

Российский Государственный
Гидрометеорологический Университет

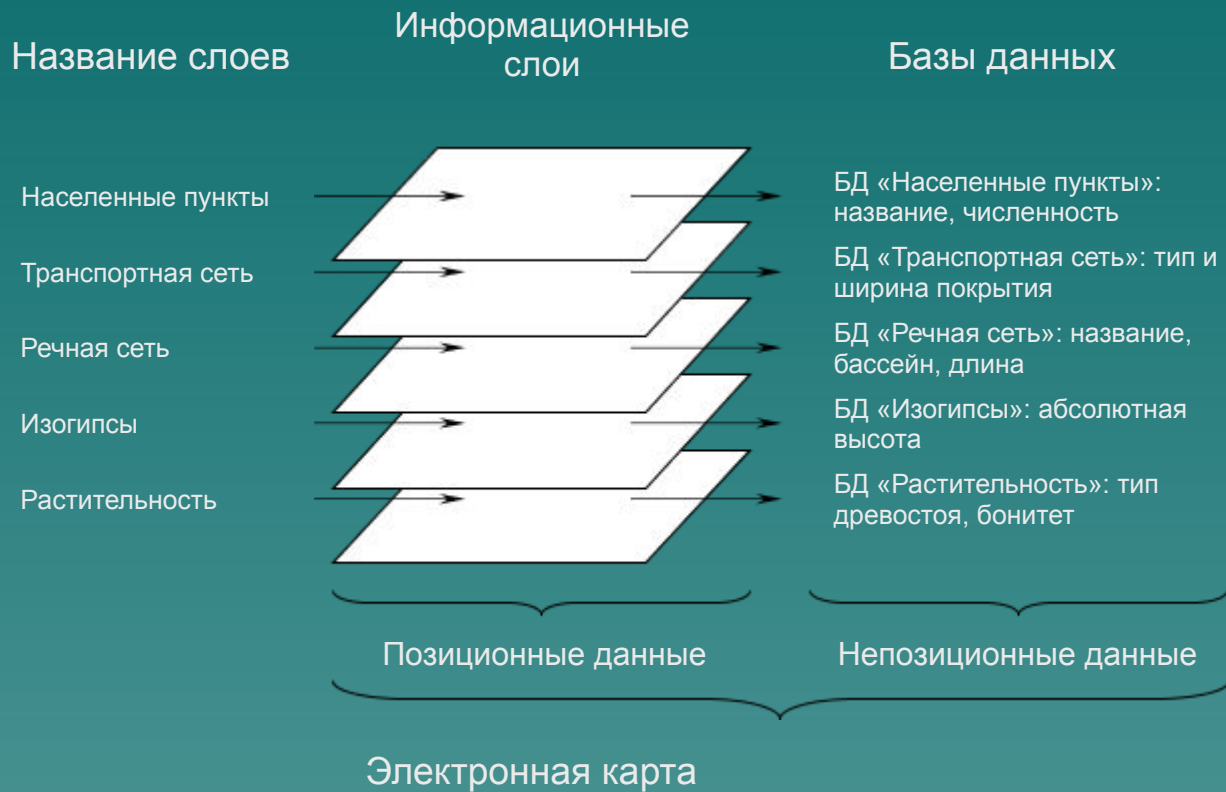
Кафедра ЭФА

***Базы данных и
управление ими***

массивы данных по какой-либо теме или

- **База данных – совокупность данных, хранящихся в цифровой форме (например, база данных о рельефе), организованных по определенным правилам, предусматривающим общие принципы описания, хранения и манипулирования данными, независимо от прикладных программ.** Формирование баз данных, доступ и работу с ними обеспечивает система управления базами данных (СУБД), которая позволяет быстро находить требуемую информацию и проводить ее дальнейшую обработку
- **База данных (БД) – совокупность массивов данных по какой-либо теме или информации, хранящихся в цифровой форме (например, база данных о рельефе, населенных пунктах и т.п.)** Формирование баз данных, доступ и работу с ними обеспечивает система управления базами данных (СУБД), которая позволяет быстро находить требуемую информацию и проводить ее дальнейшую обработку

