации от тематической классификации?

При классификации спектральные сигнатуры пикселей изображения распределяются по типам эталонных фрагментов снимка. Ими могут быть за-

ранее известные участки, занятые лесами, водоёмами, застройками и т.д. Наиболее похожие на эталоны области значений спектральных сигнатур причисляются к этим классам или, другими словами, им предсказывается соответствующее местоположение.

При кластеризации спутникового изображения автоматически группируются в кластеры. Большинство из них, учитывая явление автокорреляции будет соответствовать природным пространственным феноменам, однако каким- на это кластеризация не даёт прямого ответа. Тут уже необходимо сравнение и верификация с эталонными областями. При кластеризации вполне могут возникнуть кластеры, не имеющие на данный момент ясного смыслового толкования. Такие случаи с точки зрения методологии добычи данных описываются как пространственные выбросы, которые могут привести к обнаружению необычных сведений. Впервые это явление было писано в 1994 году /Barnett and Lewis/. Результаты их наблюдений настолько отличаются от прочих, что побуждают подозрения в фальсификации, несоблюдении условий опыта или других неожиданных факторов. Среди примеров удачной расшифровки пространственных выбросов можно привести случаи раскрытия мошенничеств с кредитными картами, контроль нарушений при голосовании, анализ спортивных достижений и роль допинга. Это побочное следствие кластеризации разумеется является важным подспорьем для добычи данных, которая вся посвящена поиску нетривиальных объяснений обычных явлений и событий.

Кластеризация- приём, который очень хорошо известен в статистике, но для целей добычи данных применяются специальные подходы, позволяющие обрабатывать массивы практически любого масштаба. Нельзя забывать, что размерность СУПБД определяется произведением числа записей на количество атрибутов в каждой из них. При этом большую роль, наряду с учётом громадных объёмов обрабатываемых информационных массивов играет и сами типы данных, которых в классической статистике принято избегать. Это двоичные, порядковые, категориальные, сеточные и другие типы данных, решающим образом влияющие на выбор применяемого алгоритма КА.

Задачу кластеризации удобно рассматривать в рамках многомерного пространства атрибутов. При наличии n объектов данных, описанных m переменными, каждый объект может быть представлен как точка в многомерном пространстве. ***При этом кластеризация сводится к обнаружению и определению высокоплотных групп из множества неравномерно распределённых точек.***

Поиском возможных групп в многомерном пространстве управляет соответствующим образом подобранный критерий сходства.

Например, по данным переписи населения можно выполнить кластеризацию субъектов Российской Федерации, скажем по четырём признакам:

- уровню безработицы;
- численности населения;

- доходу на душу населения;
- продолжительности жизни.

Области с близкими значениями этих атрибутов будут сгруппированы или объединены в кластеры. Вероятно, причины этой близости регионов могут быть рационально истолкованы.

Категории алгоритмов кластеризации

Кластерный анализ (КА)- это один из методов статистического анализа наиболее часто реализуемый в различных предметных областях. Существуют четыре крупных категории КА имеющих свои реализации в сфере пространственного анализа и добычи данных:

- 1) *Иерархические методы КА.* Начинают свою работу с того, что помещают все объекты в один-единственный кластер, а затем начинают его разбиение до тех пор, пока не будет достигнут критерий останова. Если он не задан, то разбиение будет продолжаться до тех пор, пока каждый объект не организует свой кластер. Результатом работы алгоритма является дерево кластеров, называемое *дендрограммой*. Её можно разрезать на различных уровнях и получать соответствующие кластеры, природа которых определяется соответственно уровню абстракции или гиперрастоянию между объектами пространственной выборки.
- 2) *Распределяющие алгоритмы кластеризации.* В отличие от предыдущих, изначально помещают каждый объект в отдельный кластер. После этого осуществляется циклическая перестановка (перераспределение) объектов между кластерами до тех пор, пока не будет выполнен критерий останова. При этом чаще других используется такая критериальная функция как квадратичная ошибка. В этой связи распределяющие алгоритмы стремятся к обнаружению кластеров *сферической формы*.
- 3) *Плотностные алгоритмы кластеризации.* Они предназначены для поиска кластеров, в которых плотность пространственного распределения объектов существенно выше средней. В этих алгоритмах кластеры считаются плотными сгустками объектов в пространстве данных.
- 4) *Алгоритмы кластеризации на основе решёток.* Эти алгоритмы прежде всего выполняют квантование пространства кластеризации на конечное число ячеек, а затем производят требуемые операции в квантованном пространстве. Ячейки, число объектов в которых превышает некоторое пороговое значение, рассматриваются как плотные. Плотные ячейки кластеризуются, т.е. ячейка подменяет единичный объект. Алгоритмы кластеризации на основе решёток разрабатывались для анализа больших наборов пространственных данных.

Правила ассоциации

Правила ассоциации- это шаблоны $X \rightarrow Y$. Один из наиболее известных

шаблонов добычи данных, *Пелёнки* \rightarrow *Пиво* представляют собой типичный пример правила ассоциации. Правила ассоциации являются слабой формой корреляции, не отражающей отрицательной зависимости. Например, правило- тот кто покупает валокордин, вряд ли купит кроссовки:

Áàëíéíðäèí \xrightarrow{neg} *Éðíññíáèè*

может быть истинным, но не считается правилом ассоциации. На языке теории вероятности $X \rightarrow Y$ есть выражение условной вероятности $P(Y/X)$.

Одной из самых известных попыток обобщения правил ассоциации для наборов пространственных данных являются доказательства **правил совместного размещения**.

Эти правила демонстрируют подход к отысканию правил ассоциации, при котором понятие транзакции заменяется понятием окрестности. Это необходимо в связи с тем, что транзакция является дискретным, конечным этапом запроса атрибутивных данных, в то время как пространство, куда вложены географические данные непрерывно.

Процесс поиска шаблонов совместного размещения позволяет обнаружить часто встречающиеся совместно размещённые подмножества разных событий. Например, анализ среды обитания животных и растений может выявить совместное размещение видов по модели хищник-жертва, симбиотических видов, а также места пожаров и источников возгорания.

В заключение следует сказать, что значение пространственных баз данных для решения геологических задач трудно переоценить. Это в равной степени относится к фундаментальным и прикладным геологическим исследованиям. Изложенные в данной работе концепции и примеры свидетельствуют, что пространственные БД переросли уровень информационно-поисковых систем и обладают явными признаками искусственного интеллекта. Результаты применения их в геологической практике в большей степени зависят от таланта и подготовки пользователей (в нашем случае- горных инженеров), чем от возможностей программного инструментария

4 Комплексы геологического моделирования

Обзор вопроса

Изложенные в пособии принципы и приёмы компьютерного моделирования реализованы во многих программных комплексах. Наиболее выдающиеся из них созданы в сфере нефтегазовой индустрии. Причины этого связаны, во-первых, с той огромной ролью, которую играют ископаемые углеводороды в бюджетах стран –экспортёров нефти и газа и, во-вторых, с относительно предсказуемыми условиями залегания углеводородных залежей, в основном, стратиформных по своей природе. К числу таких программных продуктов принадлежат:

- уникальный по масштабу и универсальности модульный ряд комплекса ***Schlumberger-Geoquest***;
- комплекс математического моделирования нефтегазовых залежей компании ***Landmark***;
- тщательно спроектированный и полностью интегрированный с Базой Геологических Данных комплекс ***TIGRESS*** компании PGS;
- впечатляющими возможностями трёхмерной компьютерной графики пакет прикладных программ (ППП) ***IRAP RMS***.

Перечисленные программы относятся к разряду супердорогих и реализованы на основе стандарта POSC. Они используются для создания постоянно-действующих геолого-технологических моделей углеводородных залежей и автоматизированного управления разработкой месторождений нефти и газа. Бесперебойная и продуктивная эксплуатация данных программ требует довольно мощных аппаратных средств, значительных ресурсов оперативной памяти (обычно ≥ 2 ГГБ) и высокого быстродействия процессоров.

Среди более доступных специализированных программ, пригодных для геологического моделирования выделяется комплекс ***RockWorks*** компании ***RockWare inc.***

Он удовлетворительно функционирует на рядовых IBM-совместимых компьютерах, отличается простотой и изяществом программного интерфейса. Данные для него могут храниться в электронных таблицах MS Excel, в текстовых файлах или внутренних таблицах самого комплекса, который кроме всего прочего во многих случаях поддерживает кириллические шрифты.

Несмотря на кажущуюся простоту, в ***RockWorks*** реализовано большинство алгоритмов, применяемых в дорогих ППП и, более того, разработан ряд уникальных технологий визуализации цифровых полей произвольной сложности. Пробные (*trial*) версии - ***RockWorks*** распространяются бесплатно и удовлетворительно функционируют в течении месяца. Геолог, освоивший принципы функционирования ***RockWorks*** сможет легко адаптироваться к любым комплексам моделирования, эксплуатируемым в отрасли.

В связи со сказанным, имеет смысл более подробно остановиться на

возможностях этой системы и принципах её работы, которые являются общими для всего семейства программ моделирования недр.

