



GIS: An Introduction

The GIS Technical log developed for the FAB Transport Project
by
Warren Lewington

Glossary	4
Introduction	5
GIS History	5
GIS Basic Layer Principles	5
Map Legend Principles	5
GIS Data	5
Vector Data	6
Point Features	6
Polyline Features in Detail	6
Polygon Features in Detail	6
Vector Layers, Topology and digitising data	6
Raster Data	7
Single Symbols	7
Data Capture	8
Data Analysis	8
Heads-up digitising	9
Georeferencing	9
Spatial resolution	9
Spectral resolution	9
Raster to Vector Conversion	9
Coordinate Systems	11
Coordinate Reference Systems	11
Map Projection Systems	12
Bibliography	12
Index	13

Glossary

Esri: A vector data format.

KML: Keyhole markup Language.

PPGIS: Public participation GIS. Prepared by members of the public. OpenStreetMap is part of this work.

VGI: Volunteered geographic information. Publicly generated geographic information using web applications and services, which becomes public participation GIS (PPGIS).

Introduction

Geographical Information Systems or "GIS" deal with spatial information using computers. A GIS consists of:

- Digital data that will be viewed and analysed.
- Computer hardware.
- Computer software.
- People - the GIS professionals who define the purpose, objectives and results of GIS projects (Chang pg.3).
- Organisations that have embedded GIS operations(Chang pg.3).

With a GIS application you can use digital maps to add spatial information, create spatial information and print or produce new information with the data created. The geo-spatial information enables us to be able to see trends and new patterns of information.

Fundamentally, GIS consists of geospatial data, data acquisition, data management, data display, data exploration, and data analysis (Chang pg.5).

GIS History

GIS began to be used in the 1970's. it was a field initially only accessible to large corporations and governments. Over time software packages have become more accessible like QGIS, and amateurs and professionals can begin using GIS packages relatively easily.

GIS Basic Layer Principles

Apart from the toolbars, menus and panels, GIS applications display maps as layers. Map layers are stored as files in the project file. Each layer can be a separate layer representing roads, water-courses and so on.

Map Legend Principles

Map layers are listed in the GIS applications map legend, rather than a traditional maps topographical information. The layers of a GIS map should then be related to legend information in this sense. Layers also determine, as in other software, the way in which the order of things are displayed. Rivers could be displayed over roads if desired. And yes, this I do know.

GIS Data

The purpose ultimately of GIS is to orient or link non geographical data to geographic locations.

- Latitude and longitude data is only the first way.
- Vector data is stored as 'x' and 'y' coordinates, and is used to represent points (towns), lines (rivers or roads) and areas (polygons).
- Raster data are stored as a grid of values. Raster data is composed of squares of image data drawn from satellite imagery.

Vector Data

Vector data is a way to represent real world features, effectively anything you can see on a landscape. This includes lakes, houses, roads, trees, rivers, dugouts, or trenches.

Vectors have attributes which consist of text or numerical information describing the feature being represented.

A vector is has its shape represented using geometry. The geography is made up of one or more interconnected vertices. A vertex describes a position in space using an X, Y and sometimes a Z axis. Geometries which have vertices with a z axis are often referred to as 2.5D because they describe height or depth at each vertex, but not both.

- If a geometry only has one vertex, it is a point feature .
- Where there are two or more vertices, and the first and last are not equal, it is called a polyline .
- When there are three or more vertices present, and the first vertice is equal to the last, a polygon is created.

Point Features

Points can be a matter of arbitrary discussion and also related to scale. On a certain size map a city might be a point, but expand it and it soon becomes a polygon. As mentioned, X, Y and Z coordinates depend on the Coordinate Reference System (CRS) being used.

Polyline Features in Detail

A polyline must have two or more vertices. it is a continuous path drawn through each vertex. when two vertices are joined a line is created. when two or more are joined, they form a 'line of lines' or a polyline.

We use them to use linear features such as roads, rivers, contours, footpaths, and so on. Sometimes we have special rules for polylines, so in the case of a cliff line, they should touch but never cross. Special rule can be set in a GIS system related to this type of circumstance. When polylines are digitised they will always, at some point, become straight lines, which means it is important that polylines are digitised (captured into the computer) with distances between vertices small enough for the desired scale you are working at.