Населенные пункты



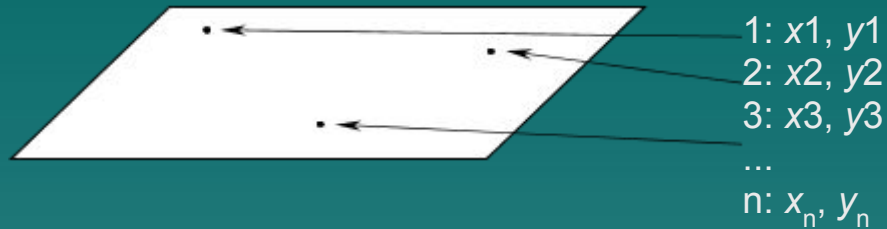
БД «Речная сеть»

Название	Длина	Порядок
Холодная	54	1
Белая	36	1
Светлая	119	2

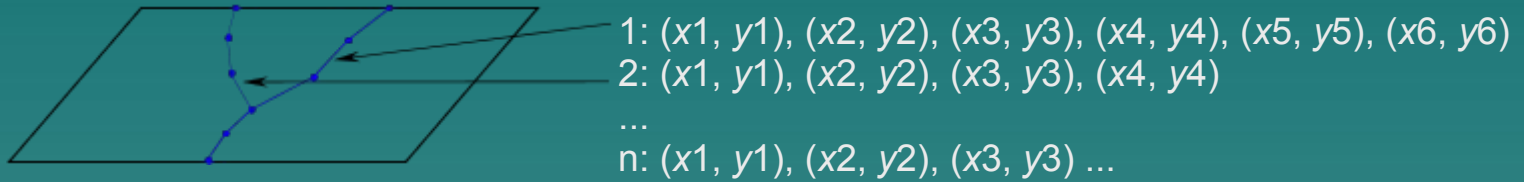
Форматы представления позиционной составляющей

Векторный формат данных

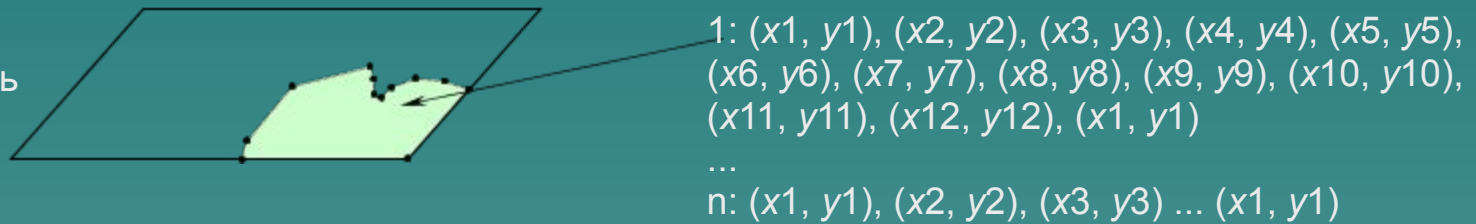
Населенные пункты



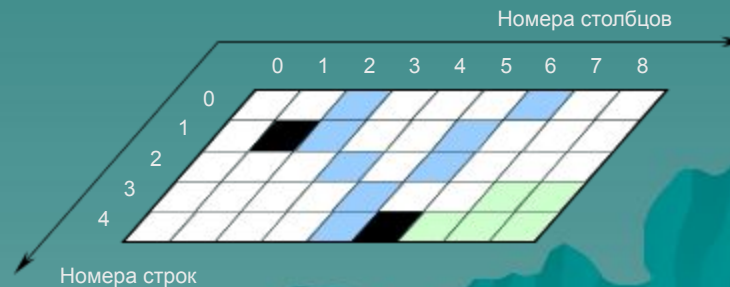
Речная сеть



Растительность



Растровый формат данных



Растр (позиционная составляющая)

Совокупность цифровых данных о пространственных объектах образует множество ***пространственных данных*** и составляет содержание ***баз географических данных (БД)***.

Требования к базе данных

База данных должна быть:

- **согласованной по времени;**
 - **полной;**
 - **позиционно точной;**
 - **достоверной;**
 - **легко обновляемой;**
- **доступной для любых пользователей.**

Проектирование базы данных.

В процессе проектирования БД обычно выделяют три основных уровня:
концептуальный, логический и физический.

