GIS Data Model

Ashutosh Kumar Jha
Scientist-'SE'
Geoinformatics Department IIRS
akjha@iirs.gov.in

Faculty Profile

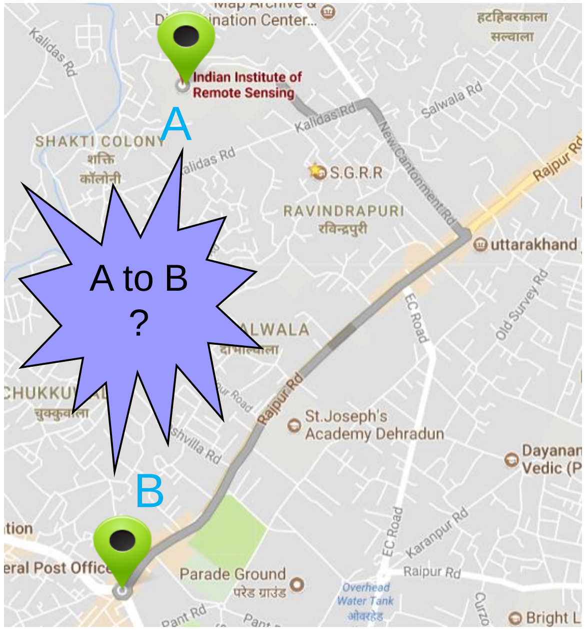
- Ashutosh Kumar Jha is scientist SE in Geoinformatics department. He holds M.Tech. in remote sensing and B.E. in Computer Engineering.
- His area of expertise is Geospatial modeling and processing optimization of raster/vector Data using High performance distributed computing.
- Currently he is working on BigGIS, Machine Learning and 3D Modeling.
- He has been actively involved in Weather Forecast and air quality application development. He has built a open source LULC dynamics modeling framework called OpenLDM. https://github.com/ashutoshkumarjha/OpenLDM
- He has been awarded Best Innovation Award in ACRS-2017 for the development of mobile application for Municipals.

Contents

- How GIS analysis is dependent upon GIS data structure?
- GIS Data Structure
- File based to DBMS based
- Spatial Relationship
- Open Standards based datamodel

GIS Analysis Example

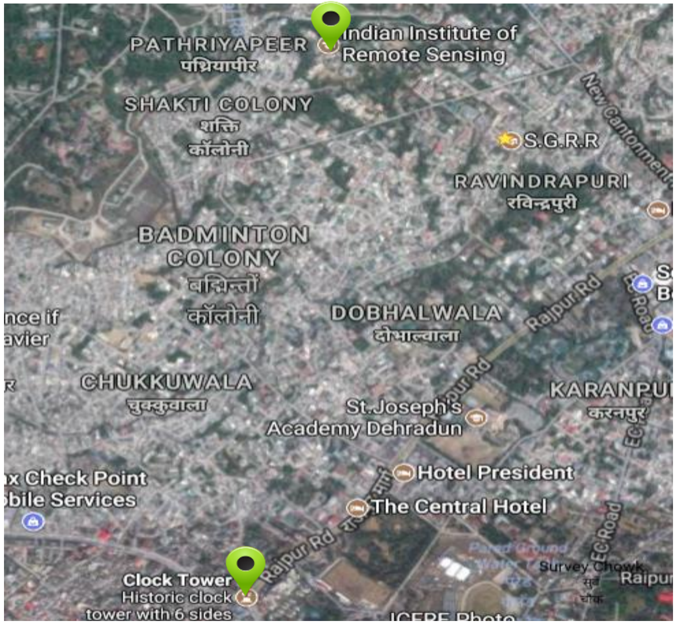
- How to reach a place from A to B



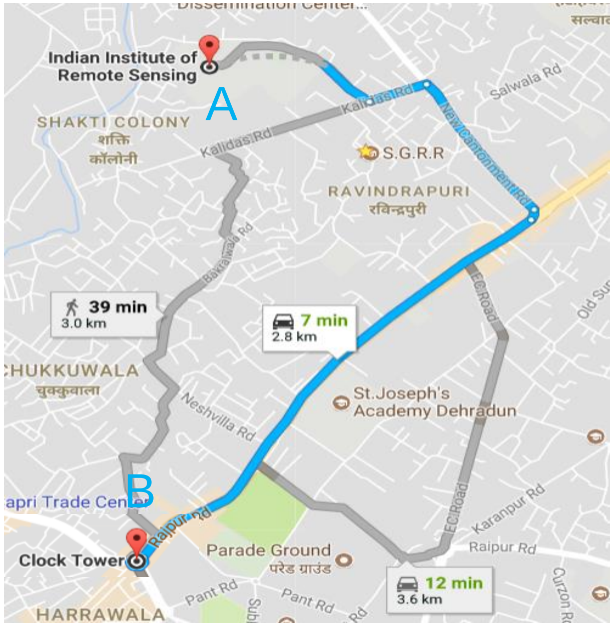
User Requirement
Or
Questions

GIS and Interaction

Reality

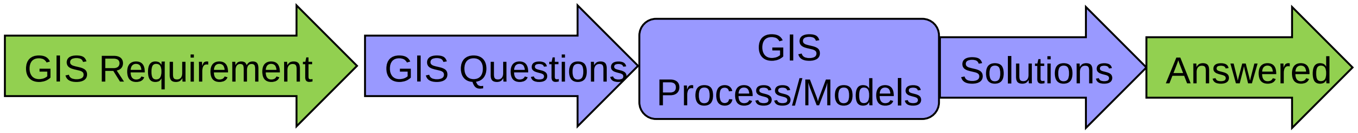


Analysis

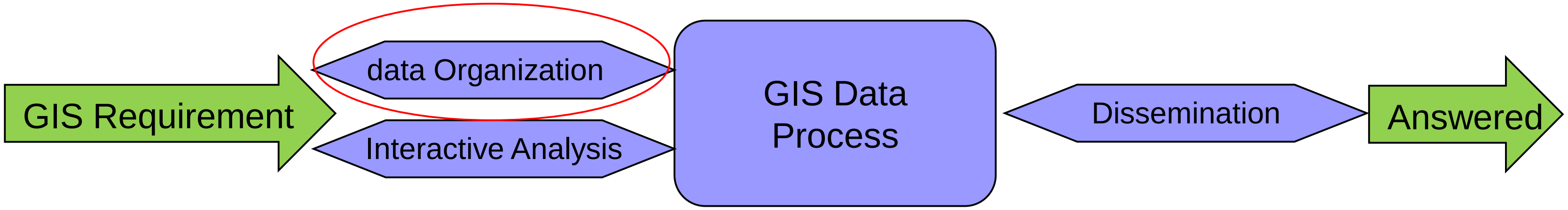


Human

Machine


GIS Requirement → GIS Questions → GIS Process/Models → Solutions → Answered

Use and Uses


GIS Requirement → data Organization / Interactive Analysis → GIS Data Process → Dissemination → Answered

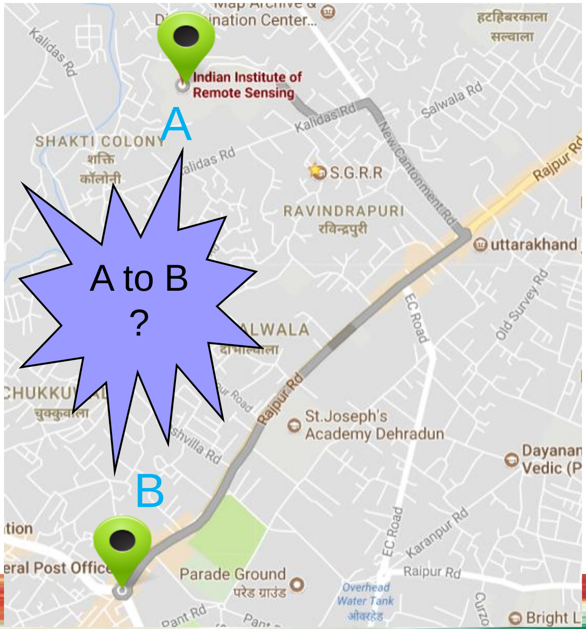
Information and Humane Analysis

Information

- A is IIRS
- B is Clock Tower
- A is at 30°20′3"N+78°02′4"E
- B is at 30°19‘3"N 78°02'30"E
- A is at Kalidas Road
- B is at Junction of Rajpur and Chakrata road
- Kalidas Road is connected to Rajpur Road from new Cant Road

Analysis

- Route A to B is IIRS via Kalidas Road to New Cant Road to Clock tower at Rajpur Road
- Time to reach 8-20 min



Machine Data

🟩 Human

🟦 Machine

Information

- A is IIRS
- B is Clock Tower
- A is at 30°20′3"N,78°02′4"E
- B is at 30°19‵3"N,78°02'30"E
- A is at Kalidas Road
- B is at Junction of Rajpur and Chakrata road
- Kalidas Road is connected to Rajpur Road from new Cant Road

Machine Data

- A is point geometry
- A and B is point geometry with coordinate 30°20′3"N,78°02′4"E and 30°19‵3"N, 78°02'30"E
- New Cant Road, Rajpur road, Kalidas road are Line geometry with intermediates **segments**
- Each line average travel time entered are 4 min, 7 min , 6min are **attributes**

