Introduction to Geospatial mapping and Analysis with…

QGIS

Hosted by:

worldpop FLOWMINDER.ORG

GeoData

UNIVERSITY OF Southampton

Email: gis-train@geodata.soton.ac.uk
Tel +44 23 80592719
www.gis-train.com

Welcome to the Course!

UNIVERSITY OF Southampton GeoData Institute

- Introductions
- Facilities
- Lunch and Refreshments
- Fire Exit

Course Overview – Day 1

1) Introduction to GIS

2) Introduction to QGIS
   Exercise 2

3) Working with the Map
   Exercise 3

4) Focus on Data
   Exercise 4

Course Overview – Day 2

5) Selections and Queries
   Exercise 5

6) Producing Maps
   Exercise 6

7) Geoprocessing and Spatial Analysis
   Exercise 7a
   Exercise 7b

Introduction to GIS – Topics

UNIVERSITY OF Southampton GeoData Institute

- What is GIS?
- Types of Questions a GIS can Answer
- Spatial Data
- GIS Project Workflow
- GIS Principles
  - Types of Geographic Information
  - GIS Data Models
  - Mapping in Layers
  - Co-ordinate Systems
- GIS Applications

1

What is a GIS?

UNIVERSITY OF Southampton GeoData Institute

Geographical Information System

- A database combined with a visual map interface
- Map and analyse objects and events that happen on the Earth
- Used with any data which can be geographically tagged
- 4 Key Elements to a GIS – HARDWARE, SOFTWARE, DATA, PEOPLE

Definition:

A system for capturing, storing, checking, integrating, manipulating, analysing and displaying data which are spatially referenced to the earth.

(Report of the Chorley Committee to the UK Department of the Environment, 1987).

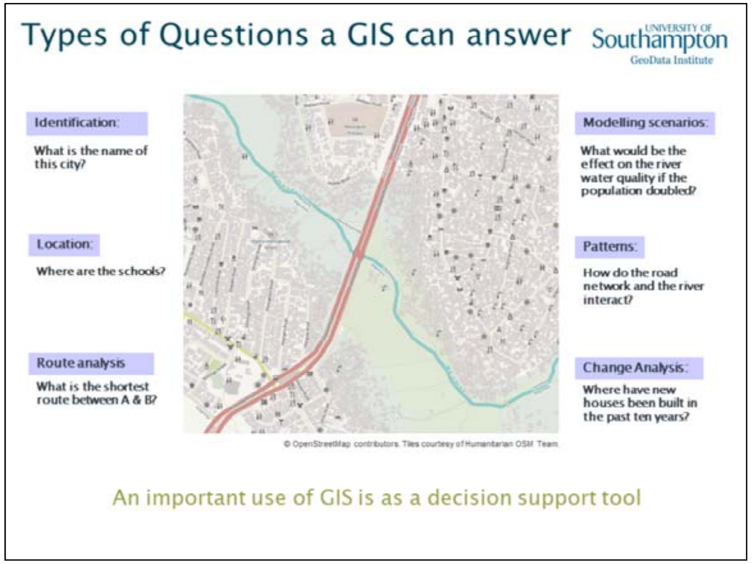
There are many different definitions of Geographical Information Systems (GIS). However, a general definition is that a GIS is a computer-based tool for mapping and analysing things that happen on the Earth. GIS technology integrates database functions such as querying of data and statistical analysis with the visual and geographical analysis benefits offered by maps.

One of the origins of GIS lay in the emergence of databases incorporating the location of objects (this could be in the form of a grid location for reference with a map, or another form of position such as an address). Much of the information used in such systems was first entered from a map, either in a paper form or on a computer screen. This information could then be used in the production of **updateable**, customised maps.

Much data involved in the management of land and property has a geographical element, in that it refers to objects or areas which exist on the earth and have a position. It is therefore sensible to organise such data by its position on a map.

A GIS gives the opportunity to display and analyse such data with respect to its geographical location, and produce customised paper maps very quickly.

☐ *GeoData Institute in*



GIS supporting decision making – 6 types of question a GIS can answer:

- **What is there …?**

  Identification - query by location

- **Where is …?**

  Location - query by attribute e.g. where is a conservation area

- **What has changed since … ?**

  Trends - needs data over several sequences over time (temporal component) e.g. boundaries changed

- **What is the best route between …?**

  Optimal path - routing questions – shortest or cheapest e.g. ambulance services

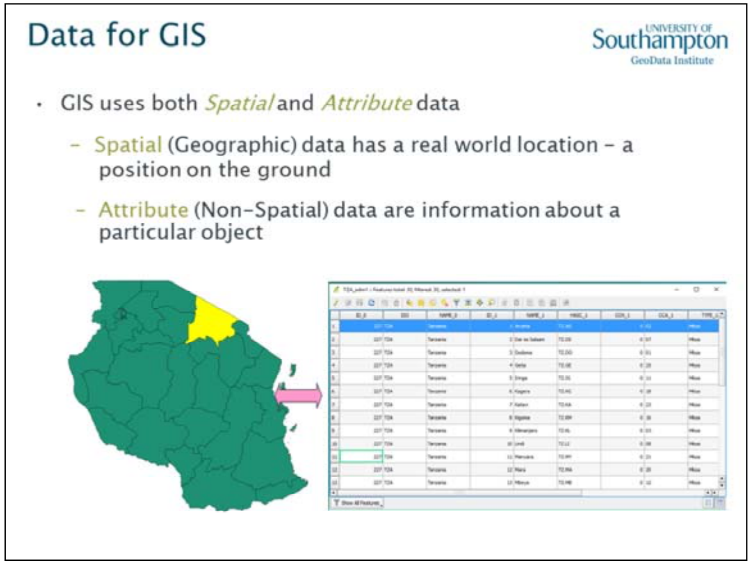
- **What relation exists between …?**

  Patterns - often several data sets e.g. factors influencing disease patterns (epidemiology), relationship between the local microclimate and location of factories & social structure of surrounding neighbourhoods

- **What if …?**

  Models - GIS is very powerful tool for investigating scenario questions / planning & forecasting activities e.g. sea level rises, roads for new houses, earthquake impacts

Ability of a piece of software to answer these types of questions distinguishes a GIS from a cartographic / design system.

