# Geographic Information Systems 2018/19
## *Week 8, Topic 1*

Mark Foley
School of Computing
DIT, Kevin Street

mark.foley@dit.ie

# *In this session...*

**Data acquisition, transformation & accuracy**

Acquisition

- Sources of spatial data

Transformation

- Line simplification
- Vector to/from Raster

Accuracy

- Standards
- Error handling

# Acquisition

Getting your GIS data into a computer

- Lots of formats both analogue and digital
- Error-prone therefore requires editing

May need to be

- Re-formatted
- Re-projected
- Generalized
- Matched/joined

# Existing GIS Data

Data already in GIS format

- Can be from many sources “official” and “unofficial”
- Trust issues – need to make value judgement, need good metadata
- Many formats, scales  and projections – may need to transform to consistent base
- May be accessed via API or may need to be downloaded and stored

# Example formats

- Shapefiles
- GPX
- GeoTIFF
- Open Geospatial Consortium standards
  - WMS
  - WFS
  - OWS
  - WCS
- PostGIS (and other database formats)
- Spatialite
- Interchange formats
  - GML
  - GeoJSON
- Web Mapping  such as
  - OpenStreetMap (data extracts easily obtained)
  - Google Maps (some functions accessible via API)

# Example sources

- OpenStreetMap
  - Data download in QGIS
  - Overpass read-only API
- Shuttle Radar Topography Mission (SRTM)
  - Height DEM in GeoTIFF or ASCII Grid
- Geonames.org
  - Placenames
- Smart Dublin
  - https://data.smartdublin.ie/
- data.gov.ie
  - https://data.gov.ie/
- Ordnance Survey
  - MapGenie
  - GeoPortal
  - GeoHive

# Open Geospatial Consortium Formats (WMS)

## Web Map Service (WMS)

A standard protocol for serving georeferenced map images over the Internet.

Requests

WMS specifies a number of different request types, two of which are required by any WMS server

**GetCapabilities** – returns parameters about the WMS (such as map image format and WMS version compatibility) and the available layers (map bounding box, coordinate reference systems, URI of the data and whether the layer is mostly opaque or not)

**GetMap** – returns a map image. Parameters include: width and height of the map, coordinate reference system, rendering style, image format

Request types that WMS providers may optionally support include:

**GetFeatureInfo** – if a layer is marked as 'queryable' then you can request data about a coordinate of the map image.

**DescribeLayer** – returns the feature types of the specified layer or layers, which can be further described using WFS or WCS requests. This request is dependent on the SLD Profile of WMS.

**GetLegendGraphic** – return an image of the map's legend image, giving a visual guide to map elements.

Map image

A WMS server usually serves the map in a bitmap format, e.g. PNG, GIF or JPEG.

# Open Geospatial Consortium Formats (WFS)

## Web Feature Service (WFS)

Provides an interface allowing requests for geographical features across the web using platform-independent calls. One can think of geographical features as the "source code" behind a map, whereas the WMS interface or online tiled mapping portals like Google Maps return only an image, which end-users cannot edit or spatially analyze. The XML-based GML furnishes the default payload-encoding for transporting geographic features, but other formats like shapefiles can also serve for transport.

The WFS specification defines interfaces for describing data manipulation operations of geographic features. Data manipulation operations include the ability to:

- get or query features based on spatial and non-spatial constraints
- create a new feature instance
- delete a feature instance
- update a feature instance

The basic Web Feature Service allows querying and retrieval of features. A transactional Web Feature Service (**WFS-T**) allows creation, deletion, and updating of features.

# Data Acquisition where data is not already in GIS format

## Geocoding / Reverse geocoding

**Geocoding** is the process of transforming a physical address description to a location on the Earth's surface (spatial representation in numerical coordinates). **Reverse geocoding**, on the other hand, converts geographic coordinates to a description of a location, usually the name of a place or an addressable location.

Geocoding relies on a computer representation of address points, the street / road network, together with postal and administrative boundaries.

A simple method of geocoding is address interpolation. This method makes use of data from a street geographic information system where the street network is already mapped within the geographic coordinate space. Each street segment is attributed with address ranges (e.g. house numbers from one segment to the next). Geocoding takes an address, matches it to a street and specific segment (such as a block, in towns that use the "block" convention). Geocoding then interpolates the position of the address, within the range along the segment.

