VIEPS GIS Shortcourse

Lecture 3: Workflows and Data types

Map Scale

- A ratio or proportion between distances measured on the map vs. that measured on the ground
- Representative fraction (RF)
  - 1:100,000 or 1/100,000
    - Map distance on the left
      - Always = 1
    - Ground distance on the right
      - Varies with scale
    - Units are always the same

Map Scale

- Statement scale
  - Similar to ratio scale
  - Ratio is expressed in words
    - 1:1000000 becomes “one cm **to** ten km”
    - 1:63360 becomes “one inch **to** one mile”
  - Useful for odd scale maps
    - imperial

Map Scale

- Bar Scale
  - Linear graphical scale
  - Allows estimation of ground distances from measurements on the map

Scale

- Scale is a very important aspect of map design
- GIS is scale independent
- Printed outputs of GIS are scale limited
- Need to decide likely scale of output early
  - Effects planning and data gathering

Map Scale

- Large scale maps
  - 1:1000 – 1:10000
  - Map small features
  - Maintain geometric shapes
- Meduim scale maps
  - 1:20000 – 1:1000000
  - Small features are gone
  - Shapes are generalised
  - Comprimises
- Small scale maps
  - 1:1000000 >
  - Symbolised features
  - Large features can be represented - geology

Map Scale

- Spatial vs Temporal scale
  - Temporal – measurements over time
    - Hydraulic head in a bore
    - Land surface morphology
  - Temporal scale
    - Refers to the number of measurements of an event over a period of time
      - Higher measurement frequency
      - Higher temporal resolution
      - Finer temporal scale

Map Classification

- Topographic vs thematic
  - Topographic / General
    - Not designed for a specific application
    - Accurate and consistent
    - Location based
    - Geographic reference for thematic maps and data integration

Map Classification

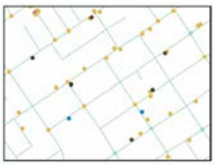
- Topographic vs thematic
  - Thematic Maps
    - Application specific features
      - Geology
      - Structure
      - Geophysics
    - various scales based on topographic maps
    - Geographically referenced biophysical or socioeconomic data
    - Graphical overlay

GIS Analysis

- Showing the geographic distribution of data
- Querying GIS data
- Identifying what is nearby
- Overlaying different layers
- Doing a complex analysis

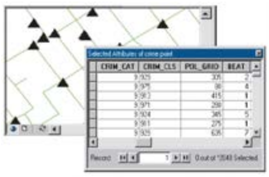
Showing the geographic distribution of data

- The simplest form of GIS analysis is presenting the geographic distribution of data
- Sticking pins in map on the wall
- Pattern detection



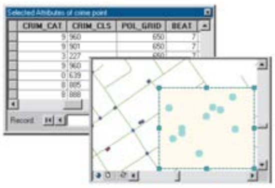
Querying GIS data

- Select data in the database.
- Identify and focus on a specific set of features.
- *Attribute queries*, also called aspatial queries, find features based on their attributes.



Querying GIS data cont.

- Location queries, also called spatial queries, find features based on where they are.
- Eg. all geochemical data from Broken Hill



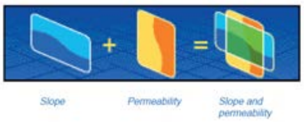
Identifying what is nearby

- Creating a buffer around the feature

- The output of one procedure can be used in another.

Overlaying different layers

- You can create new information when you overlay one set of features with another.
- There are several types of overlay operations, but all involve joining two existing sets of features into a single new set of features.



GIS project elements

- Project design (objective)
- Project design (structure)
- Data input (capture)
- Data editing (preprocessing)
- Data management (organization)
- Data exploration & analysis
- Data display (presentation)

- Metadata, data quality, standards

Identify the objective

- Consider the following questions when you are identifying your objectives:
  - What is the problem to solve? How is it solved now? Are there alternate ways to solve it using a GIS?
  - What are the final products of the project - reports, working maps, presentation-quality maps?
  - Who is the intended audience of these products - the public, technicians, planners, officials?
  - Will the data be used for other purposes? What are the requirements for these?
- This step is important because the answers to these questions determine the scope of the project as well as how you implement the analysis.

