

ГЕОФИЗИЧЕСКИЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ





ТЕМА № 5.

Хранение и выборка данных



Подсистема хранения и выборки данных

Подсистема хранения и выборки данных обеспечивает организацию разнородных данных и пространственных моделей в единую логически непротиворечивую модель, которую в дальнейшем можно будет эффективно применять в различных технологиях анализа и управления.

Основные структуры компьютерных файлов

1. Неупорядоченный массив записей – простейшая структура файла :
 - удобная для ввода данных;
 - неудобная для поиска нужной информации.
2. Последовательно упорядоченный файлы:
 - используют буквы алфавита или числа для сортировки данных;
 - эффективны при выполнении операции поиска.
3. Индексированные файлы

Таблица

ID	Фамилия	Имя	Отчество	Должность
1	Терентьев	Владимир	Петрович	инженер	...
2	Лопухин	Сергей	Николаевич	техник	...
3	Пастухов	Илья	Ильич	техник
4	Пастухов	Кирилл	Ильич	рабочий	...
5	Хузин	Ренат	Вагизович	инженер	...
...



Индексный файл

Фамилия	ID
Лопухин	2
Пастухов	3, 4
Терентьев	1
Хузин	5
...



База данных

База данных (БД) - организованный набор взаимосвязанных файлов данных .

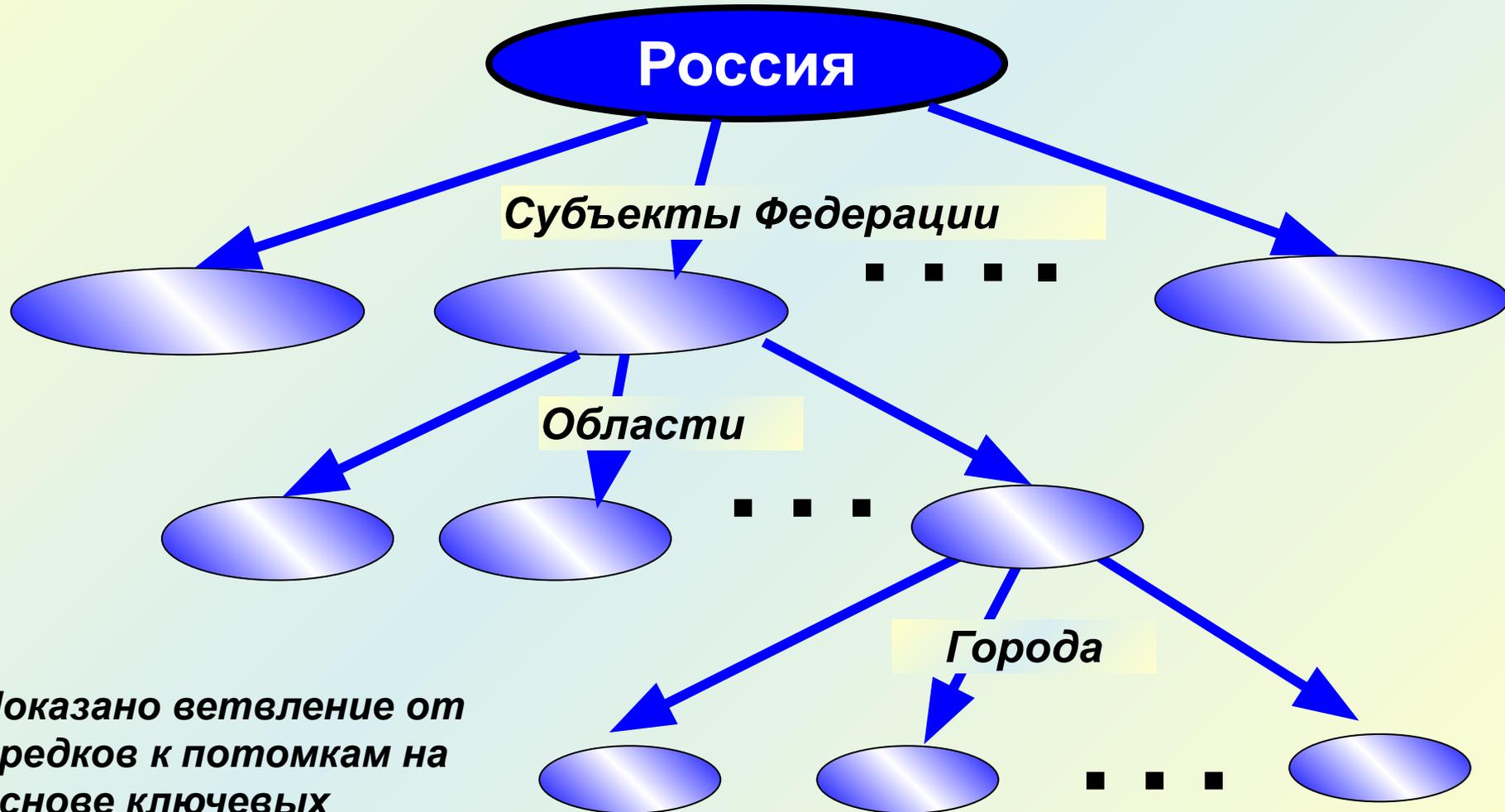
Назначение – организация данных по специальным правилам и принципам, позволяющим осуществлять:

- однократную запись информации;
- централизованное безопасное хранение;
- выборку данных, удовлетворяющим определенным условиям;
- многократное свободное или санкционированное (по паролю) обращение к данным.

Организацию и управление БД реализует **система управления базой данных (СУБД).**



1. Иерархическая



Показано ветвление от предков к потомкам на основе ключевых атрибутов



3. Реляционная

Логический элемент	Элемент базы данных
Объект	Строка (запись)
Атрибут	Столбец (поле)
Класс	Таблица

Rowid	ZONE_CODE	DESCRIPTION
1	000	NODATA
2	AGR	Agricultural
3	AIR	Airport
4	COM	Commercial
5	FLD	Flooded
6	IND	Industrial
7	INS	Institutional
8	OS	Open Space
9	RES	Residential
10	SDP	Special Development Plan

запись

Первичный ключ

Rowid	ZONE_CODE	DESCRIPTION
1	000	NODATA
2	AGR	Agricultural
3	AIR	Airport
4	COM	Commercial
5	FLD	Flooded
6	IND	Industrial
7	INS	Institutional
8	OS	Open Space
9	RES	Residential
10	SDP	Special Development Plan