# Data Acquisition where data is not already in GIS format

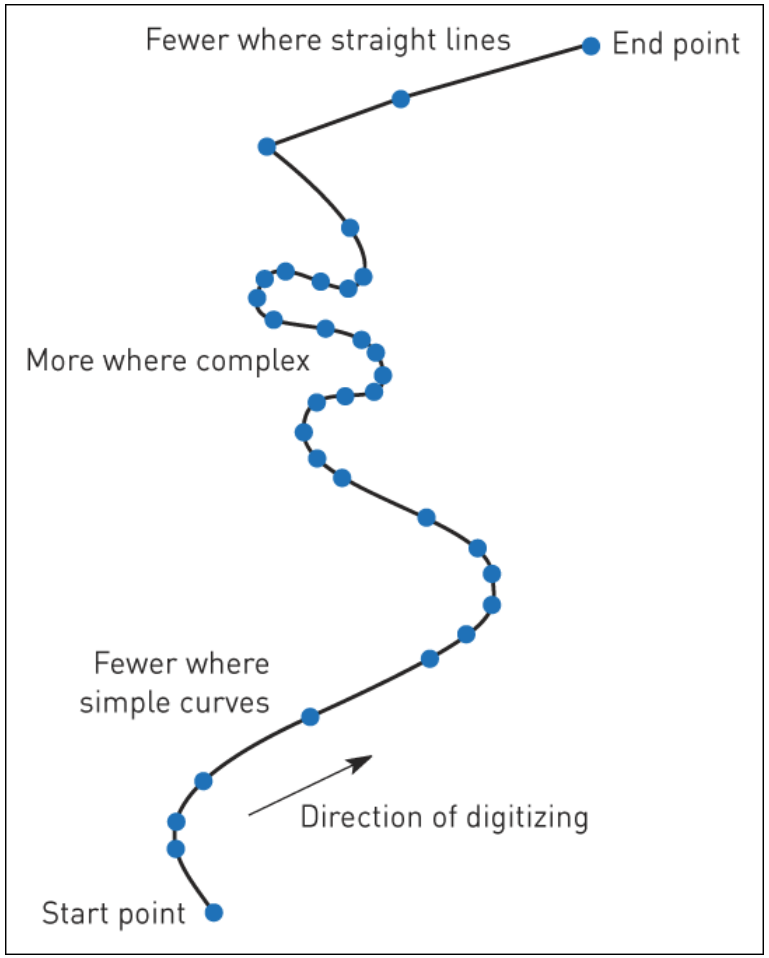
## Georeferencing an image

Georeferencing is the process of assigning real-world coordinates to each pixel of the raster. Many times these coordinates are obtained by doing field surveys - collecting coordinates with a GPS device for few easily identifiable features in the image or map. In some cases, where you are looking to digitize scanned maps, you can obtain the coordinates from the markings on the map image itself. Using these sample coordinates or GCPs (Ground Control Points), the image is warped and made to fit within the chosen coordinate system.

# Create data manually

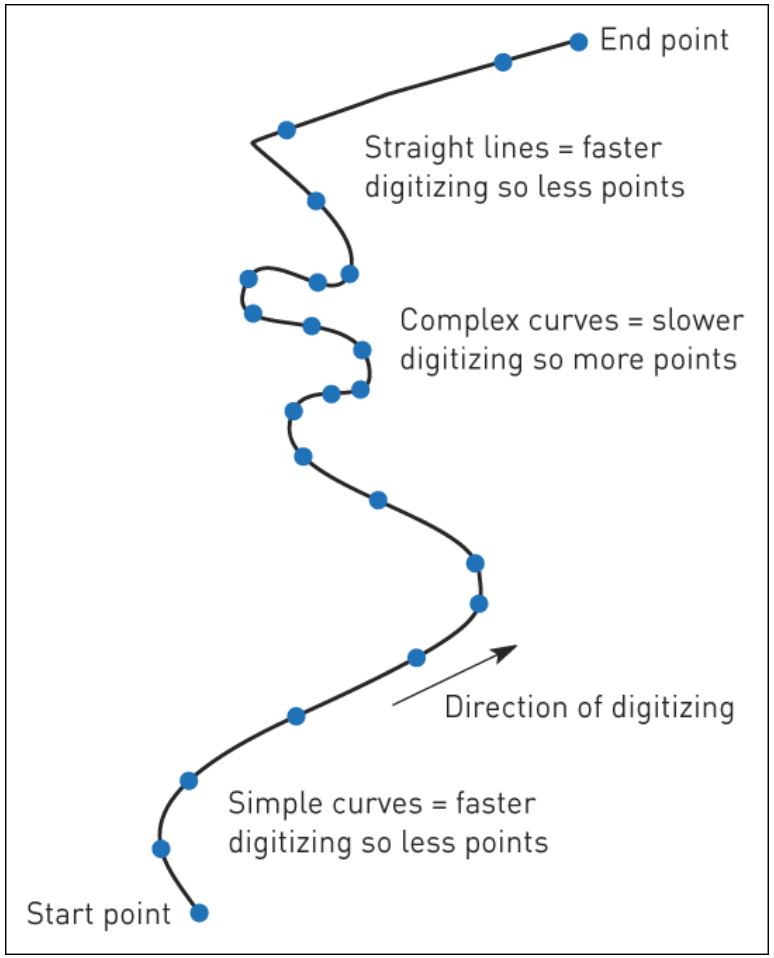
- Manual digitizing
  - Acquiring data from paper maps
  - Registration
  - Digitizing features
  - Adding attributes
  - “Field Papers” example
- Major source of positional errors