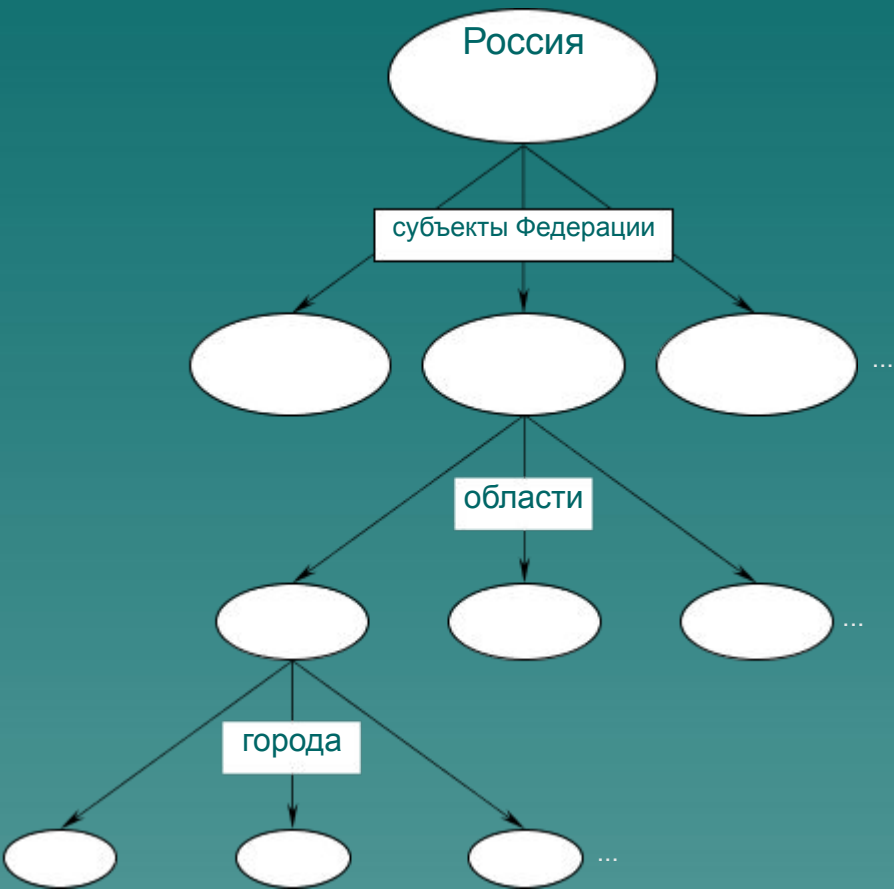
Концептуальный уровень не зависит от имеющихся аппаратных и программных средств. Для БД ГИС он связан с **концептуальной моделью географических данных** и включает:

- описание и определение рассматриваемых объектов;
- установление способа представления объектов в базе данных;
- выбор базовых типов пространственных объектов (точки, линии, полигоны, ячейки раstra);
- способ представления размерности взаимосвязей реального мира в БД.

Логический уровень определяется имеющимися программными средствами и практически не зависит от технического обеспечения. Он включает разработку логической структуры элементов базы данных в соответствии с системой управления базами данных (СУБД), используемой в программном обеспечении.

Наиболее распространёнными логическими структурами – моделями БД и их СУБД – являются **иерархическая, сетевая, реляционная.**