Внешний ключ

FID	Shape	AREA	PERIMETER	ZONE#	ZONE-ID	ZONE_CODE
29	Polygon	139761.1	3436.182685761	29	31	RES
30	Polygon	19311.13	1227.994790069	30	25	AIR
31	Polygon	1394.393	269.1558402356	31	35	IND
32	Polygon	10618.05	433.2512163686	32	33	RES
33	Polygon	9529.783	418.2222455404	33	34	RES
34	Polygon	16141.88	812.9035032412	34	38	000
35	Polygon	44579.73	879.9199925836	35	36	IND
36	Polygon	74082.59	1254.269129168	36	37	SDP
37	Polygon	11033.96	439.7286407905	37	39	RES
38	Polygon	9639.264	420.0301261116	38	41	RES
39	Polygon	62731.4	4448.708737922	39	40	000

- Общие поля для объединения или связывания таблиц: **Первичный ключ; Внешний ключ.**

- Записи с одинаковыми значениями полей сопоставляются

Совпадающие записи

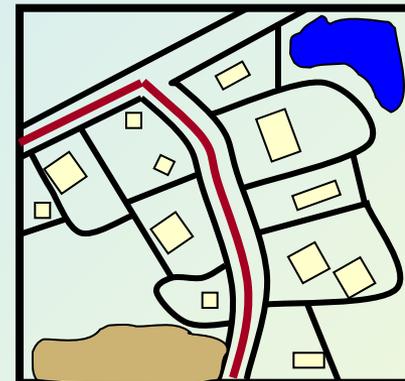
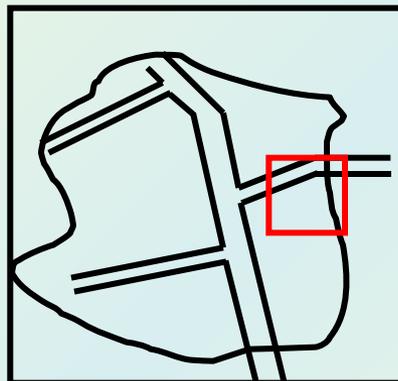
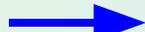


4. Объектно-ориентированная

Область изучения

Выбранная область

Содержание



Юго-западный район

квартал

Владения

Адреса

Площади

Код

налогообложения

.....

Частная
собственность

Дома

Водные объекты

Улицы

ж/дороги

Земельные участки

.....

Общественные
парки

Транспорт



Требования к базе данных геоинформационных систем

База данных должна быть:

- **согласованной по времени,**
- **полной,** достаточно подробной для предполагаемого создания ГИС или картографического произведения,
- **позиционно точной,** абсолютно совместимой с другими данными, которые могут добавляться в нее,
- **достоверной,** правильно отражающей характер явлений,
- **легко обновляемой,**
- **доступной** для пользователей.

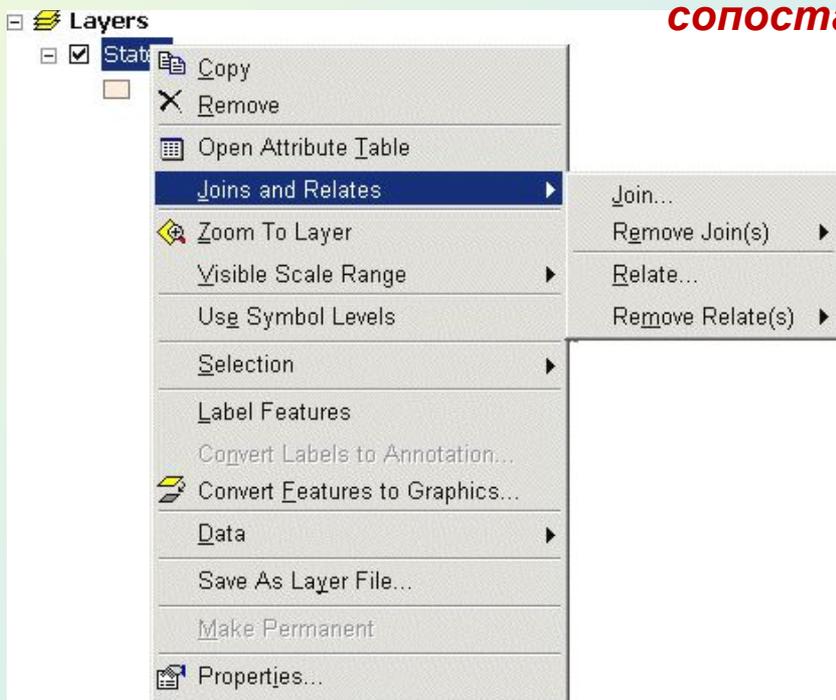
Большинство геоинформационных систем, в том числе и ArcGIS, ориентируются на использование реляционных баз данных.



Преимущества реляционных баз данных

- **Простая структура данных, позволяющая осуществлять быструю выборку и обновление данных;**
- **Возможность установления связей между таблицами.**

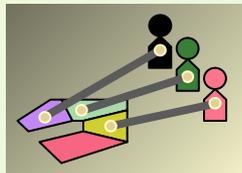
В ArcGIS реализованы 2 метода сопоставления таблиц по ключевому полю:



- **Соединение (Join)** присоединяет атрибуты одной таблицы к другой таблице;
- **Связь (Relate)** определяет отношения между двумя таблицами.

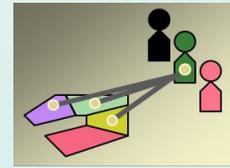
Соединение таблиц (Join)

- Объединяет атрибуты двух таблиц;
- Исходный файл остается существовать отдельно;
- Предполагает кардинальность **один-к-одному** или **многие-к-одному**.