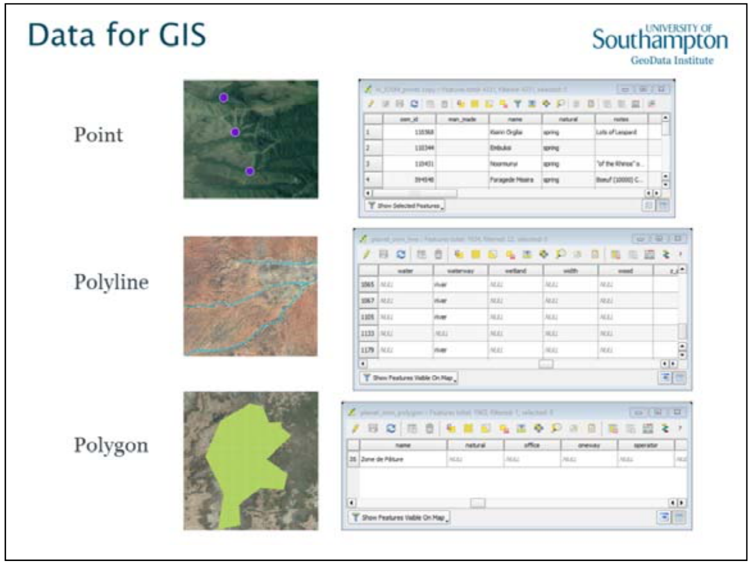


Spatial, or geographic data are data which has a real world position or location. For example a set of points drawn on a map to indicate the locations of features such as gates or bridges would be a spatial data set.

Attribute, or non-spatial data are data which does not in itself have a location in the real world. It can be thought of as "information".

GIS data always has a geographical element (it always has a location on the ground). GIS data usually has "attributes" linked to it - they are data which does not concern the location of an object, but rather the "information" related to an object. For example, if a car park is marked on a GIS map as a point (the spatial data), the attributes might be information such as the car park type (multi-storey, underground etc.), number of spaces, opening hours and charges.

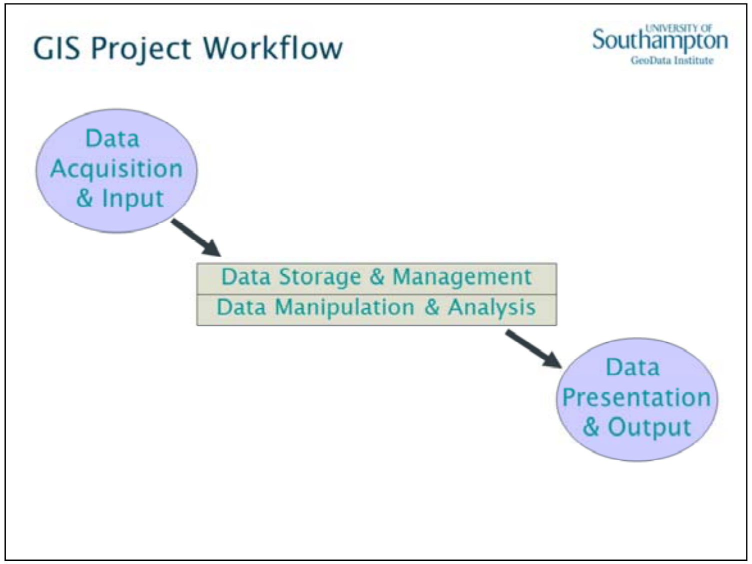
The power of GIS lies in the ability to link the attribute data of an object to its position in the real world, and perform analysis based on both the location of objects and their attributes.



Spatially Referenced Data – one of the key ideas / distinguishing features of a GIS

Data which represents different forms of real world features. It describes the position of such features by using either a single or a series of grid references (or another co-ordinate system). They could be single point locations (e.g. boreholes, survey sites), linear features (e.g. roads, rivers, cables, pipelines) or bounded areas (e.g. administrative areas, fields, conservation zones or ecosystems). These different types of spatially referenced data are known as **Points**, **Lines** and **Polygons** (Areas); a line and a polygon being made up of a series of points.

Spatially referenced data can also represent the attributes of such features, i.e. a value or characteristic of an object that may or may not be unrelated to its position, e.g. soil type or elevation, by storing them as data at that location.



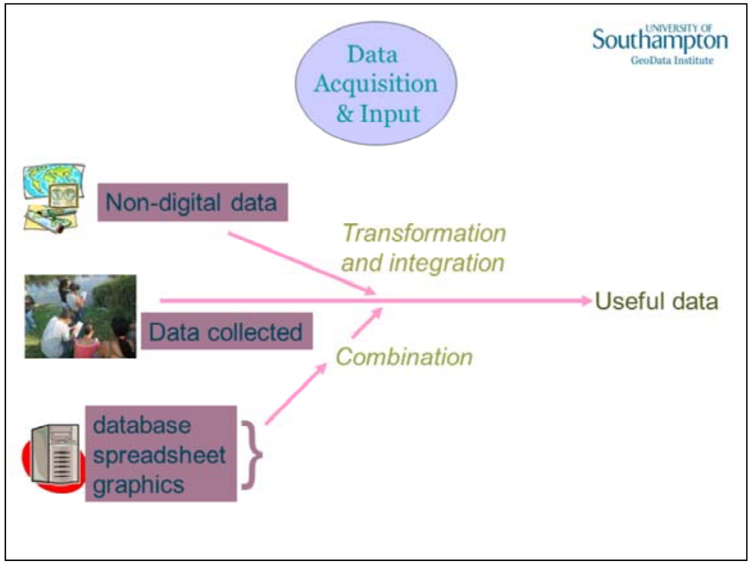
The main GIS tasks can be grouped into the following 4 categories.

1 - Data Acquisition & Input

2 – Data Storage and Management

3 – Data Manipulation & Analysis

4 – Data Presentation & Output



1 - Data Acquisition and Input

This could include the input of data from non-digital sources (such as paper maps or records), the collection and entry of completely new data, or the integration of data from other digital sources (such as Databases, Spreadsheets or other GIS software).

Examples:

-Digitising historical maps so that you can identify changes

-Location of Listed Buildings – collected by mobile GIS for example

Before any display or analysis of data is possible within a GIS, they must first be gathered and converted into formats accessible for analysis.