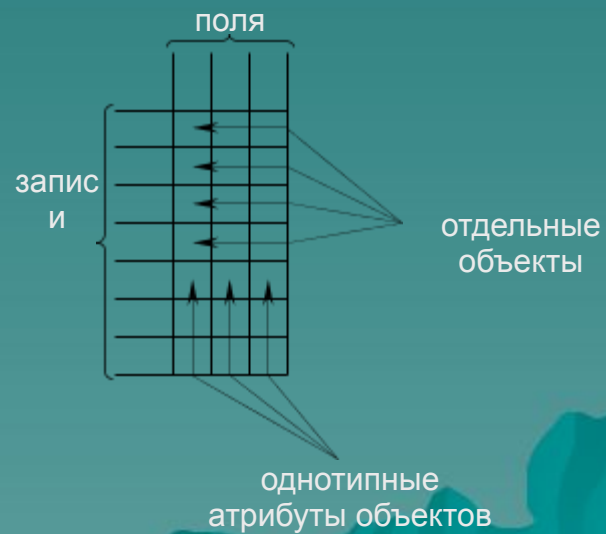
Иерархическая



Сетевая



Реляционная



a



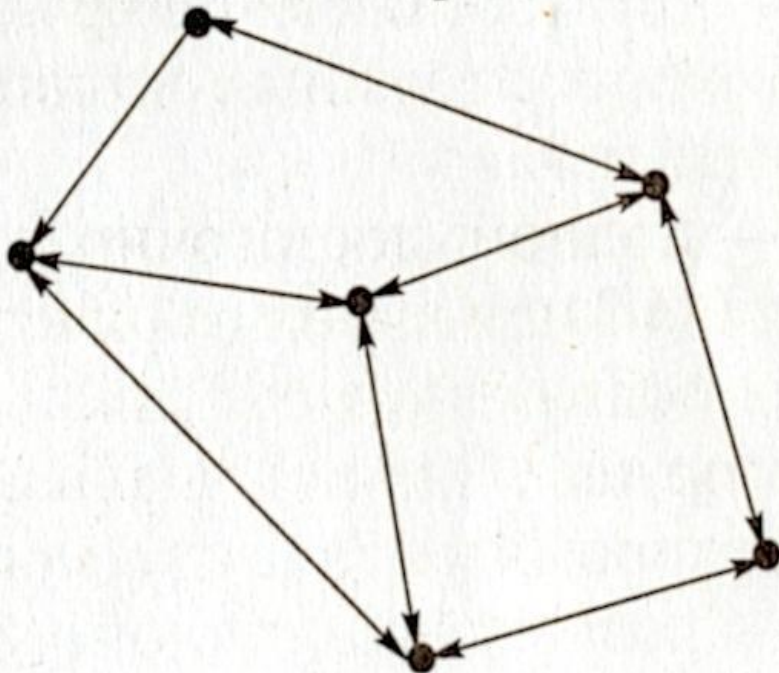
В иерархической модели

(рис. а) записи данных образуют древовидную структуру, при этом каждая запись связана только с **одной** записью, находящейся на более высоком уровне. Доступ к любой записи осуществляется по строго определённым «веткам» и узлам такого «дерева».

Иерархические модели хорошо подходят для задач с явно выраженной иерархически соподчинённой структурой информации и запросов. Они обладают низким быстродействием, трудно модифицируемы, но эффективны с точки зрения организации машинной памяти.

б

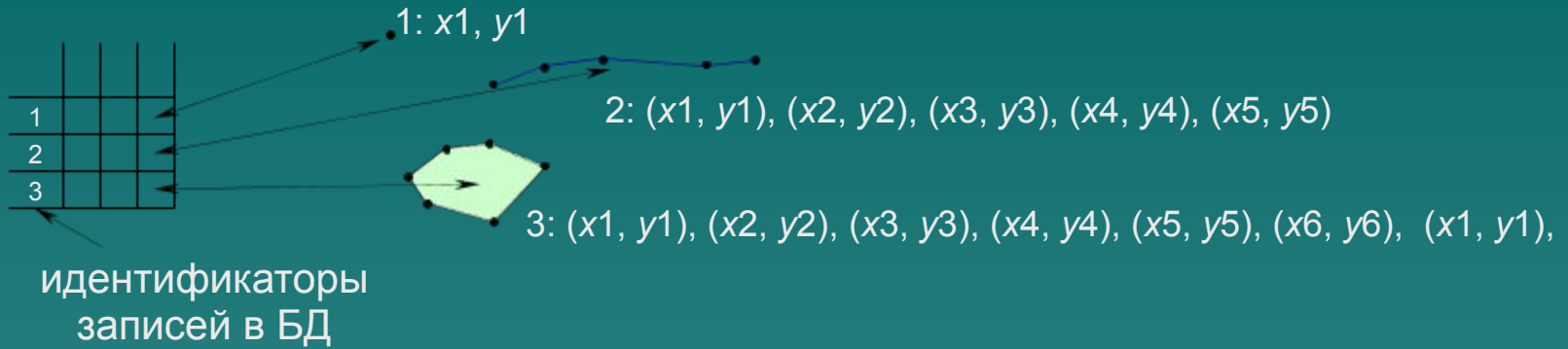
Областной центр



В сетевых моделях (рис.б) каждая запись в каждом из узлов сети может быть связана с несколькими другими узлами; кроме данных записи содержат в себе указатели, определяющие местоположение других записей, связанных с ними.