Analysis

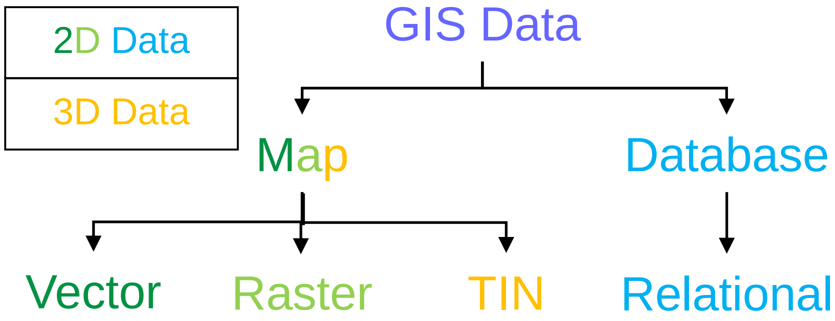
- Time to reach 7 min now

Find the Path from location A (IIRS) to location B (Clock tower)

Machine Data Organisation

Machine Data

- A is point geometry
- A and B is point geometry with coordinate 30°20’3"N+78°02’4"E and 30°19‘3"N 78°02'30"E
- New Cant Road, Rajpur road, Kalidas road are Line geometry with intermediates segments
- Each line average travel time entered are 4 min, 7 min , 6min



GIS Spatial Data

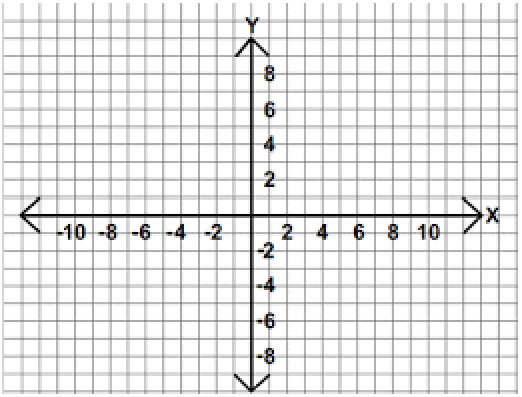
- **Geometry**
  - □ **Points**
    - **#pId, Location,**
  - □ **Line**
    - **#lid, Points on Line Segments, Line Segments**
- **Attribute**
  - □ **Points**
    - **#pId , Name**
  - □ **Line**
    - **#lId Road Name, Length, travel time**

Extra Auxiliary Data

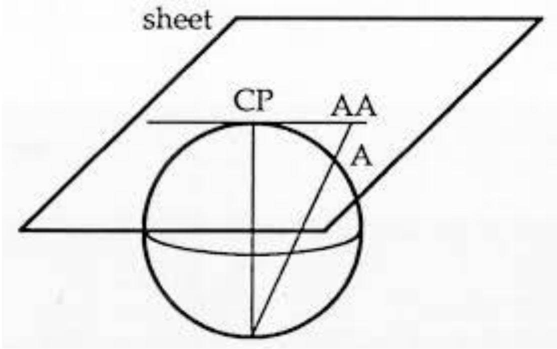
- Coordinate Systems, Extent, scale

GIS Spatial Data Structure

Coordinate System



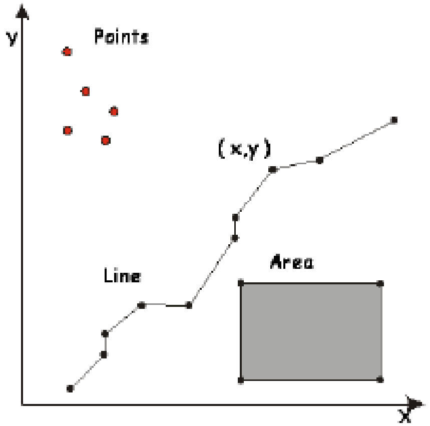
Location Reference Plane



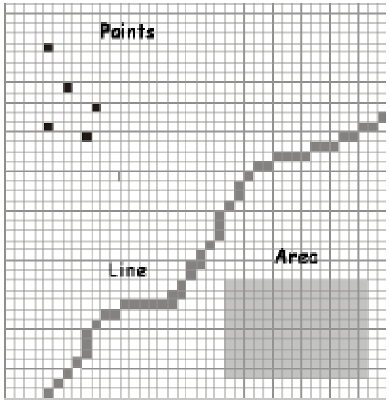
3D globe to 2D
(Spherical to Cartesian)

Spatial Data Format

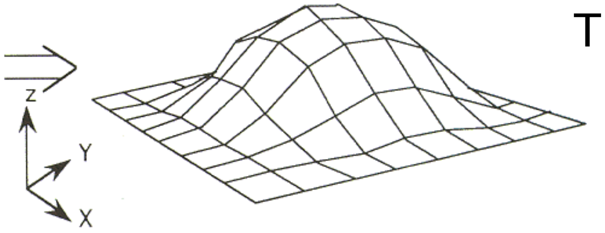
Vector



Raster

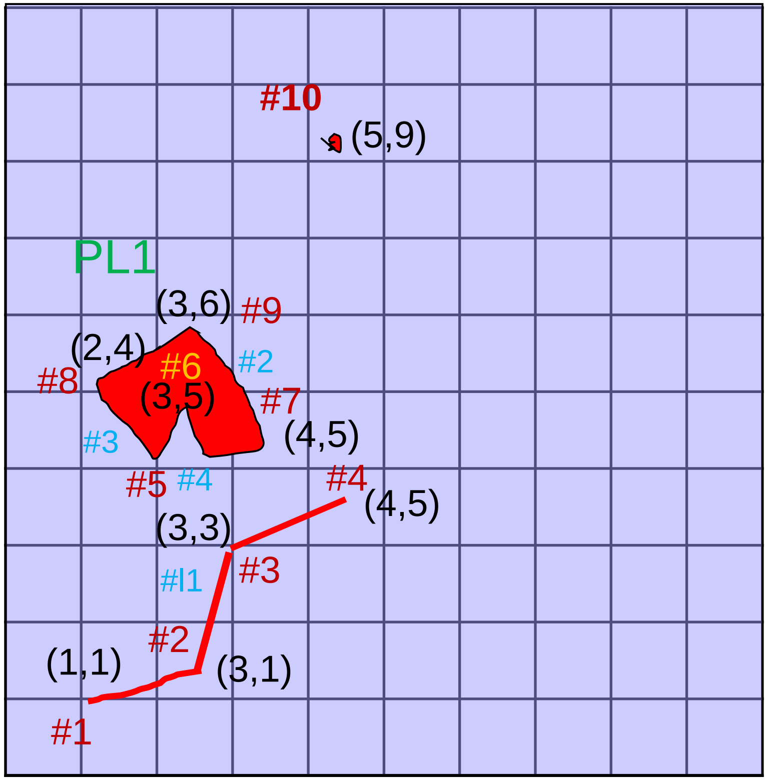


TIN



Map Data Structure

Basic Vector Data Structure



Reference Coordinate System

(Projected or GCS)

Point ID	X	Y
#1	1.xx	1.xx
#2	3.xxx	1.xx
…..	3.xx	3.xx
#10	5.xx	9.xx

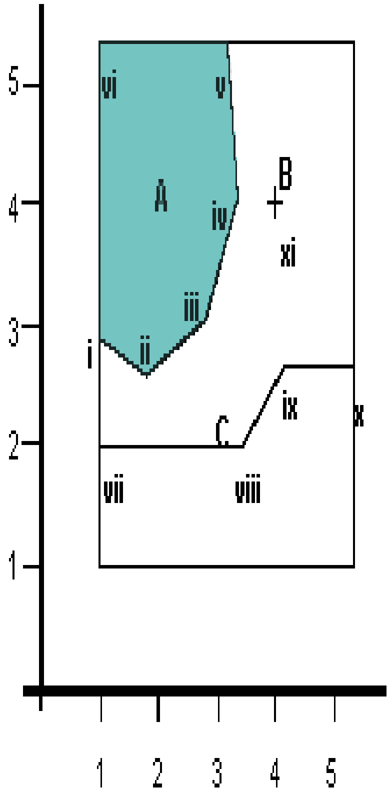
Line ID	Begin Node	End Node	Node List
#1	#1	#4	#1,#2,#3,#4
#2	#7	#9	#7,#8,#9
#4	#5	#4	#5,#6,#4

Polygon ID	Lines
PL1	#2,#3,#4

Tabular Structure

Vector Data Model: Spaghetti

"spaghetti“ Model



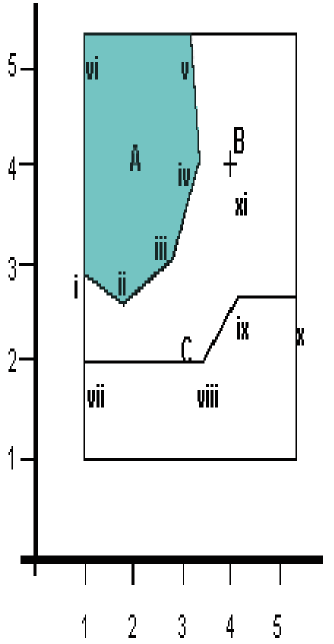
A, 6 (identifier of polygon and number of vertex)
1, 3 (coordinates of the first vertex)
1.8, 2.6
2.8, 3
3.3, 4
3.2, 5.2
1, 5.2
1, 3 (coordinates of the first vertex again)
B, 1 (identifier of the point and number of vertex)
4, 4
C, 4 (identifier of the line and number of vertex)
1, 2
3.5, 2
4.2, 2.7
5.2, 2.7

A: Polygon
B: Point
C: Line

- simple
- easy to manage
- no topology
- lots of duplication, hence need for large storage space
- very often used in CAC (computer assisted cartography)

