**3.4** **Secondary GIS data methods**

**Reading material**

|  |
| --- |
| In this lesson, you are going to learn about secondary data capturing methods.  Secondary data collection refers to gathering information that is already available. The data was previously collected, has undergone necessary statistical analysis and is not owned by the researcher. This data is usually one that was collected from primary sources and later made available for everyone to access. |
| **Scanning & Georeferencing of Analogue Maps**  Scanning is one of the simplest ways to convert physical documents into digital formats. When you scan a document, the scanner uploads a digital image of the page. Like all images, scanned documents will appear in raster format.  Maps are often scanned in order to:   * Use digital image data as a background for other (vector) map data. * Convert scanned data to vector data for use in a vector GIS.   Scanning requires that the map scanned be of high cartographic quality, with clearly defined lines, text and symbols; be clean and have lines of 0.1mm width or wider. |
| Scanning comprises two operations:   * scanning, which produces a regular grid of pixels with grey-scale levels (usually in the range 0-255) * binary encoding – to separate the lines from the background using automated feature recognition techniques. |
| Editing of scanned data can include pattern recognition of shapes and symbol candidates; line thinning and vectorization; error correction; supplementing missing data and forming topology.   * Once maps are scanned, they will then be subjected to ***Georeferencing process***  which is a process of associating features on the scanned image with real world x and y coordinates. It is therefore the process through which geographical information systems use the positions of some reference points to fix the geographic location of each topographic feature so that it can be viewed, queried, and analyzed with other geographic data. Georeferencing may involve shifting, rotating, scaling, skewing, and in some cases warping, rubber sheeting, or orthorectifying the data. |
| **Digitizing and vectorization – vector data capture**  Digitizing is the transformation of information from analogue format, such as a paper map, to digital format, so that it can be stored and displayed with a computer. Digitizing can be manual, semi-automated (automatically recorded while manually following a line), or fully automated (line following).  Manual digitizing involves an operator using a digitizing table (or tablet) (known as heads-down digitizing), or with the operator using a computer screen (heads-up digitizing). |
| The two most commonly used digitizing methods are :   * Point-Mode digitizing : A method of digitizing in which the digitizer selects particular points, or vertices, to encode. * Stream mode digitizing : a method of digitizing in which, as the cursor is moved, points are recorded automatically at preset intervals of either distance or time. With stream digitizing you set a tolerance, and points will be added only when you move the specified number of pixels in either the x or y direction. A large tolerance leads to fewer points, while a small tolerance leads to smoother curves.     Source : Muvunyi Germain, INES Ruhengeri (lecture notes) |
| **Topology** expresses the spatial relationships between connecting or adjacent vector features (points, polylines and polygons) in a GIS. Topological or topology-based data are useful for detecting and correcting digitising errors (e.g. two lines in a roads vector layer that do not meet perfectly at an intersection). Topology is necessary for carrying out some types of spatial analysis, such as network analysis. |
| The process of Digitization involves ***topology rules and errors.***  **Topological errors:**   * **Dangles or dangling nodes** are lines that are not connected but should be connected. With dangling nodes, gaps occur in the linework where the two lines should be connected. Dangling nodes also occur when a digitized polygon doesn’t connect back to itself, leaving a gap where the two end nodes should have connected, creating what is called an open polygon. * **Switchbacks, Knots, and Loops:** These types of errors are introduced when the digitizer has an unsteady hand and moves the cursor or puck in such a way that the line being digitized ends up with extra vertices and/or nodes. In the case of switchbacks, extra vertices are introduced and the line ends up with a bend in it. With knots and loops, the line folds back onto itself, creating small polygon like geometry known as [weird polygons](https://www.gislounge.com/what-are-weird-polygons/). * **Overshoots and Undershoots:** Similar to dangles, overshoots and undershoots happen when the line digitized doesn’t connect properly with the neighbouring line it should intersect with. During digitization a snap tolerance is set by the digitizer. The snap tolerance or snap distance is the measurement of the diameter extending from the point of the cursor. Any nodes of neighbouring lines that fall within the circle of the snap tolerance will result in the end points of the line being digitized