There are 2 stages of data acquisition.

The **primary** stage involves the collection of data in the field. This may involve ground based studies of topography, geology and vegetation, coastal surveys, airborne surveys using satellites and aeroplanes, and socio-economic studies that could involve interviews and transcription of written documents.

The **secondary** stage involves the conversion of this data into a format useful in GIS, for example digitising the features from a paper map, deriving land use classification polygons from remotely sensed imagery or plotting points from coordinates.

GIS Data

Data Acquisition & Input

University of Southampton GeoData Institute

- Data capture and import is usually the most costly aspect of a GIS project
- If the data exists in a non-digital form it will need to be captured or digitised in some way before use, e.g.

Paper maps > transformed using a process known as *Georeferencing*

Field data or data in documents > reformatted into table-based data

Creating a GIS database is a complex process which may involve data capture, verification and structuring processes. The options you may wish to consider are:

•Acquire data in digital format from data supplier.

•Digitise existing data.

•Carry out your own survey of geographic phenomena.

The first stage of digital data capture requires the inputting of spatial information into the GIS. When dealing with paper maps this necessitates tracing information from the source to create a vector file. **Digitising** can be carried out using a digitising tablet to which the paper is attached, or on-screen, using a scanned map as a traceable background. Before digitising the source image must be **georeferenced**. A paper map or scanned image must contain spatial referencing information. This allows the scale, rotation and translation of the data to be calculated, in addition to identifying any problems in it (for example, distortion introduced in the scanning process).

Once georeferenced the source objects can be digitised; attributed points, lines and polygons can be drawn to match the objects seen on the map. Because the scale and orientation have been defined the relative positions in real-world co-ordinates can be calculated by the offset distances on-screen or on the digitising tablet. Digitising is commonly carried out in a native GIS format, or in a format exportable to GIS, such as DXF (Drawing Exchange Format). Georeferencing will be covered later in the course.

GIS Data

Data Acquisition & Input

- Sources of digital data:
  - Commercial data suppliers
  - Open-source data (usually web-hosted)
  - Obtain from others working on similar projects
- Before embarking on a major data capture exercise, it is worth checking to see if anyone already holds the data in digital format!

A variety of government agencies and academic and commercial organisations are involved in collecting and disseminating geographic data.

The types of data that are collected range from the general to the highly specific, but include topographic data, satellite imagery, aerial photography, land use, natural resources, population and settlement data. Much of this data will be subject to copyright agreement, and can only be acquired for a price, but it is important to be aware of what is available.

Once acquired the data may require geo-rectification and further data capture or processing to convert it to a format suitable for analysis.

GIS Data…..a cautionary note..

Datasets may exist… but are they accurate?

Information may be:

- out of date
- at an inappropriate scale
- spatially incorrect
- just plain wrong!

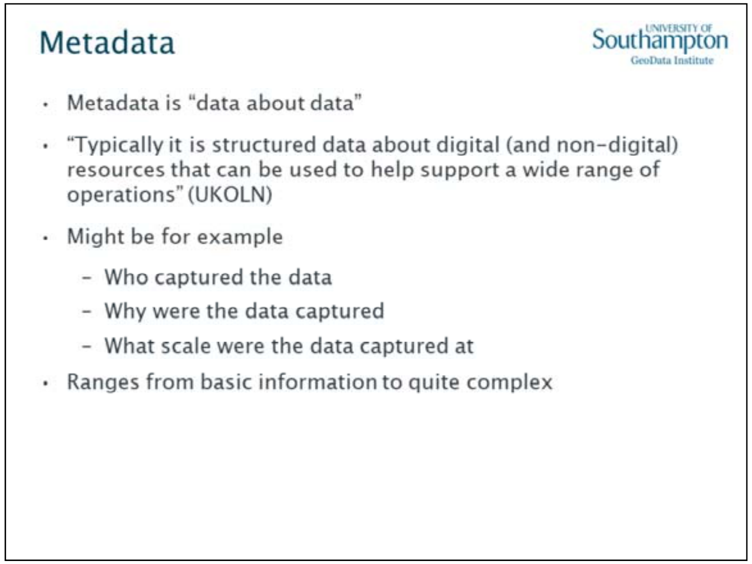
It is important to recognise that a GIS is not a “miracle machine”

GIGO >> “Garbage in….Garbage out!”

GIS is not a universal panacea for all problems but it is a very powerful tool.

To get a meaningful result out of a GIS analysis – need sensible model of real world features to begin with i.e. Garbage IN Garbage OUT.

Consider data quality and assumptions underpinning data



Good site to look at: UK Office for Library Networking
http://www.ukoln.ac.uk/metadata/

There are a variety of different standards available to use, most currently used are based on the ISO 19115 standard:

http://www.dcc.ac.uk/resources/metadata-standards/iso-19115

For example:

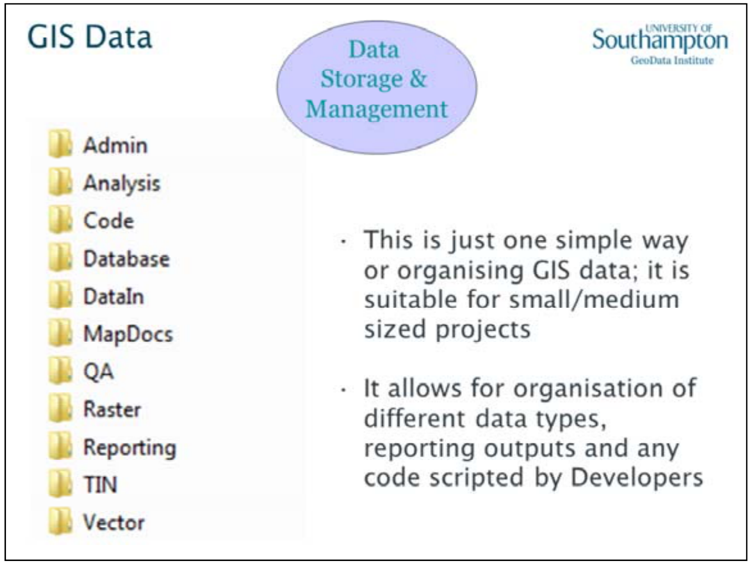
Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata http://www.fgdc.gov/metadata/csdgm/