4.1 Функциональность и входные данные

Основная проблема исследователя- дефицит данных, который ощущается практически во всех случаях. Любая достоверная информация о недрах может быть востребована при моделировании и программные системы должны предоставлять эту возможность. Чаще всего используются три типа источников данных:

- результаты различных опробований и испытаний, проводимых вдоль стволов буровых скважин;
- данные площадного опробования;
- результаты поверхностных геофизических исследований (сейсмическая, гравиметрия и т.п.).

При этом важно, чтобы информация всех типов была пространственно и аналитически увязана между собой, а система располагала соответствующим инструментарием для этого.

Комплекс **RockWorks** специально адаптирован для работы с различными данными и функционально разделён на два крупных модульных ряда:

- **Borehole Manager** (Управляющий скважинами):
- **Geological Utilities** (Геологические утилиты).

Первый ряд (именно его возможности освещаются в данном пособии) обеспечивает хранение и обработку скважинных данных, которая включает:

- 1) ввод, просмотр и выбор данных по скважинам;
- 2) построение разных типов контурных карт и карт фактического материала;
- 3) построение разнообразных разрезов и колонок и планшетов по скважинам;
- 4) автоматизированное создание геологических и стратиграфических профилей разных типов;
- 5) создание 3D (трёхмерных моделей) цифровых полей;
- 6) создание схем трещиноватости;
- 7) построение гидрогеологических разрезов;
- 8) инструментарий конвертации файлов Географических Информационных Систем (*.shp) в формат **RockWorks**;
- 9) строительство триангуляционных схем, монтаж псевдообъёмных образов из пространственных примитивов и многое другое.

Вторая компонента комплекса- **Geological Utilities** (Геологические утилиты) включает в себя ряд полезных программ обработки геологических данных, в том числе и площадного типа. Поэтому, здесь также представлены

инструменты построения карт и разрезов. Кроме этого имеются утилиты вычисления объёмов, используемые при расчёте запасов полезных ископаемых, пакет статистики, программы обработки геодезических исследований и пересчёта координат между картографическими проекциями.

Такой гибкий подход позволяет интегрировать в одной базе данных материалы, характеризующие конкретный блок литосферы, полученные в ходе разномасштабных исследований и разных лет. Достаточно серьёзный географический инструментарий позволяет монтировать исходные данные в единый проект, свойствами которого исследователь может управлять самостоятельно, без привлечения штата программистов.

В различных программных системах, предназначенных для геологического моделирования, существует понятие т.н. *проекта*. Почти во всех случаях он понимается в 2-х аспектах, которые можно обозначить как информационный и пространственный аспекты, соответственно. В этом смысле комплекс *RockWorks* не составляет исключения

В информационном плане проекты *RockWorks* представляют собой обыкновенные папки файловой системы ОС Windows NT/2000/XP (рисунок 18).

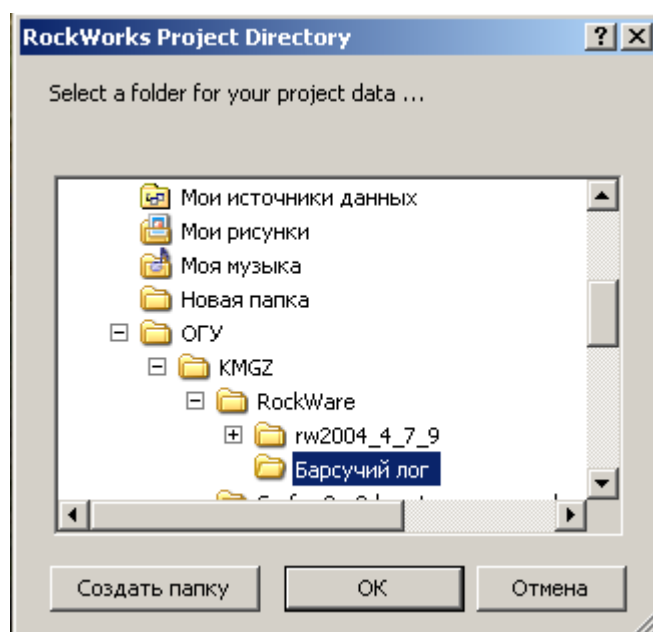


Рисунок 18- Структура проекта *RockWorks*

Путь до них, как и название определяются пользователем. В этих папках сохраняется вся информация по текущему проекту в специализированных форматах комплекса. При желании в проекте можно хранить файлы любого типа, имеющие отношение к проводимым исследованиям. Такое решение вопроса организации хранилища данных представляется довольно примитивным, но только на первый взгляд.

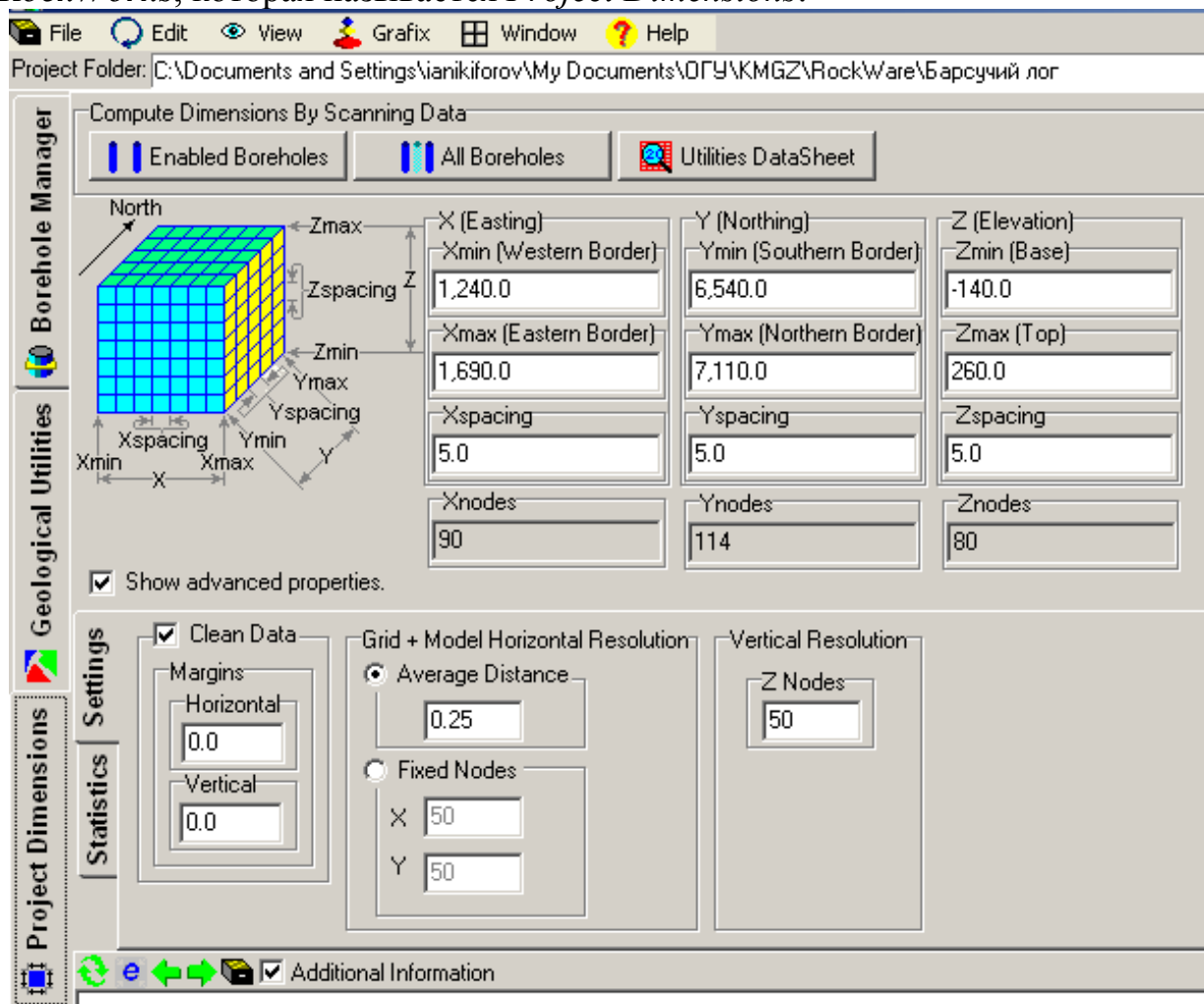
Во-первых, организовать проект можно различными способами. Можно, например, создавать сетевые проектные каталоги и использовать обычные средства защиты данных на уровне администратора локальных сетей.

Во-вторых, простота организации информационного хранилища привлекает специалистов по геологическому моделированию, которые могут

строить геологические архивы по своему усмотрению, без помощи системных администраторов. Справедливости ради надо сказать, что и в большинстве суперсистем из списка, приведённого в начале раздела, информационные хранилища организуются сходным образом. Лишь приобретение крайне дорогостоящих лицензий, на использование отраслевых баз данных, даёт возможность создания проектов средствами СУБД (см. раздел 3) с присутствующими им возможностями и сервисом.

В пространственном аспекте проекты рассматриваются как трёхмерные сетки (в дальнейшем - *гриды* от английского *grid*) в случае **RockWorks**- регулярные. В узлах этих воображаемых решёток хранятся числа, соответствующие характеристикам исследуемого цифрового поля. Обычно они возникают как результат интерполяции значений из точек наблюдения в узлы грида. Таким образом, объём компьютерной памяти, требуемой для хранения этих данных зависит от размаха грида по каждому из 3-х измерений, от размера ячеек вдоль каждой оси и от типа хранимой информации.