У одного участка - один владелец

1:1



У нескольких участков - один владелец

N:1

Участки (до Соединения)- целевая таблица

OBJECTID*	SHAPE*	PARCEL_ID*	ZONE_CODE*	SHAPE_Length	SHAPE_Area
1	Polygon	67508	601	512.602492	13042.492751
2	Polygon	67246	601	372.992656	6203.424403
3	Polygon	67247	603	353.692046	5446.766292
4	Polygon	67253	603	313.013884	5380.550025
5	Polygon	67254	603	401.035888	7320.703588

таблица-источник

OBJECTID*	ZONE*	DESCRIPTION
1	601	Commercial
2	602	Institutional
3	603	Residential
4	604	Office

Многие-к-одному

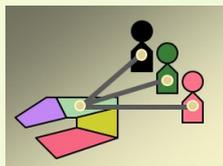
Участки (виртуальная таблица после Соединения)

OBJECTID	SHAPE	Parcel.PARCEL_ID	ZONE_CODE	SHAPE_Length	SHAPE_Area	OBJECTID	ZONE	ZoneCodeDesc.DESCRPTION
1	Polygon	67508	601	512.602492	13042.492751	1	601	Commercial
2	Polygon	67246	601	372.992656	6203.424403	1	601	Commercial
3	Polygon	67247	603	353.692046	5446.766292	3	603	Residential
4	Polygon	67253	603	313.013884	5380.550025	3	603	Residential
5	Polygon	67254	603	401.035888	7320.703588	3	603	Residential
6	Polygon	67256	603	376.675717	2622.991768	3	603	Residential



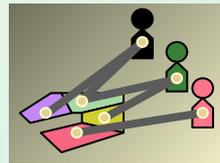
Связывание таблиц (Relate)

- Отношения между двумя таблицами **один-ко-многим, многие-ко-многим**;
- Таблицы остаются независимыми, поля не добавляются к целевой таблице;
- Выборки отображают «связанную» информацию.



У нескольких участков - один владелец

1:N



У нескольких участков - несколько владельцев

N:N

0073	115
0073	117
0073	117
007602	620

- Find & Replace...
- Select By Attributes...
- Select All
- Clear Selection
- Switch Selection
- Add Field...
- Related Tables**
- Create Graph

1) Создайте выборку

Attributes of Blocks	
KEYFIELD	OBJECTID
06.071.0073 .101	1 F
06.071.0073 .111	2 F
06.071.	
06.071.	
06.071.	
06.071.	

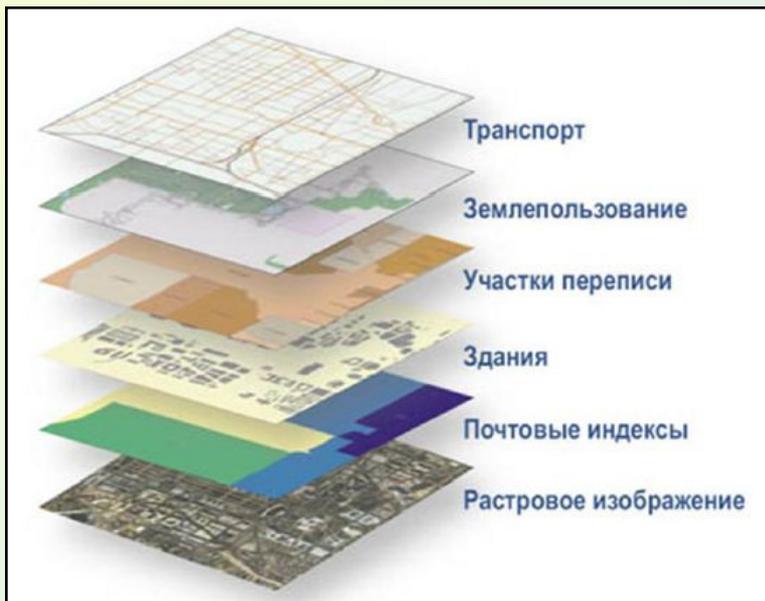
Attributes of Blk_Dmg			
OBJECTID*	STATEFP	CNTY	TRAC
	12 06	071	0086
	13 06	071	0073
	14 06	071	0078
	15 06	071	0078
	16 06	071	0078
	17 06	071	0078

2) Откройте связанную таблицу



Принципы организации данных

1. Послойный принцип организации данных



- Пространственные данные о территории организуются (раслаиваются) в виде *набора тематических слоев*.
- *Слой состоит из однородных данных*, объединяемых общей тематикой.
- Изучаемая область реального мира представляется *набором слоев взаимосвязанных данных*.

2. Объектно-ориентированный принцип организации данных

Объектно-ориентированный принцип организации данных в ГИС фокусирует внимание не столько *на общих свойствах объектов* (моделируемых через деление на слои), сколько на их *положении в какой-либо сложной иерархической схеме классификации и на взаимоотношениях между объектами*.

При этом группировка объектов соответствует их логическим взаимосвязям. В силу этого удобно отображаются различные родственные и генетические отношения между объектами, отношения соподчиненности, функциональные связи между объектами.

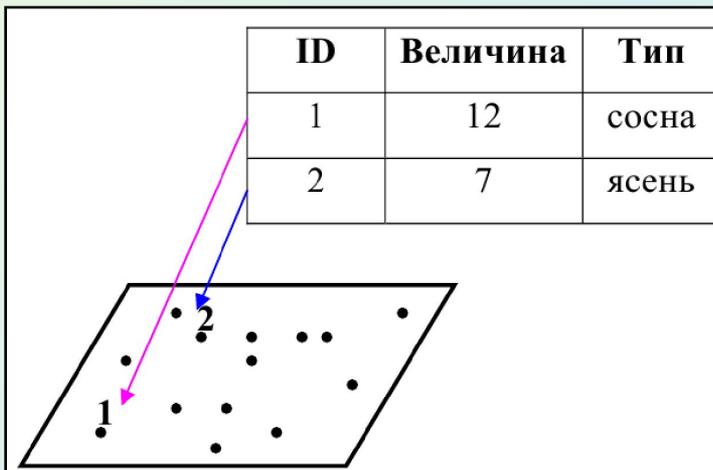


1. Геореляционная модель организации данных

Это *классическая модель* организации векторных моделей пространственных данных.

Сущность модели заключается в *раздельном хранении* значений *координат* и *атрибутивных данных*.

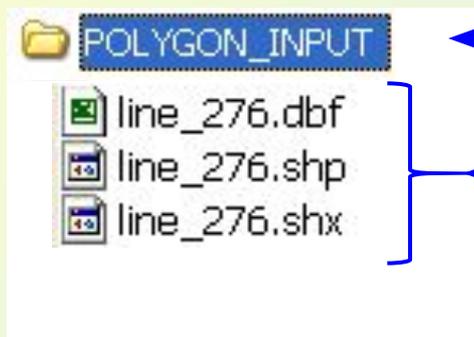
- Координаты каждого объекта хранятся в двоичных файлах.
- Атрибутивные значения и описание топологии хранятся в таблицах реляционной СУБД.
- Между записями в таблицах пространственных данных и записями в таблице атрибутов устанавливается отношение "один-к-одному".
- Связь между географическим объектом и записью в таблице атрибутов поддерживается через единственный уникальный номер – идентификатор объекта.
- Идентификатор хранится в двух местах: в файлах географических объектов, содержащих пары координат X, Y , и в соответствующих записях таблицы атрибутов географических объектов.