Vector Data Model:DIME

vertex dictionary



file 1

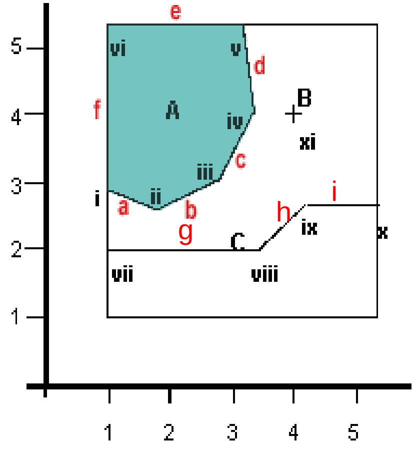
vertex	X	Y
i	1	3
ii	1.8	2.6
iii	2.8	3
iv	3.3	4
v	3.2	5.2
vi	1	5.2
vii	1	2
viii	3.5	2
ix	4.2	2.7
x	5.2	2.7
xi	4	4

file 2

polygon A: i, ii, iii, iv, v,
point B: xi
line C: vii, viii, ix, x

- No duplication
- No topology

Dual Independent Map Encoding (DIME) format



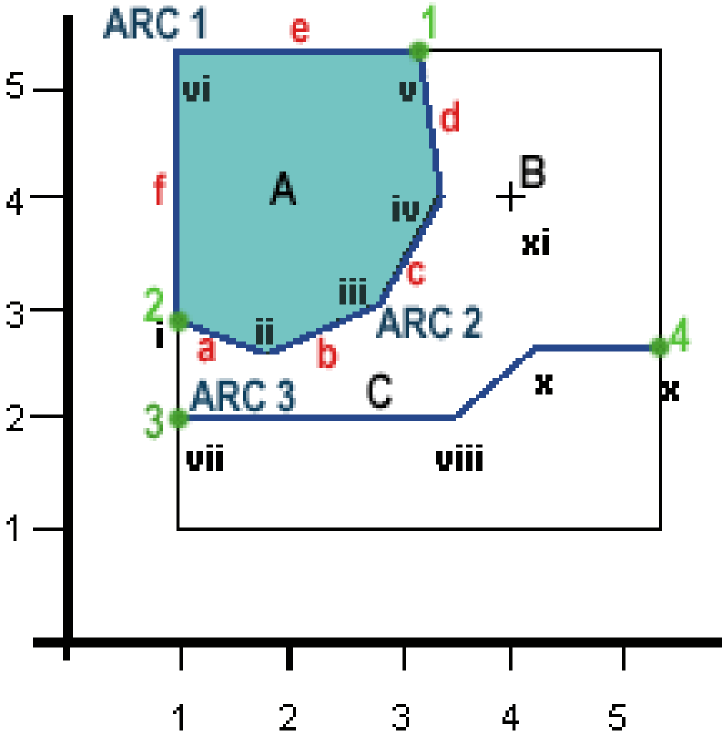
file 1

vertex	X	Y
i	1	3
ii	1.8	2.6
iii	2.8	3
iv	3.3	4
v	3.2	5.2
vi	1	5.2
vii	1	2
viii	3.5	2
ix	4.2	2.7
x	5.2	2.7
xi	4	4

segment	right polygon	left polygon	from vertex	to vertex
a	external	A	i	ii
b	external	A	ii	iii
c	external	A	iii	iv
d	external	A	iv	v
e	external	A	v	vi
f	external	A	vi	i
g	external	external	vii	viii
h	external	external	viii	ix
i	external	external	ix	x

polygon	segments
A	a, b, c, d, e, f

Vector Data Model:ARC



File 1. Coordinates of nodes and vertex for all the arcs

ARC	F_node	Vertex	T_node
1	3.2, 5.2	1, 5.2	1,3
2	1,3	1.8,2.6 2.8,3 3.3,4	3.2, 5.2
3	1,2	3.5,2 4.2,2.7	5.2,2.7

File 2. Arcs topology

ARC	F_node	T_node	R_poly	L_poly
1	1	2	External	A
2	2	1	A	External
3	3	4	External	External

File 3. Polygons topology

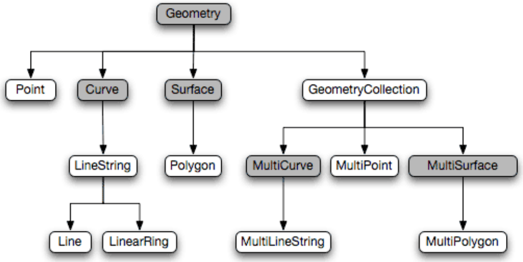
Polygon	Arcs
A	1, 2

File 4. Nodes topology

Node	Arcs
1	1,2
2	1,2
3	3
4	4
5	5

ARC / NODE structure or POLYVRT

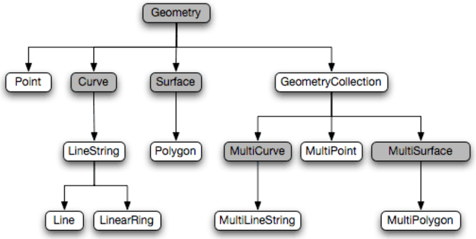
OGC Simple Feature Model



- The *Geometry* class
  - □ Geometry is the root class of the hierarchy. Geometry is an abstract class.
  - □ The subclasses of Geometry are restricted to 0, 1 and 2 dimensional geometric objects that exist in two-dimensional coordinate space.
  - □ All geometry classes are defined so that valid instances of a geometry class are topologically closed (i.e. all defined geometries include their boundary).

OGC Simple Feature Model

The *Geometry* class



Geometry

Point

Curve

Surface

GeometryCollection

LineString

Polygon

MultiCurve

MultiPoint

MultiSurface

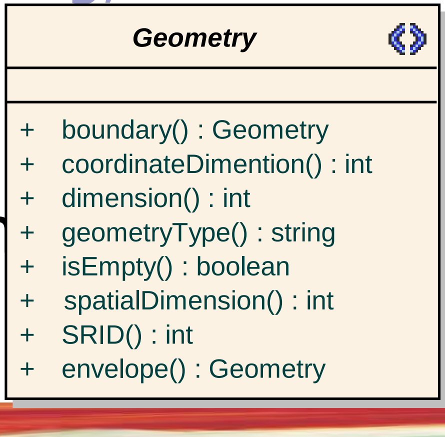
Line

LinearRing

MultiLineString

MultiPolygon

- □ Any geometry has a spatial reference system: the yardstick against which units are measured (use one SRS per project!)
- □ Plain geometries are: Point, LineString, Line, LinearRing, Polygon
- □ Geometric collection types: MultiPoint, -LineString, -Polygon.
- □ Besides the methods of the class, there exist methods for testing Spatial Relations between geometric objects and Spatial Analysis.

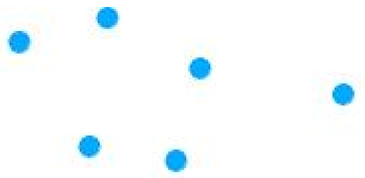


Geometry

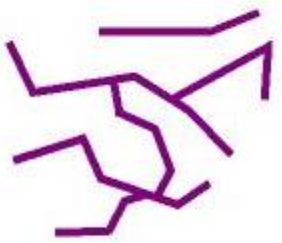
+ boundary() : Geometry
+ coordinateDimention() : int
+ dimension() : int
+ geometryType() : string
+ isEmpty() : boolean
+ spatialDimension() : int
+ SRID() : int
+ envelope() : Geometry

Collection Feature Model

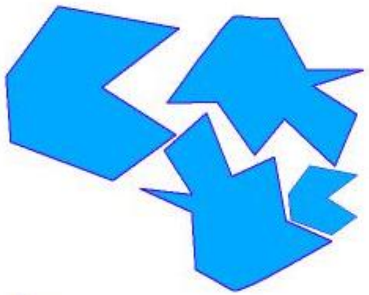
- Single type
  - □ Point, LineString and Polygon
- Homogeneously typed multipart
  - □ **MultiPoint**: a value is a set of points
  - □ **MultiLineString** : a value is a set of linestrings.
  - □ **MultiPolygon** : a value is a set of polygons
- **GeometryCollection**: a value is a hybrid set of singular geometries



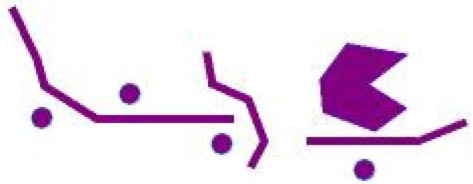
Rest Location



Cycle Path



Department



Department Area

Raster Grid Coordinate System

Col Number(X) →

Line number(Y) ↓

0	1	2	3	4	5	6	7	8	9
1	■	■	■	■	■	■	■	■	■
2	■	■	■	■	■	■	■	■	■
3	■	■	■	■	■	■	■	■	■
4	■	■	■	■	■	■	■	■	■
5	■	■	■	■	■	■	■	■	■
6	■	■	■	■	■	■	■	46	■
7	■	■	■	■	■	■	■	■	■
8	■	■	■	■	■	■	■	■	■
9	■	■	■	■	■	■	■	■	■

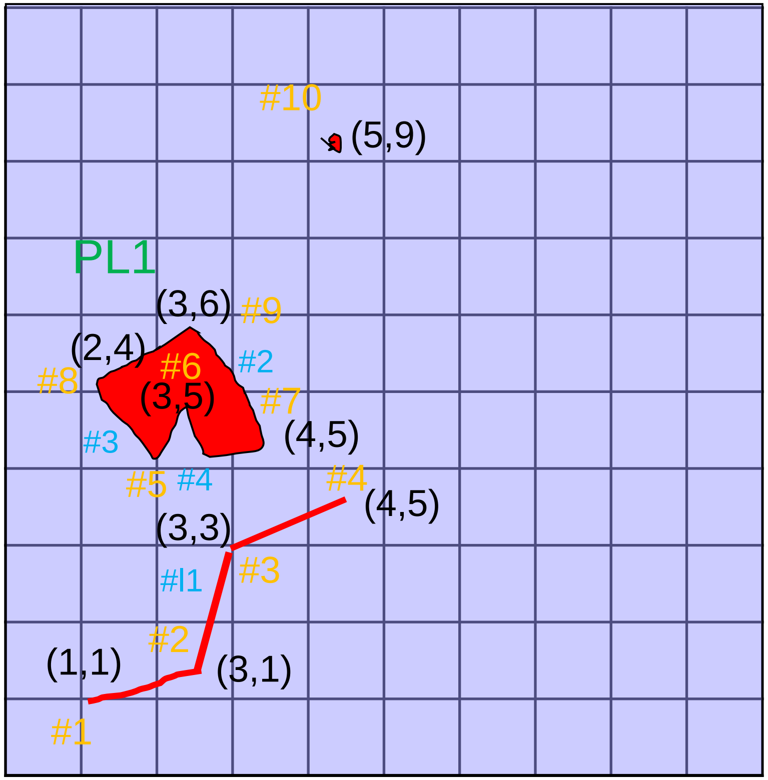
Pixel (8,-4) Center

f(8,-6)=46

- Spatial Arrangement is called Grid
- Value in cell is attribute
- Fixed topology on grid value