Attributes of a polyline describe its properties or characteristics.

Polygon Features in Detail

Polygons can have shared geometry. They can also have limits and boundary divisions set as attributes. This can be controlled using topology.

Vector Layers, Topology and digitising data

Scale of the map and drawing polygons on the map is critical to capturing the level of detail you require. Likewise, layers are the only way to create vector data, and you edit vector data in layers. You can customise vector layers using a range of symbols, colours for the polygons and vector lines and so on. There are two main types of vector data:

- Overlay: This combines the geometries and attributes of input layers.

- Buffering: This creates buffer zones from select features.

We can use vector data in a GIS the same way as in a paper map.

Vector data must be accurate to be useful. It must be up to date. Errors in topology work include:

- Slivers are caused when the edges of adjacent polygons don't align.
- Overshoots can occur when a line like a road does not meet another feature exactly at an intersection, and continues onwards.
- Undershoots can occur when a line feature it should be connected to do not connect because it is too short.

These errors break the relationships between the features. hence any such errors need to be fixed in order to improve the accuracy of the vector map being produced. Remember this data may well be used again for other studies, which means rigor must be applied to the data creation. Topology is not automatically enforced for accuracy. The user needs to prepare and be rigorous with their data preparation.

Vector data models can be georelational or object-based, with or without topology, simple, or composite. A georelational model stores geometries and attributes of spatial features in separate systems (Chang pg.6). Also within this range of modeling is the triangulated irregular network (TIN) which approximates terrains with a set of non-overlapping triangles (analogous to the really cool digitised topographies games and so on may use). Another model is dynamic segmentation which combines linear measurements with two dimensional projected coordinates, like mapping distances along highways between major cities (Chang pg.7).

Attributes are also important for spatial analysis. This analysis draws information from the attributes stored in layers and the geometric spatial information of each of the vector objects. Attributes are stored in a table, usually in some kind of relational database.

Raster Data

Raster data is a package of information made up of rows and columns - its the way a jpg file is composed of a mosaic of coloured pixels. This sort of data is very useful in situations where gradient lots of information which are not homogeneous are required for analysis, or are intended to be analysed. There are four tools for Raster data:

- Local
- Neighbourhood
- Zonal
- global operations.

Single Symbols

Point features fall into this category. you can set their marker colour and the shape of the marker (square, circle or star). You cannot force the GIS to draw features based on one of its properties. TO so this you need a graduated, continuous, or unique symbol. line and polygon layers allow colour and line style settings.

Before the layers on a map are built, decide what attributes and symbology will be used for each item and each layer. Know what attributes are needed and know what the symbology will be for each as required.

Ultimately the goal of collecting attribute data is to analyse and interpret spatial information. how this is done depends on the questions you are trying to answer.

Data Capture

GIS systems generally store data in their 'shape file'. It is a combination of three different files that work together. It is called a table_shapefile.

Extension	Description
.shp	The geometry of vector features are stored in this file.
.dbf	The attributes vector features are stored in this file.
.shx	This file is an index that helps the GIS application to find features more quickly.

If the files related to each layer were required during a transfer of map information, you would, say for a layer related to trees, find trees.shp, trees.shx, trees.dbf, trees.prj, and trees.qml. These shapefiles are the basic point of entry and data storage, and don't require a database, which is more complex. Given I will be using a database, it seems sensible to get the basics then move on. Plan what you intend to achieve before you start:

- Before creating a new layer, determine what vector information will be on it (point, polyline, polygon?).
- Make up a list of potential layers even if you might not need them.
- Create a new shape_file.
- This starts with selecting a new vector layer.
- Add the fields to the attribute table.
- Indicate the nature of the fields inputs (strings, integer, floating point and so on). This is effectively creating a database table for each layer.

To capture data:

- The next step is to capture a point.
- Capture the point on the layer you are working on.
- To capture data from another source, you can overlay a raster layer. This is called "heads-up digitising" and you can trace the features and points using the rasterised layer onto the vector layer.
- You could also use a digitising table, which is pretty trick, used by professionals, and requires expensive equipment. the device used is a 'puck', which traces features from the paper map, into a data file connected to the computer.

