Spatial Data

Learning outcomes

- Explain the difference between data and information
- Describe the main characteristics of spatial data
- Give examples of map projections and why they are important
- Provide details of different methods of spatial referencing
- Define topology
- Explain the thematic characteristics of spatial data
- List the main sources of spatial data
- Explain why data standards are an important issue in GIS

Introduction

Simplified / filtered view of the World

Data and Information

- Data – raw facts
- Information – data with meaning and context added

Modes

- Spatial, Temporal, Thematic

A piece of data is 'spatial' only if it is referenced to location

Data sources

- Primary – captured specifically for use in your project
- Secondary – created for some other purpose but useful in your project

Representations

Are needed to convey information

Fit information into a standard form or model

Almost always simplify the truth that is being represented

Representations can rarely be perfect

- Details can be irrelevant, or too expensive and voluminous to record

It’s important to know what is missing in a representation

- Representations can leave us uncertain about the real world

The Fundamental Problem

Geographic information links a place, and often a time, with some property of that place (and time)

- “The temperature at 34 N, 120 W at noon local time on 12/2/99 was 18 Celsius”

The potential number of properties is vast

- In GIS we term them *attributes*
- Attributes can be physical, social, economic, demographic, environmental, etc.

The number of places and times is also vast

- Potentially infinite

The more closely we look at the world, the more detail it reveals

- Potentially *ad infinitum*
- The geographic world is infinitely complex

Humans have found ingenious ways of dealing with this problem

- Many methods are used in GIS to create representations or *data models*

Discrete Objects and Continuous Fields

Two ways of conceptualizing geographic variation

- The most fundamental distinction in geographic representation

Discrete objects

- The world as a table-top
- Objects with well-defined boundaries

Continuous Fields

- Properties that vary continuously over space

Discrete Objects

Points, lines, and areas

Countable

Persistent through time, perhaps mobile

Biological organisms

- Animals, trees

Human-made objects

- Vehicles, houses, fire hydrants

Continuous Fields

Properties that vary continuously over space

- Value is a function of location
- Property can be of any attribute type, including direction

Elevation as the archetype

- A single value at every point on the Earth's surface
- The source of metaphor and language
  - Any field can have slope, gradient, peaks, pits

Spatial entities - Vector

Points

- Single point (0 dimension), (x,y)

Lines

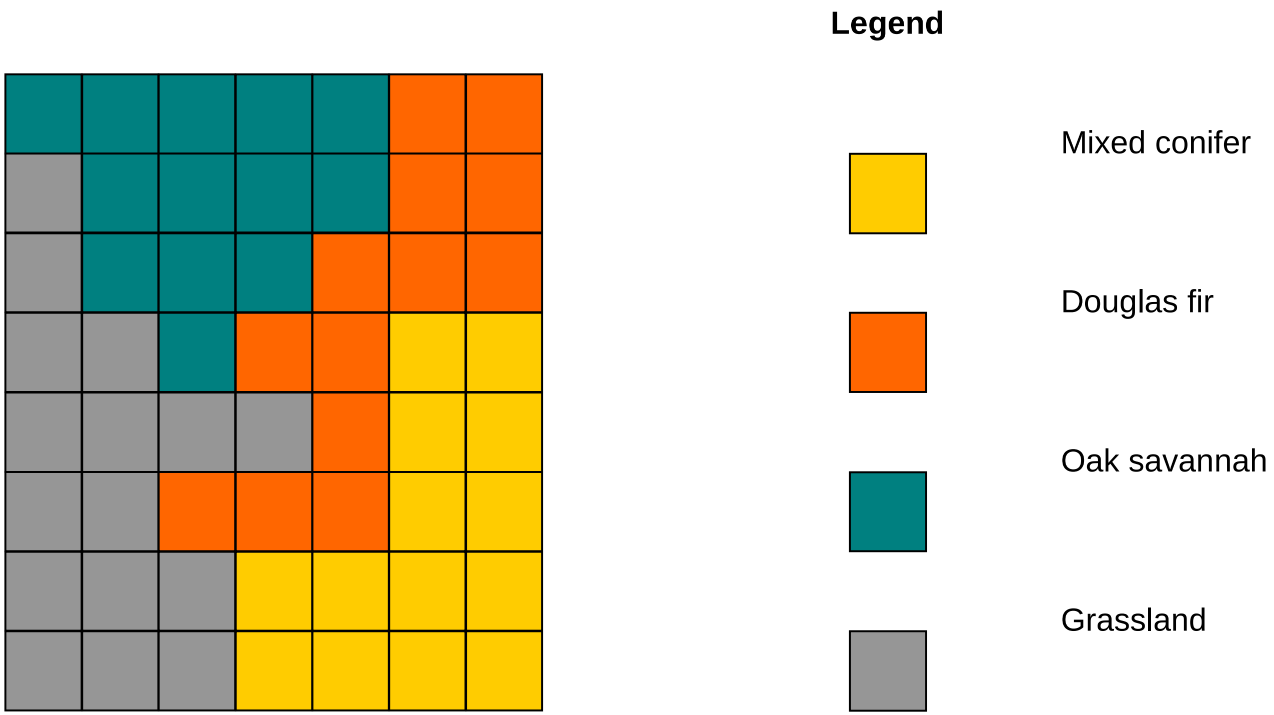
- Drawn between two point locations (1 dimension)
- Polyline – multiple line segments – list of point locations – (x1y1, x2y2 ... xnyn)

Polygons (Areas)

- As polyline but first & last points are equal thus ensuring closure
- Represents area (2 dimensions)

Note chosen representation varies with scale

Raster



Raster representation. Each color represents a different value of a nominal-scale field denoting land cover class.