Рассмотрим подробнее способы управления этими важными параметрами модели. Все они сосредоточены в одной вкладке интерфейса **RockWorks**, которая называется *Project Dimensions*.



Рисунок–19- Интерфейс *Project Dimensions*

Размеры пространственных блоков, охватываемых гридами, могут настраиваться в соответствии интересам исследователя.

Если он желает создать сеть, охватывающую буквально все скважины проекта, следует нажать среднюю кнопку в верхнем ряду панели *Project Dimensions* (рисунок 19). В этом случае возникнет самый обширный грид из всех возможных вариантов. Данный выбор называется *All Boreholes* (все скважины).

Иногда геолог сосредоточен исключительно на изучении группы скважин, предварительно выделенных в проекте, т.е. строить грид для всего проекта целиком не имеет смысла. В этом случае для генерации пространственной сети, охватывающей только выделенные скважины следует нажать первую кнопку верхнего ряда. Данный выбор называется *Enabled Boreholes*.

При эксплуатации модульного ряда *Geological Utilities*, размеры грида будут автоматически настраиваться в соответствии с координатами точек наблюдений в электронной таблице информационного хранилища проекта. Комплекс сканирует точки наблюдений по координатам X , Y , Z . Во избежание неоднозначности толкования координат, их названия дублируются пояснениями: *X-Easting* (восток), *Y-Northing* (северное направление)*.

Размерности проектируемого грида отражаются в девяти числовых полях в единицах измерения, используемых в электронной таблице информационного хранилища проекта.

Самым важным моментом настройки грида являются действия, влияющие на число ячеек по каждому из трёх его пространственных измерений. Возможность этих манипуляций появляется после установки флажка *Show advanced properties*. После этого открывается дополнительная информационная панель, где можно задать число ячеек по каждому измерению проекта непосредственно (кнопка *Fixed Nodes*) или с помощью весьма изощрённого эвристического алгоритма (кнопка *Average Distance*). В общем случае увеличение числа ячеек ведёт к более тонкому и точному отображению цифровых полей и, в итоге к качественному моделированию геологической обстановки. Однако, при этом стремительно (в кубе) возрастает потребность в компьютерной памяти, а это может себе позволить не каждый. Следовательно, для продуктивной работы необходимо достичь разумного компромисса между желаемым и возможным.

Average Distance

Алгоритм рассчитывает размер ячейки грида, исходя из среднего расстояния между точками наблюдений, указываемого пользователем. Например, если это число равно 0.25, то между двумя точками возникнет четыре узла грида. Возможно, после этого производительность *RockWorks* при работе, связанной с пространственным моделированием, станет недопустимо низкой и придётся увеличить *Average Distance*. Тогда грид будет содержать в

· Как известно, в системе Гаусса-Крюгера (используемой на территории бывшего СССР) принято обратное толкование пространственных осей координат.

себе уже меньше ячеек и дефицит оперативной памяти исчезнет. В противном случае операцию по увеличению *Average Distance* придётся повторить.

Fixed Nodes

Пользователь имеет возможность непосредственно задать число ячеек по каждому пространственному измерению грида. Достоинство этого метода состоит в том, что геолог может учесть априорно известные представления об изменчивости геологического пространства на данном участке. Например, если изучается месторождение осадочного генезиса, то нет необходимости делать размеры ячеек в плоскости XY такими же мелкими как по разрезу (измерение- Z).

В том или ином виде все описанные в данном параграфе процедуры повторяются во всех развитых комплексах геологического моделирования.

4.2 Приложение Borehole Manager

Borehole Manager одна из двух главных программных компонент комплекса *RockWorks*. Она разработана для поддержки и обработки данных, распределённых вдоль стволов буровых скважин. Список всех скважин проекта расположен в самом левом окне. Скважины, помеченные птичками, являются активными, т.е. учитываются при геологических построениях и расчётах. Ввод новых скважин возможен двумя способами- интерактивным и пакетным, с помощью импорта из внешних источников данных. В простейшем случае ввод осуществляется:

- 1) выбором пункта меню ***File/New Log***. Здесь и далее такая запись означает, что из меню File надо выбрать элемент меню New Log;
- 2) вводом нового имени скважины, например "142A". Это имя становится именем файла -142A.BH, который будет храниться в папке текущего проекта;
- 3) нажатием кнопки ОК. Удаление скважин производится командой ***File/Erase Log***.

Информация по скважинам вводится в электронную таблицу, состоящую из вкладок, специфических для определённого типа данных (рисунок 20).

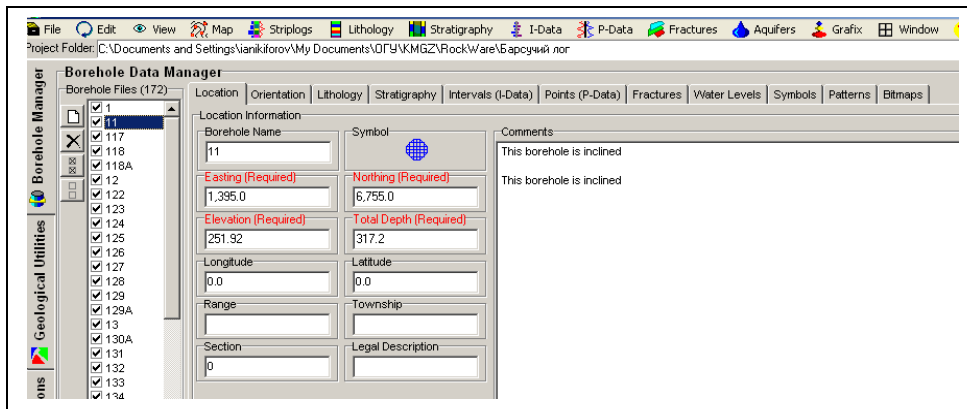


Рисунок 20 – Главная панель *Borehole Manager*

Например, вкладка *Lithology* предназначена для хранения поинтервальной информации о литологическом разрезе, а геофизические данные хранятся во вкладке *Points (P-Data)*. Это означает, что результаты данных исследований приурочены к точкам ствола скважины и однозначно привязаны по глубине, в отличие от интервальных данных, которые привязаны к интервалам глубин.

Разумеется, геолог сам может выбирать- какие данные в какой вкладке хранить. Например, данные результатов бороздового опробования керна логично накапливать во вкладке *Intervals (I-Data)*, но не возбраняется сохранять их в виде точечных данных, по аналогии с данными каротажа.

Рассмотрим организацию вкладок *Borehole Manager* более подробно.

Главная панель содержит 11 вкладок разной структуры, посредством которых осуществляется управление данными всего проекта:

№	Название	Содержание	Описание
1	<i>Location</i>	Месторасположение	Координаты заложения скважины
2	<i>Orientation</i>	Направление	Ориентация невертикального отрезка ствола
3	<i>Lithology</i>	Литология	Литологическое описание
4	<i>Stratigraphy</i>	Стратиграфия	Стратиграфическое описание
5	<i>Intervals (I-Data)</i>	Интервальные данные	Цифровые интервальные данные
6	<i>Points (P-Data)</i>	Точечные данные	Цифровые точечные данные
7	<i>Fractures</i>	Трещиноватость	Описание трещиноватости
8	<i>Water Levels</i>	Уровень воды	Зеркало подземных вод
9	<i>Symbols</i>	Условные знаки	Обозначения разных типов скважин
10	<i>Patterns</i>	Штриховки	Условные обозначения геологического разреза
11	<i>Bitmaps</i>	Растровые картинки	Файловый путь до растровых картинок и фото, иллюстриру-

№	Название	Содержание	Описание
			Ющих разрез скважины

4.2.1 Ввод данных

Вкладка Location (поля обязательные к заполнению пронумерованы)

Используется для ввода:

- 1) наименования скважины (*ID*);
- 2) координаты заложения *X (Easting)*;
- 3) координаты заложения *Y(Northing)*;
- 4) альтитуды (*Surface elevation*);
- 5) глубины скважины (*Total depth*).

Имеются поля для комментариев и дополнительных параметров привязки скважины в географических координатах (Longitude, Latitude), которые могут использоваться при пересчёте в прямоугольные координаты. Здесь также можно выбрать соответствующий условный знак для скважины.

Вкладка Orientation

Используется для ввода данных по искривлённым участкам ствола скважины:

- 1) глубина (*Depth*);
- 2) азимут (*Bearing*);
- 3) зенитный угол (*Inclination*) Особенностью является то, что значение 0° принадлежит горизонтальным фрагментам ствола; вертикальное (вниз) направление обозначается как -90° , а вертикальное (вверх) как $+90^\circ$.

Для вертикальных скважин эту вкладку можно не заполнять.

Вкладка Lithology (Литология)

Предназначена для ввода данных по литологическим разновидностям, подсечённым скважиной. Вкладка связана с библиотекой литологических типов пород и соответствующими образцами штриховок. Они поставляются вместе с программой, но могут быть созданы самостоятельно.

Данные по литологии отображаются на 2D и 3D планшетах, разрезах, блок-диаграммах, картах и пространственных моделях. Поля вкладки:

- 1) глубина до кровли (*Depth to Top*);
- 2) глубина до подошвы (*Depth to Base*);
- 3) название (*Keyword*);
- 4) описание (*Description*).