- Raster resolution to be chosen: ½ of smallest object
- Geo-referencing point is only to the origin

Raster Data Representation



				1					
	2	2							
	2	2	2						
			3	3					
		3							
	3	3							

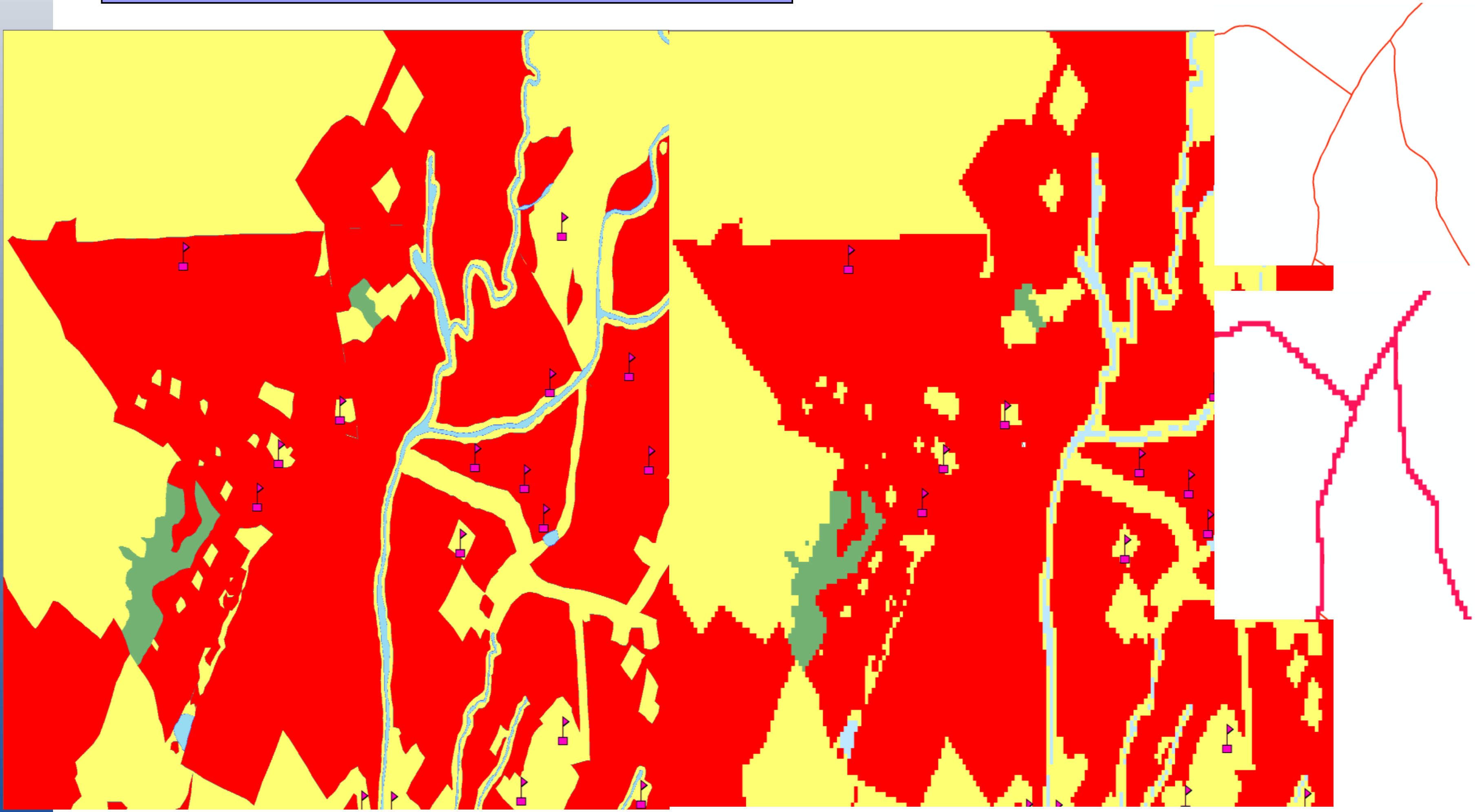
Point

Polygon

Line

Rasterization Effect

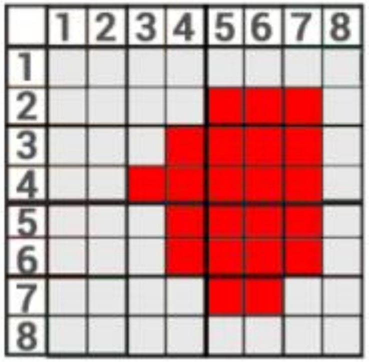
Smoothness of geometry is lost



Raster Data Compression

Run Length Encoding

AAAAAABBBBCCCCCCCCC = 6A4B9C

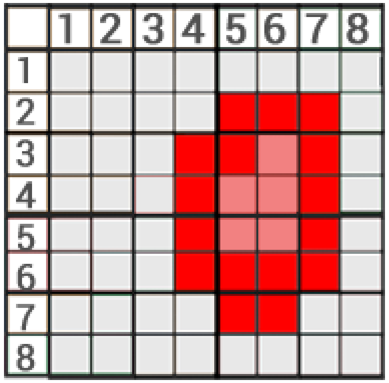


(8, 8, 1)

(0,8)
(0,4) (1,3) (0,1)
(0.3) (1,4) (0,1)
(0,2) (1,5) (0,1)
(0,3) (1,4) (0,1)
(0,3) (1,4) (0,1)
(0,4) (1,2) (0,2)
(0,8)

Grouping Rows of Data

Chain Coding

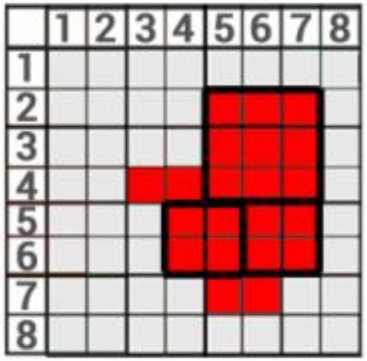


(5,2)

(E3, S4, W1, S1, W1, N1, W1, N3, E1, N1)

Defining the Exterior Boundary

Block Coding



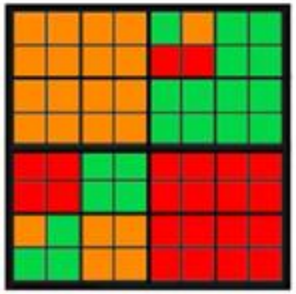
Block Size: 9
Count: 1
Coordinates: 5,2

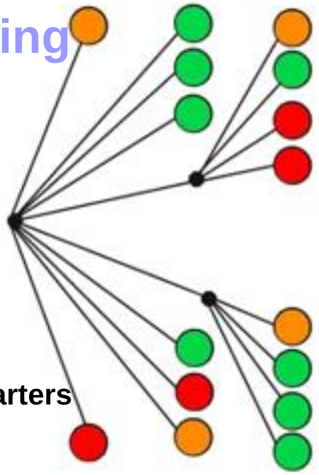
Block Size: 4
Count: 2
Coordinates: (4,5) (6,5)

Block Size: 1
Count: 4
Coordinates: (3,4) (4,4) (5,7) (6,7)

Grouping Blocks of Data

Quadtree Encoding



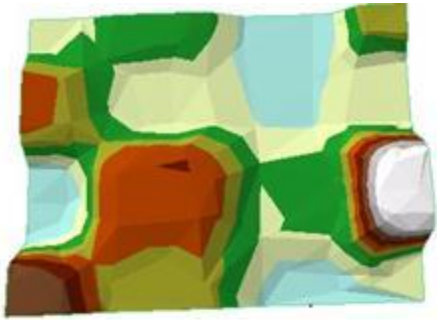


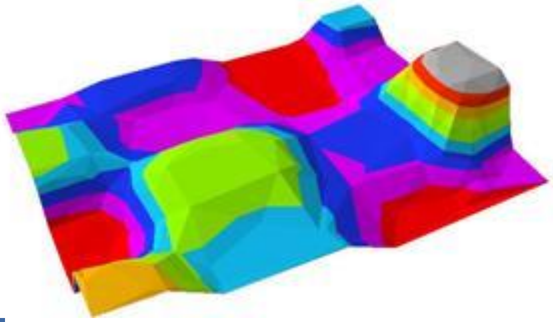
Subdividing Data Into Quarters

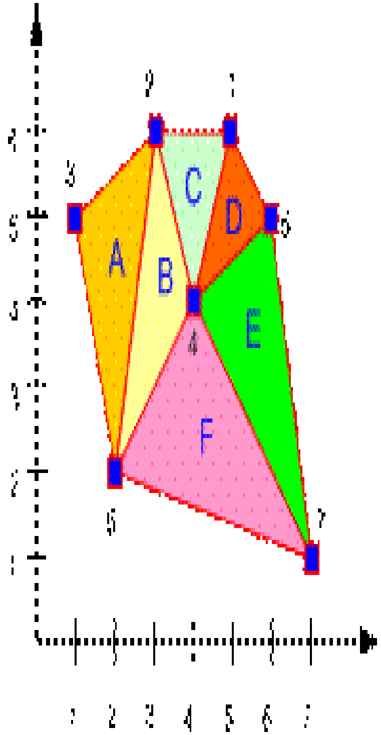
Vector Vs Raster

Vector	Raster
Based on object model .e.g each feature is a bounded	Cell based Modal : Full area field cover
Position of points can be in fraction	X,Y location will always be number
Takes less space	Takes more space
Object based either point, line or polygon	Basic object cell
Geometry based analysis	Algebra based analysis
All Points are referenced	Grids Origin is referenced. Cell location are computed relative to origin
File Type: arcInfo (.e00), shape(.shp), KML, KMZ, OpenSteetFormat(.osm) , Autocad (.dwg and .dxf) , Bently Microsation (*.dgn)	File Type: Geotiff(.tiff), Erdas Imagin (*.img), Scientic Format (.hdf,.nc) etc