Maps and their influence on the character of spatial data

- Traditional method of storing & analysing spatial data
- Shapes the way we think about space
- Purpose
  - Thematic
  - General purpose / topographic
  - Prerequisite for judging quality / accuracy / completeness of data

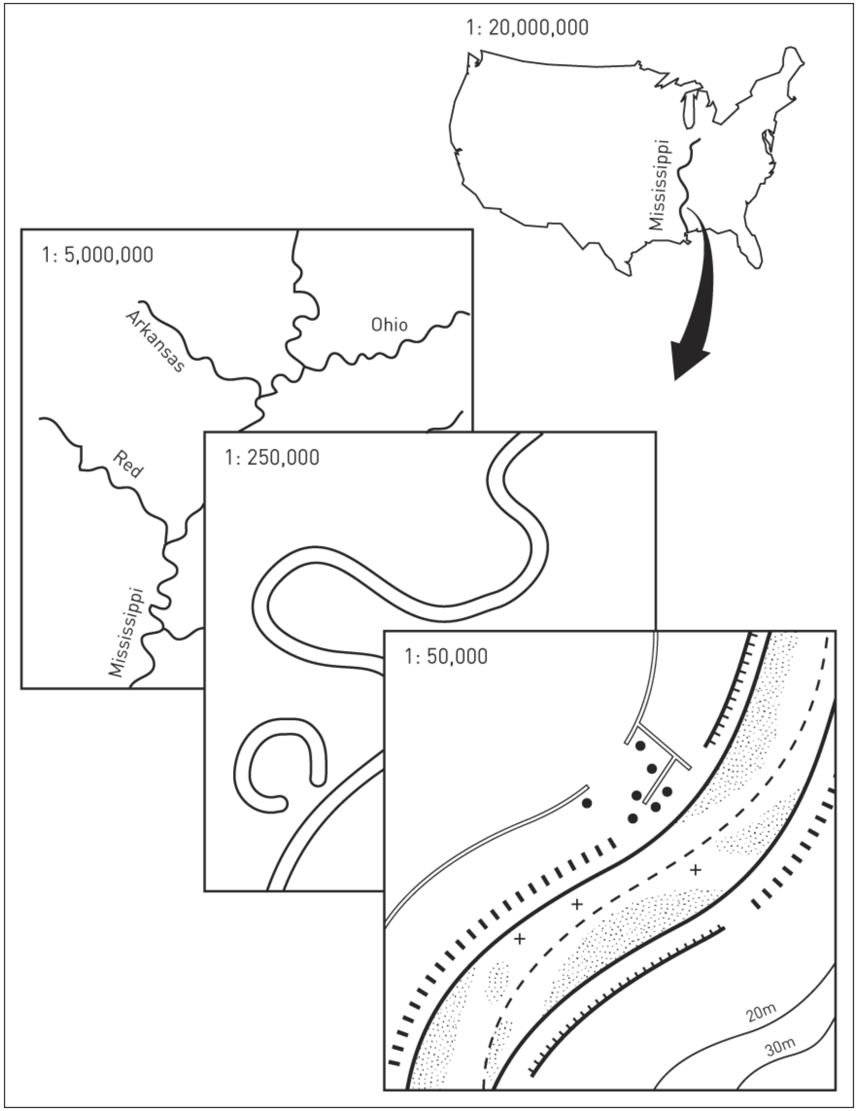
Scale

Ratio of distance on the map to distance on the ground

Large scale => small area & vice versa

Ratio	1:5000	1:1,000,000
Verbal (nominal)	1 cm represents 50 m	1 cm represents 10 km
Graphical	0 100 200 km	0 10 20 30 40 km

Scale-related generalization



Generalization issues

To avoid clutter some compromises must be made

Selection

Simplification

Displacement

Smoothing & enhancement

Projections

Problem: how to flatten an (approximate) sphere (the earth) onto a 2d surface

There are many reasons for wanting to project the Earth's surface onto a plane, rather than deal with the curved surface

- The paper used to output GIS maps is flat
- Flat maps are scanned and digitized to create GIS databases
- Rasters are flat, it's impossible to create a raster on a curved surface
- The Earth has to be projected to see all of it at once
- It's much easier to measure distance on a plane

Distortions

Any projection must distort the Earth in some way

Two types of projections are important in GIS

- *Conformal* property: Shapes of small features are preserved: anywhere on the projection the distortion is the same in all directions
- *Equal area* property: Shapes are distorted, but features have the correct area
- Both types of projections will generally distort distances

Cylindrical Projections

Conceptualized as the result of wrapping a cylinder of paper around the Earth

The Mercator projection is the best-known cylindrical projection

- The cylinder is wrapped around the Equator
- The projection is conformal
  - At any point scale is the same in both directions
  - Shape of small features is preserved
  - Features in high latitudes are significantly enlarged

Conic Projections

Conceptualized as the result of wrapping a cone of paper around the Earth

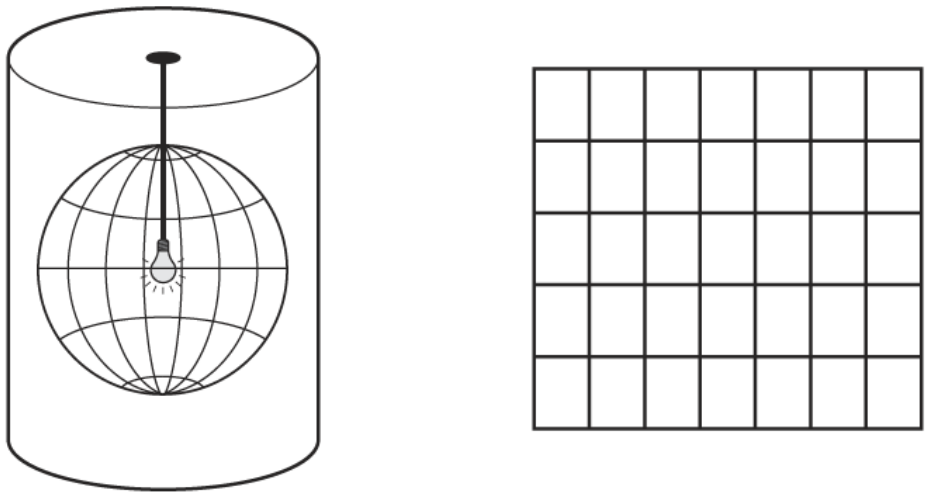
- Standard Parallels occur where the cone intersects the Earth
- The Lambert Conformal Conic projection is commonly used to map North America
- On this projection lines of latitude appear as arcs of circles, and lines of longitude are straight lines radiating from the North Pole