automatically snapping to the nearest node. Undershoots and overshoots occur when the snap distance is either not set or is set too low for the scale being digitized. Conversely, if the snap distance is set too high and the line endpoint snaps to the wrong node. In a few cases, undershoots and overshoots are not actually errors. One instance would be the presence of cul-de-sacs (i.e., dead ends) within a road GIS database. * **Slivers** are gaps in a digitized polygon layer where the adjoining polygons have gaps between them. Again, setting the proper parameters for snap tolerance is critical for ensuring that the edges of adjoining polygons snap together to eliminate those gaps. Where the two adjacent polygons overlap in error, the area where the two polygons overlap is called a sliver.     Source : <https://www.gislounge.com/what-are-weird-polygons/> |
| **Topology rules**  Fortunately, many common errors that can occur when digitising vector features can be prevented by topology rules that are implemented in many GIS applications. Except for some special GIS data formats, topology is usually not enforced by default. Many common GIS, like QGIS, define topology as relationship rules and let the user choose the rules, if any, to be implemented in a vector layer. |
| The following list shows some examples of where topology rules can be defined for real world features in a vector map:   * Area edges of a municipality map must not overlap. * Area edges of a municipality map must not have gaps (slivers). * Polygons showing property boundaries must be closed. Undershoots or overshoots of the border lines are not allowed. * Contour lines in a vector line layer must not intersect (cross each other).   GIS data collection  Source : <https://www.agiratech.com/introduction-to-topology-gis> |
| **Photogrammetry – vector data capture**  Photogrammetry is the process of determining measurements from overlapping images, essentially creating 3D models by analysing multiple photos of a single area. This complex science is surprisingly affordable and can result in very detailed topographic visualizations.  One situation where vectorization may be necessary is when using an aerial photograph to create a property map. In this case, the process could look like the following:  Step 1. Scan the aerial photograph  Step 2. Upload the scanned photo to GIS software  Step 3. Vectorize the photograph (converting it from pixels to polygons)  Step 4. Create a map layer with the vector data |
| **Volunteered Geographic Information (VGI)**  VGI refers to a new application through which, every single person with a smartphone, can share spatial data and upgrade the information we have for any place. This happens through social media sites. For example, when the user takes a picture, there is an automated question which asks the user, “where is that photo taken?” The user then can answer this question and improve the geodata. In this way, there is a bidirectional help because we can share, as well as collect data. |
| Copyright, © Qiming Zhou GEOG3600. Geographical Information Systems  Establishing and Maintaining Geographical Databases. - ppt download  Source: Qiming Zhou, Geographical Information Systems Establishing and Maintaining Geographical Databases. |
| **Choice of data acquisition methods**    Source: Guy Picton Phillipps, Brent PCT, https://www.csp.org.uk/system/files/documents/2019-05/gis-basics-surfaces.pdf |
| **Capturing attribute data**  Attribute can be obtained from various sources including:   1. **Remote sensing image interpretation and processing**     Source: Remote Sensing Image Classification: A Comprehensive Review and Applications by Maryam Mehmood et al (March,20220), <https://doi.org/10.1155/2022/5880959> |
| **2. Field investigations & Site Visits:** field investigations are sometimes conducted using field questionnaires or other documents. In this case, thematic/attribute data can be collected by sampling method.    Source: <https://www.shutterstock.com/image-photo/businessman-lost-field-using-map-23639869> |
| 3. Attributes can be entered by **direct data loggers**, **manual keyboard entry, optical character recognition (OCR)** or, increasingly, **voice recognition.** An essential requirement for separate data entry is a common identifier (also called a key) that can be used to relate object geometry and attributes together following data capture. |

**Exercise materials and tasks**

**Quiz questions**

You finished this session! Well done!

Now, please answer the following questions to test your knowledge:

1. Please complete the blank space with the correct term:

\_\_\_\_\_\_\_\_\_ (**Scanning**/Printing) is one of the simplest ways to convert physical documents into digital formats. When you scan a document, the scanner uploads a digital image of the page. Like all images, scanned documents will appear in \_\_\_\_\_\_\_ (vector/**raster**) format.

In what ways can we capture attribute data?

1. **Field investigation**
2. **direct data loggers**
3. **Site visits**
4. **Remote sensing image processing**

Photogrammetry is the process of taking reliable measurements from photographs.

1. **True**
2. False