Triangulated Irregular Network Model



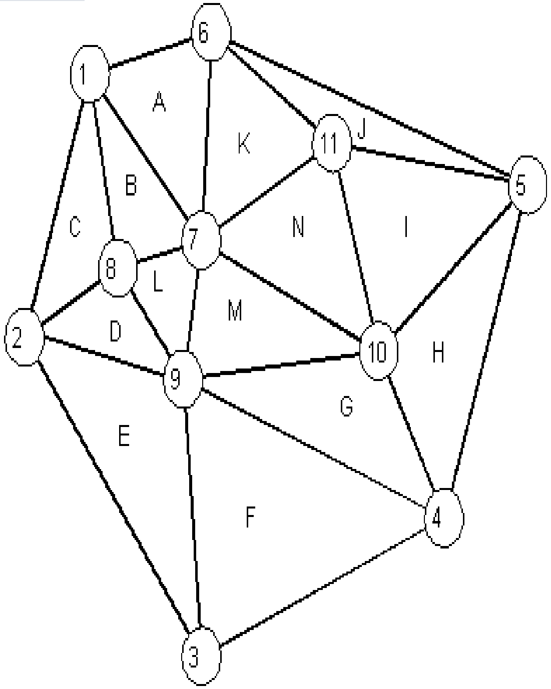




Node No	X	Y	Z
1	5	6	3
2	3	6	5
3	1	5	6
4	4	4	4
5	6	5	3
6	2	2	2
7	7	1	8

Triangle	Node Sequence	Neighbors
A	3,2,6	-,B,-
B	2,4,6	A,C,F
C	2,1,4	-,D,B
D	1,5,4	C,-,E
E	5,7,4	D,-,F
F	4,7,6	B,E,-

TIN Model



X-Y Coordinates	
node#	coordinates
1	x1, y1
2	x2, y2
3	x3, y3
. . .	. . .
11	x11, y11

Z Coordinates	
node#	z_value
1	z1
2	z2
3	z3
. . .	. . .
11	z11

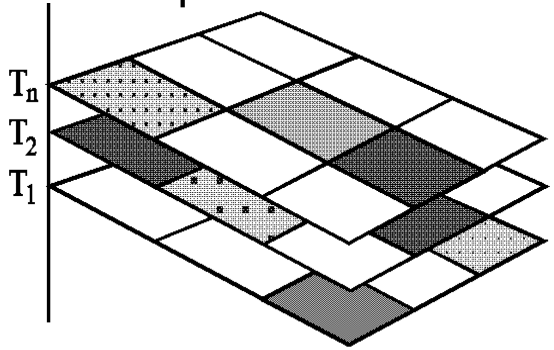
EDGES	
△	adjacent △
A	B, K
B	A, C, L
C	B, D
D	C, E, L
E	D, F
F	E, G
G	F, H, M
H	G, I
I	H, J, N
J	I, K
K	A, J, N
L	B, D, M
M	G, L, N
N	I, K, M

NODES	
△	node#
A	1, 6, 7
B	1, 7, 8
C	1, 2, 8
D	2, 8, 9
E	2, 3, 9
F	3, 4, 9
G	4, 9, 10
H	4, 5, 10
I	5, 10, 11
J	5, 6, 11
K	6, 7, 11
L	7, 8, 9
M	7, 9, 10
N	7, 10, 11

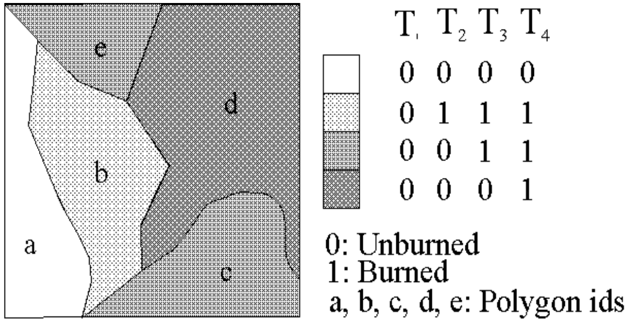
- 3D Data model
- Triangle based
- Delaunay Triangle based
- On triangular height can be computed using interpolation
- Node height are stored

Representing spatiotemporal information Time-stamping

Snapshot Model



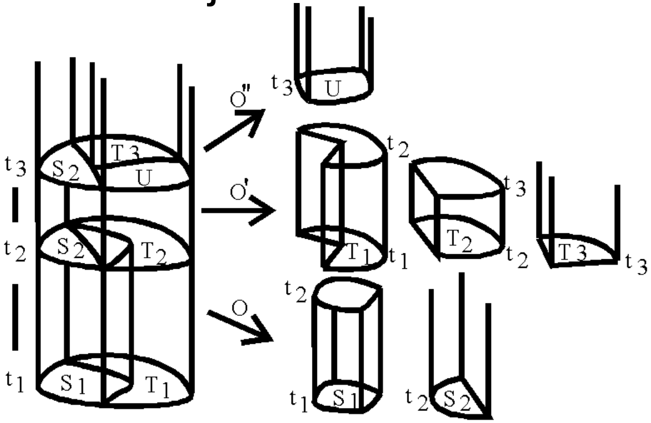
Space-Time Composites (STC)



	T_1	T_2	T_3	T_4
□	0	0	0	0
░	0	1	1	1
▒	0	0	1	1
▓	0	0	0	1

0: Unburned
1: Burned
a, b, c, d, e: Polygon ids

ST-Object model



ST-objects modeling regional change

Decomposition of ST-objects into ST-atoms

- Time Dimension : Can be Irregular
- Discrete Change

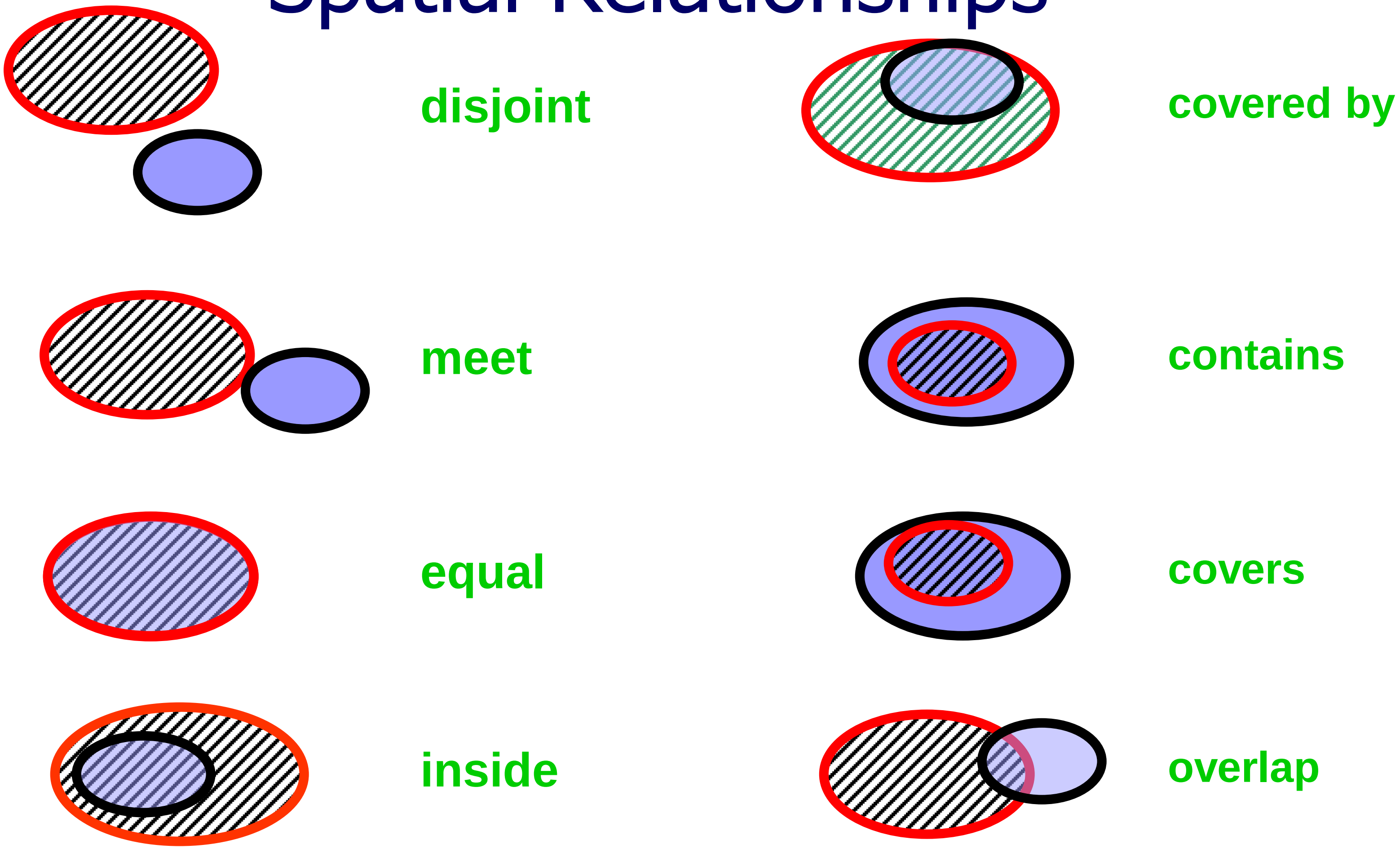
Src: http://loi.sscc.ru/gis/data_model/may.html