ГИС осуществляет совместное согласованное управление целостной информацией объектов, распределяемой между файловой системой и базой данных.



Шейп-файл



← **Папка шейп-файла в файловой системе**

← **Компоненты линейного шейп-файла**

<имя>.shp - пространственные данные

<имя>.dbf - атрибутивные данные

<имя>.shx - индексный файл

<имя>.prj - пространственная привязка

<имя>.shp.Xml – файл метаданных

<имя>.snb

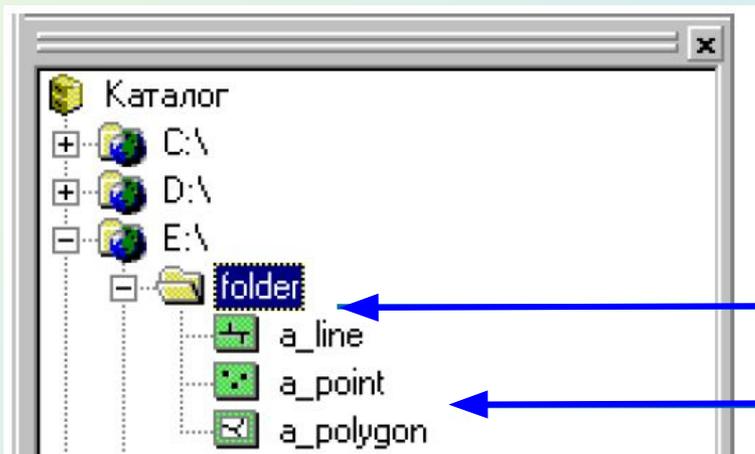
<имя>.snx

<имя>.ain

<имя>.aix

← **дополнительные файлы**

**Шейп-файл содержит
объекты только одного
геометрического типа**



← **Папка в ArcCatalog**

← **Шейп-файлы в ArcCatalog**



1. Объектно-реляционная модель организации данных

Особенность модели :

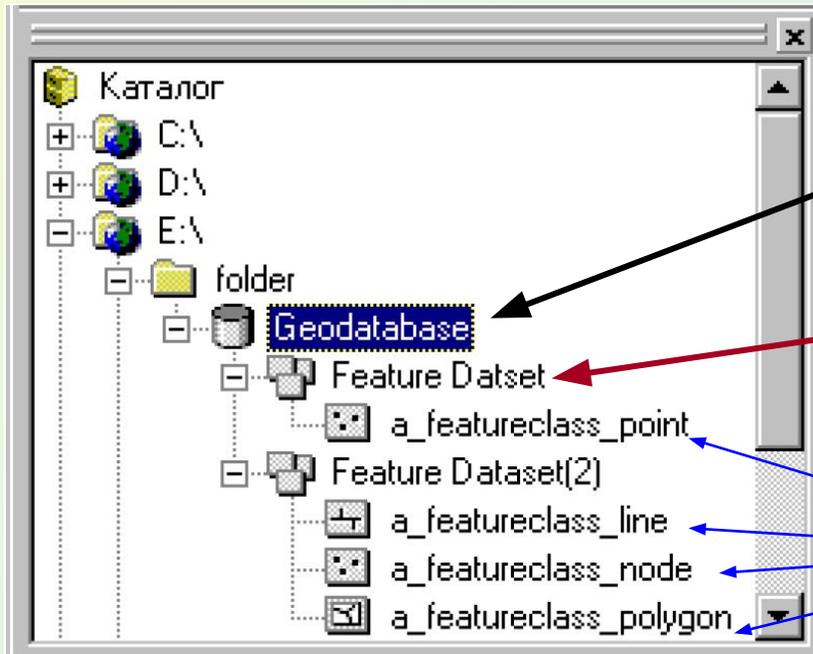
- *координаты* объектов и *атрибутивные данные* хранятся вместе в *таблицах* реляционной базы данных;
- *управление* пространственными данными осуществляет *СУБД*.

Модель реализована в виде **базы геоданных**, разработанной ESRI.





База геоданных



База геоданных в ArcCatalog

Набор классов объектов

Классы пространственных объектов

-  **Folder** — Папка с базой геоданных в файловой системе
-  **Geodatabase.ldb** — Файлы базы геоданных
-  **Geodatabase**



База геоданных: пространственная привязка

Пространственная привязка состоит из трех компонент:

Система координат задает картографическую проекцию и ее параметры.

Координатный домен ограничивает диапазон значений x, y, m, z .

Масштаб определяет, сколько целочисленных единиц соответствует единице карты и определяет точность координат.



База геоданных

Набор классов

Класс пространственных объектов

Класс пространственных объектов

Класс пространственных объектов

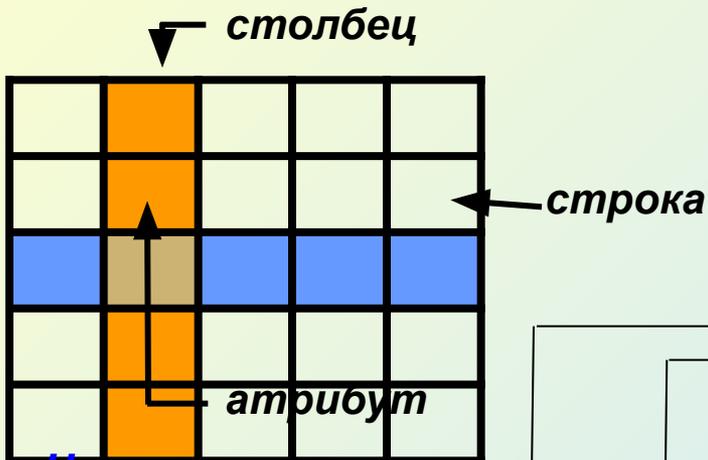
Пространственная привязка

Пространственная привязка



Таблицы

Таблица - это набор строк.
 Строка – это набор атрибутов.
 Столбцы представляют однотипные атрибуты.
 Поле – это описание столбца.
 Все строки таблицы имеют одинаковый набор полей.