Такие модели очень трудно редактировать, например, удалять записи, так как вместе с данными нужно редактировать и указатели. Подобные модели хорошо работают в случае решения сетевых, коммуникационных задач.

Реляционная БД



ID	Название	Длина, км	Площадь водосбора, км ²	Сток в мировой океан
1	Волга	3530		Нет
2	Дон	1870		Да
3	Днепр	2200		Да
4	Амур	2824		Да

Название поля

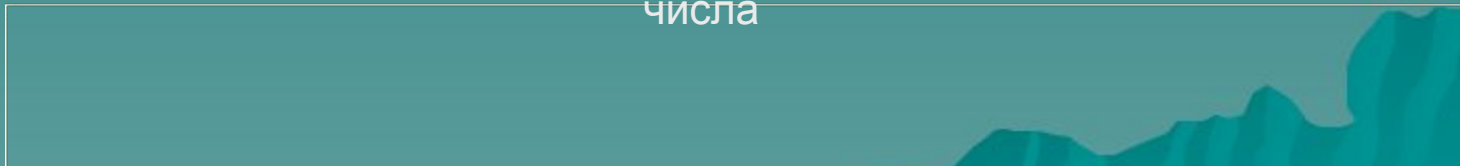
Отдельные объекты на карте

Идентификатор

Символьный тип данных

Числовой тип данных: целые и действительные числа

Бинарный тип данных



- Реляционные базы данных обеспечивают возможность и удобство выполнения следующих операций:
 - • ввод текущих значений полей записи;
 - • выборка – получение данных из базы данных, соответствующих одному или нескольким критериям;
 - • включение – добавление новых записей в базу данных;
 - • удаление – исключение существующих записей из базы данных;
 - • обновление – редактирование существующих данных записей;
- объединение – генерация новой базы данных на основе двух или более имеющихся с записями из прежних баз или рассчитываемыми по некоторым условиям.

Город	Численность населения, чел.	X-координата	Y-координата	Страна
Вена	1 875 000	16,320990	48,202120	Австрия
Лондон	11 100 000	-0,177998	51,487910	Великобритания
Амстердам	1 860 000	4,894833	52,373040	Нидерланды
Осло	720 000	10,712310	59,937930	Норвегия
Москва	13 100 000	37,938250	55,764230	Россия
Вашингтон	3 221 400	-76,953830	38,890910	США

Реляционные СУБД свободны от всех ограничений, связанных с организацией хранения данных и спецификой запоминающих устройств. Эти модели имеют табличную структуру (рис. в): строки таблицы соответствуют одной записи сведений об объекте, а столбцы – поля – содержат однотипные характеристики всех объектов. Всевозможные способы индексации данных существенно сокращают время поиска и запроса к данным.

Наиболее известные СУБД реляционного типа:

dBASE, Clipper, Foxbase, Paradox, ORACLE, Microsoft Access

Физический уровень

связан с аппаратными и программными средствами

Определяются объемы хранимой в БД информации и необходимые объемы памяти ЭВМ (оперативной и долговременной), рассматриваются вопросы о структурировании файлов на диске или других носителях информации для обеспечения программного доступа к ним и т.п.

Представление точечных, линейных и площадных объектов в базе данных и на цифровой карте

В БД ГИС картографические источники и итоговые карты представляются в виде ***цифровых карт*** .

Любая БД состоит из цифровых представлений дискретных объектов. Содержание карты можно хранить в БД в виде цифровой карты, описав объекты карты объектами базы данных.

Географические объекты, моделируемые с помощью карты или ГИС, имеют три формы представления:

- **объект в действительности;**
- **объект, представленный в базе данных;**
- **знак, который используется для показа объекта (предмета) на карте или на другом графическом изображении.**

Основные элементы базы данных

Для цифрового представления типов реальных объектов необходимо выбрать подходящую форму объектов (например, множество точек для представления множества городов). Пространственные типы объектов БД могут группироваться в **слои**, именуемые также **покрытиями** или **темами**. Один **слой** представляет один тип объектов или группу концептуально взаимосвязанных типов объектов.