Data Analysis

This is related to using the vector and raster data to find answers to questions related to the GIS mapped and located points, areas, lines, and so on. Buffering, overlay, local, neighbourhood, zonal,

and global operations, are all part of preparing analysed information useful to the end-user. Another form of analysis is spatial interpolation, which uses points with known values to estimate values at other points. It is a way of creating surface data from sample points in a GIS map, one of which is kriging, which can predict and estimate errors in the sample and point data (Chang pg.9).

Geocoding is a way of turning items like landmarks or postal addresses into points. These can then be located on the 'X' and 'Y' coordinate data system. this is important to Danny's work, and he often mentions this (Chang pg.9).

Another method which would be very useful for transport systems is Least-cost path analysis. This is also observed in car-mounted GPS units. The purpose of this is to establish paths with the shortest paths between stops or points (Chang pg.9). Note that this type of work is done with raster data and creates or works with virtual paths. The difference here is that vector data uses network analysis, and works with existing road network datasets.

All of these tools can be used with all the usual highly sophisticated statistical modeling tools, regression models, process models, anything really. QGIS is equipped with an 'R' Python coding extension for this purpose.

Heads-up digitising

The critical path if you are using a rasterised image as a heads-up digitising tool is that your geospatial data is properly georeferenced (Chang pg.5 & QGIS.org), and properly displays in the correct position in the map view based on the GIS applications internal model of the earth. A properly georeferenced raster layer will overlay properly above the map, and allows good transfer of data onto the map vector layer. Properly geo-registering the map features onto the map layers (matching the vector map with the digitised raster layer) means a perfect overlap. Using the correct scale is also crucial to this work.

Georeferencing

Georeferencing is defining exactly where on the earth's surface a raster data set is meant to be located. It is stored with the digital version of the aerial photo or map. This is fine but this project uses aerial photographs and maps produced before GIS was even contemplated, and computing machines were as advanced as the slide rule and abacus. So in this case, the project will require careful raster geo-referencing and verification that the pgs chosen are actually correctly used.

Spatial resolution

This is related to the fixed pixel size within each raster file. It is usually of a quality representing the sensor or system that created the information.

Spectral resolution

This is related to the R, G, B, colour bands. it also covers all the wavelengths on the light spectrum, and those not visible to the human eye. Non-visible light raster images are called multi-spectral images. Measuring infra-red light can be useful for measuring water bodies for example. Images containing only one band of light are often called grey-scale, while raster images made of false colouring can be called pseudocolour images.

Raster to Vector Conversion

Some GIS applications offer the ability to draw raster features from a raster image and turn them into vector data. The processing is performed using contrast changes to determine where topographical

features might be, and calculating the vector data relevant to the feature as points, polylines, or polygons.

Coordinate Systems

The fundamental principle in GIS is that maps used together must be correctly aligned spatially (Chang pg.20). Every time I use a map in this project, I will have to convert them to a common spatial reference system. Map sets available on the Internet and via other such sources can be measured using longitude and latitude, while others are, as mentioned earlier, lying on different projected coordinate systems. We process or match these sets of maps by projection or reprojection, effectively taking a dataset from one coordinate system and converting it mathematically into another coordinate system.

The geographic coordinate system is latitude (parallels) and longitude (meridians).

Coordinate Reference Systems

Maps are a two dimensional rendering of a three dimensional topography. Therefore a coordinate reference system must be used which defines how the two dimensional, projected map in the GIS relates to the real world. The important point to also remember is the earth is actually an ellipsoid, not a sphere, meaning that mathematical correction is not uniform, but varies according to the tangential relationship touching at the ground surface level at a specific latitude (up or down towards the poles), and of course, by longitude (longitude lines converge from the centre width to the narrow poles). This makes for a complex mathematical relationship when considering conversion of a two dimensional paper map into three dimensional spatial dataset, and maintain extreme accuracy. The type of information being developed has a direct bearing on the CRS chosen and the map projection that goes with it. Any place on earth can be specified by a set of three numbers, called coordinates. Many projected coordinated reference systems are available, an example being the Universal Transverse Mercator (UTM) grid system, which divides the Earth's surface between 84° north and 80° south into 60 zones (Chang pg. 5) - making the map on paper look like a zig-zagged or oddly cut rectangle! CRS can be divided into:

- projected coordinate reference systems or Cartesian or rectangular coordinate reference systems.
- Geographic coordinate reference systems , using degrees of latitude and longitude and sometimes a height value to describe altitude locations above earth. The most common is WGS 84.