Класс пространственных объектов

Независимая
таблица

код	описание
1	Шоссе
2	Магистралы
3	Главные дороги
4	Дороги в жилом массиве
5	Дороги без покрытия

fid	geom	shp_len	тип	поверхность	ширина	рядов	название
101	polyline	4507.2	2	асфальт	85.3	4	Шоссе Олд-Тасс
102	polyline	3401.1	1	бетон	45.1	2	Калле-Мейя
103	polyline	2321.8	3	асфальт	75.1	4	Дорога Каунти
104	polyline	689.2	5	гравий	35.2	2	Дорога Макс-Дэниел
Предопределенные поля			Пользовательские поля				

- Идентификатор объекта
- Геометрия
- Поле слежения за геометрией
- Кодированное значение
- Описательная строка
- Непрерывное числовое значение
- Дискретное числовое значение
- Текст для названий

Атрибут можно использовать для связывания таблиц



Типы полей

Название	Заданный диапазон/ длина/формат	Применение
Короткое целое (Short integer)	от -32768 до 32767	Целые числа внутри заданного диапазона, кодированные значения
Длинное целое (Long integer)	от -2147483648 до 2147483647	Большие целые числа внутри заданного диапазона
С плавающей точкой одинарной точности (Float)	от -3.4×10^{-38} до 1.2×10^{38}	Числовые значения с дробными величинами внутри заданного диапазона
С плавающей точкой двойной точности (Double)	от -2.2×10^{-308} до 1.8×10^{308}	Числовые значения с дробными величинами внутри заданного диапазона, координаты
Текст (Text)	до 64000 символов	Имена и другие текстовые значения
Дата (Data)	mm/dd/yyyy hh:mm:ss AM/PM	Дата и/или время
BLOB	любое	Изображения, мультимедиа

Управление атрибутами с помощью доменов

Атрибутивные домены

Интервальный домен

Давление

от 2000 до 14000

Домен кодированных значений

Тип породы

«глина», «песчаник», «алевролит»

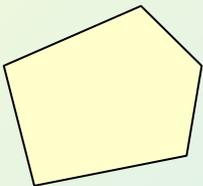
Значения по умолчанию

Зонирование земель

R-4

Правила разделения и объединения

Разделение пространственных объектов



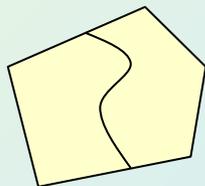
Владелец	Зона	Значение
Гоулд	R-4	25000

Дублирование

Значение по умолчанию

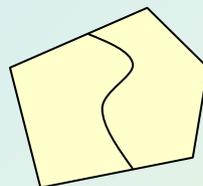
Геометрическая пропорция

Правило разделения



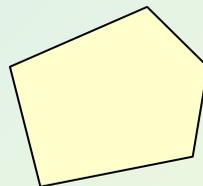
Владелец	Зона	Значение
Гоулд	R-4	14000
Гоулд	R-4	11000

Объединение пространственных объектов



Округ	Урожай	Собрано, %
Лейквью	24000	35
Риверсайд	45000	47

Правило объединения



Значение по умолчанию

Суммирование

Взвешенное среднее

Округ	Урожай	Собрано, %
Монтейн	69000	43



Простое поведение с использованием подтипов

Класс объектов может иметь специальный атрибут, называемый подтипом. Подтип используется для наиболее значимой классификации объектов в пределах класса.

Подтипы позволяют контролировать целостность данных.

Каждому подтипу можно задавать особое простое поведение, реализуемое значениями по умолчанию, атрибутивными доменами, правилами связности и правилами отношений.

FID	Geom	подтип	ширина	число рядов	название
101		асфальт	85.3	4	Шоссе Чимайо
102		бетон	45.1	2	Изабель
103		асфальт	75.9	4	Калле-Петра
104		гравий	35.2	2	Дорога Максимилиан

Подтип «асфальт»

Реализация простого поведения

FID	Geom	подтип	ширина	число рядов	название
101		асфальт	85.3	4	Шоссе Чимайо
103		асфальт	75.9	4	Калле-Петра

Значения по умолчанию	Атрибутивные домены	Правила объединения/разделения	Правила связности	Правила отношений
Новой асфальтовой дороге по умолчанию дается ширина 35 футов.	Допустимая ширина - 35, 40, 45. Допустимое число рядов - 1,2,4.	Асфальтовая дорога при объединении получает число рядов по умолчанию	Двухрядная асфальтовая дорога может соединяться только с другой двухрядной.	Асфальтовая дорога может быть связана с пересечениями в виде эстакад или тоннелей.



Топология в базе геоданных

Топология создается для набора классов пространственных объектов и хранится как класс базы геоданных.



Класс **Топология** включает :

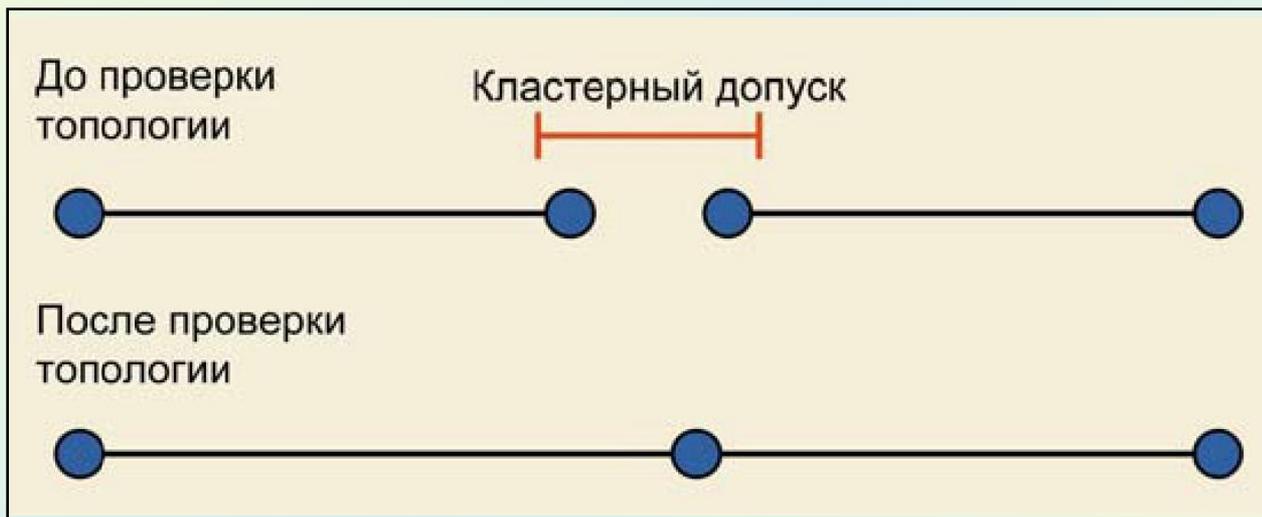
1. перечень классов пространственных данных (подтипов), участвующих в топологических отношениях;
2. топологические параметры (кластерный допуск, ранги и правила);
3. слой пространственных объектов, который содержит измененные области, ошибки и исключения .