Modeling: spatial relationships

- *Topological* relationships: e.g. adjacent, inside, disjoint. Are invariant under topological transformations like translation, scaling, rotation
- *Metric* relationships: e.g. Distance

A way to represent relative relationship between spatial object invariant to any transformation

Spatial Relationships



disjoint

covered by

meet

contains

equal

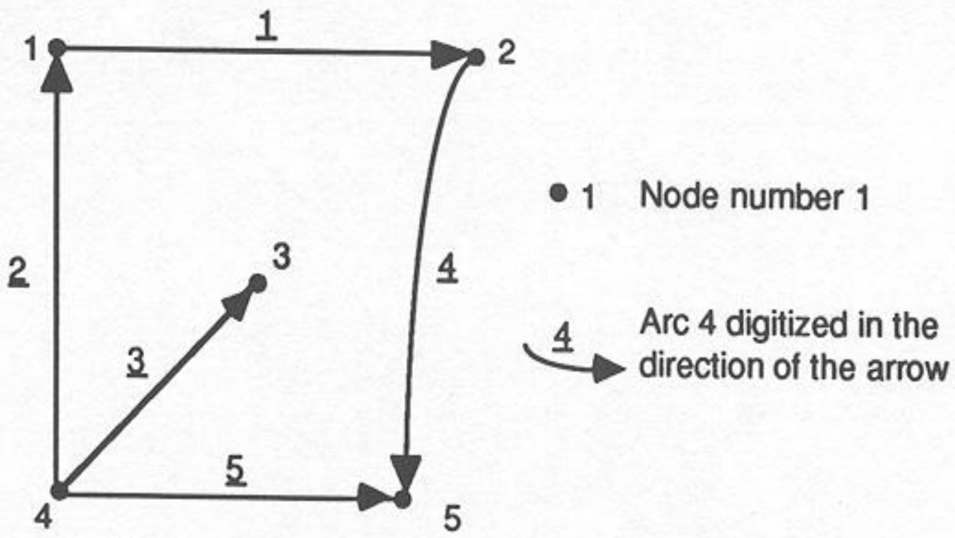
covers

inside

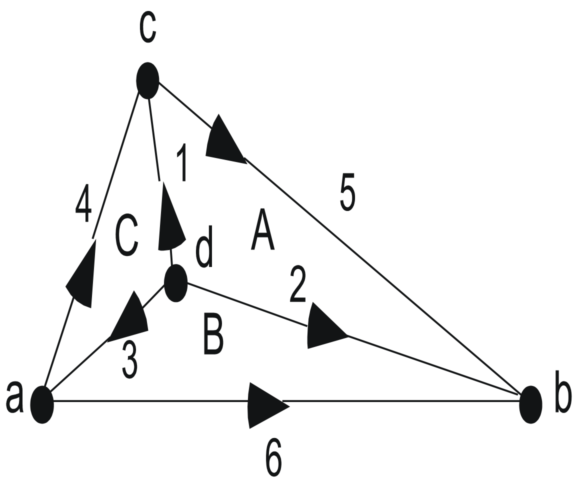
overlap

valid topological relationships between two simple regions (no holes, connected)

Modeling: Building topology

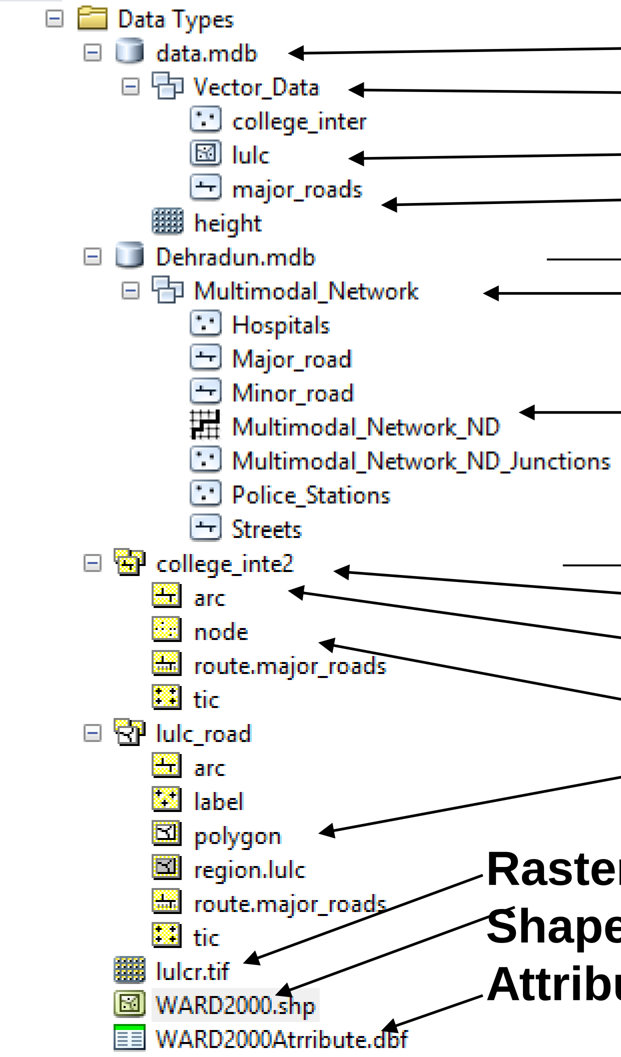


ARC	FNODE	TNODE
1	1	2
2	4	1
3	4	3
4	2	5
5	4	5



arc	from	to	left	right
1	d	c	C	A
2	d	b	A	B
3	d	a	B	C
4	a	c	O	C
5	c	b	O	A
6	a	b	B	O

Spatial File Formats View



- Data Types
  - data.mdb
    - Vector_Data
      - college_inter
      - lulc
      - major_roads
    - height
  - Dehradun.mdb
    - Multimodal_Network
      - Hospitals
      - Major_road
      - Minor_road
      - Multimodal_Network_ND
      - Multimodal_Network_ND_Junctions
      - Police_Stations
      - Streets
  - college_inte2
    - arc
    - node
    - route.major_roads
    - tic
  - lulc_road
    - arc
    - label
    - polygon
    - region.lulc
    - route.major_roads
    - tic
  - lulcr.tif
  - WARD2000.shp
  - WARD2000Atrribute.dbf

Personal Geodatabase

Feature data set

Feature class (feature type = polygon)

Feature class (feature type = arc)

In a gdb, feature class can have only one feature type.

Coverage (= feature class)

Network Data (= feature class)

A coverage can have multiple feature types-now viewed as a shortcoming.

Coverage (= feature class)

Feature type (arc)

Feature type (point)

Feature type (polygon)

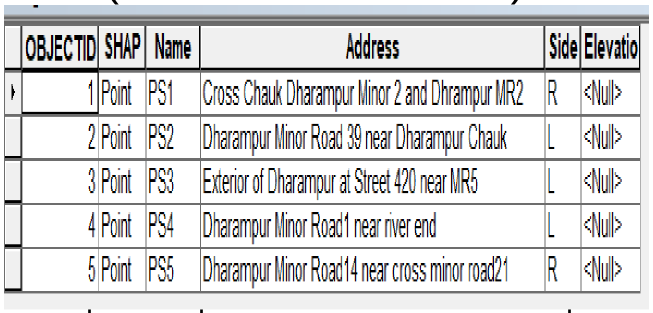
Raster

Shapefile

Attribute (table)

Tracts feature class table (attributes in columns)

Features (rows)



OBJECTID	SHAP	Name	Address	Side	Elevatio
1	Point	PS1	Cross Chauk Dharampur Minor 2 and Dhrampur MR2	R	<Null>
2	Point	PS2	Dharampur Minor Road 39 near Dharampur Chauk	L	<Null>
3	Point	PS3	Exterior of Dharampur at Street 420 near MR5	L	<Null>
4	Point	PS4	Dharampur Minor Road1 near river end	L	<Null>
5	Point	PS5	Dharampur Minor Road14 near cross minor road21	R	<Null>