Вкладка Stratigraphy (Стратиграфия)

Предназначена для ввода данных по стратиграфической или формационной принадлежности единиц геологического разреза скважины.

Вкладка связана с пополняемой библиотекой стратиграфических наименований и соответствующими образцами условных знаков. Они поставляются вместе с программой, но могут быть созданы самостоятельно.

Данные по стратиграфии используются при структурном картировании, создании карт изопахит, разрезах, блок-диаграммах и др. Поля вкладки:

- 1) глубина до кровли (Depth to Top);
- 2) глубина до подошвы (Depth to Base);
- 3) название (Formation).

Вкладка Intervals (Интервальные данные)

Используется для хранения данных, полученных при исследованиях, охватывающих определённый интервал ствола скважины, например- геохимическом опробовании. Эти данные впоследствии могут быть отображены в виде столбчатых диаграмм на разрезах скважин, при построении вертикальных профилей, разрезов, карт и 3D-моделей.

Поля вкладки:

- 1) глубина до кровли (Depth to Top);
- 2) глубина до подошвы (Depth to Base);
- 3) наименование i-го признака, значение которого представлено в соответствующей ячейке электронной таблицы.

Вкладка Points (Точечные данные)

Используется для хранения данных, полученных при исследованиях, приуроченных к определённой точке ствола скважины, например- каротаже. Эти данные впоследствии могут быть отображены в виде кривых на разрезах скважин, при построении вертикальных профилей, разрезов, карт и 3D-моделей.

Поля вкладки:

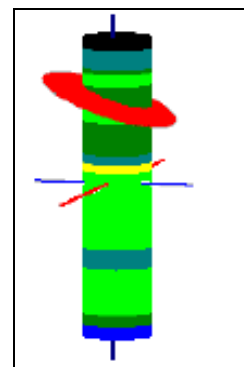
- 1) глубина (Depth);
- 2) наименование i-го признака, значение которого представлено в соответствующей ячейке электронной таблицы.

Вкладка Fractures (Трещиноватость)

Используется для хранения данных о трещиноватости горных пород, которые предполагается отобразить на каротажных диаграммах или разрезах скважин, картах и т.п.

Поля вкладки:

- 1) глубина (Depth) Вводится глубина очередной трещины (разлома, зоны и т.д.). Это значение должно быть положительным;
- 2) направление (Direction) Означает азимут падения трещины (0-360°);
- 3) угол падения трещины (Dip Angle) Значение угла должно быть приведено к следующей шкале: 0= горизонталь; минус 90= отвесно вниз; 90= вертикально вверх;
- 4) цвет (Color) Дважды щёлкнув «мышкой» по столбцу можно выбрать цвет воображаемого диска, представляющего трещину. Этот диск может быть показан на разрезах и каротажных диаграммах;
- 5) радиус (Radius) Щёлкнув на ячейке можно ввести значение для радиуса влияния трещины в единицах, принятых в данном проекте (метры, футы, дюймы ..);
- 6) раскрытие (Aperture) Вводится величина раскрытия трещины в тех же единицах что и Радиус.



Вкладка Water Levels (Уровень подземных вод)

Предназначена для ввода и хранения данных по уровням подземных вод, замеренных вдоль ствола скважины в различных водоносных горизонтах. Эта информация используется при построении разрезов, отображается на каротажных диаграммах, картах, блок-диаграммах и т.д.

Поля вкладки:

- 1) дата замера (Date);
- 2) глубина до кровли водоносного горизонта (Depth to Top);
- 3) глубина до подошвы водоносного горизонта (Depth to Base).

Вкладка Symbols (Условные обозначения для ствола скважины)

Используется для задания специальных условных обозначений для отрезков ствола скважины.

Поля вкладки:

- 1) глубина (Depth);
- 2) символ (Symbol). Условный значок для данной глубины. Выбирается из таблицы, возникающей, после щелчка по данному полю;
- 3) пояснение (Caption). Слово или поясняющая фраза для выбранного значка.

Вкладка Patterns (Геологические условные обозначения)

Используется для задания условных обозначений (обычно литология) по интервалам разреза, вскрытого данной скважиной. Эта информация отображается на разрезах, отображается на каротажных диаграммах, картах, блок-диаграммах и т.д.

Поля вкладки:

- 1) глубина до кровли (Depth to Top);
- 2) глубина до подошвы (Depth to Base);
- 3) штриховка (Pattern). Чтобы выбрать вариант штриховки разреза, необходимо дважды щёлкнуть по этому столбцу. Появится таблица штриховок. Если желаемого образца штриховки нет, то его можно создать самостоятельно;
- 4) пояснение (Caption). Слово или поясняющая фраза для выбранного значка.

Вкладка Pictures (Картинки)

Используется для хранения растровых картинок, связанных с определёнными интервалами ствола скважины. Это могут быть, например фотографии образцов керна.

Поля вкладки:

- 1) глубина до кровли (Depth to Top);
- 2) глубина до подошвы (Depth to Base);
- 3) шапка растра (Header Base). Количество строк пикселей на поясняющую информацию, записанную в верхней части растра. Оно будет пропущено при визуализации;
- 4) подпись растра (Footer Top). Количество строк пикселей на поясняющую информацию, записанную в нижней части растра. Оно будет пропущено при визуализации;
- 5) имя графического файла типа (BMP, JPG, TIF, EMF, WMF, PNG, PCX, TGA и др), в котором хранятся растровые образы.

4.2.2 Обмен данными с внешними хранилищами

Обычно сбор и подготовка геологической информации происходит вне среды *RockWorks* на компьютерах, оснащённых стандартным программным обеспечением. Это могут быть таблицы MS Excel или текстовые файлы формата ASCII. Модуль *Borehole Manager* имеет возможность экспорта данных из своих электронных таблиц в названные выше информационные структуры и импорта сведений из них в своё хранилище.

Экспорт данных

Экспорт информации из *Borehole Manager* происходит при выборе пункта меню: **File / Export / From Borehole Manager**. После этого экспортеру предлагается на выбор два доступных формата- Excel и ASCII.

При экспорте в формат Excel, создаётся книга, каждый лист которой имеет название соответствующей вкладки электронной таблицы *Borehole Manager* (рисунок 21). Можно выбрать вариант экспорта для всех или только активных (помеченных птичками) скважин проекта.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
3	1	2.5	135.2	-73														
4	1	10	163	-77														
5	1	20	169	-78														
6	1	114.5	175.6	-75.6														

Рисунок 21 – Результат экспорта Borehole Manager в формат Excel

При экспорте в формат ASCII создаётся или один текстовый файл, в котором содержится вся информация из вкладок *Borehole Manager* или несколько, по одному на каждую вкладку. Как и в предыдущем случае можно выбрать вариант экспорта для всех или только активных скважин проекта.

Названия файлов будут складываться из имени, задаваемого пользователем и наименования вкладки, присоединённого к имени через знак «_».

Например, если имя – ABC, то наименования файлов будут: **ABC_Location.txt, ABC_Orientation.txt, ABC_Lithology.txt** и т.д..

Импорт данных

Возможности импорта данных из внешних источников у *Borehole Manager* несколько выше, чем для экспорта. Допускается импорт и трансформация следующих типов данных:

Импорт данных в ASCII формате. Осуществляется выбором **File / Import / Into Borehole Manager / ASCII**

Импорт данных в формате Excel. Осуществляется выбором **File / Import / Into Borehole Manager / Excel**. В обоих случаях файловая структура импортированных данных должна в точности соответствовать экспортной. Другими словами файлы *.XLS должны быть организованы как показано на выше приведённом рисунке. В случае отступления от этих стандартов появится сообщение об ошибке и импорта не произойдёт.

Borehole Manager позволяет обмениваться данными с другим рабочим модулем *RockWorks-Geological Utilities*. В частности имеется возможность:

– трансформации и передачи данных о местоположении скважин из

Geological Utilities во вкладку «Locations» *Borehole Manager*,

- трансформации и передачи данных по стратиграфическим отбивкам н из *Geological Utilities* во вкладку «Stratigraphy» *Borehole Manager*.

Перечисленные возможности позволяют отделить процесс подготовки данных от их обработки и анализа. При этом не только не нарушается привычный ход камеральной службы, но и экономятся средства на лицензию *Rockworks*.

4.2.3 Функциональность Borehole Manager

Создание двумерных карт

Карты созданные в среде Borehole Manager могут включать:

Уникальные метки и условные знаки для каждой скважины, трассы наклонных скважин, различные границы и фоновые рисунки. Кроме того, вы имеете возможность расчёта 2-мерной регулярной сети, позволяющей строить изолинии как в виде контуров, так и закрашенных областей.