Топологические параметры

Кластерный допуск

Кластерный допуск - это минимальное расстояние между несовпадающими вершинами.



При проверке топологии пространственные объекты, находящиеся на расстоянии меньше кластерного допуска, совмещаются в один объект.



Ранги координатной точности

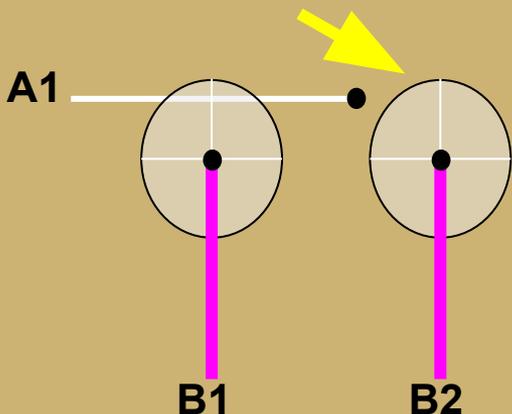
Ранги определяются для каждого класса (подтипа) объектов, участвующих в топологических отношениях. Количество рангов 1-50.

Проверяемая геометрия

Равные ранги

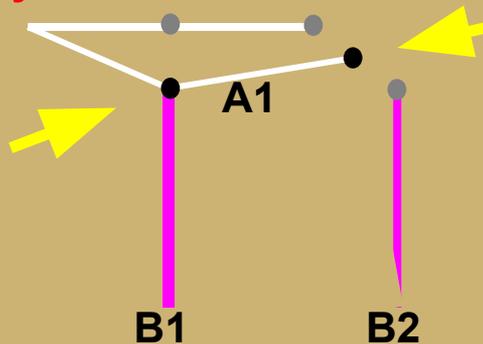
Ранг $A > B$

Кластер

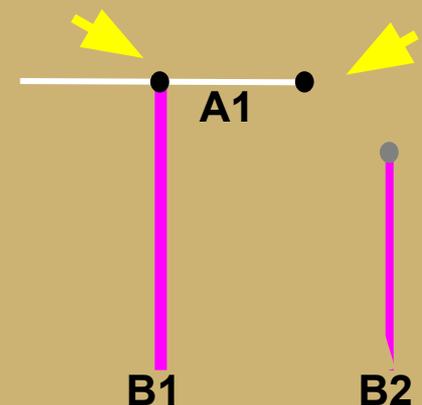


Вершина к узлу

середина



Низкий к высокому



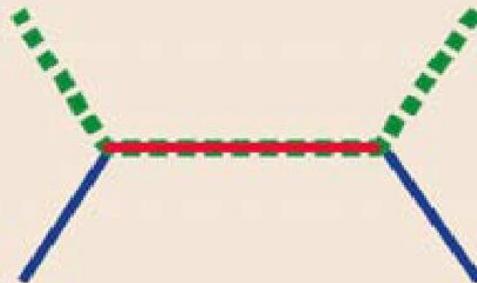
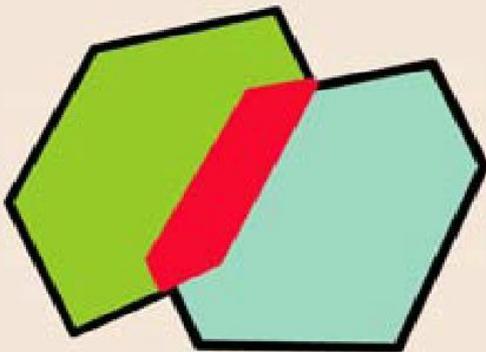
Ранги определяют, какие объекты могут быть перемещены при совмещении вершин в результате проверки топологии.



Правила топологии

Правила определяют допустимые пространственные отношения между объектами и контролируют отношения:

- *между пространственными объектами внутри одного класса объектов,*
- *между объектами разных классов или между подтипами пространственных объектов.*



Пример правил "Не должны перекрываться"

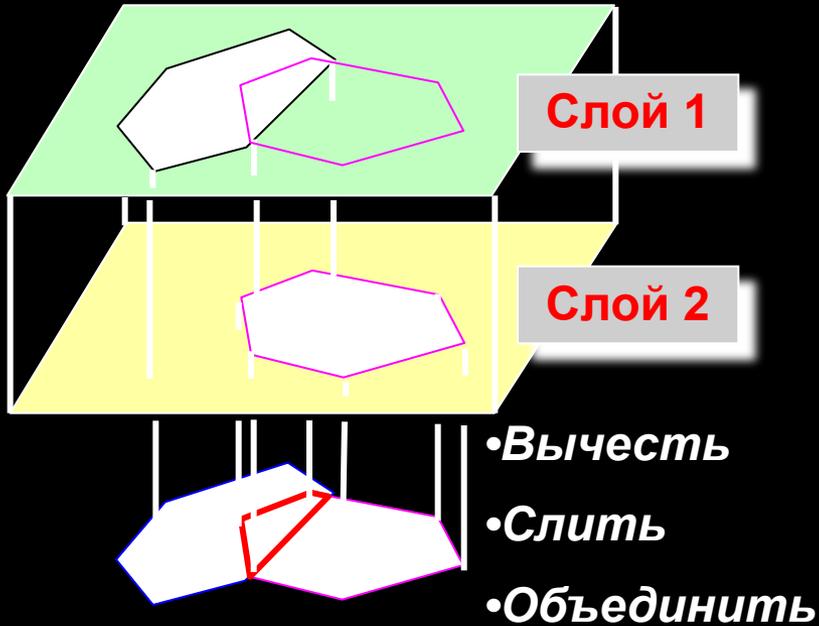
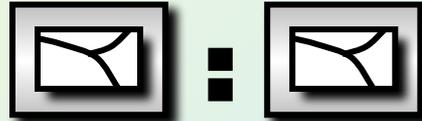
Красные полигон и линия показывают те участки, где произошло отклонение от правил



Примеры правил

- Правило для полигонов:
«Не должны пересекаться»