Latitude, running parallel to the equator, divides the earth into 180 equally spaced sections, 1° from north to south apart. in the northern hemisphere, latitude is measured from the equator at 0° to the north pole at 90°. the southern hemisphere is from the equator at 0° to -90° at the south pole. the distance between each line of latitude is thus 60 nautical miles.

Longitude, has always been a problem because longitude lines don't have geometric conformity, they running perpendicular to the equator converge at the north and south poles. Longitude lines are thus curves, which means as latitude rises, their mathematical or geo-spatial relationship changes. This is the fundamental problem with mapping on paper. Starting at Greenwich or 0°, the lines of longitude are measured 180° east or 180° west. In digital mapping applications, lines west are assigned negative values (120° west on a map would be -120° in a GIS). Thus only at the equator is 1° of longitude equal to 1° of latitude.

The further division of the lat and long polygons is using minutes and seconds. There are sixty minutes in a degree, and sixty seconds in a minute (3,600 minutes in a degree). So at the equator one second of latitude or longitude equals 30.87624 metres.

Map Projection Systems

Map Projection is a way of projecting the spherical shape of the earth in two dimensions and is directly related to the coordinate referencing systems mentioned above. Globes typically have a scale detail of 1:100 Million, where a GIS can supply detail down to 1:250,000. Maps, it must be remembered are merely 2 dimensional representations of a curved, spherical reality which is infused with varied topography, from deep oceans to high mountain ranges. Each map projection has advantages and disadvantages, and ultimately the best projection for a map depends on the scale, and the intended purpose for its use.

There are three families of map projections.

- a. Cylindrical projections.
- b. Conical projections.
- c. Planar projections.

These projections are never absolutely accurate representations of the spherical earth. Distortion of angular conformity, distance and area are common. Compromised projection examples are the Winkel Tripel projection and the Robinson projection.

Maintaining correct angular properties that conform to the 90° angle relationships between north, east, south and west are called conformal or orthomorphic projection. Use these map types when the preservation of angular relationships is necessary. This is the Mercator projection or Lambery conformal Conic projection.

To accurately measure distances, a map called an equidistant projection is required. These maps require that the scale is kept constant. It is considered correct when it correctly represents distances from the centre of the projection to any other place on the map. these maps are used for seismic mapping, navigations, and radio mapping. the Plate Carree Equidistant Cylindrical and the Equirectangular projection are examples while the Azimuthal Equidistant projection is used as the United Nations logo map.

Maps with equal areas are ones where all mapped areas have the same proportional relationship to the areas on the earth that they represent. They are useful for calculating areas. The Alber's equal area, Lambert's equal area and Mollweide Equal Area Cylindrical projections are examples of these types of map.

Bibliography

http://docs.qgis.org/2.18/en/docs/user_manual/working_with_projections/working_with_projections.html provides the detail related to workign with projections.

http://docs.qgis.org/2.18/en/docs/user_manual/working_with_raster/raster_properties.html

Working raster data mentions that Proj 4.0 data scripts may be usable

http://www.qgistutorials.com/en/docs/georeferencing_basics.html

Index

c

Cartesian 11
 Coordinate Reference System 6
 Coordinate Reference Systems 3, 11
 CRS 6

d

Data Analysis 3, 8
 data capture 3, 8
 data capture process 8
 distortion of angular conformity 12

g

Geographic coordinate reference systems 11
 Georeferencing 3, 9
 geocoding 9
 georeferenced 9

h

heads-up digitising 3, 9
 Raster layers 8

k

kriging 9

l

Latitude 5
 least-cost path analysis 9
 longitude 5

m

Map projection 12

maps representations of reality 12

o

Overshoots 7

p

point feature 6
 polygon 6
 polyline 6
 polyline features 3, 6
 projected coordinate reference systems 11
 pseudocolour images 9

r

Raster 5
 Raster data 3, 7
 rectangular coordinate reference systems 11

s

Slivers 7
 Spatial resolution 3, 9
 Spectral resolution 3, 9
 spatial interpolation 9

u

Undershoots 7

v

vector data 6

w

WGS 84 11