Одни и те же географические явления можно представить в разных масштабах и с разной точностью. Переход от одного представления к другому достаточно сложен, например переход

Системы управления базами данных в ГИС

Как правило, ГИС создаются на основе уже существующих систем управления базами данных (СУБД).

Приобретение или аренда СУБД составляет основную часть затрат на программное обеспечение системы.

СУБД выполняет множество функций, которые в противном случае следовало бы программировать в ГИС.

Пути использования СУБД в ГИС

- Выполнение ГИС-процедур полностью через СУБД, тогда доступ ко всем данным осуществляется только через СУБД и все данные должны удовлетворять требованиям, заложенным при ее разработке.
- Некоторые данные (обычно таблицы атрибутов и их отношений) доступны через СУБД, поскольку они вполне соответствуют модели, а к некоторым данным (обычно пространственно локализованным) доступ прямой, так как они не удовлетворяют требованиям модели СУБД

Функции СУБД

- управление данными во внешней памяти;
- управление буферами оперативной памяти;
- операции над БД;
- обеспечение надежности хранения данных в БД;
- поддержка языка управления БД.

Управление данными во внешней памяти

Эта функция обеспечивает организацию структуры внешней памяти как для хранения данных, входящих в БД, так и для служебных целей, например для убыстрения доступа к данным.

В некоторых СУБД используются возможности файловых систем, в других работа производится на уровне функционирования устройств внешней памяти. В любом случае пользователи СУБД не обязаны знать, какая структура используется или как организованы файлы.

Обычно в СУБД создается собственная система именования объектов БД.

Управление буферами оперативной памяти

СУБД обычно работают с БД значительного размера, существенно большего доступного объема оперативной памяти.

Чтобы СУБД не зависела от скорости работы устройств внешней памяти, используется организация собственных наборов буферов оперативной памяти с определенными правилами замены и обновления буферов.

Операции над БД

Последовательность операций над БД, рассматриваемых СУБД как единое целое, называется транзакцией.

При выполнении **транзакции** СУБД либо фиксирует во внешней памяти изменения в БД, произведенные этой транзакцией, либо не производит никаких изменений.

Понятие **транзакции** важно для сохранения логической целостности БД, особенно в многопользовательских СУБД. Каждая **транзакция** начинается при целостном состоянии БД и оставляет это состояние целостным после своего завершения

Обеспечение надежности хранения данных в БД

Одним из основных требований к СУБД является надежность хранения данных во внешней памяти, т.е. СУБД должна обладать **способностью восстановления последнего согласованного состояния БД после любого аппаратного или программного сбоя.**

Возможны два вида аппаратных сбоев:

«мягкие» сбои - приводят к внезапной остановке работы компьютера

«жесткие» сбои - характеризуются потерей информации на носителях внешней памяти.

Журнал — это особая часть БД, недоступная пользователям СУБД и поддерживаемая с особой тщательностью, в которую поступают записи обо всех изменениях основной части БД. Самая простая процедура обеспечения надежности восстановления БД — откат транзакции, выполненной пользователем, для чего все записи от одной транзакции связывают обратным списком от конца к началу

Поддержка языков управления БД

Для работы с базами данных используются специальные языки, называемые **языками баз данных**

В современных СУБД обычно поддерживается единый интегрированный язык, содержащий все необходимые средства для работы с БД

Стандартным языком наиболее распространенных в настоящее время реляционных СУБД является язык **SQL (Structured Query Language)** Он позволяет определять схему реляционной БД и манипулировать данными.

Базовые понятия реляционных баз данных

В преобладающем большинстве ГИС используются **реляционные базы данных**, поддерживаемые такими СУБД, как **dBase**, **INFO**, **ORACLE**, **INFORMIX** и т.п.

Такие БД позволяют разработчикам ГИС разделить проблему управления пространственными данными на две части:

- как представлять геометрию объектов и топологию пространственных объектов (вектор или растр);
- как работать с атрибутами этих объектов.

Управляемые реляционными СУБД модели данных называют **геореляционными** моделями.