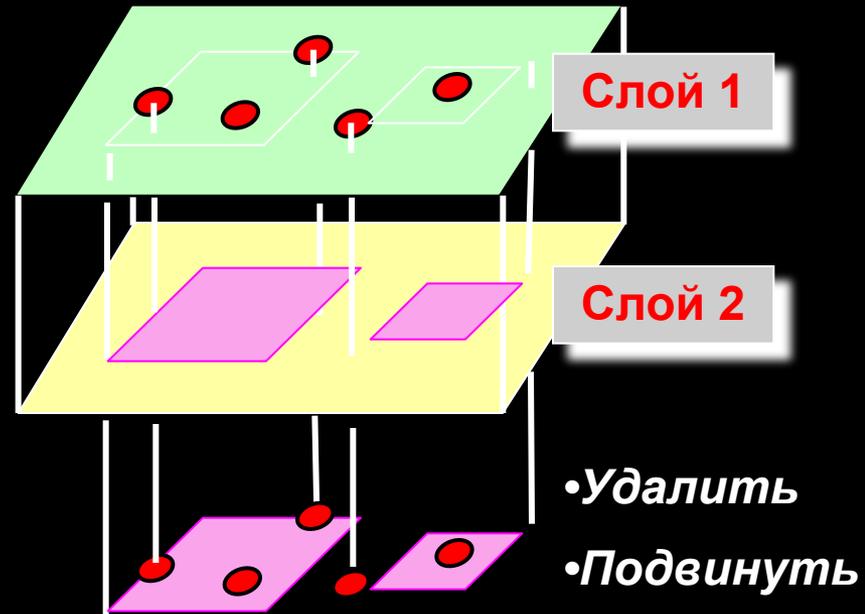
Растительность и гидрография



Варианты исправления

- Точки и полигоны:
«Должны быть полностью внутри»

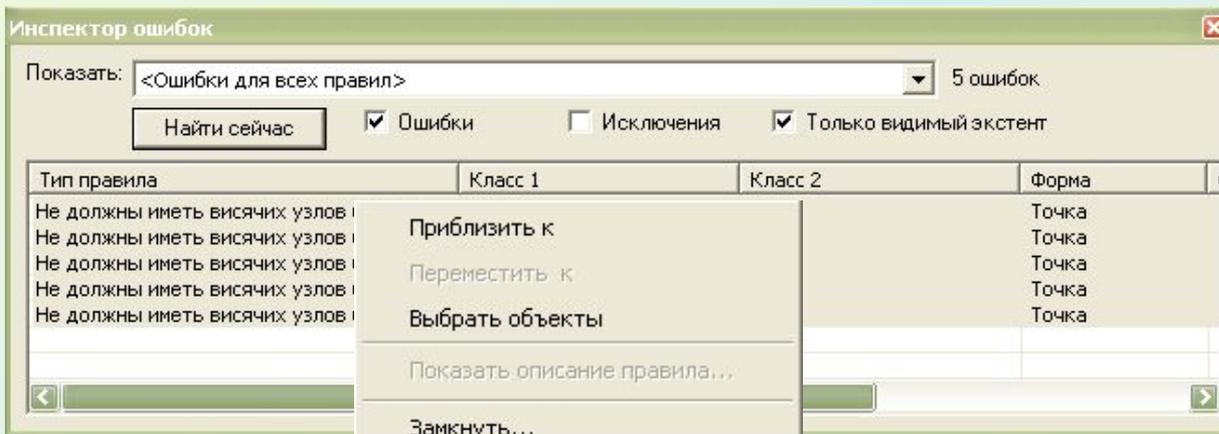
Участки и точки адресов



Варианты исправления

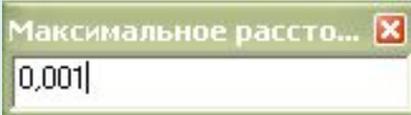
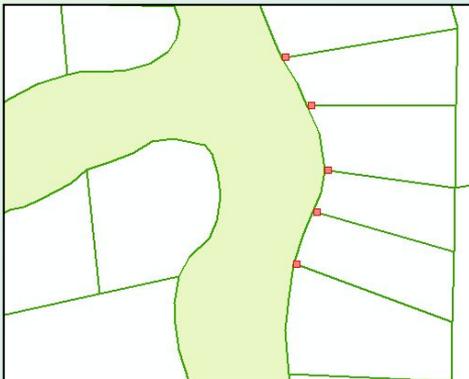


Проверка и исправление ошибок в ArcMap

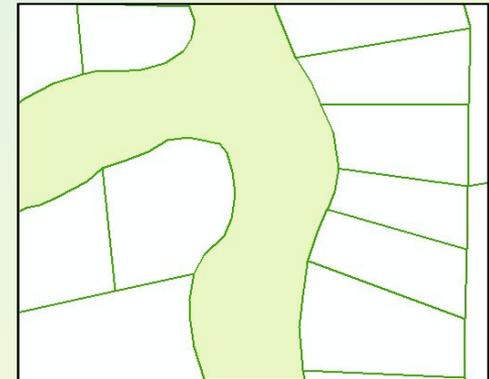


**Выбираем
ошибки**

**В меню
нажимаем
Растянуть**



**Указываем
максимальное
расстояние**

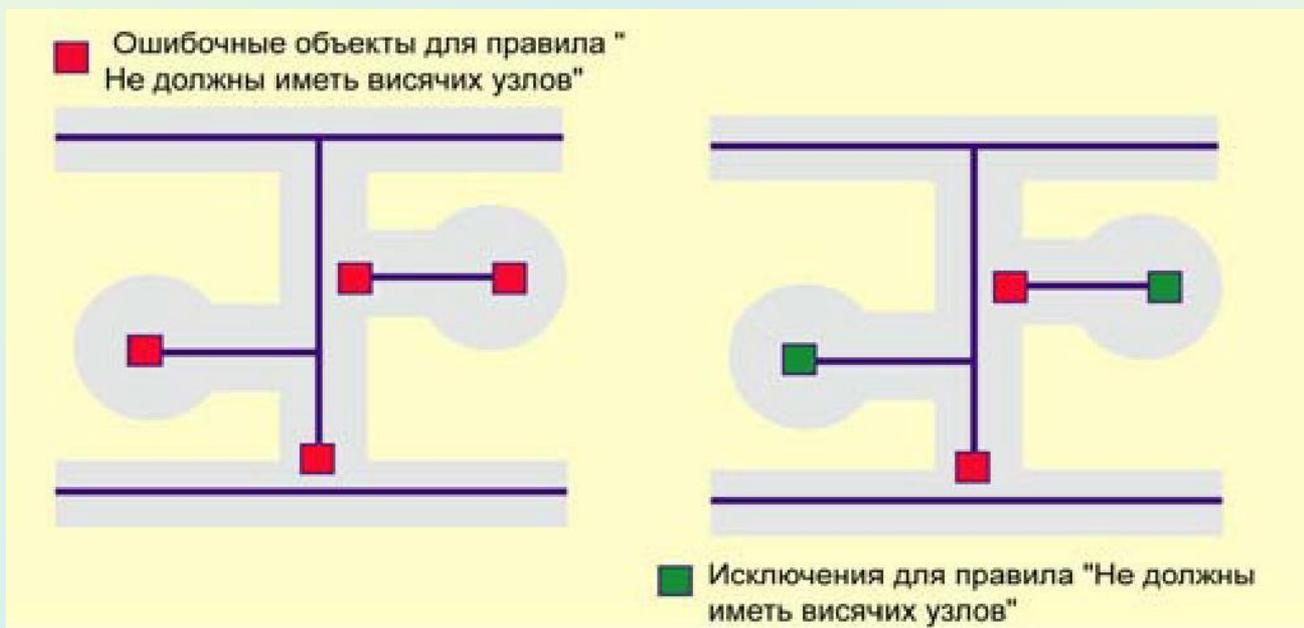




Выявленные при проверке топологии ошибки **нельзя**
удалить.