Задача: Генерация карты фактического материала (Рис. 23)

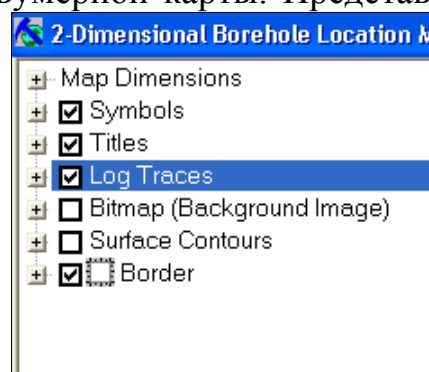
Дано:

- 1) 172 трассы буровых скважин.
- 2) Координаты устья хранятся во вкладке «Location» заданы в условных прямоугольных координатах.
- 3) Сведения по инклинометрии сосредоточены во вкладке «Orientation»

Решение:

Выбор пункта меню *Map\Borehole Locations...*

Возникает кнопочная панель параметров двумерной карты. Представленная панель обладает полной функциональностью. Это значит, что с её помощью можно строить не только карты размещения точек наблюдений и скважин, но и рассчитывать grids поверхностей и отображать их в виде изолиний разного типа. Однако, для данной задачи следует выбрать усечённый граф обработки, который оформляется путём простейшей настройки представленной панели. Для доступа к нужным параметрам следует щёлкнуть по соответствующей кнопке, выбрать нужное значение из списка или впечатать желаемое в поля ввода. В развёрнутом виде заполненная панель (рисунок 22), с помощью которой была сгенерирована карта фактического материала (рисунок 23), показана ниже.



Пояснения по пунктам:

Опция	Пояснение	Выбранный вариант	Пояснение
Map Dimensions	Расчёт размеров карты	Project Dimensions	Размеры проектного параллелипипеда
Symbols	Условный знак	Size 0.5 Thickness 1	<ul style="list-style-type: none"> - Размер равен 0.5% от ширины карты - Толщина линии- тонкая
Titles	Подписи скважин	Font	<ul style="list-style-type: none"> - Name- Arial - Size- 0.7 - Color- цвет выбирается щелчком по полю <p>Angle-угол поворота шрифта=90°</p>
Log Traces	Трассы скважин	Line Type	Толщина и цвет линии выбираются щелчком по полю
Border	Подписи рамок карты	North/Top	Подписывается только верхняя рамка (допускается кириллица)
Axes	Оформление рамок карты	West/Left East/Right South/Base North/Top	Отображаются все 4 рамки чёрным цветом
Tick-Marks	Координатные риски	Position Labels Decimal Places	<ul style="list-style-type: none"> - Inside- обращены внутрь - Font- параметры шрифта - Кол-во десятичных позиций в подписях координат определяется автоматически
Dividers	Координатная сетка	Horizontal Vertical Line Style	<ul style="list-style-type: none"> - Линии координатной сетки в 2-х направлениях - Стиль линий- пунктир

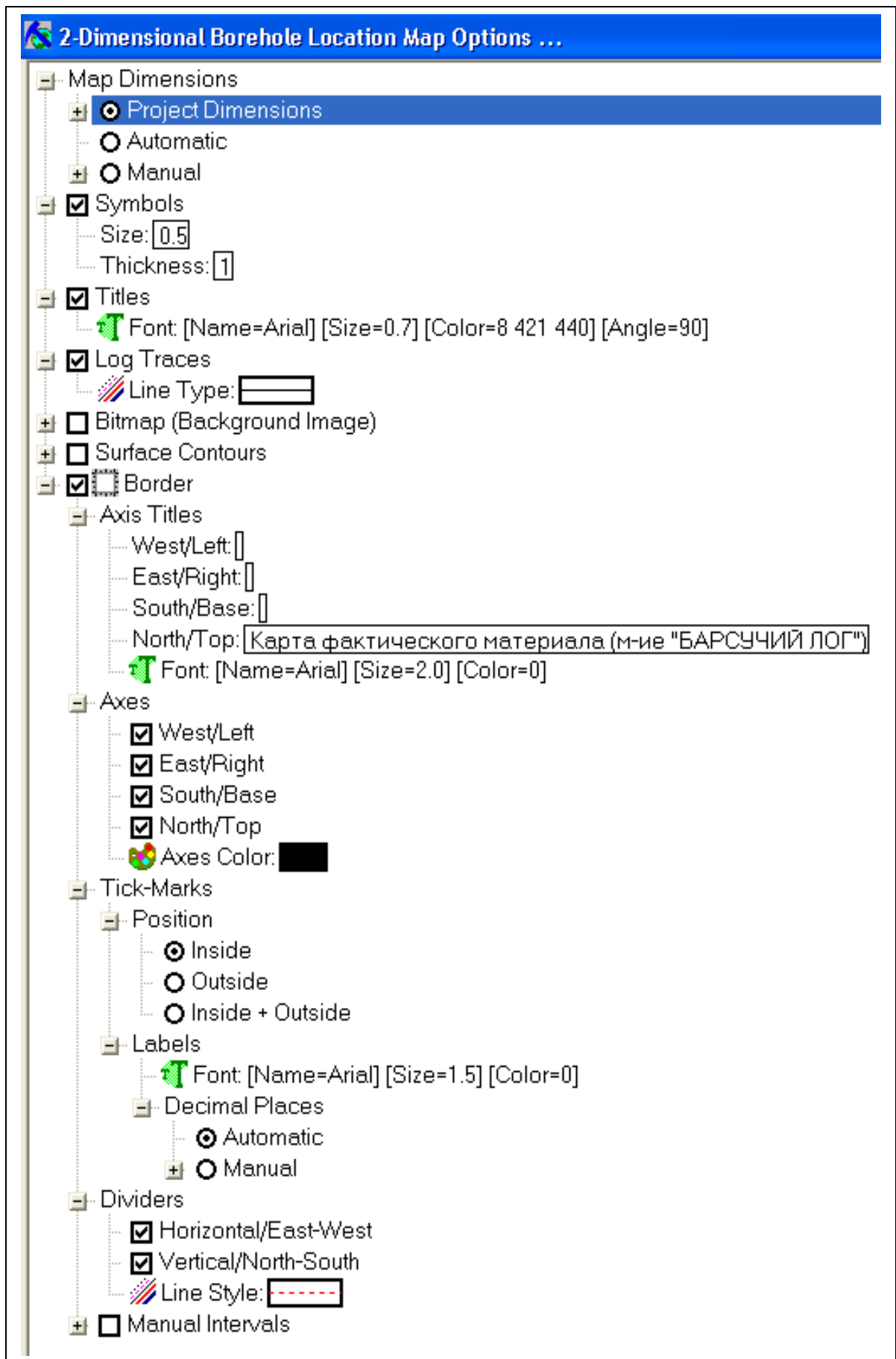


Рисунок 22 – Развёрнутая панель Borehole Manager для генерации двумерных карт

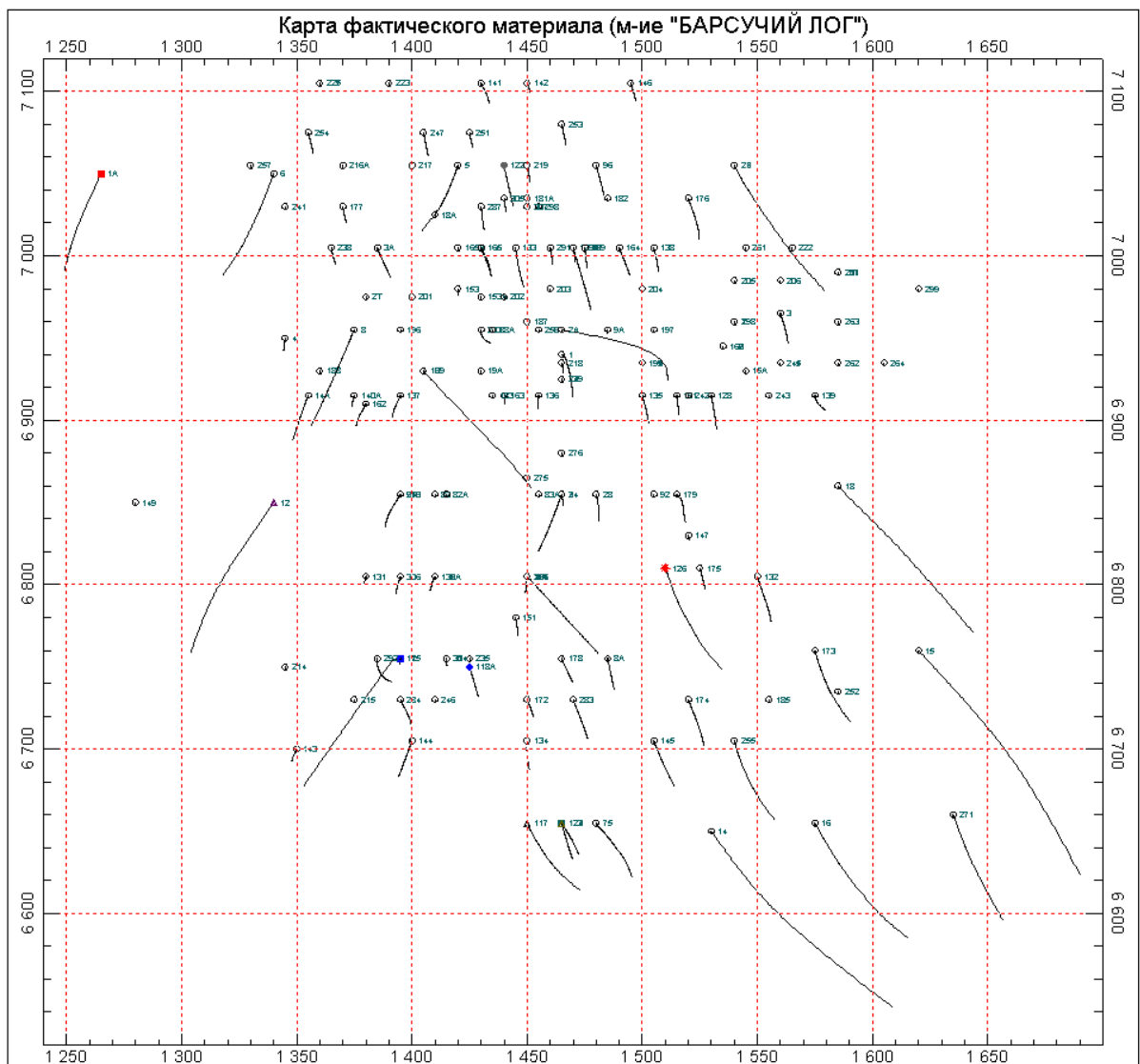


Рисунок 23- Вариант карты фактического материала

Задача: Генерация карты структурной поверхности (Рис. 24)

Дано:

- 1) 43 буровые скважины;
- 2) координаты устья хранятся во вкладке «Location» заданы в условных прямоугольных координатах;
- 3) сведения по инклинометрии сосредоточены во вкладке «Orientation»;
- 4) интервалы отбивок стратиграфических подразделений хранятся во вкладке «Stratigraphy».