Structural design

- identifying the spatial data you will need based on the requirements of the analysis
- determining the required feature attributes
- setting the study area boundary
- choosing the coordinate system to use

Data Capture

- The process of converting data (spatial, e.g, maps and attribute, e.g., field survey) into a form usable by a GIS
- E.g., digitizing, scanning, typing, downloading, converting from other digital sources

Data Capture methods

- Digitizing – the process of converting line or area features on paper into digital vector format
- Scanning – images on paper to digital raster format
- Database entry from keyboard – Microsoft Access, Oracle, FileMaker
- Survey data entered from total station files
- GPS locational data
- Aerial photos can be digital or scanned
- Satellite images always digital

Data Preprocessing or Editing

- Making spatial and attribute data usable, checking for errors
- Includes:
  - creating/building attribute tables
  - creating topology (spatial relationships)
  - checking and correcting errors
  - adding missing data

Data Management

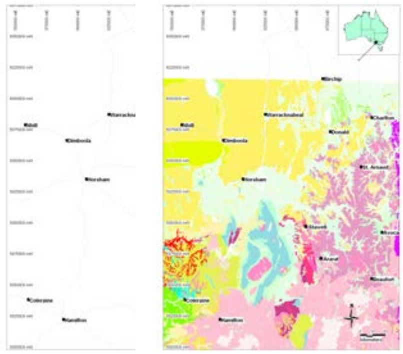
- How data are stored, retrieved, and how it is related within the system
  - what format data is stored in
  - how to use the information stored within the database
  - how is the spatial data related to the non-spatial attributes
- Think about how you would write an essay
  - this stage is like creating an outline, deciding on what part goes where, etc.

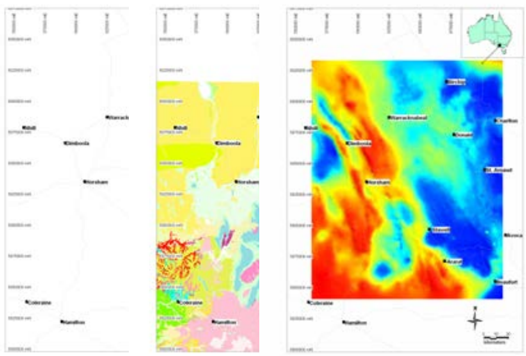
Data Analysis

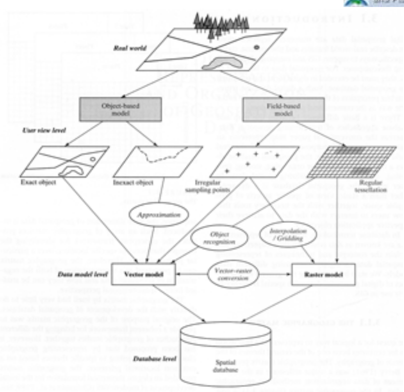
- Using the data to answer questions posed by the user (analysis)
  - Simple mapping
  - Geometric modeling functions - calculating distances, generating buffers, and calculating areas and perimeters.
  - Coincidence modeling functions - overlaying datasets to find places where values coincide.
  - Adjacency modeling functions - allocating, pathfinding, and redistricting.
  - Statistical analysis
  - Attribute manipulation

Data Presentation

- Any output from a GIS
  - Softcopy – displayed as a graphic image on display screen to be seen by another human
  - Hardcopy – reports, charts, or maps printed on any durable media (paper)
  - Computer Compatible – tapes or disks for archival storage or transfer to another system
- Your final product should effectively communicate your findings to your audience.
- In most cases, the results of a GIS analysis can best be shown on a map.
  - Charts
  - Reports
  - These can be printed separately, embedded in documents, or placed on your map.