See also:

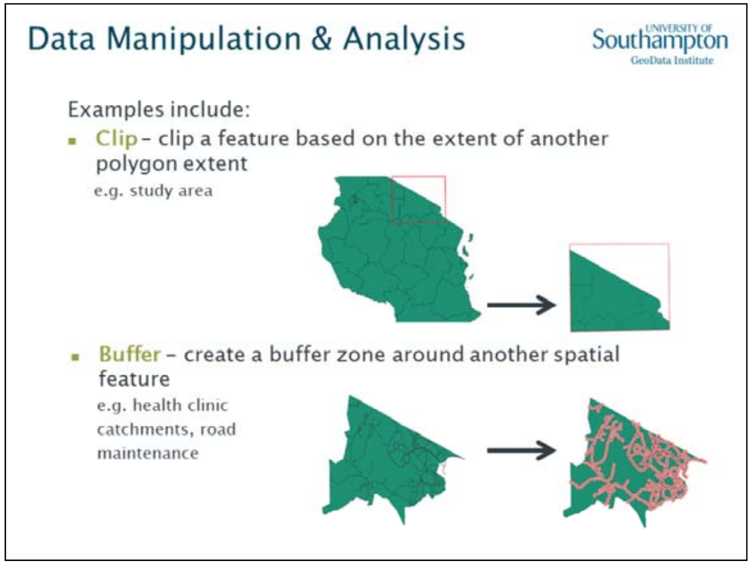
UK Location – for UK/INSPIRE-compliant metadata standard:

http://data.gov.uk/location/uk-location-infrastructure#metadataexplained

Introduction to GIS



It is important when beginning any new GIS project to organise your data. You will come to learn that when using a GIS many outputs are generated, which may result in confusion if a proper Data Management procedure is not followed.



3 - Data Manipulation and Analysis

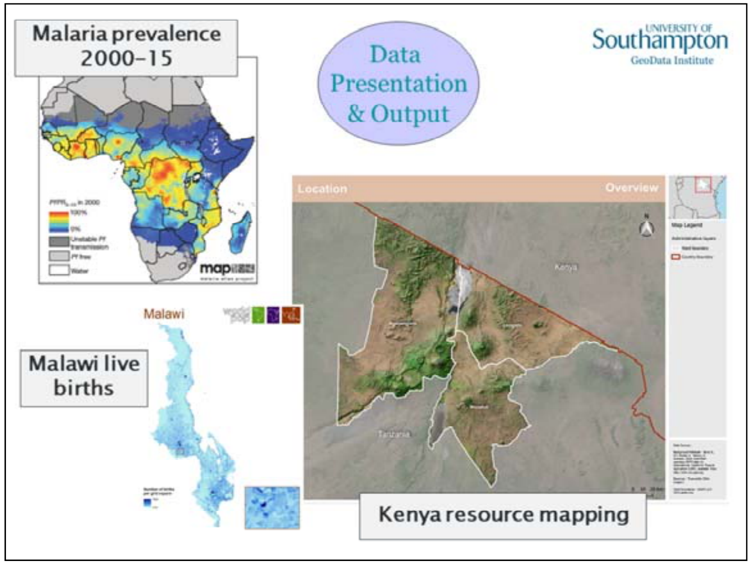
The ability to edit and analyse data as required, and the ability to carry out queries on the data to discover particular features or trends.

Examples:

-You may wish to find features above a certain size or of a certain type.

-You could find all the bank ATMs within a certain study area

-You could locate all properties that fall within a buffer zone



4 - Data Visualisation, Presentation and Output

The ability to produce output from the GIS, in the form of maps and reports which show the results of analysis carried out in the GIS. Data output could take the form of images on a computer screen, computer files or printed maps and reports produced by a printer or plotter.

Unlike paper maps electronic GIS maps can be very easily revised, corrected and updated.

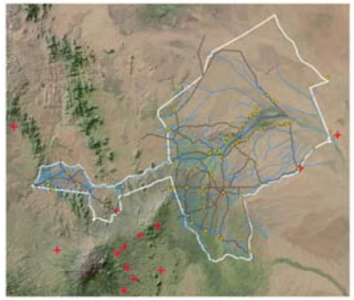
Introduction to GIS – Topics

- What is GIS?
- Types of Questions a GIS can Answer
- Spatial Data
- GIS Project Workflow
- GIS Principles
  - Types of Geographic Information
  - GIS Data Models
  - Mapping in Layers
  - Co-ordinate Systems
- GIS Applications

Types of Geographic Information

Two types of geographic information which relate to distinct geographic phenomena:

- Continuous data can be measured anywhere (e.g. elevation, temperature, population estimates, disease risk) and forms a continuous surface
- Discrete data describes individual features and where they are located; represented as point, line or polygon features (e.g. settlements, health facilities, roads, administrative areas).



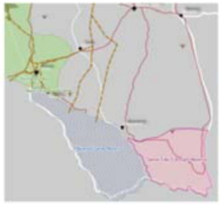
Continuous data occurs across all geographic space (e.g. depth, elevation, temperature, salinity). The value can be measured at any given point and therefore forms a surface of data values.

Discrete data describes individual objects with finite dimensions which are typically mapped as point, line or area features (e.g. wrecks, islands, borders).

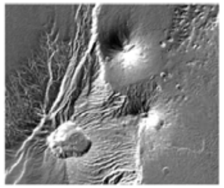
GIS Data Models

Southampton GeoData Institute

- Vector



- Raster



Data models are digital data formats which *model* the reality previously described. The GIS user can choose which model to best represent their geographic study. Ideally data models and structures should be chosen in order to:

- minimize information loss when creating a digital representation of reality i.e. preserve detail / characteristics from real world features that are most important for our particular needs - key attributes
- maximize the analytical and presentational options given the data – best suited to intended purpose of data

The two main types of data structure for representing spatial features within a GIS are Vector and Raster

They are discussed on the next few slides with examples.