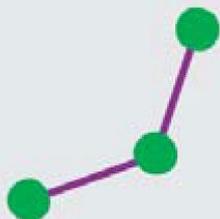
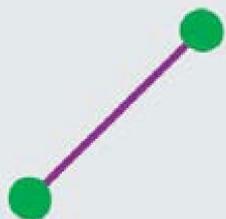
Варианты действий:

- **Игнорировать** – это не мешает использовать данные.
- **Обозначить ошибку как исключение** из правила.
- **Исправить** ошибку предлагаемыми методами.

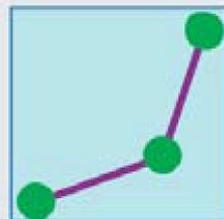




Результаты проверки топологии: **измененные области, ошибки и исключения** – записываются в специальные вспомогательные топологические слои и хранятся в классе Топология.



Редактирование объекта



Созданная измененная область

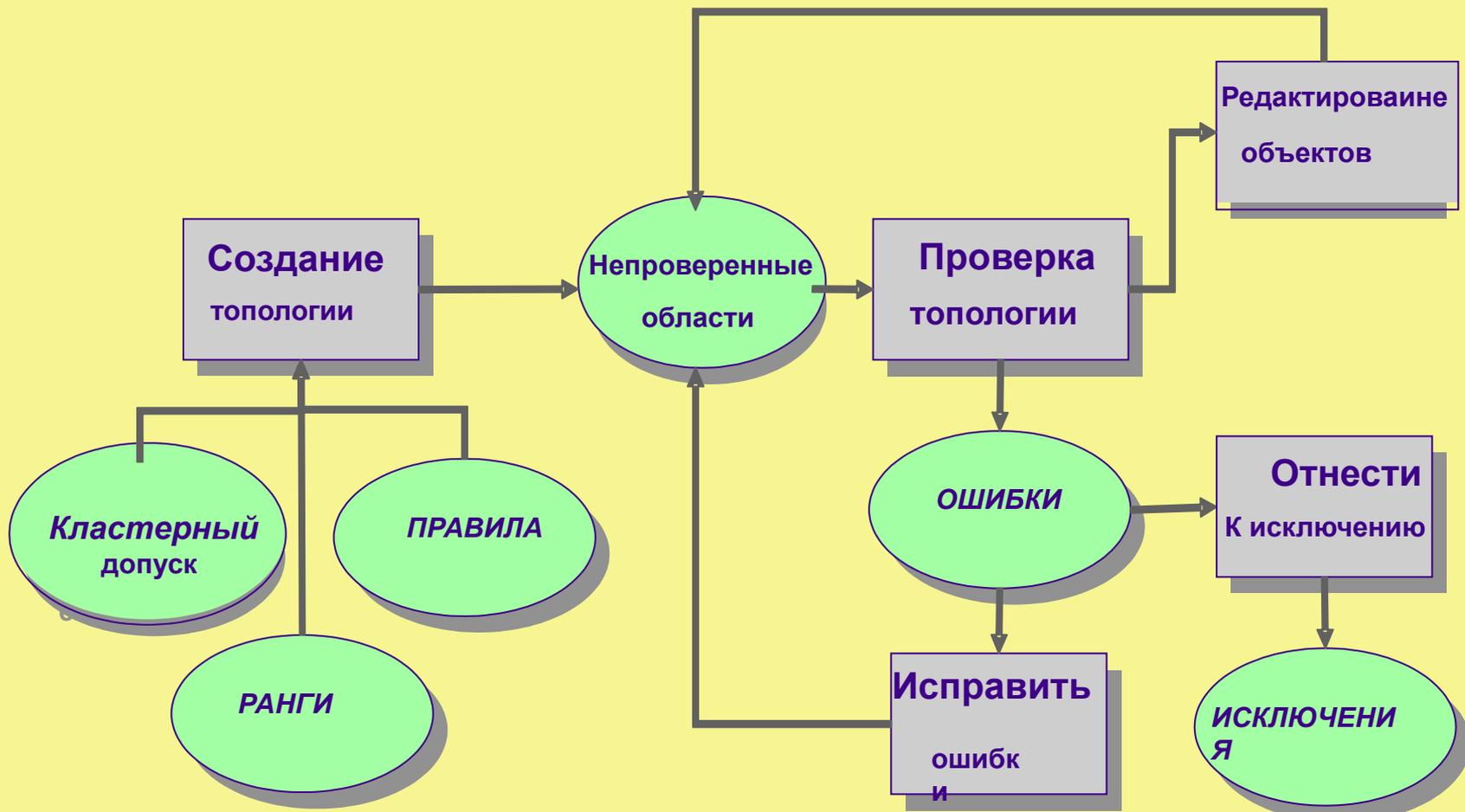
При редактировании объектов создается **измененная область**, чтобы ограничить территорию, в пределах которой необходимо провести вновь проверку топологии

Топологические правила требуют, чтобы здания (синий) не заходили за границу участков (розовый), **Ошибки топологии** отображаются фиолетовым цветом.





Создание и работа с топологией в БГД





Типы баз геоданных

Тип базы геоданных	СУБД	Примечания
Персональная база геоданных	Microsoft Jet Engine (Access)	<ul style="list-style-type: none">· Однопользовательское редактирование,· Размер до 2 GB,· Нет поддержки версий.
Многопользовательская база геоданных	<ul style="list-style-type: none">· Oracle,· Oracle с Spatial или Locator,· IBM DB2,· IBM Informix,· Microsoft SQL Server.	<ul style="list-style-type: none">· Требуется шлюз ArcSDE,· Многопользовательское редактирование,· Рабочий процесс с версиями,· Размер и количество пользователей зависят от СУБД.