Raster Data

- Geospatial information represented as a surface that is divided into a regular grid of cells
    - Grid paper
    - TV screen
    - Digital images
- The x,y coordinates of at least one corner of the raster are known, so it can be located in geographic space.
- Useful for storing and analyzing data that is continuous across an area.
    - Each cell contains a value that can represent membership in a class or category, a measurement, or an interpreted value
- Images and grids
    - Images, such as an aerial photograph, a satellite image, or a scanned map, are often used for generating GIS data

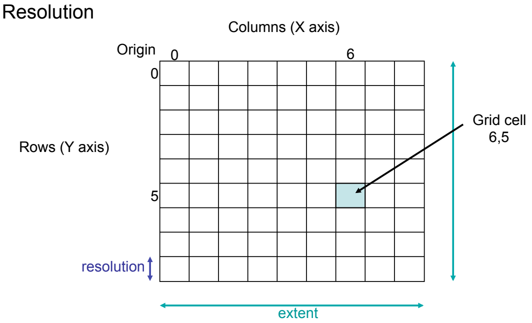
- Grids represent derived data
    - Used for analysis and modeling
    - Created from sample points, such as for a surface of chemical concentrations in the soil
    - Classification of an image, such as for a land cover grid. Grids can also be created by converting vector data.
- Continuous values
    - elevation surface
    - density
- Categories
    - rock type
    - additional attributes about each category
        - numeric code
        - name of the rock type
        - a habitat suitability rating for certain wildlife species, and a general type code. This is unlike feature data, where attributes are stored for each individual feature.

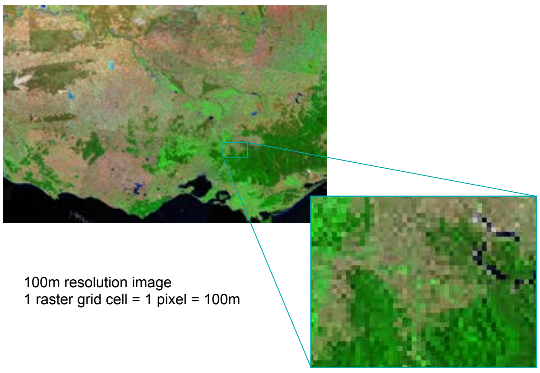
Raster data

- Cell size controls resolution
  - decreasing the cell size to store higher resolution data substantially increases the total volume of data that must be stored.
  - Raster grid cells are also sometimes referred to as pixels, which stands or *picture element*
    - Usually square
    - Word “pixel” - usually reserved for computer photography or satellite imagery processing
    - Preferred term In GIS - “raster grid cells”

Grid Structure

- Extent
- Resolution





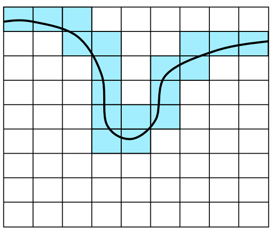
100m resolution image
1 raster grid cell = 1 pixel = 100m

Raster data

- Advantages
  - many data sets available
  - easy to overlay multiple themes
  - able to represent multiple continuous surfaces
  - different file formats readily inter-converted
  - fast computer lookup and display
- Disadvantages
  - poor representation of discrete objects
  - constant resolution throughout region modeled
  - exact boundary location difficult
  - difficult to change projection or coordinate system
  - generates very large data sets

Raster data

- Encoding issues
  - Every cell must have a value
    - 1-255 - 8-bit image
  - Every cell can have only *one* value
  - Diagonal cells may be counted as connected or disconnected
  - Points and lines have to move to the center of grid cells
    - Lines may get fat
    - Areas may need separately coded edges (different color than the area itself) to smooth out the transition



Encoding and compression

- Binary raster grid
  - 2 values

0	0	0	0	0	0	0	0	0
0	0	0	1	1	0	0	0	0
0	0	1	1	1	0	0	0	0
0	1	1	1	0	0	0	0	0
0	1	1	0	0	0	0	0	0
0	0	0	0	0	0	1	1	0
0	0	0	0	0	0	1	1	0
0	0	0	0	0	0	1	1	1
0	0	0	0	0	0	0	1	1

Encoding and compression

- 10 X 10 grid = 100 bytes
- Need to compress data due to large file sizes
- Many compression algorithms
- Run-length encoding
- Value point encoding
- Quadtree compression