# Point and stream mode digitizing




**Point mode** – person digitizing decides where to place each individual point such as to most accurately represent the line within the accepted tolerances of the digitizer. Points are placed closer together where the line is most complex and where the line changes direction. Points are placed further apart where the line is less complex or made up of straight line segments.

# Point and stream mode digitizing

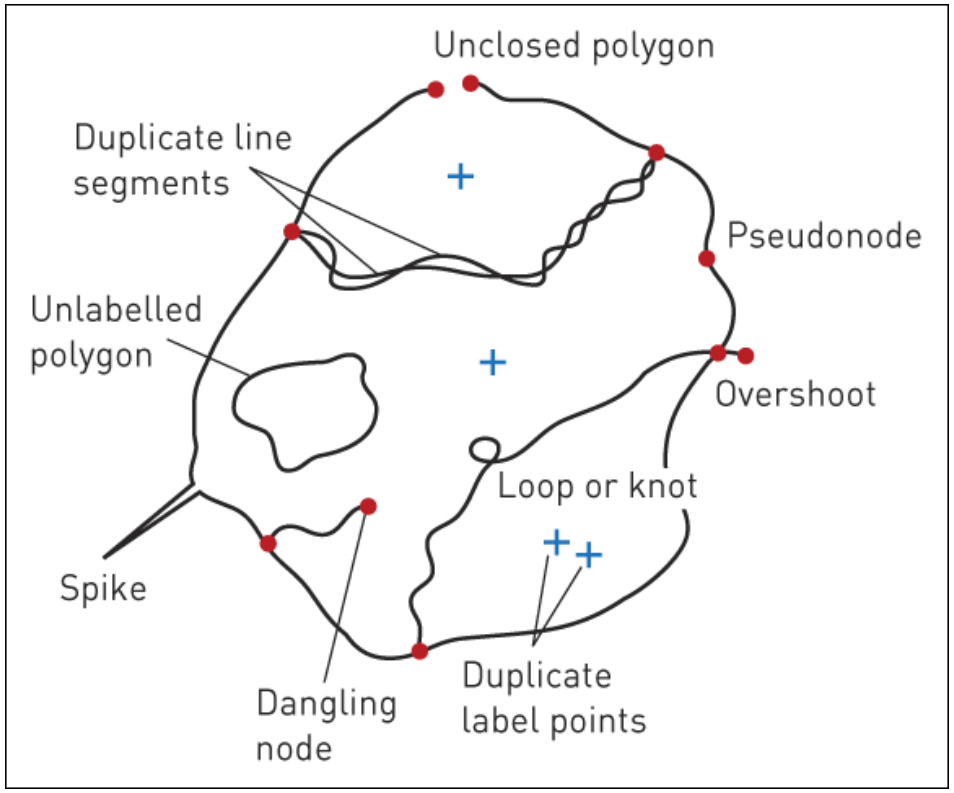



**Stream mode** – person digitizing decides on time or distance interval between the digitizing hardware registering each point as the the person digitizing moves the cursor along the line. Points are placed closer together where the line is most complex only as the person digitizing slows the movement of the cursor down to more accurately follow the line. Points are placed further apart where the line is less complex or made of straight line segments allowing the person digitizing to move the cursor more quickly

# Automatic digitizing

- Automatic digitizing
  - Scanning
  - Appropriate where raster is required
  - Automatic line following
- Problems with scanning
  - Optical distortion
  - Unwanted information (coffee stains, annotations etc.)
  - Selection of scanning tolerances to ensure important data encoded and background ignored
  - File formats produced
  - Amount of editing required

# Detecting and correcting errors




3 main sources of error

- Errors in source data
- Errors introduced during encoding
- Errors propagated during data transfer and conversion

Common errors

- Missing entities
- Duplicate entities
- Mislocated entities
- Missing labels
- Duplicate labels
- Under/overshoots, loops & spikes etc.

# *Coming next...*

- Transformation
  - Line simplification
  - Vector to/from Raster
- Accuracy
  - Standards
  - Error handling