Решение:

Выбор пункта меню **Stratigraphy\Structural Elevations\2-Dimensional...**

Как и в предыдущем случае на экране появится универсальный кнопочный интерфейс (УКИ). Его настройка обеспечит генерацию карты кровли любого стратиграфического подразделения, сведения по которому сосредоточены во вкладке «Stratigraph». Учёт кривизны скважин и поправки заложения будут учтены автоматически.

Пояснения:

- в качестве картируемой стратиграфической единицы выбрана условная формация *Bedrock*;
- картироваться будет поверхность кровли (*Top*);
- создаваемая регулярная сеть будет называться: *Strat.grd*;
- размерность сеточной модели будет совпадать с проектной;
- в качестве алгоритма расчёта грида выбран вариант *Inverse Distance* (обратно взвешенных расстояний);
- возникшая пространственная сетка (*Smooth Grid*) будет сглажена с применением стандартных параметров: Filter Size- 2 и Iterations- 3;
- размеры карты (*Map Dimension*) определяются автоматически;
- подпись карты проставляется вдоль верхнего её края. В этом качестве используется фраза «Структурная карта кровли Иреньского горизонта».

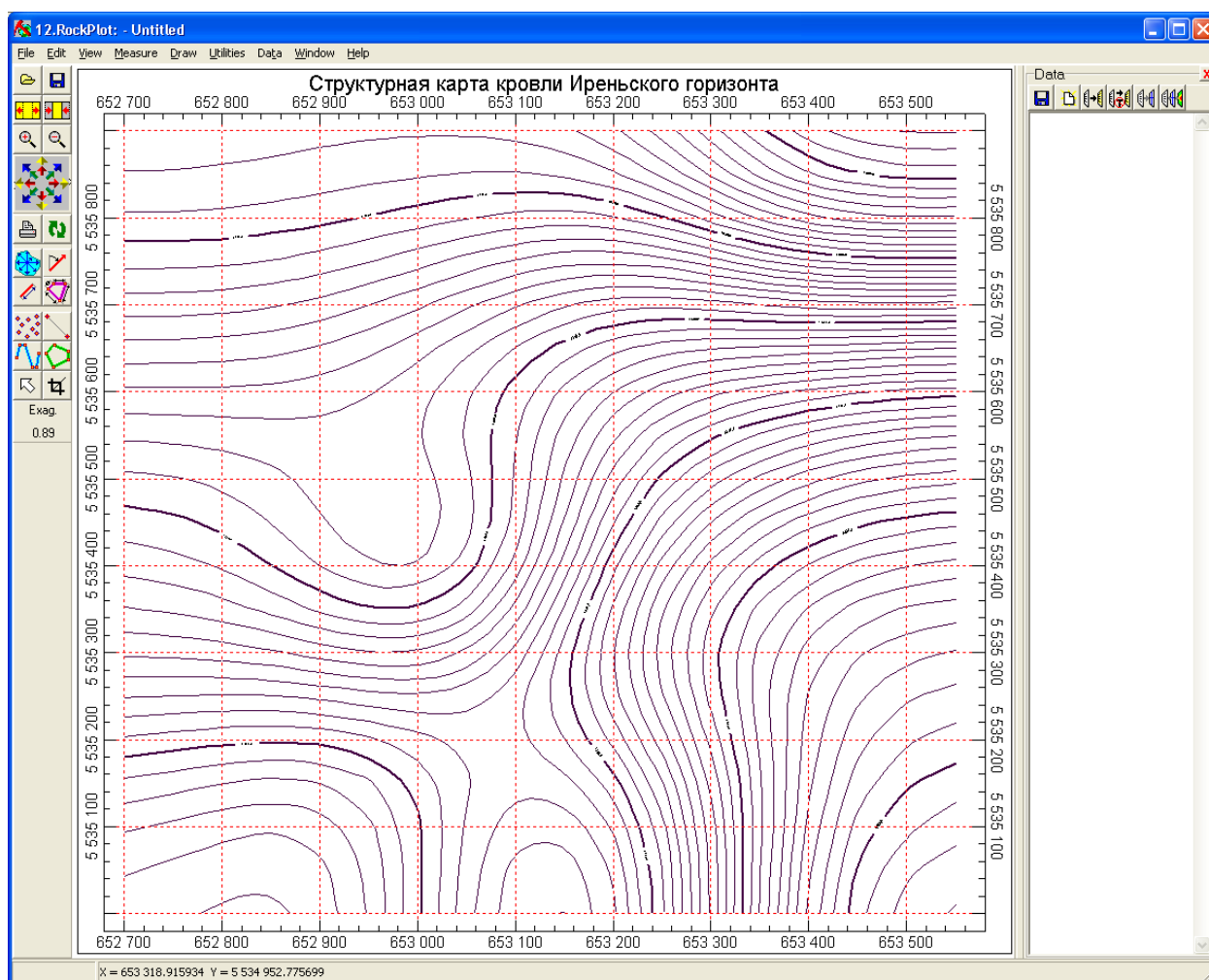


Рисунок 24– Вариант структурной карты для выбранных параметров УКИ

Задача: Генерация структурной блок-диаграммы (Рис.25)

Дано:

- 1) 43 буровые скважины;

- 2) координаты устья хранятся во вкладке «Location» заданы в условных прямоугольных координатах;
- 3) сведения по инклинометрии сосредоточены во вкладке «Orientation»;
- 4) интервалы отбивок стратиграфических подразделений хранятся во вкладке «Stratigraphy».

Решение:

Выбор пункта меню *Stratigraphy\Model...*

Настройка УКИ, в ходе которой определяются источники данных, параметры модели, внешний вид и характеристики, создаваемой блок диаграммы.

Ввиду того, что базовые принципы настройки функциональности Borehole Manager полностью унифицированы, упоминавшиеся ранее пункты УКИ здесь подробно не рассматриваются. Это касается главным образом способов настройки размерности модели, выбора алгоритма расчёта 3-х мерного грида. Они в точности такие же как в двух изученных ранее случаях.

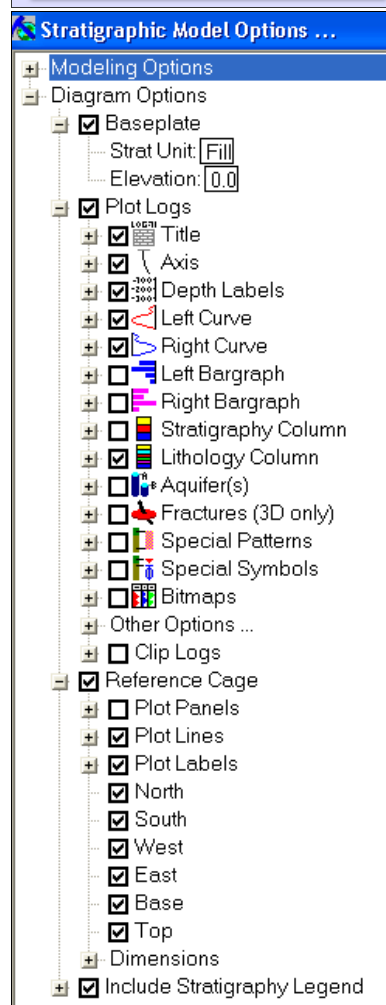
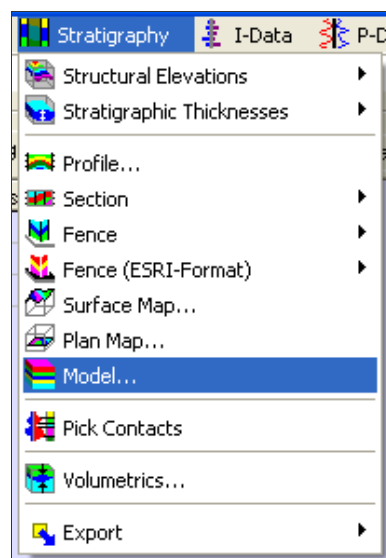
Пояснения:

В создаваемой блок-диаграмме, выбирается стратиграфическое подразделение учебной базы данных (параметр- Strat Unit: *Fill*).