A	A	A	A	B	B	B	B	B	B
A	A	A	A	A	B	B	B	B	B
A	A	A	A	A	B	B	B	B	B
A	A	A	A	A	A	B	B	B	B
A	A	A	A	A	D	D	D	D	B
C	C	C	C	D	D	D	D	D	D
C	C	C	C	D	D	D	D	D	D
C	C	C	C	C	D	D	D	D	D
C	C	C	C	C	D	D	D	D	D
C	C	C	C	D	D	D	D	D	D

Run-length encoding

- The value of the attribute
- The number of cells in the run
- The row number

A	A	A	A	B	B	B	B	B	B
A	A	A	A	A	B	B	B	B	B
A	A	A	A	A	B	B	B	B	B
A	A	A	A	A	A	B	B	B	B
A	A	A	A	A	D	D	D	D	B
C	C	C	C	D	D	D	D	D	D
C	C	C	C	D	D	D	D	D	D
C	C	C	C	C	D	D	D	D	D
C	C	C	C	C	D	D	D	D	D
C	C	C	C	D	D	D	D	D	D

Value	Length	Row
A	4	1
B	6	1
A	5	2
B	5	2
A	5	3
B	5	3
A	4	5
D	4	5
B	1	5
C	4	6
D	6	6
C	4	7
D	6	7
C	5	8
D	5	8
C	5	9
D	5	9
C	4	10
D	6	10

3 x 19 = 57 bytes

Value point encoding

- The position number for the end of each run is stored in the point column
- The value for each cell is stored in the value column

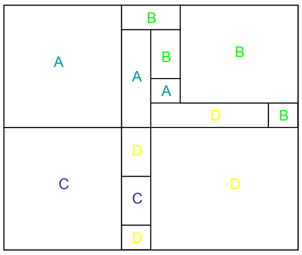
A	A	A	A	B	B	B	B	B	B
A	A	A	A	A	B	B	B	B	B
A	A	A	A	A	B	B	B	B	B
A	A	A	A	A	A	B	B	B	B
A	A	A	A	A	D	D	D	D	B
C	C	C	C	D	D	D	D	D	D
C	C	C	C	D	D	D	D	D	D
C	C	C	C	C	D	D	D	D	D
C	C	C	C	C	D	D	D	D	D
C	C	C	C	D	D	D	D	D	D

21 X 2 = 42 bytes

Value	Point
A	4
B	10
A	15
B	20
A	25
B	30
A	36
B	40
A	45
D	49
B	50
C	54
D	60
C	64
D	70
C	75
D	80
C	85
D	90
C	94
D	100

Quadtree compression

- A data structure that provides a more compact raster representation by using variable-sized grid cells
- Instead of dividing an area into same-sized cells, finer resolution is obtained by using finer cell subdivisions

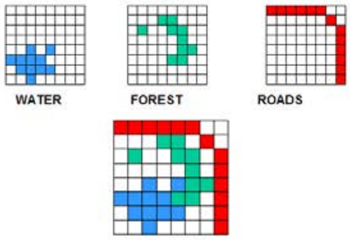


Lossy vs Lossless compression

Raster data

- Raster file types
  - Bitmap (.bmp)
  - Tagged interchange format (.tif)
    - some maps are provided as tifs
  - Graphics interchange format (.gif)
  - Photoformat (.jpg)
  - .dem
  - ArcINFO GRID format
  - Many RS formats (.BIL,. BSQ, .BIP, etc.)

Raster overlays

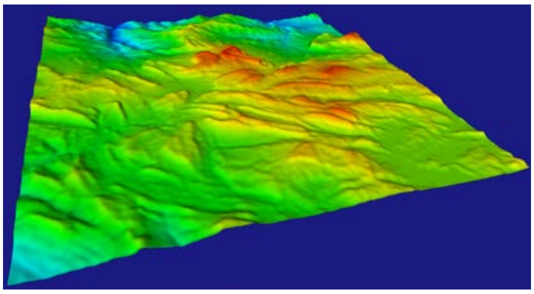


Decide how to deal with coincident pixels

Raster data sources

- Remote Sensing
  - AVHRR 1.1. Km resolution
  - LANDSAT 30 m resolution
  - SPOT 20 m resolution
- Geophysical datasets (processed)
- Air photos – TIFF or GeoTIFF formats
- Scanned topographic maps
- DEMs
  - 7.5-minute 30 m horizontal resolution, 7-30 m vertical
  - 30-minute
  - 1-degree
  - GTOPO 30 (~1 km horizontal resolution)