The Vector Data Model

Southampton GeoData Institute

Individual coordinate pairs are combined to represent features as either:

- Points
- Lines
- Polygons

Vectors are suited to represent discrete features

In a **vector** system the locations, courses or boundaries of features are stored as a series of co-ordinate pairs which are connected to describe the map objects. Each object is defined by a series of co-ordinates which describe the location and shape of the object. A point on the map would be represented by a single pair of coordinates and a line or polygon would be represented by a series of coordinate pairs, linked together to form the shape. Polygons start and end at the same location to close the polygon. Attribute information is stored separately from spatial data and **joined** (i.e. linked) through **feature IDs**. The GIS is built upon Geometric rules or **TOPOLOGY** which explicitly describes the features, their elements and also their relationships with other features. Topology states which order the points of the shape (or nodes) have been digitised. Topological information is utilised when creating networks (e.g. rivers or roads) because one can tell which way the lines are 'flowing'.

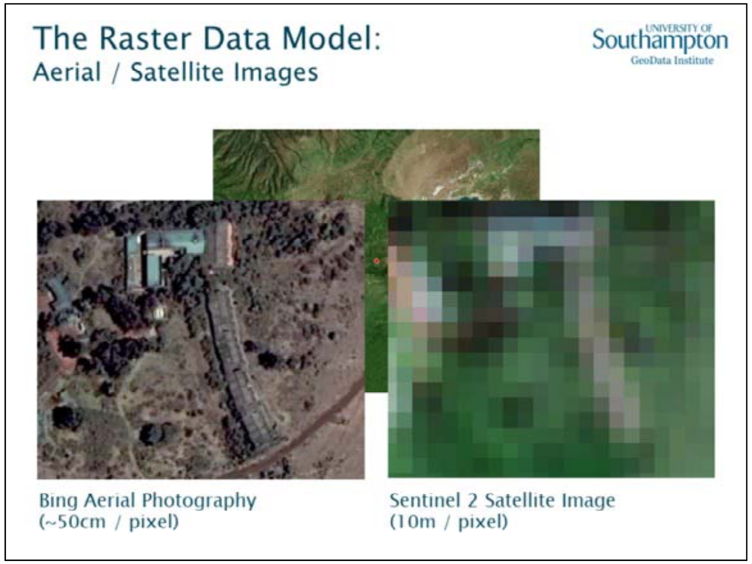
Choice of geometry type:

Points - These are single markers indicating the location of a particular object or feature. They differ from regions in that the size and shape of the marker does not represent the size and shape of the object on the ground. They are only meant to locate where the object actually is. For example, a church may be represented as a point on a map, with a particular symbol.

Lines - These are objects which cover a given distance, but are not concerned with area. They can be a single straight line, or lengths of line made up of several linked sections (polylines). An object such as a path or road could be represented as a line on a map. The line would indicate the position and length of the path, but would give no indication of the area of its surface.

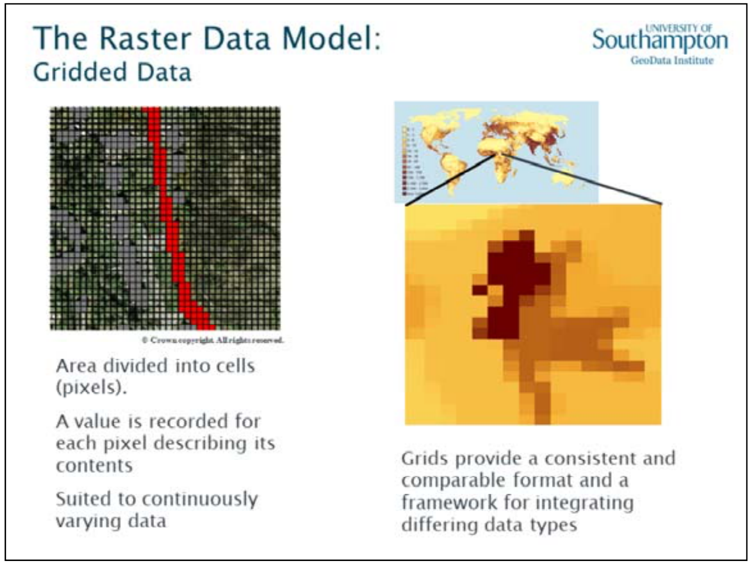
Polygons - These are enclosed areas within a defined boundary marking the extent of the area covered by the feature. For example, a forest would usually be drawn on a map as an *area* of forest.

Note that objects could be represented by more than one type of object. For example a Car Park could be represented as a point showing its location or a polygon giving both its location and area covered. Often, the scale of the map would dictate the choice of object used.



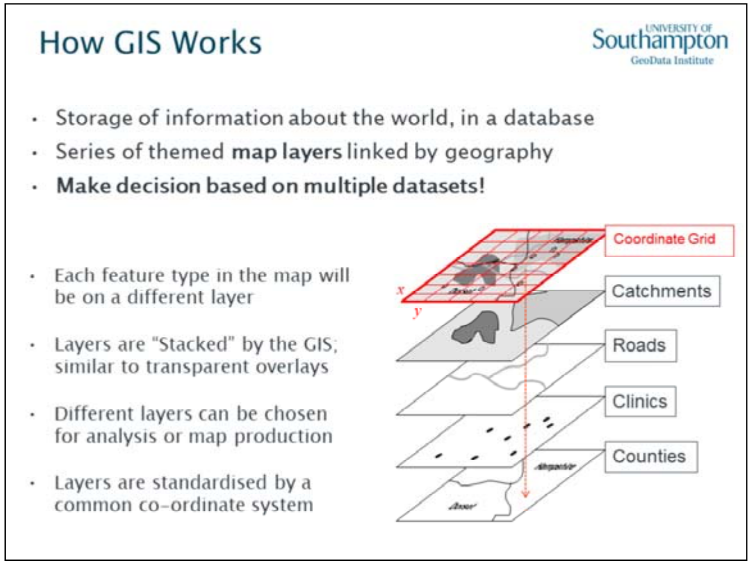
Together with base maps, one of the most common types of raster data used in a GIS is digital aerial photography,

In digital photos, the size of the cells or **RESOLUTION** of the raster image determines what features can be seen. For example, in a dataset with a pixel resolution of 1m it is generally not possible to see objects unless they are bigger than this. It may be possible to see *something* if the object covers more than half of the pixel. However, it will be poorly defined.