The “Unprojected” Projection

Assign latitude to the *y* axis and longitude to the *x* axis

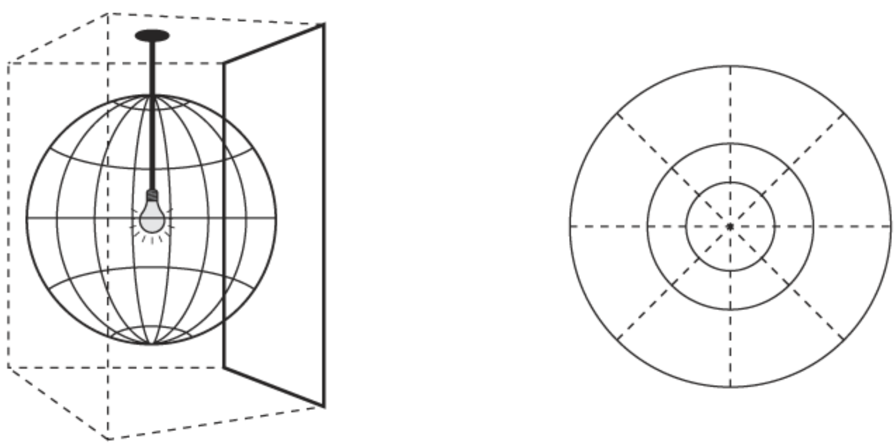
- A type of cylindrical projection
- Is neither conformal nor equal area
- As latitude increases, lines of longitude are much closer together on the Earth, but are the same distance apart on the projection
- Also known as the Plate Carrée or Cylindrical Equidistant Projection

Projections: (a) cylindrical; (b) azimuthal



- Continuous picture of the Earth
- Countries near the equator in true relative positions
- Distance increases between countries located towards top and bottom of image
- The view of the poles is very distorted
- Area for the most part is preserved

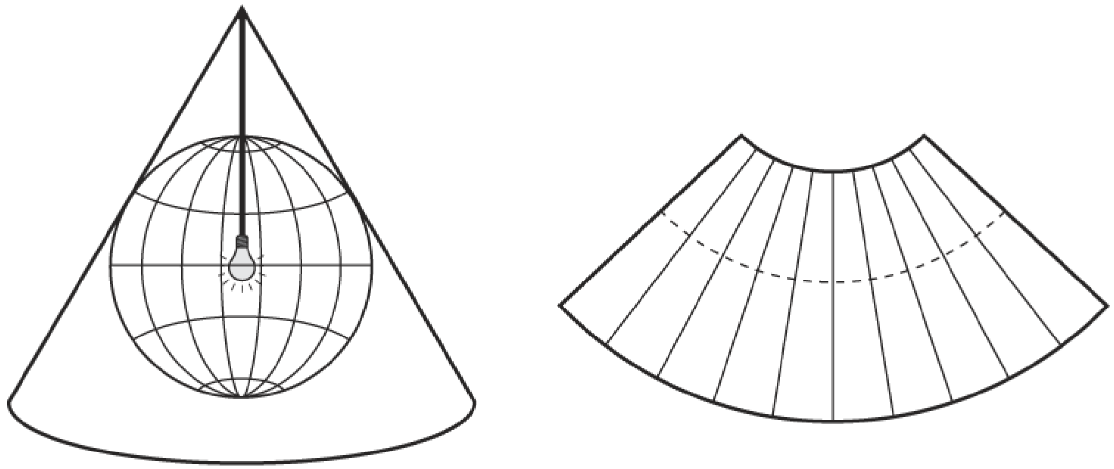
(a) Cylindrical projection *(light in a circular room analogy)*



- Only a part of the Earth's surface is visible
- The view will be of half the globe or less
- Distortion will occur at all four edges
- Distance for the most part is preserved

(b) Azimuthal projection *(light in a square room with flat walls analogy)*

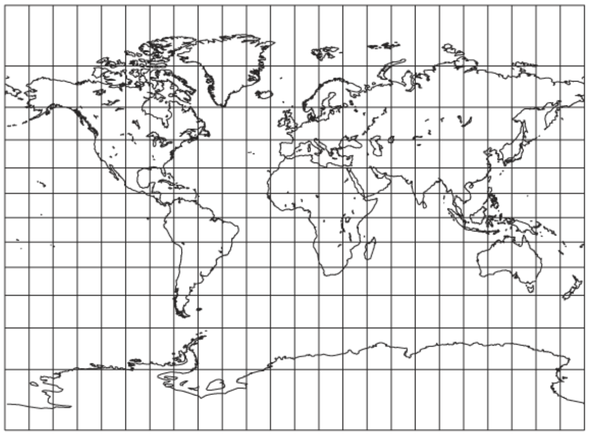
Projections: (c) conic



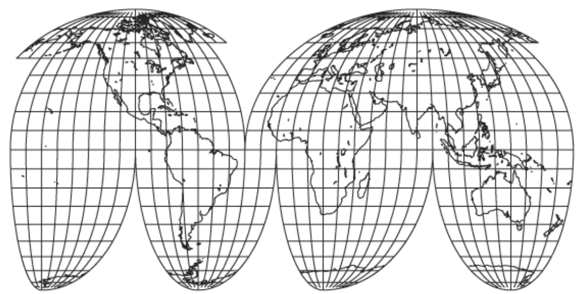
- Area is distorted
- Distance is very distorted towards the bottom of the image
- Scale for the most part is preserved