Все остальные построения вышеза-легающих толщ стратиграфического разреза будут производиться относительно формации *Fill*, которая принимается за базовую. Она представляется горизонтальной плоскостью с отметкой (Elevation: 0- абсолютная отметка от уровня моря)

На диаграмме будут представлены трассы скважин (*Plot Logs*) с уточняющими параметрами, а именно:

- Наименования скважин: *Title*
- Собственно трассы: *Axis*
- Отметки глубин: *Depth Labels*
- Каротажные диаграммы из вкладки **P-Data**
- Вспомогательная разграфка и подписи модели (*Reference Cage*) по странам света и осям координат (*North, South, West, East, Base, Top*)
- Заявлен вывод на блок-диаграмму стратиграфической легенды.



Такой вариант заполнения УКИ инициирует генерацию следующей блок диаграммы (рисунок 25):

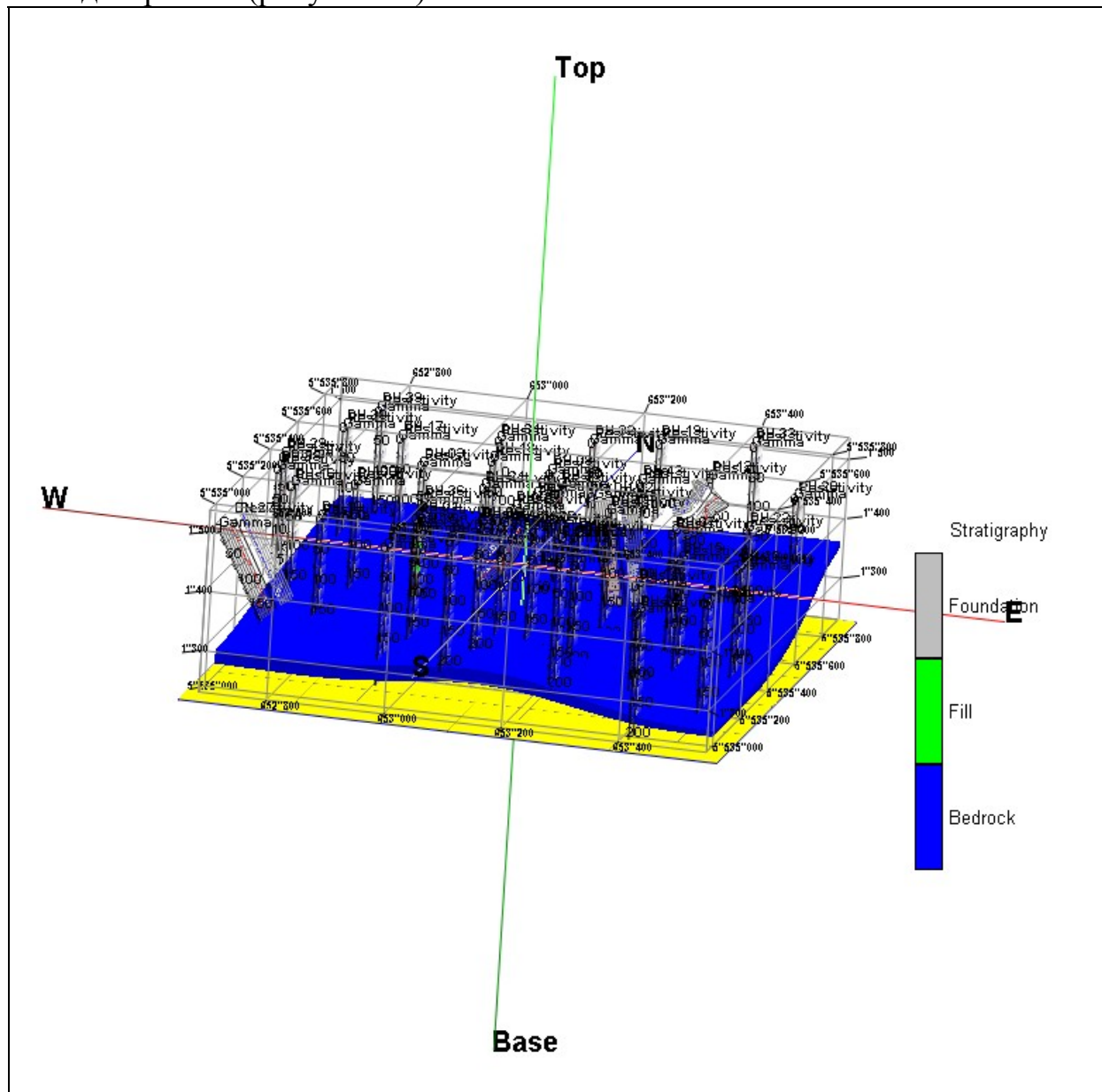


Рисунок 25- Вариант структурной блок-диаграммы для выбранных параметров УКИ

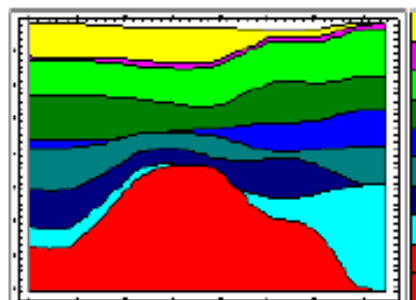
Перечисленные случаи геологического картопостроения представляют далеко не полный перечень возможностей модуля **Borehole Manager**. Почти каждая информационная вкладка связана со своей картографической реализацией, но их описание выходит за рамки данного пособия. С помощью **Borehole Manager** легко создаются карты мощностей, а также весьма сложные макеты геологических карт. При этом используется известный метод пересечения структурных моделей заданными поверхностями несогласия.

Профили и разрезы

В данном пособии рассматриваются способы автоматизированного построения вертикальных разрезов через слоистые толщи (вкладка «*Stratigraphy*»). При этом используются два полярных подхода, условно названные «модельным» и «прямым».

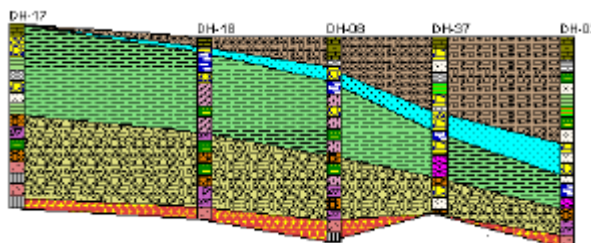
Модельный подход.

Позволяет строить разрез любой сложности и в любом направлении не придерживаясь профилей бурения. Такие возможности объясняются предварительным расчётом стратиграфических поверхностей несогласия (гридов) по всему вскрытому разрезу. Поскольку данные о стратиграфии уже индуцированы в узлы трёхмерного грида во время интерполяции, построение разреза становится формальной процедурой. Следы стратиграфических подразделений просто проецируются на вертикальную плоскость разреза.

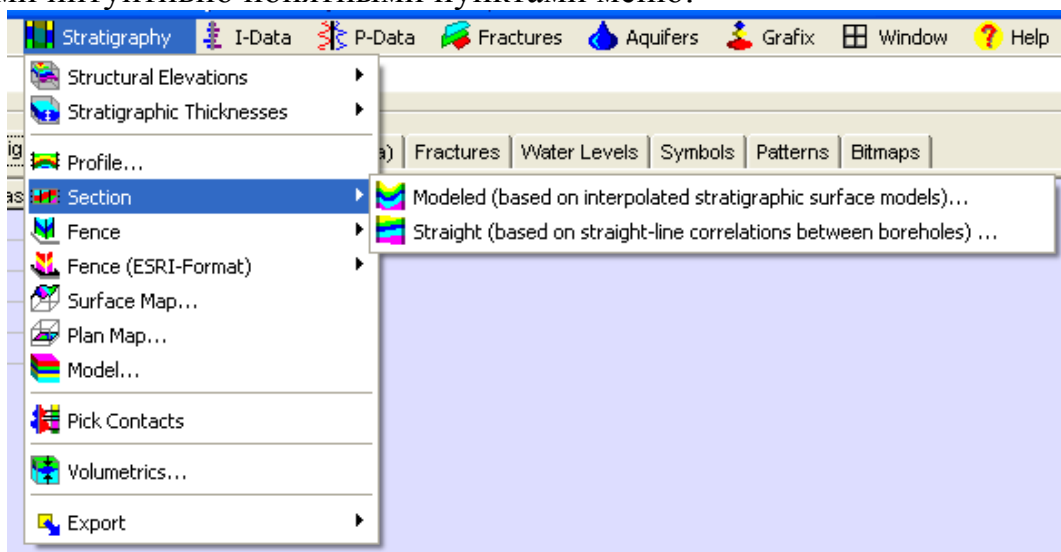


Прямой (фактографический) подход

Разрезы строятся традиционным способом проходя между буровыми скважинами.



Для выбора того или иного метода необходимо воспользоваться соответствующими интуитивно понятными пунктами меню:



Процедура настройки задания через УКИ не отличается от описанных ранее случаев.

Трёхмерные модели цифровых геологических полей

Наиболее мощные возможности модуля *Borehole Manager* связаны с использованием интервальных и точечных данных для построения пространственных моделей недр.