Together with base maps, one of the most common types of raster data used in a GIS is digital aerial photography,

In digital photos, the size of the cells or **RESOLUTION** of the raster image determines what features can be seen. For example, in a dataset with a pixel resolution of 1m it is generally not possible to see objects unless they are bigger than this. It may be possible to see *something* if the object covers more than half of the pixel. However, it will be poorly defined.



GIS works by storing information about real world objects, and linking it to the location of the object on a map. The data about each object is stored in a database (an organised set of data, stored on a computer). It is usual to store information about different feature types in separate databases or spreadsheets. For example, if you were collecting data on the location of vegetation types and also in a different survey were collecting information on number of pedestrians using a path, you would store the different data in different databases rather than mixing them all in one!

This is sensible not only to aid viewing of the data, but also because the different data sets would have different attributes.

The same general rule applies in GIS. Different data sets are stored in different databases, and different map object types are stored in different *layers*, rather than being mixed together in one.

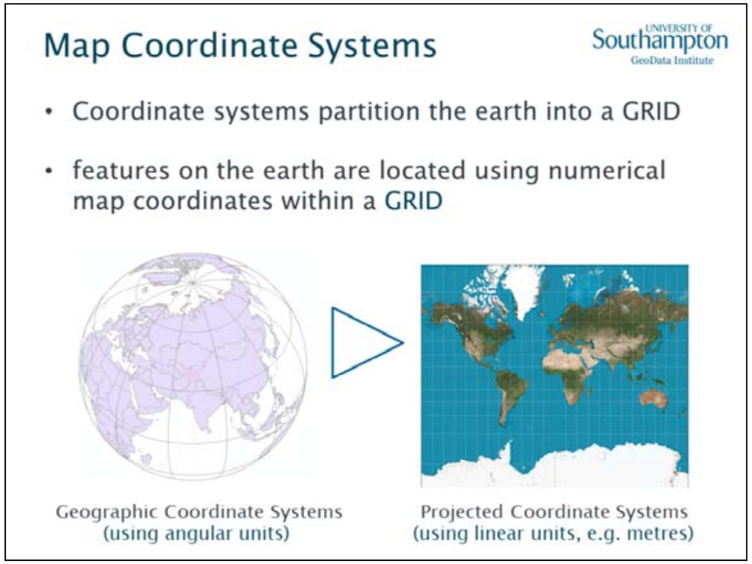
To produce a map which is a composite of several different layers, the different layers can all be opened and viewed in the GIS at the same time.

This has the same effect as would be achieved by drawing out each map layer on a piece of transparent film, and stacking the films on top of each other.

To produce a customised map for a particular purpose, the particular layers required would be displayed together to produce the final composite map.

A series of different "Thematic" layers can be used (each layer contains data for a different feature type e.g. roads, car parks, bridges, railways) to build up a complete map consisting of several layers (e.g. transport network). This allows you to ask spatial questions e.g. How many samples in Dorset?

The Co-ordinate system is the key principle in a GIS system



Before getting into the detail of coordinate systems and projections, it is helpful to think about map coordinates themselves. Coordinates are used to locate features on the Earth's Surface or to work out your location. In navigation, the most common form of coordinate system is that of longitude and latitude, as shown on this map of the globe. This treats the Earth as a sphere and subdivides the surface into a series of gridlines called a ***graticule***.

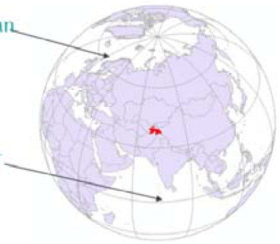
An alternative type of coordinate reference system are *projected systems*, where the spherical model is projected to a 2D plane and coordinates are calculated in linear values, metres, yards etc.

Geographic Coordinate Systems

University of Southampton GeoData Institute

This is a **Geographic Coordinate System (GCS)**

A GCS uses a three-dimensional spherical grid to define locations on the earth



Vertical lines = Longitude
Horizontal lines = Latitude

Measured in Degrees & Minutes (seconds)

e.g. Dushanbe, TAJIKISTAN: 38o34’ N, 68o46’ E

Datum describes the shape, size & orientation of the grid….

Vertical lines represent the **longitude** and horizontal lines represent the **latitude** and the units of measure are degrees (°), Minutes (‘) and seconds (‘’). This allows one to use the coordinate grid as an ***index*** for locating features and places. For example, the coordinates for Dushanbe, the capital of Tajikistan in Central Asia are shown above. All coordinates have a grid upon which measurements are based. The most well known latitude line is the equator at 0^o latitude. Latitude positions are given relative to the equator – that is, are they North or South of it? Longitude is (usually) referenced East or West of Greenwich, UK, although other prime meridians have also been used (e.g. Paris meridian).

The Datum describes the shape, size & orientation of these grid lines. Grids (or ellipsoids) can be tailored to more accurately model specific parts of the Earth. So whilst WGS84 is a global ‘best fit’, others such as are NAD27 are used to model North America.

Projected Coordinate Systems

University of Southampton GeoData Institute

Map Projection Principles…

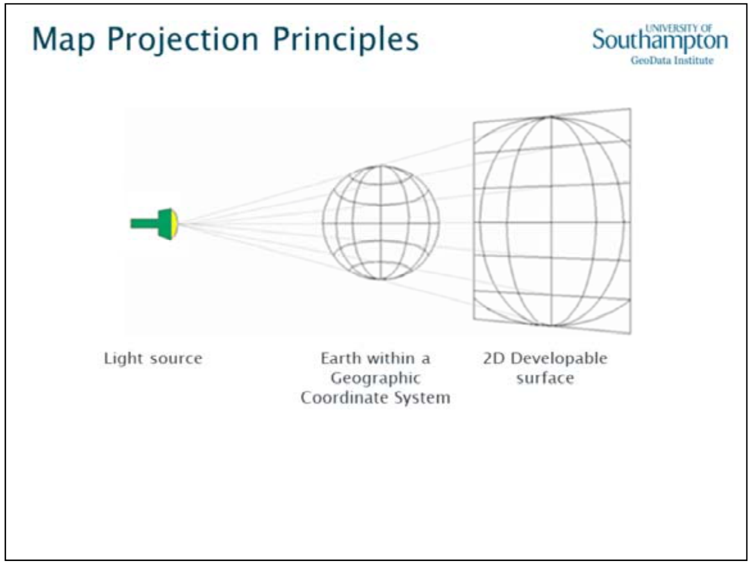
- How best to cut it?
- How flat can you get it?
- However you do it, you will distort the map in some way

Moving beyond the orange principle…

Using a Geographic coordinate system for linear measurements are problematic because you have to allow for the curvature of the earth, chief among them being calculations of area. This presented a challenge from the very earliest days of map-making. Still to this day we work in a 2D context (though advances in 3D visualization and virtual reality may change this). So how to convert from a 3D globe to a 2D surface?