Основные преимущества геореляционных моделей

- нет необходимости хранить атрибуты с пространственными данными, но они всегда могут содержаться где-нибудь в системе или поставляться, например, по сети;
- атрибуты могут быть изменены или удалены без изменения пространственной БД;
- коммерческие реляционные СУБД стандартны и могут управляться стандартными запросами;
- хранение атрибутивных данных в реляционных БД не противоречит основным принципам слоев в ГИС;
- атрибуты могут быть привязаны к пространственным единицам и представлены разными способами.

Качество данных и контроль ошибок.

Представления о качестве данных, их точности и оценке погрешности становятся чрезвычайно важными при создании баз и банков данных ГИС.

Существует практически всеобщая тенденция забывать об ошибках в данных, если последние представлены в цифровой форме. Все пространственные данные до некоторой степени неточны, но в цифровой форме они обычно представляются с высокой точностью, определяемой параметрами памяти компьютера. Необходимо каждый раз рассматривать два вопроса:

- насколько правильно представляемые в БД цифровые структуры отражают реальный мир?
- насколько точно алгоритмы позволяют рассчитать истинное значение результата?

Показатели качества данных определяются стандартами

Основные из них: позиционная точность и точность атрибутов объектов, а также логическая непротиворечивость, полнота, происхождение, относящиеся к базе данных в целом.

Позиционная точность данных и типы ошибок

Позиционная точность определяется **как величина отклонения измерения данных о местоположении (обычно координат) от истинного значения**

При ее определении, как правило, исходят из масштаба исследования или первичного материала, например в данных о природных ресурсах стремятся достичь точности карты заданного масштаба.

Обеспечение большей точности требует более качественных исходных материалов, но всегда следует задаться вопросом, оправданы ли дополнительные затраты задачами исследования.

- Точность координат определяется по-разному в растровом и векторном представлении.
- Точность растра зависит от размера ячеек сетки
- Для избежания потери информации можно использовать ячейки меньшего размера, с тем, например, чтобы показать искусственные объекты, но следует оценить, что будет представлять выбранная ячейка в заданном масштабе
- В большинстве случаев неясно, относятся ли координаты, представленные в растровом формате, к центральной точке ячейки или к одному из ее углов; точность привязки, таким образом, составляет $1/2$ ширины и высоты ячейки.

Точность атрибутов определяется **как близость их к истинным показателям (на данный момент времени)**

В зависимости от природы данных точность атрибутов может быть проанализирована разными способами.

Для непрерывных атрибутов, представляющих модель поверхности, например ЦМР, точность определяется как погрешность измерений по этой модели

Для атрибутов объектов, выделяемых в результате **классификации**, точность выражается в оценках **соответствия, определенности или правдоподобия**

В общем случае для оценки точности атрибутов полезно составить матрицу ошибок классификации. Для этого нужно взять несколько случайных точек, определить их категорию по базе данных, затем на местности определить истинный класс и заполнить матрицу классификации

Если, например, число классов 4, а число обследованных точек 100, из них на местности определено 25 точек класса А, 18 точек — В, 24 — С и 33 — D

Матрица классификации

Местность	Класс в БД				
	А	В	С	Д	Всего
А	12	7	3	3	25
В	3	10	3	2	18
С	3	5	15	1	24
Д	4	4	4	21	33
Всего	22	26	25	27	100

Логическая непротиворечивость, полнота, происхождение

Эти элементы качества данных относятся к базе данных в целом, а не к объектам, атрибутам или координатам.

Логическая непротиворечивость связана с внутренней непротиворечивостью структуры данных, с топологическим представлением данных, что означает наличие исчерпывающего списка взаимоотношений между связными геометрическими представлениями данных без измерения хранимых координат пространственных объектов.

- Полнота связана со степенью охвата данными множества объектов, необходимых для представления реальности или отображения на результирующей карте (все ли соответствующие объекты включены в базу данных?).
- Она зависит от правил отбора объектов или явлений, генерализации и масштаба.

- Происхождение включает сведения об источниках данных, времени сбора данных, точности источников и цифровых данных, организации, которая их собирала, об операциях по созданию базы данных (как кодировались данные и с какого исходного материала, как происходила их обработка)
- Обычно эта информация содержится в специальных файлах **метаданных**