(c) Conic projection *(light in a tepee analogy)*

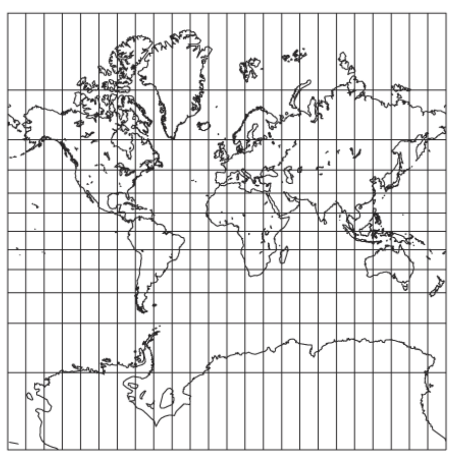
Commonly used projections



(c) Miller Cylindrical



(d) Mollweide



(e) Mercator

(b) Lambert conformal conic

Spatial referencing

Placenames

Postal addresses and postal codes
Linear referencing systems
Cadasters
Latitude and longitude
Projections and coordinate systems

Spatial referencing

- Geographic co-ordinate systems
  - Latitude/Longitude
- Rectangular co-ordinate systems
  - Grid
- Non co-ordinate systems
  - Postal addresses
  - Postcodes

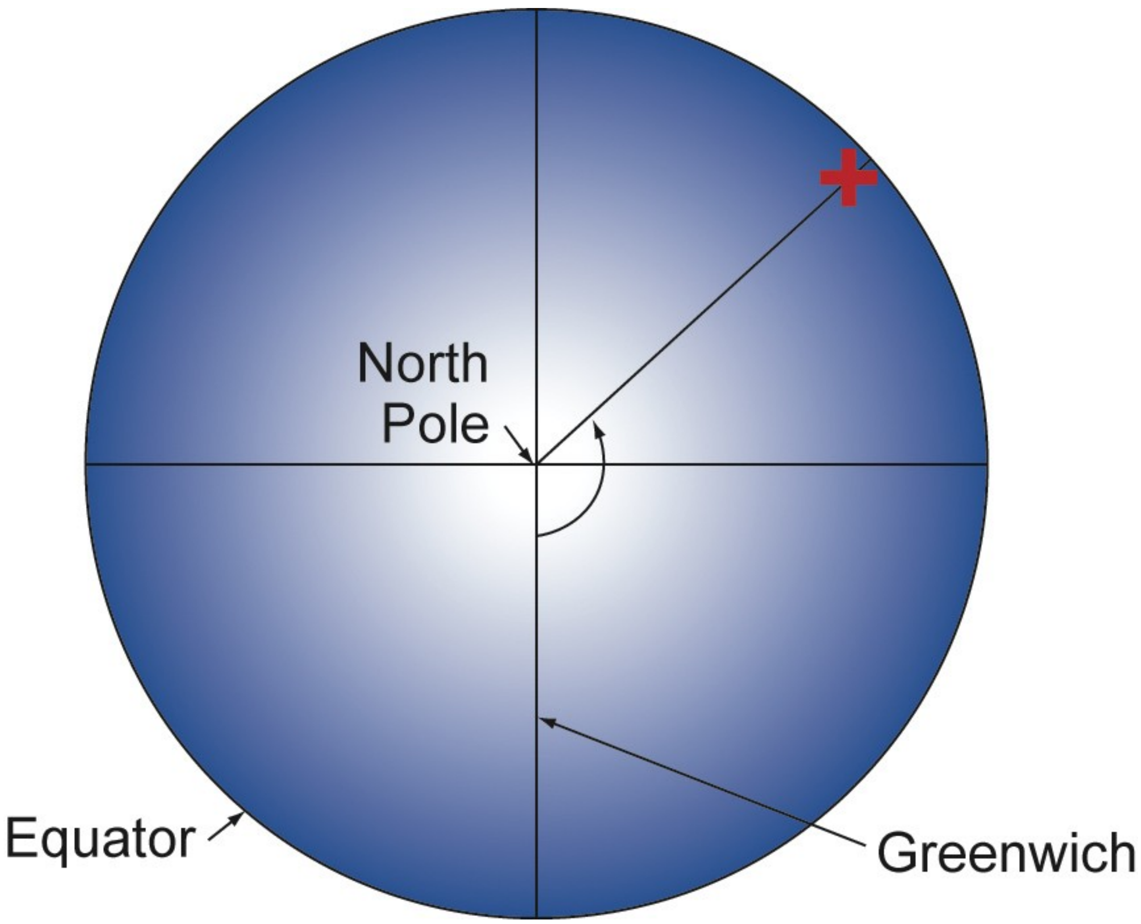
Latitude and Longitude

The most comprehensive and powerful method of georeferencing

- Metric, standard, stable, unique

Uses a well-defined and fixed reference frame

- Based on the Earth's rotation and center of mass, and the Greenwich Meridian



Definition of longitude. The Earth is seen here from above the North Pole, looking along the Axis, with the Equator forming the outer circle. The location of Greenwich defines the Prime Meridian. The longitude of the point at the center of the red cross is determined by drawing a plane through it and the axis, and measuring the angle between this plane and the Prime Meridian.