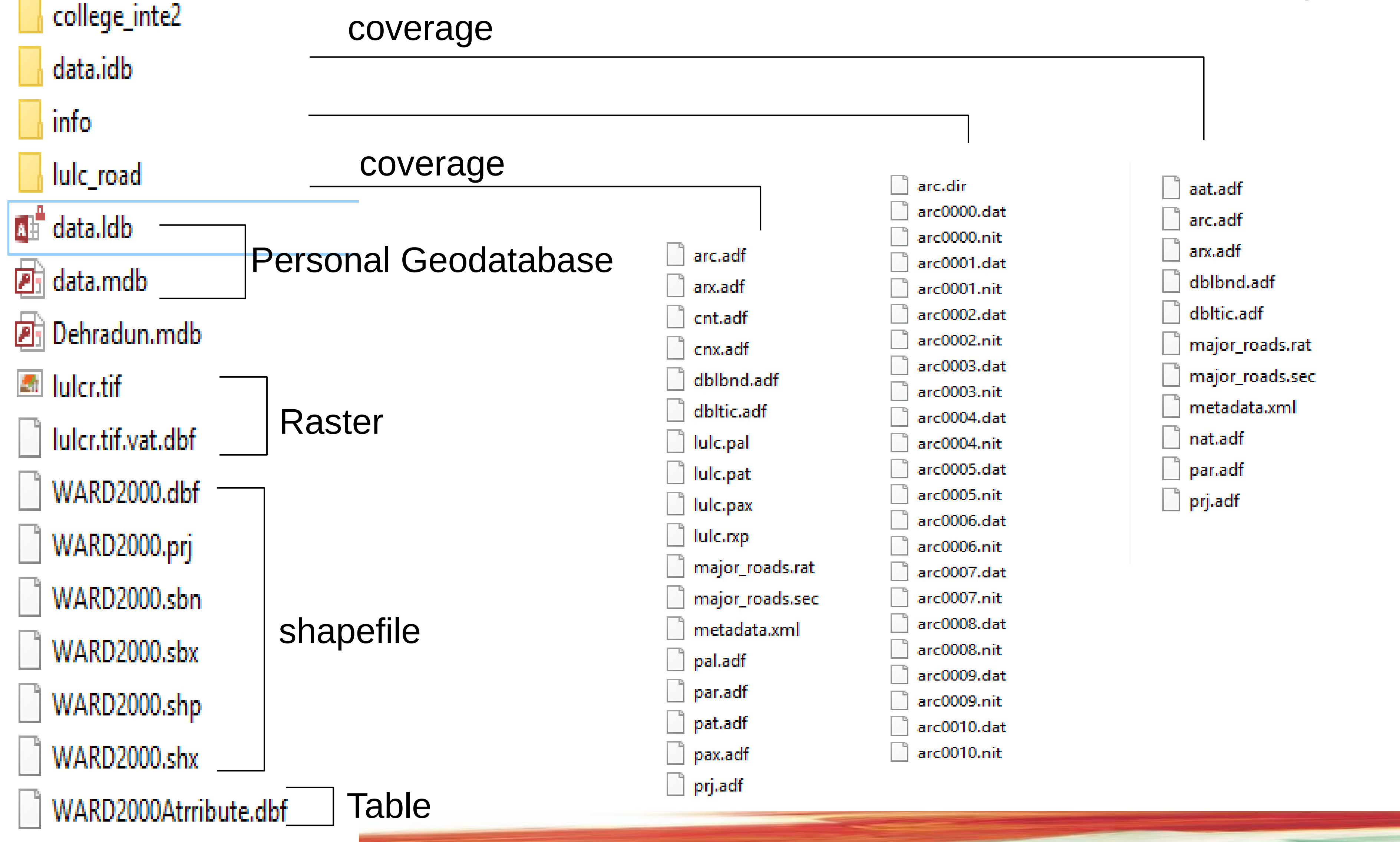
Feature ID (key field)

Feature type

Secondary or Foreign key

Spatial File Formats

Info 'master' folder for AVCAT workspace



- college_inte2 — coverage
- data.idb
- info
- lulc_road — coverage
- data.ldb — Personal Geodatabase
- data.mdb
- Dehradun.mdb
- lulcr.tif — Raster
- lulcr.tif.vat.dbf
- WARD2000.dbf — shapefile
- WARD2000.prj
- WARD2000.sbn
- WARD2000.sbx
- WARD2000.shp
- WARD2000.shx
- WARD2000Atrribute.dbf — Table

lulc_road:

- arc.adf
- arx.adf
- cnt.adf
- cnx.adf
- dblbnd.adf
- dbltic.adf
- lulc.pal
- lulc.pat
- lulc.pax
- lulc.rxp
- major_roads.rat
- major_roads.sec
- metadata.xml
- pal.adf
- par.adf
- pat.adf
- pax.adf
- prj.adf

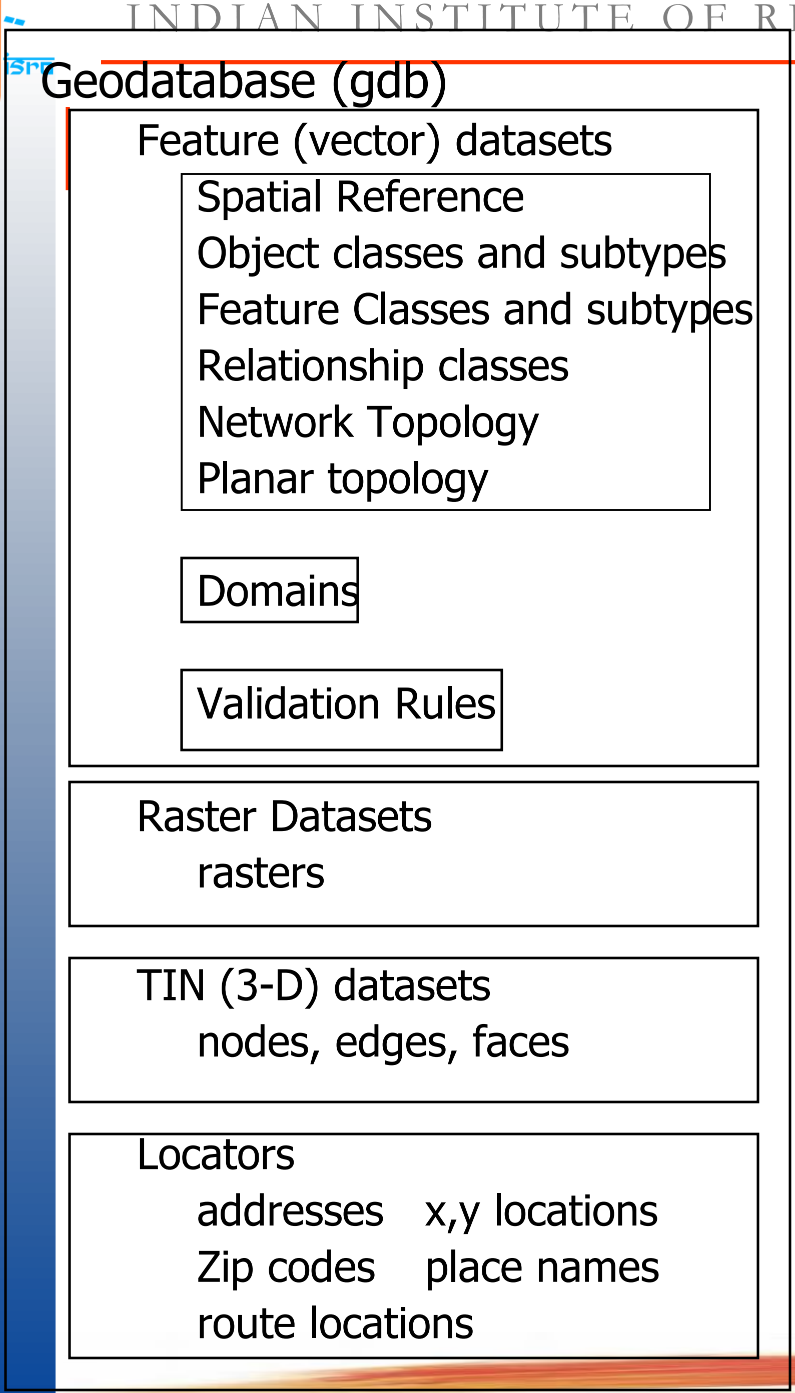
info:

- arc.dir
- arc0000.dat
- arc0000.nit
- arc0001.dat
- arc0001.nit
- arc0002.dat
- arc0002.nit
- arc0003.dat
- arc0003.nit
- arc0004.dat
- arc0004.nit
- arc0005.dat
- arc0005.nit
- arc0006.dat
- arc0006.nit
- arc0007.dat
- arc0007.nit
- arc0008.dat
- arc0008.nit
- arc0009.dat
- arc0009.nit
- arc0010.dat
- arc0010.nit

college_inte2:

- aat.adf
- arc.adf
- arx.adf
- dblbnd.adf
- dbltic.adf
- major_roads.rat
- major_roads.sec
- metadata.xml
- nat.adf
- par.adf
- prj.adf

Anatomy of a Geodatabase



Geodatabase (gdb)

Feature (vector) datasets

Spatial Reference
Object classes and subtypes
Feature Classes and subtypes
Relationship classes
Network Topology
Planar topology

Domains

Validation Rules

Raster Datasets
rasters

TIN (3-D) datasets
nodes, edges, faces

Locators
addresses x,y locations
Zip codes place names
route locations

Geodatabases may contain: feature datasets, raster datasets, TIN datasets, locators

Feature datasets contain vector data

All data in a single feature dataset share a common *spatial reference* system

Similar *Objects* (e.g. Jane Blow, land owner) are instances of *object classes* (e.g. land owners) and have **no** spatial form.