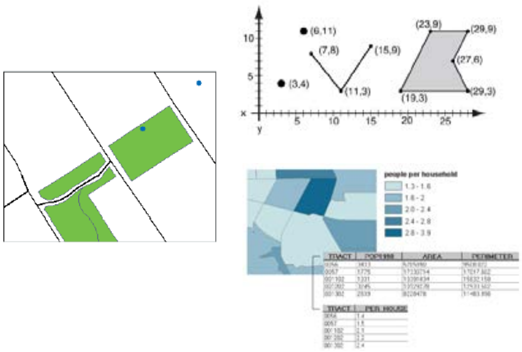
Raster data



Vector data

- What is vector data?
  - Represents geographic phenomena with points, lines, and polygons
  - Vector models are useful for representing and storing discrete features such as buildings, pipes, or parcel boundaries
  - Points are pairs of x,y coordinates
  - Lines are sets of coordinates that define a shape
  - Polygons are sets of coordinates defining boundaries that enclose areas
  - Coordinates are most often pairs (x,y) or triplets (x,y,z, where z represents a value such as elevation)
  - The coordinate values depend on the geographic coordinate system the data is stored in.
  - ArcGIS stores vector data in feature classes and collections of topologically related feature classes. The attributes associated with the features are stored in data tables.
  - ArcGIS uses three different implementations of the vector model to represent feature data: coverages, shapefiles, and geodatabases.

Vector data

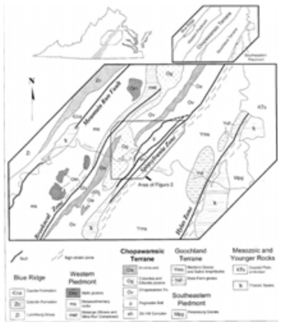


Vector data

- Points
  - Cultural Data
    - Towns
    - Houses
    - Gates
  - Geological Data
    - Sample localities
    - Measurements

Vector data

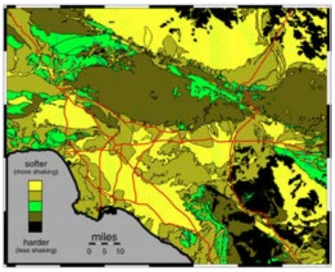
- Lines
  - Cultural data
    - Roads
    - Fence lines
    - Drainage – rivers, creeks
  - Geological data
    - Structural form lines
    - Faults



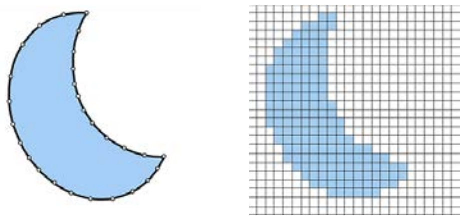
Spears and Bailey, 2002

Vector data

- Polygons – polyregions
  - Cultural Data
    - Town extents
    - Buildings
    - Lakes
    - Ocean
  - Geological data
    - Rock types
    - Alteration types
    - Structural zones
    - etc



Vector data



Vector data is much more efficient than raster data

Vector data

- Advantages
  - More accurate locations of points and lines
    - Precision measurement
  - More compact storage than rasters
  - Can have topology
    - Enables certain kinds of analysis, e.g., Networks
  - Work well with pen plotters and digitizers
- Limitations
  - Harder to implement inside a computer
  - More difficult model for some to understand
  - Difficult to overlay
  - Cannot store continuous-tone data, e.g., photographs or satellite images

Attribute data

- Geographic data is attributed
  - Information
  - Attribute stored in tables
  - Relational databases
  - Object oriented databases

Relational Database

- 1960s/early 1970s - specialised data management software appeared
  - either hierarchical (tree) or network (CODASYL) databases
  - not relational or object-oriented-
- very complex and inflexible which made life difficult when it came to adding new applications or reorganising the data

Relational Databases

- Solution was relational databases
  - based on the concept of normalisation
  - the separation of the logical and physical representation of data
    - Very flexible
    - SEQUEL, SQL
- Set of 2-dimensional tables which are known as "relations"
- Each table has rows and columns
- The relationships between the tables is defined by one table having a column with the same ***meaning*** (but not necessarily value) as a column in another table.