Definition of Latitude

Requires a model of the Earth's shape

The Earth is somewhat elliptical

- The N-S diameter is roughly 1/300 less than the E-W diameter
- More accurately modeled as an ellipsoid than a sphere
- An ellipsoid is formed by rotating an ellipse about its shorter axis (the Earth's axis in this case)

The History of Ellipsoids

Because the Earth is not shaped precisely as an ellipsoid, initially each country felt free to adopt its own as the most accurate approximation to its own part of the Earth

Today an international standard has been adopted known as WGS 84

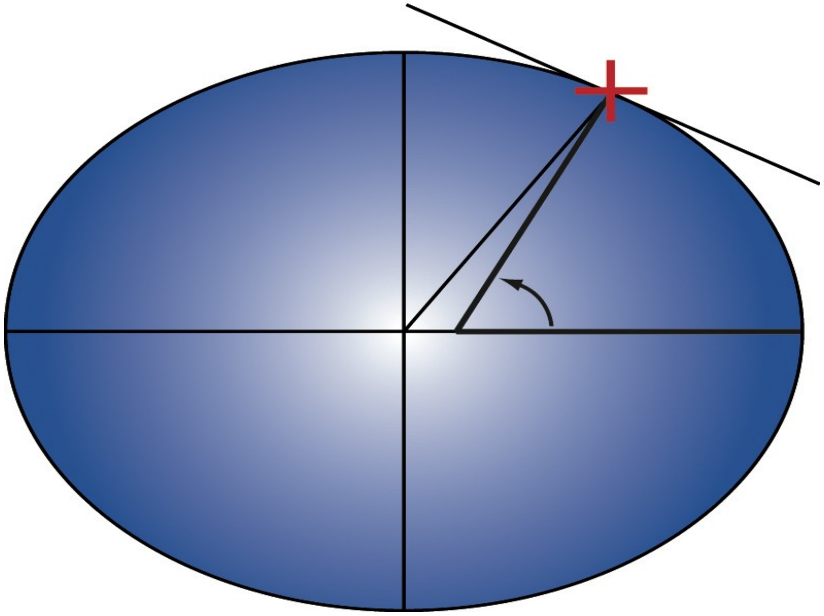
- Its US implementation is the North American Datum of 1983 (NAD 83)
- Many US maps and data sets still use the North American Datum of 1927 (NAD 27)
- Differences can be as much as 200 m

Latitude and the Ellipsoid

Latitude (of the blue point) is the angle between a perpendicular to the surface and the plane of the Equator

WGS 84

- Radius of the Earth at the Equator 6378.137 km
- Flattening 1 part in 298.257



Georeferencing

Is essential in GIS, since all information must be linked to the Earth's surface

The method of georeferencing must be:

- Unique, linking information to exactly one location
- Shared, so different users understand the meaning of a georeference
- Persistent through time, so today's georeferences are still meaningful tomorrow

Uniqueness

A georeference may be unique only within a defined domain, not globally

- There are many instances of Springfield in the U.S., but only one in any state
- The meaning of a reference to London may depend on context, since there are smaller Londons in several parts of the world

Georeferences as Measurements

Some georeferences are metric

- They define location using measures of distance from fixed places
    - E.g., distance from the Equator or from the Greenwich Meridian