Features and feature classes are **spatial** objects (e.g. land parcels) which are similar and have same spatial form (e.g. polygon)

Object (or feature) classes are the tables, and *objects* (or *features*) are the rows of the table

Attributes are in the columns of the table

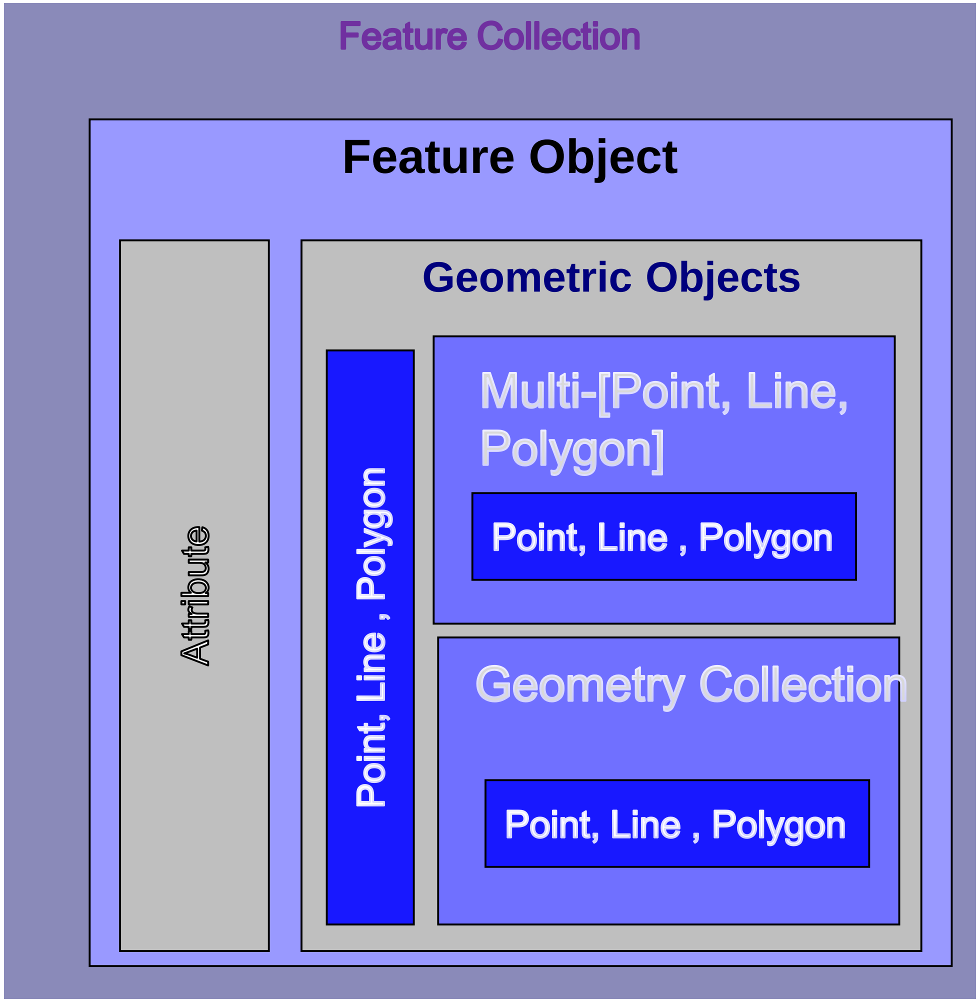
Subtypes are an alternative to multiple object (or feature) classes (e.g. 'concrete', 'asphalt', 'gravel' road subtypes): think of subtype as the most significant classification variable (attribute) in the class table

Domains define permitted data values.

Topology is saved as a relationship between the feature classes in the feature dataset.

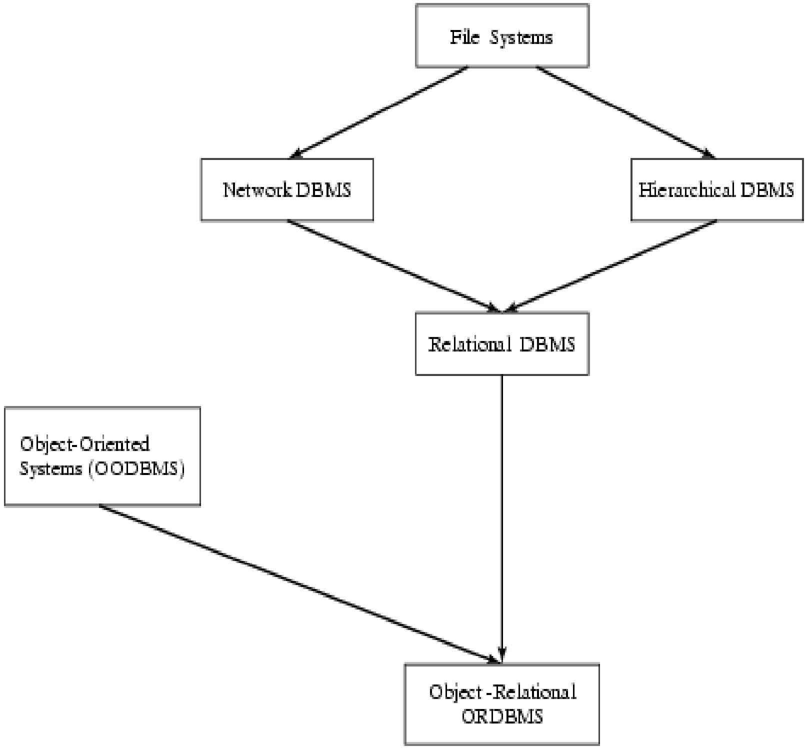
Geojson

- **FeatureCollection**
- **Features**
  - □ **Attribute**
  - □ **Geometric Objects**
    - Point
    - MultiPoint
    - LineString
    - MultiLineString
    - Polygon
    - MultiPolygon
    - GeometryCollection

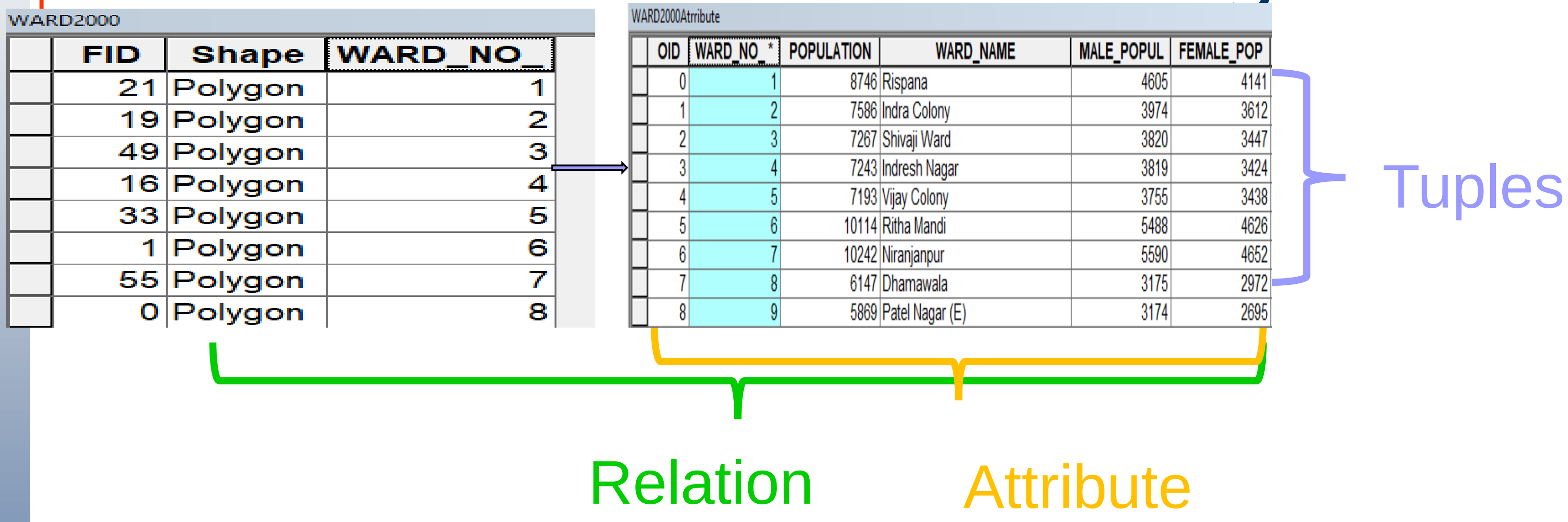


Evolution of DBMS technology

- Flat file Model
- Hierarchical Model
- Network Model
- Relational Model
- Object-Relational



Attribute Data Relational Organisation



WARD2000

	FID	Shape	WARD_NO_
	21	Polygon	1
	19	Polygon	2
	49	Polygon	3
	16	Polygon	4
	33	Polygon	5
	1	Polygon	6
	55	Polygon	7
	0	Polygon	8

WARD2000Attribute

	OID	WARD_NO_*	POPULATION	WARD_NAME	MALE_POPUL	FEMALE_POP
	0	1	8746	Rispana	4605	4141
	1	2	7586	Indra Colony	3974	3612
	2	3	7267	Shivaji Ward	3820	3447
	3	4	7243	Indresh Nagar	3819	3424
	4	5	7193	Vijay Colony	3755	3438
	5	6	10114	Ritha Mandi	5488	4626
	6	7	10242	Niranjanpur	5590	4652
	7	8	6147	Dhamawala	3175	2972
	8	9	5869	Patel Nagar (E)	3174	2695

- *Table Based*
- *One-One,One-Many,Many-One Relatioship*
- SQL(Strucure based query for tuples to build new relatioship or filter relationship
- Relational Alzebra based data organisation for consistency using normalisation

Thank You

Contact Details of the Faculty:

Email- akjha@iirs.gov.in
Tel- 0135-2524134

Question

- Vector Data has georeferenced information for each point (**Y**/N)
- Which of these data structure is based on tessellation?
  - □ 1. Simple Vector Point
  - □ 2. Raster
  - □ **3. TIN**
  - □ 4. Polygon
- Which of these geometrical models have more than one features?
  - □ 1. Multipoint
  - □ 2. Multiline String
  - □ 3. Multipolygon
  - □ 4**. Geometry Collection**
- Which of these is used for raster data format?
  - □ **1. Run length encoding Model**
  - □ 2. Spaghetti Model
  - □ 3. Dictionary Model
  - □ 4.Dime Model
- How many topological neighbours are there of a vector polygon?
  - □ 1. 8
  - □ 2. 4
  - □ **3. n-numbers**