Relational Databases

- For example consider a database with 2 tables:

```
emp(id        number
    ,name       varchar(30)
    ,job_title varchar(20)
    ,dept_id   number)

dept (id    number
     ,name varchar(30))
```

- There is an implied relationship between these tables because *emp* has a column called *dept_id* which is the same as the *id* column in *dept*
- called a foreign-key relationship
- prevents values being stored that are not present in the referenced table

Tables

Emp

Id	Name	Dept Id
1	Bill Smith	3
2	Mike Lewis	2
3	Ray Charles	3
4	Andy Mallory	4
5	Mandy Randall	6
6	Allison White	1

Dept

Id	Name
1	HR
2	IT
3	Marketing
4	Sales
5	Finance

Look out the window…

- Topological view
  - a collection of points, lines & areas in geometric relation to each other
- Object-oriented view
  - sidewalks, buildings, trees, people …

Topology

- Peter has 2 arms (left and right) that are both attached to his body
  - This does not change whether he eats, sleeps or dances
  - Nick sits next to Eliza on her right, while Emma sits next to Eliza to her left.
    - Who sits in between Nick and Emma?
- In GIS, maps can be stored as spagetti or topologically.
  - In the latter case, spatial relationships between connecting or adjacent geographic features such as points, lines, and polygons are explicitly stored – good for analysis

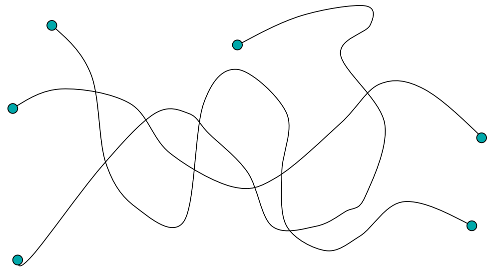
Topology

- The relative location of geographic phenomena independent of their exact position
  - In digital data, topological relationships such as connectivity, adjacency and relative position are usually expressed as relationships between nodes, links and polygons.
  - For example, the topology of a line includes its from- and to-nodes, and its left and right polygons.
  - Topology is useful in GIS because many spatial modelling operations do not require coordinates, only topological information. For example, to find an optimal path between two points requires a list of the lines or arcs that connect to each other and the cost to traverse each line in each direction. Coordinates are only needed for drawing the path after it is calculated.

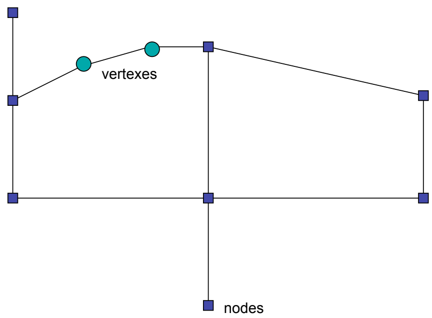
Vector file types

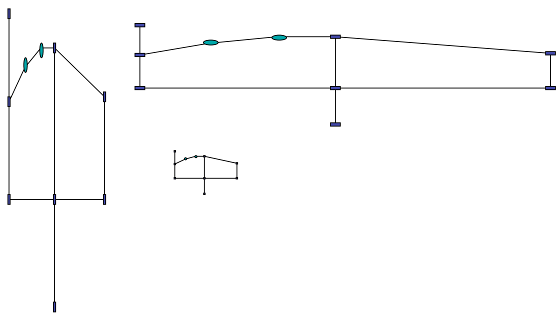
- SPAGHETTI (CAD style)
  - No relationships stored
  - Small size
  - Poor for analysis
- TOPOLOGICAL (Arc/INFO style)
  - Stores internal relationships
  - Bigger, more complex files
  - Good for analysis