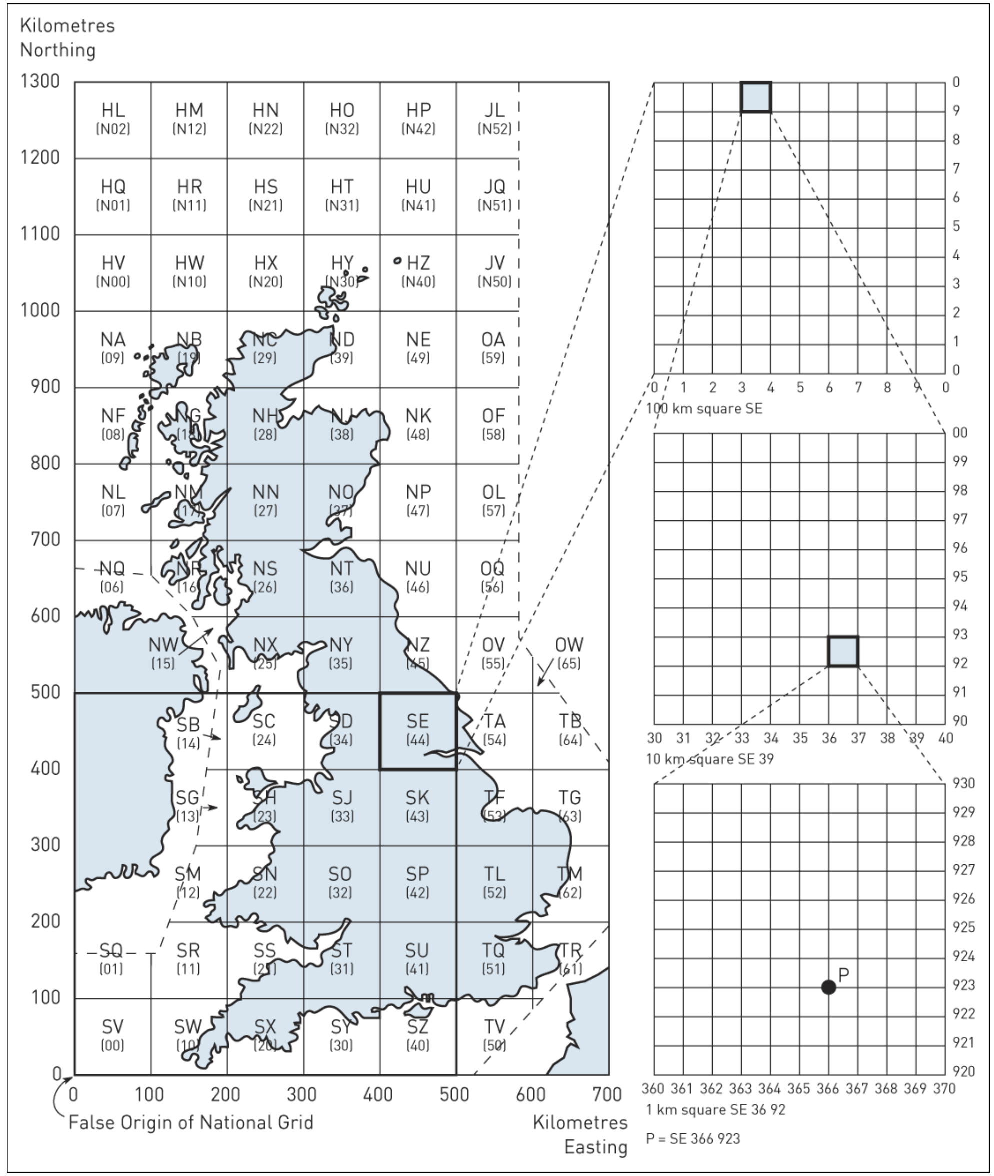
Others are based on ordering

- E.g. street addresses in most parts of the world order houses along streets

Others are only nominal

- Placenames do not involve ordering or measuring

UK National Grid



Placenames

The earliest form of georeferencing

- And the most commonly used in everyday activities

Many names of geographic features are universally recognized

- Others may be understood only by locals

Names work at many different scales

- From continents to small villages and neighborhoods

Names may pass out of use in time

- Where was Camelot?

Postal Addresses and Postcodes

Every dwelling and office is a potential destination for mail

Dwellings and offices are arrayed along streets, and numbered accordingly

Streets have names that are unique within local areas

Local areas have names that are unique within larger regions

If these assumptions are true, then a postal address is a useful georeference

Where Do Postal Addresses Fail as Georeferences?

In rural areas

- Urban-style addresses have been extended recently to many rural areas

For natural features

- Lakes, mountains, and rivers cannot be located using postal addresses

When numbering on streets is not sequential

- E.g. in Japan

Postcodes as Georeferences

Defined in many countries

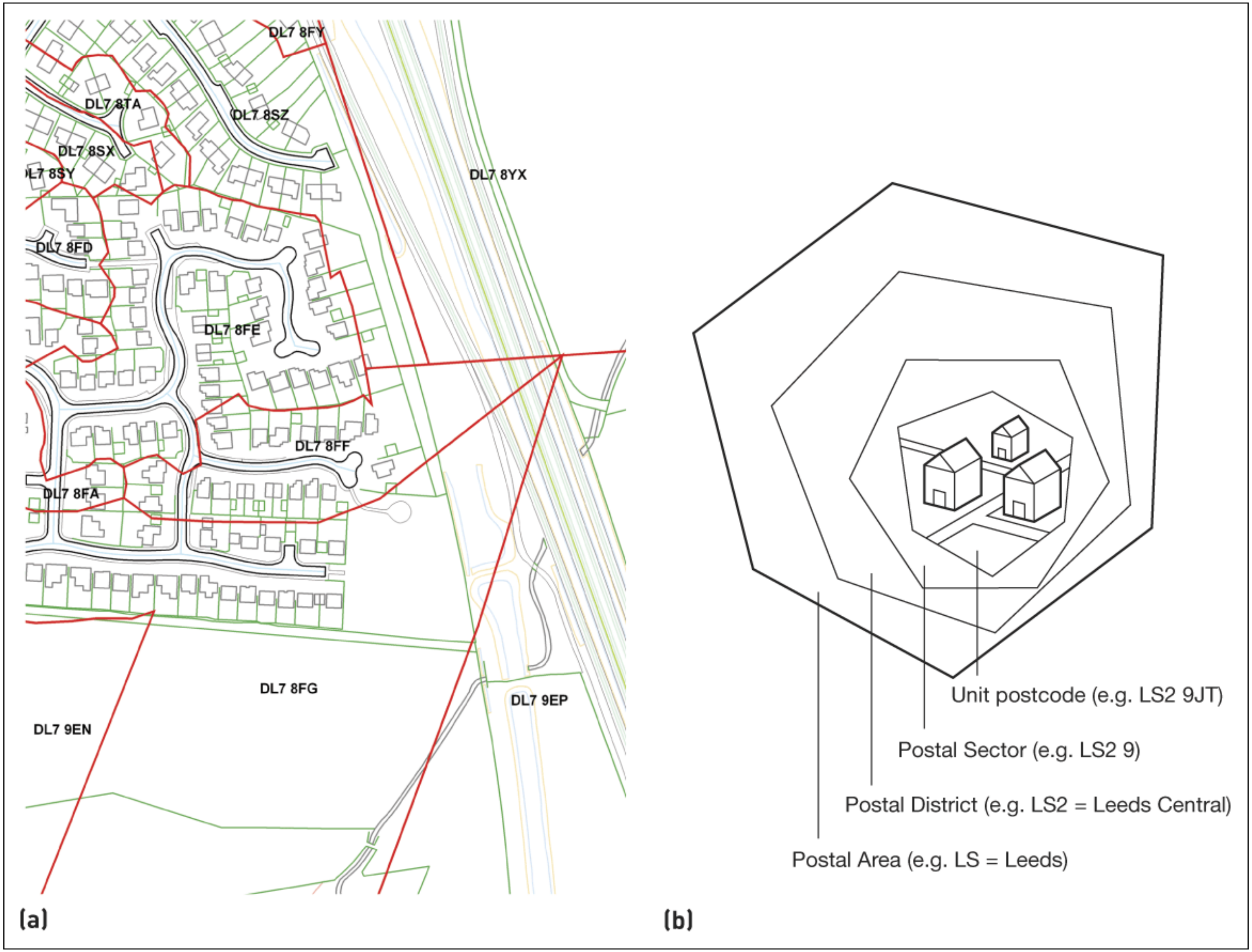
- E.g. ZIP codes in the US

Hierarchically structured

- The first few characters define large areas
- Subsequent characters designate smaller areas
- Coarser spatial resolution than postal address