Imagine the globe to be represented by an orange! So if we had to peel those countries away and lie them flat on a table – it’s not possible without ripping or stretching the peel. It’s the same with our 3D grid, we are forced to distort the grids proportions somehow. Many ways to cut and arrange the peel, all of which cause some form of distortion. Distortion of distances, country shapes, areas and angles or some combination of all four. The larger any single section, such as a continent, the more difficult it is to flatten without major distortions

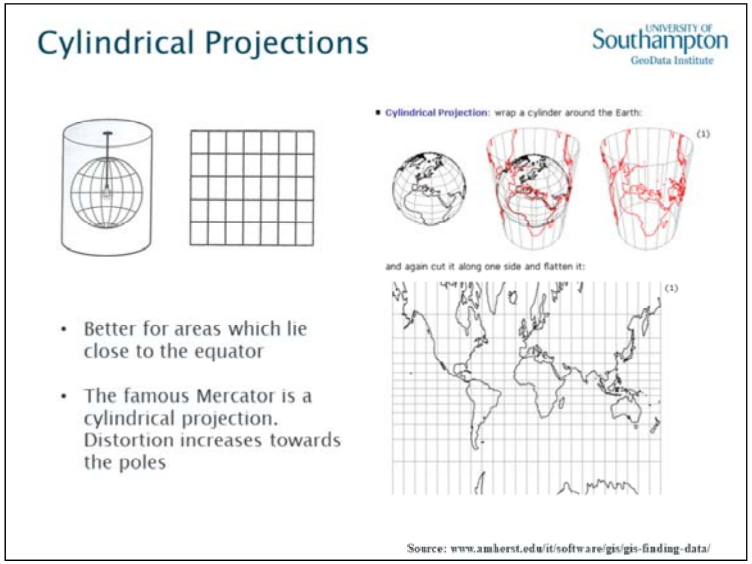
The orange analogy is just one way of understanding map projections…….



The term map projection comes from the theoretical concept of projecting light from a source through the earth's surface, maintaining the GCS (spheroid grid) onto a two-dimensional surface (known as a developable surface)

you could then trace the shapes of the surface features onto the two-dimensional surface. This two-dimensional surface would be the basis for your map.

There are many different ways to arrange these 3 elements.

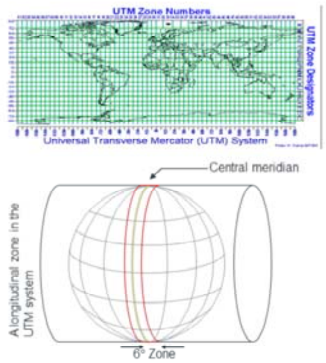


Conceptually, cylindrical projections are created by wrapping a cylinder around a globe and projecting light through the globe onto the cylinder. A conformal projection.

The most famous of all map projections—the Mercator—is a cylindrical projection - most topographic maps you see use a form of mercator. Its virtues are that it is conformal and that any straight line drawn on it is a true compass bearing. This property has made the Mercator (and still makes it) an invaluable nautical chart.

The UTM System

Southampton UNIVERSITY OF GeoData Institute



- UTM divide the world into 60 x 6° longitudinal zones (from 80°S to 84°N)
- Each zone is divided "vertically" into 20 quadrilaterals
- High level of accuracy +/- 3.5° from central meridian of each zone

1. In UTM (from 80°S to 84°N)

   The globe is divided into 60 longitudinal zones each has a width of 6°

   Each zone is further divided "vertically" to give 20 quadrilaterals, each of a height of 8° except the top one which is 12° high

2. UTM zones are designated by a combination of zone number and letter

   Numbers: 1 – 60 in which the central meridian of Zone 1 coincides with the 180th meridian

   Letters: C-X, omitting I and O and leaving A, B, Y and Z for UPS in the polar regions

   e.g. Tasmania (outlined in red in the upper diagram) is located in Zone 55G

3. A scale factor (SF) of 0.9996 is applied to the central meridian of the zone in order to minimise the distortion toward the edges

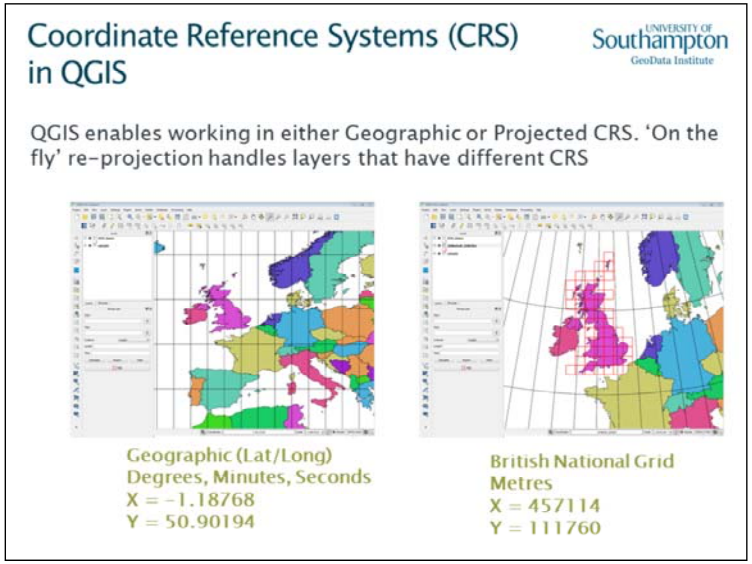
Zones designed to work well with data falling within +-3.5 degrees of their central meridians. This provides a one-degree overlap between neighbouring zones. Because UTM is based on a Mercator projection, however, the distortion becomes exponentially greater as one moves away from the central meridian.