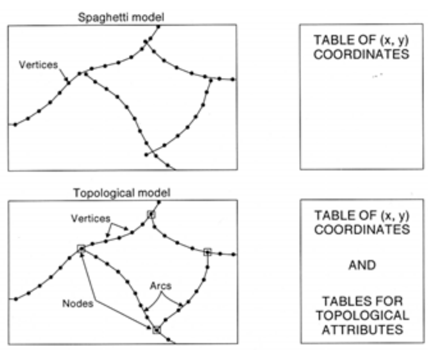
Spaghetti



Topological



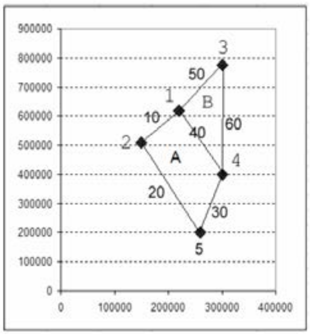


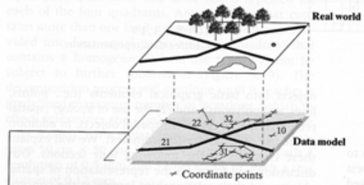


PointID	X-value	Y-value
1	200101.22	620112.3
2	150000.00	510000.4
3	300100.10	777000.1
4	300000.00	400000.00
5	260000.00	200000.00

LineID	FromNode	ToNode
10	2	1
20	2	5
30	5	4
40	4	1
50	1	3
60	4	3

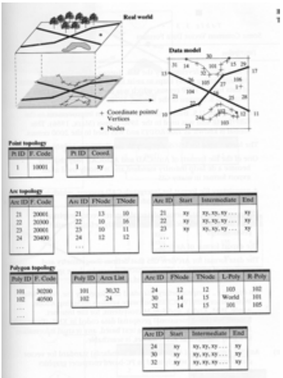
PolygonID	LineID sequence
A	20, 10, 40, 30
B	50, 60, 40

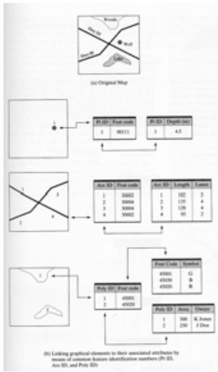


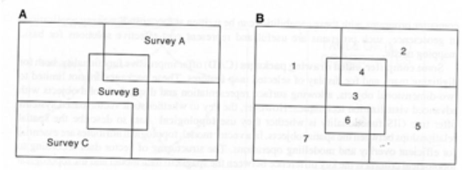


Stored data

Feature ID	Feature type	Location
10	Point	xy
21	Line	$x_1y_1, x_2y_2, x_3y_3 \dots x_ny_n$ (string)
22	Line	$x_1y_1, x_2y_2, x_3y_3 \dots x_ny_n$ (string)
31	Polygon	$x_1y_1, x_2y_2, x_3y_3 \dots x_1y_1$ (closed loop)
32	Polygon	$x_1y_1, x_2y_2, x_3y_3 \dots x_1y_1$ (closed loop)







POLYGON	NONE	SURVEY A	B	C
1	1	0	0	0
2	0	1	0	0
3	0	0	1	0
4	0	1	1	0
5	1	0	0	0
6	0	0	1	1
7	0	0	0	1

Vector data

- Triangulated meshes
  - TIN – triangulated irregular network
  - Geospatial information is represented as a network of linked triangles drawn between irregularly spaced points with x, y, and z values.
  - Efficient way to store and analyze surfaces
  - Heterogeneous surfaces that vary sharply in some areas and less in others can be modeled more accurately than with a raster
    - Many points can be placed where the surface is highly variable, and fewer points can be placed where the surface is less variable.

Victorian Institute of Earth and Planetary Sciences

Draping

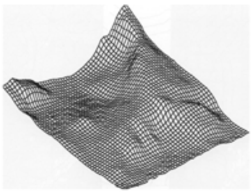


Plate 1. A wire frame diagram for part of the Snowdonia National Park, Wales (the grid interval is 20m).

Plate 2. The wire frame model in Plate 1 draped with orthorectified aerial photography of the same area.