Useful for mapping

UK Postcode System



Converting Georeferences

GIS applications often require conversion of projections and ellipsoids

- These are standard functions in popular GIS packages

Street addresses must be converted to coordinates for mapping and analysis

- Using *geocoding* functions

Placenames can be converted to coordinates using *gazetteers*

Topology

Geometric characteristics of ojbects which do not change under transformation

Three elements

- Adjacency
- Containment
- Connectivity

Useful in spatial analysis, routing etc.

Thematic characteristics of spatial data

Entities and attributes

Entity -> something in which we have an interest

Attribute -> characteristics of an entity

How does a line know it's a road as opposed to a river?

Scales of measurement

Types of Attributes

Nominal, e.g. land cover class

Ordinal, e.g. a ranking

Interval, e.g. Celsius temperature

- Differences make sense

Ratio, e.g. Kelvin temperature

- Ratios make sense

Cyclic, e.g. wind direction

Cyclic Attributes

Do not behave as other attributes

- What is the average of two compass bearings, e.g. 350 and 10?

Occur commonly in GIS

- Wind direction
- Slope aspect
- Flow direction

Special methods are needed to handle and analyze

Sources of spatial data

Apart from maps ...

- Census and survey data
- Aerial photographs
- Satellite imagery
- Field survey and GPS
- ...
- Anything that can record location

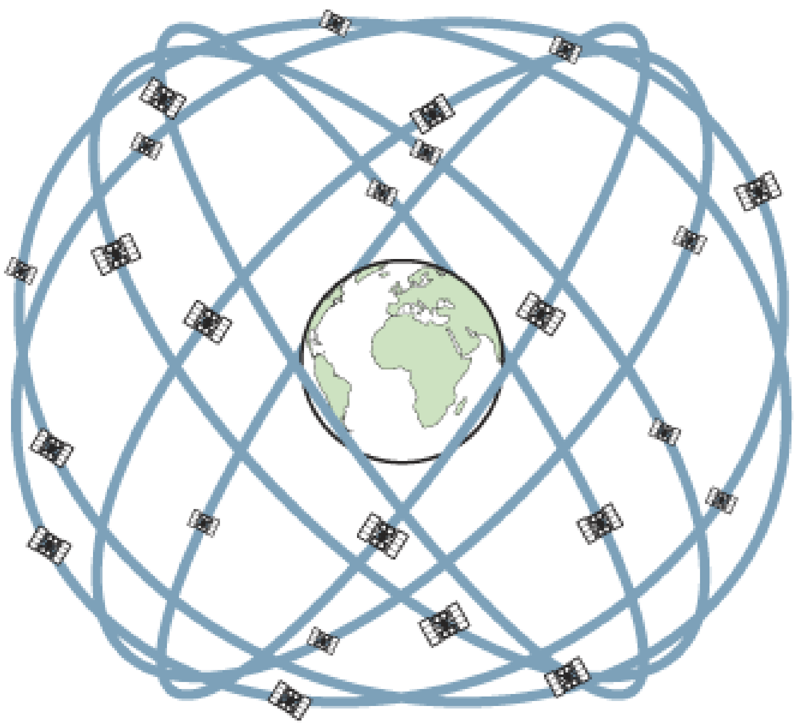
The Global Positioning System

Allows direct, accurate measurement of latitude and longitude

Accuracy of 10m from a simple, cheap unit

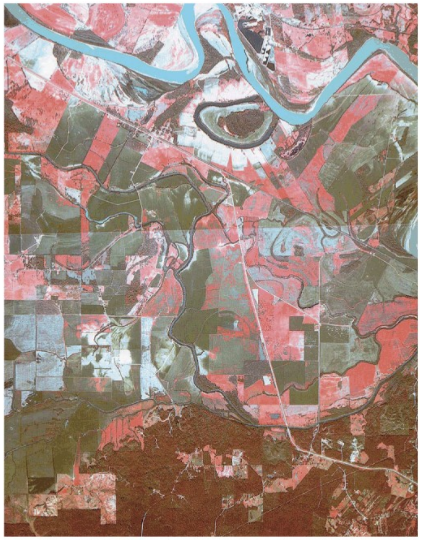
- Differential GPS capable of sub-meter accuracy
- Sub-centimeter accuracy if observations are averaged over long periods