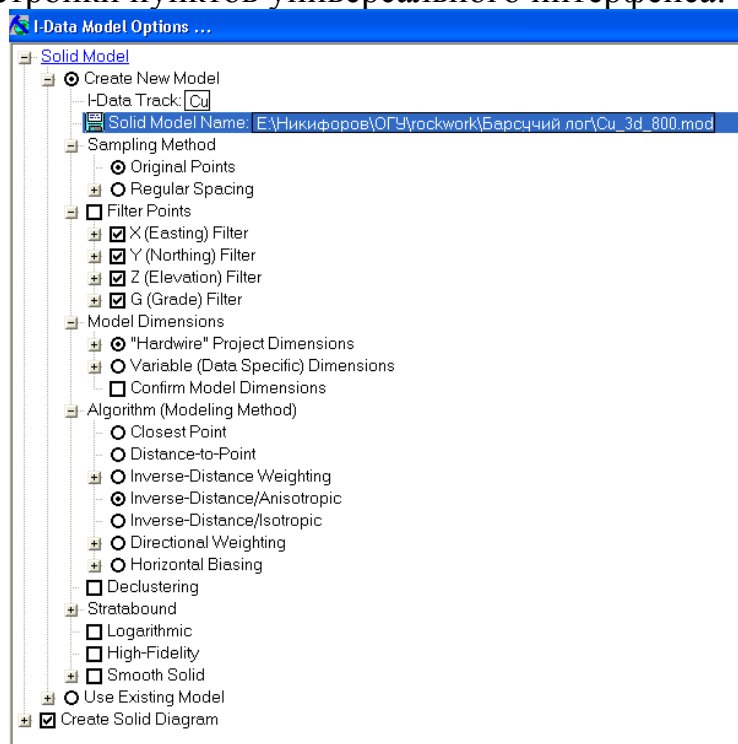
В качестве примера рассмотрим учебную базу данных «Барсучий лог», в которой собраны результаты полуколичественного спектрального анализа около 8000 бороздовых проб. Требуется построить 3-D модель первичного геохимического ореола *Cu* для порогового значения свыше 0.8% (рисунок. 26)

Необходимая информация сосредоточена во вкладках:

<i>Location</i>	-	сведения по местоположению устьев и глубин разведочных скважин;
<i>Orienta-tion</i>	-	ведения по инклинометрии наклонных скважин;
<i>I-Data</i>	-	результаты анализа геохимических проб на содержание химического элемента с условным названием- <i>Cu</i> .

Запуск программы моделирования производится через следующие пункты меню: *I-Data Model*:

После настройки пунктов универсального интерфейса:



Возникает информационная конструкция, допускающая визуализацию в произвольном ракурсе и цветовой гамме:

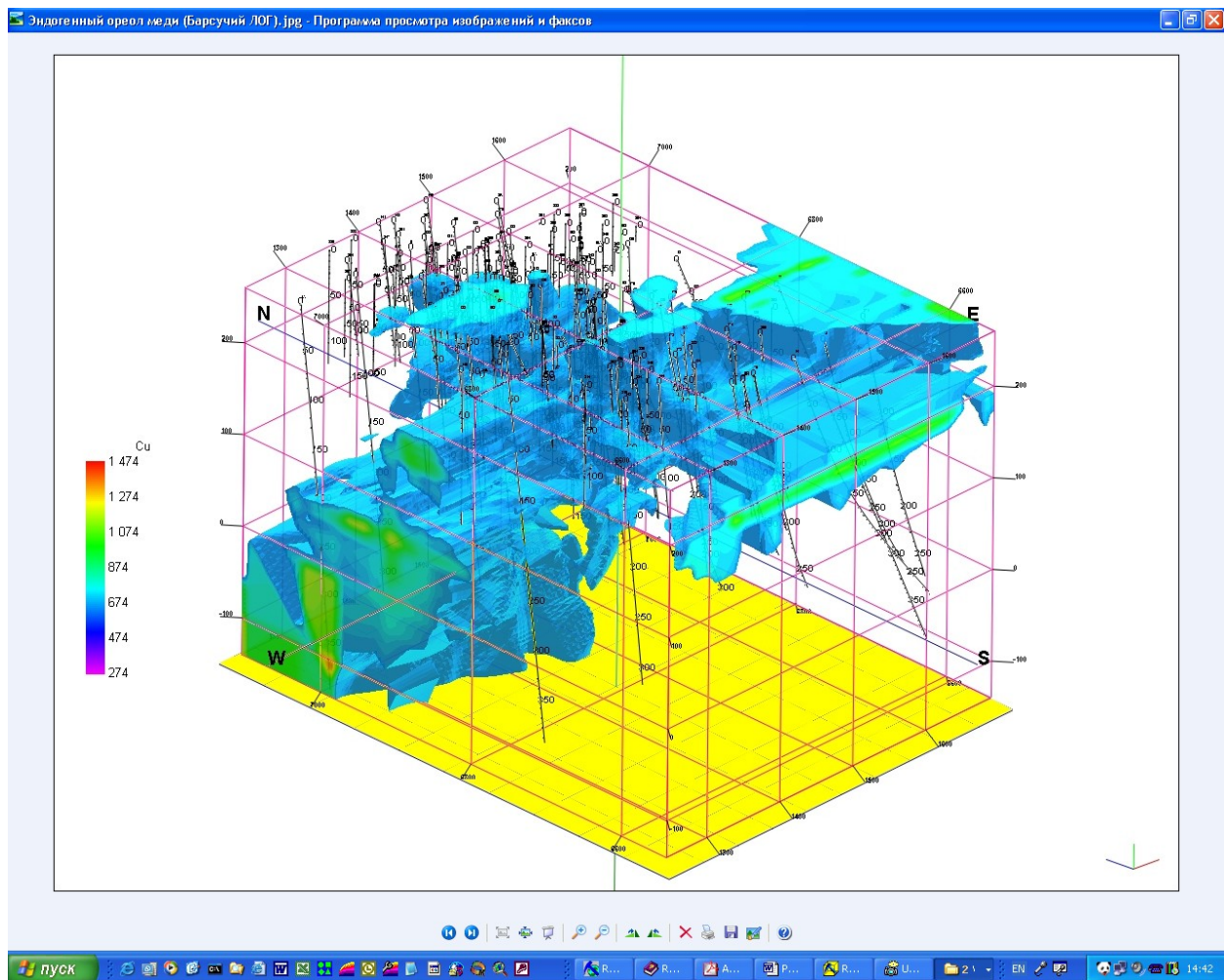


Рисунок 26 – Вариант трёхмерной модели геохимического ореола для выбранных параметров УКИ

В рамках данного пособия возможности комплекса *RockWorks* не могут быть изложены достаточно подробно. Они чрезвычайно обширны и позволяют пользователю самостоятельно разрабатывать встроенное программное обеспечение. Арсенал информационных методов комплекса складывался, во многом, исторически, охватывая не только геологические предметные области, но и проблематику смежных дисциплин. Так, в *RockWorks* включены функции поддержки географических информационных систем, модули обработки геодезической съёмки, пересчёты и трансформации систем координат. С его помощью можно осуществлять подсчёт запасов полезных ископаемых, обрабатывать данные сейсморазведки и многое другое.

По-видимому, в настоящий момент комплекс *RockWorks* является наиболее универсальным, доступным и интуитивно понятным программным инструментом геологического моделирования недр.

Список использованных источников

- 1 **Короновский, Н.В.** Общая геология: учеб. для вузов/ Н.В. Короновский.- М.:МГУ, 2002,- 448 с. -ISBN 5-211-04490-8.
- 2 **Мэйндоналд, Дж.** Вычислительные алгоритмы в прикладной статистике/ Дж. Мэйндоналд.- М.: Финансы и статистика, 1978. - 350 с.- ISBN 5-279-00057-4.
- 3 **Хагер Н.** Этапы формирования моделей: сб. "Эксперимент, Модель. Теория"-М.: Наука, 1982. - С. 64.
- 4 **Digital Geologic Maps Data Model:** [Электронный ресурс]: Интернет-страница Ассоциации Американских Геологов. - Доклад о Национальном проекте базы данных геологических карт. -Режим доступа <http://ncgmp.usgs.gov/ngmdbproject/>. - свободно по чтению.
- 5 **Шаши, Ш.** Основы пространственных баз данных/ Шаши Ш., Санжей Ч.- М.: Кудиц-образ, 2004. - 309 с.- ISBN5-93378-091-X.
- 6 **OGIS (1999):** [Электронный ресурс] : Интернет-страница консорциума OGC.- Библиотека электронных документов по использованию стандарта OGIS в электронной картографии. - Режим доступа: <http://opengis.org/press/>. – свободно по чтению.
- 7 **Barclay, T.** A spatial data warehouse/ Barclay T., Gray J. : [Электронный ресурс]: Microsoft Terraserver.- MSR-TR-99-29. - Режим доступа: <http://research.microsoft.com/pubs/>. – свободно по чтению.
- 8 **Tobler, W.** Cellular geography: [Электронный ресурс]: Интернет публикация –Режим доступа: <http://www.geog.ucsb.edu/tobler/publication/>. - свободно по чтению.
- 9 **Baumann, P.** Management of multidimensional discrete data/ Baumann P// The VLDB Journal, -1994. -№3. –С. 25-78.