Coordinate Systems

Southampton UNIVERSITY OF GeoData Institute

A classic coordinate system error! **Datum shift**

This is an all too common error in GIS operations – Spatial referencing error due to an incorrectly defined datum.



Two types of Coordinate Systems:

1) <u>Geographic coordinate system</u>

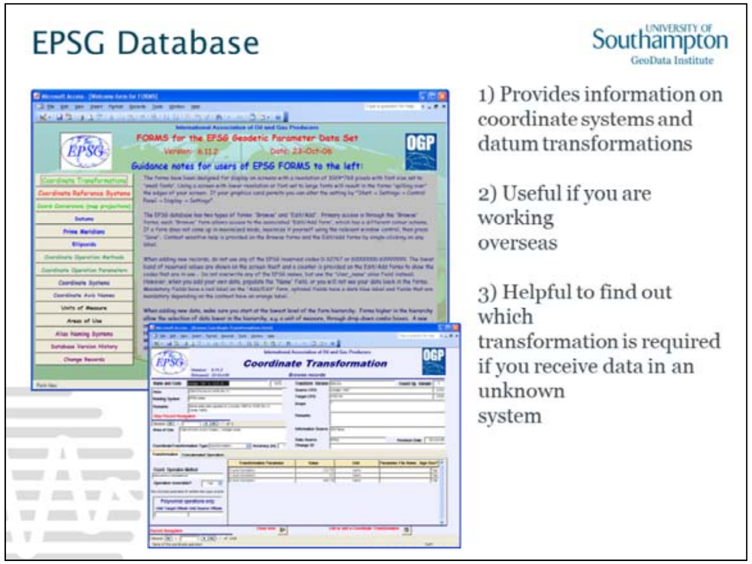
A reference system that uses latitude and longitude to define the locations of points on a spherical model of the earth's surface.

2) <u>Projected coordinate system (rectangular)</u>

Projected coordinate systems use a mathematical conversion to transform latitude and longitude coordinates that fall on the earth's three-dimensional surface to a two-dimensional surface. A projected coordinate system is defined by a **geographic coordinate system**, a **map projection**, any **parameters** needed by the map projection, and a **linear unit of measure**.

A projected coordinate system is a system that defines locations on a flat map and is based on x,y coordinates

The coordinate system (geographic or projected) defines the location of the spatial data on the earth.



EPSG database can be downloaded from the internet at: http://www.epsg.org/databases/Discv7_4.html

Introduction to GIS – Topics

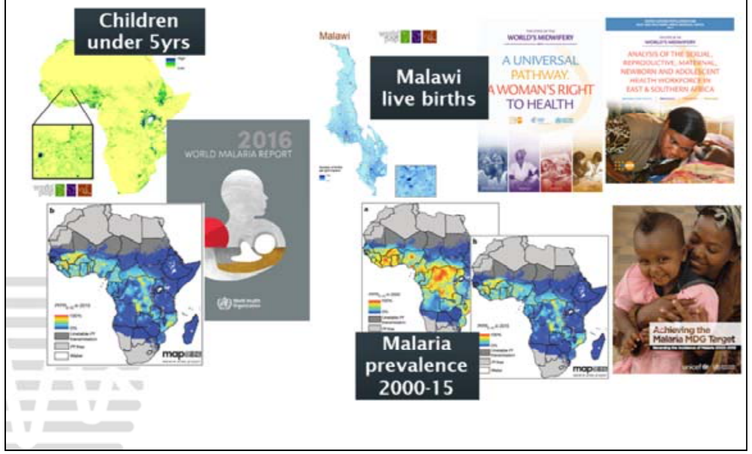
University of Southampton GeoData Institute

- What is GIS?
- Types of Questions a GIS can Answer
- Spatial Data
- GIS Project Workflow
- GIS Principles
    - Types of Geographic Information
    - GIS Data Models
    - Mapping in Layers
    - Co-ordinate Systems
- GIS Applications and Impact

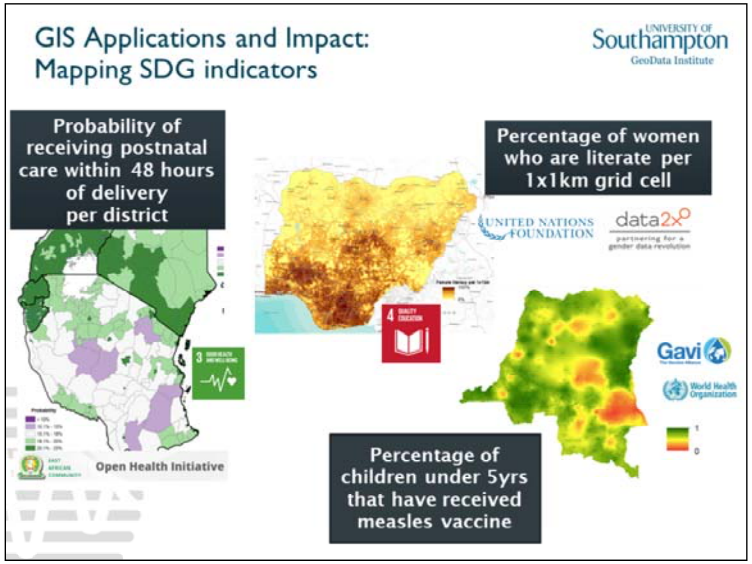
1

GIS Applications and Impact: Health Metrics

University of Southampton GeoData Institute



Are all of these new datasets and methods just crazy academic ideas? No, they're already being used to produce health metrics, situation reports, provide development indicators. A small subset of these examples are provided here, with census disaggregation gridded population datasets used to provide the numbers presented in the World Malaria report, humanitarian needs assessments and World Bank reports. Maps of births and pregnancies were used to provide statistics in recent UNFPA midwifery reports, and mapping from GPS-located survey data was used to measure and map changing malaria prevalence in a recent UNICEF-WHO report.



Examples of maps produced in similar ways for metrics relating to all of the targets highlighted previously – in each case highlighting heterogeneities and hotspots of populations being left behind that are masked by national or provincial-level aggregate indicators. These can be produced with (using small area estimation techniques) or without the use of census data, and each has accompanying validation statistics and uncertainty maps.