GPS satellite constellation



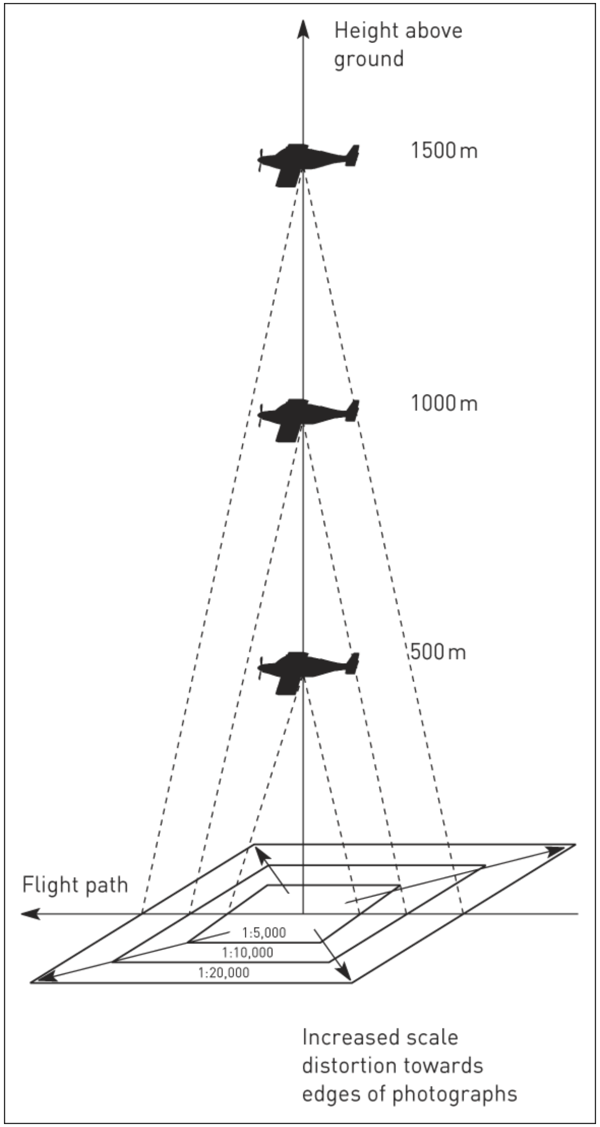
(a) Satellite configuration

Aerial photographs

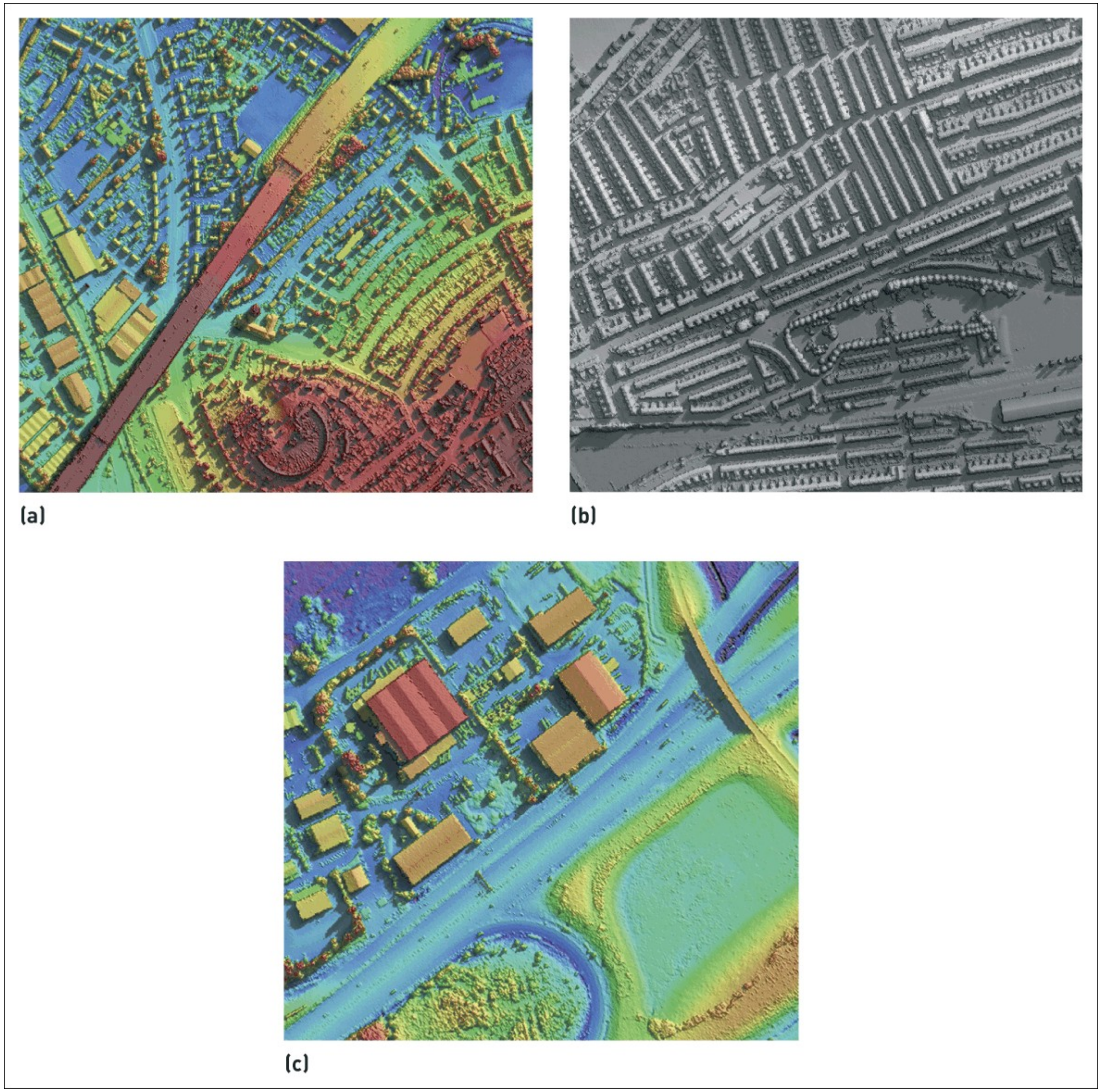

(a) Infrared vertical aerial photograph


(b) Vertical colour aerial photographs showing archaeological remains


(c) Vertical black and white aerial photograph



LIDAR



Data issues

- Data quality
  - GIS nearly always relies on exterlayy sourced data – unlike other IT systems
  - Need to be confident in data source
  - Need to be aware of characteristcis of data – provenance, age, method of collection, resolution, known error
- Quality => fitness for purpose – comes at a cost
- Data standards
  - Open Geospatial Consortium (OGC)
  - Technical standards – file formats, database representations etc
  - Interchange standards - XML-based