Lecture Notes on

GIS DATA SOURCES, DATA INPUT METHODS, and DATA QUALITY

Brighton Chamunorwa

Contents

Unit 1 GIS DATA SOURCES 1
1.1 Definitions: 1
1.2 Classification of data sources 1
1.3 Suitability/appropriateness of data sources 2
1.3.1 Cost 3
1.3.2 Time 3
1.3.3 Accuracy 3
1.4 Geoportals 3
1.4.1 Types of Geoportals 3
1.4.2 Obtaining data from geoportals 4
1.4.3 Geoportals in South Africa 5
1.4.4 Metadata 6
1.5 Master input data list (MIDL) 6
1.6 REMOTE SENSING AS A SOURCE OF GEOGRAPHICAL DATA 7
1.6.1 Benefits of Remote Sensing as a source of geographical data 7
1.6.2 Image interpretation and analysis 8
Unit 2 DATA INPUT METHODS 11
2.1 General remarks 11
2.2 Digitising 12
2.2.1 Tablet Digitising 13
2.2.2 On screen digitising 15
2.2.3 Data Elements Created 16
2.2.4 Digitising curved lines 17
2.2.5 Advantages of manual digitising 18

2.2.6 Disadvantages of manual digitising 18
2.3 Scanning 19
2.3.1 Flat-bed scanner 19
2.3.2 Drum Scanner 20
2.3.3 Requirements for scanning 21
2.3.4 Scanning – limitations 21
2.4 Input of existing digital data 22
2.4.1 Problems associated existing digital data. 22
2.5 Keyboard entry 23
2.6 Attribute data entry 23
Unit 3 DATA QUALITY 24
3.1 Basic definitions 24
3.2 Data source errors 24
3.2.1 Field data 24
3.2.2 Remote sensing 25
3.2.3 Maps 25
3.3 Data input errors 26
3.3.1 Scanning maps or images 26
3.3.2 Georeferencing 27
3.3.3 Digitising 29
3.3.4 Attribute Errors 31

Unit 1 GIS DATA SOURCES

1.1 Definitions:

GIS DATA SOURCES refers to the origin of data required to meet the needs of a specific GIS application.

Data is commonly defined as raw or unorganized form (such as alphabets, numbers, or symbols) that refers to, or represents, conditions, ideas, or objects. In GIS data may be textual or numeric values that refer to, or represent type, condition of real world geographical phenomenon. For example, the number of floors in an academic building, size of a parking lot and dominant rock type of a geological formation.

1.2 Classification of data sources

Primary data sources are datasets obtained directly from the real world for a specific GIS application. The most important part is that data on real world features is obtained by direct measurements on the feature itself. That is when the number of floors in an academic building or dominant rock type of a geological formation is known by measuring or observing directly out in the real world.

Examples of primary data sources:

The following are various ways in which data on geographical phenomenon can be obtained by measuring or observing directly in the real world.

- Remote-sensing.
    - Digital satellite images.
    - Digital aerial photographs.
- GPS measurements.
- Survey measurements.

In addition to measurements in the real, it is possible to know about the floors in an academic building by using university records, reading an article about the academic building on internet. **Secondary data sources** are digital and analogue datasets that were originally captured for another purpose and need to be converted into a suitable digital format for a specific GIS application.

Examples of secondary data sources

- **Maps.**
  - Hardcopy, scanned maps.
- **Tabular data** may be standard lists, such as census reports or business marketing information, typically in printed form. The data might be typed into the GIS or copied by a scanner, an instrument that transfers copied information directly into the computer.
- **Textual data.** Text discussions may not be easily reduced to GIS format, but the information can be important, thereby requiring translation by a user. Sometimes, copies of the text may be scanned and stored as an associated part of the database
- **Digital products** which are processed data sets - sometimes complete GIS databases and coverages compiled by another organisation. Digital products are becoming a major medium (vehicle) for GIS data acquisition because of their efficient storage, ease of transfer into computers, and convenience of update.

1.3 Suitability/appropriateness of data sources

Since data can be obtained from primary sources or secondary sources. One may be confronted with a situation to choose one of the two data sources. There are three primary dimensions that can be used to evaluate the suitability of a data source for a particular GIS project. Scenarios to consider for each dimension are discussed below. It is important to note that here these dimension are considered separately. When giving explanations you are not expected to express how accuracy of the data may have influence on its cost for instance.

1.3.1 Cost

The cost of obtaining data should be confined to the budget limits. Thus in order to choose a data source it is crucial to consider the costs associated by primary or secondary data sources. For example determine whether it is cheaper to conduct a field survey or obtain secondary data.

1.3.2 Time

GIS projects are conducted within a specified duration or time frames. The choice of data source should allow project completion within these pre-specified times. Hence when selecting data source it is necessary to consider time required to gather the data. For example consider the time required to conduct a field survey or procurement and delivery of secondary data.

1.3.3 Accuracy

Accuracy is very paramount if the data is to be used to create useful information products. In order to choose between primary and secondary data sources one should evaluate the accuracy of already available secondary data. Field surveys and other primary data sources may be necessary if the available data does not meet the current project accuracy requirements.

1.4 Geoportals

A geoportal is a type of web portal used to find and access geographic information via the Internet. Geoportals provide a single point of access for searching and downloading GIS data from a multitude of sources. They provide capabilities to query metadata records for relevant data and services, and then link directly to the on-line content services themselves.

1.4.1 Types of Geoportals

There are three basic types of spatial portal:

- Catalogue portals.

- Application portals.
- Enterprise portals.

1.4.1.1 Catalogue portals

They create and maintain indexes or 'catalogues' of metadata that describe the nature and location of resources. Resource owners (or 'service providers') register their services at the portal and supply metadata descriptions. The portal arranges metadata records from service providers into a consistent, searchable catalogue and makes this available to users. Through the catalogue users can search for services coming from any of the registered service providers. In most cases providers continue to host their own service and the portal simply connects users to the service(s) in which they are interested.

1.4.1.2 Application portals

In addition to the catalogue portal's generic search tools, application portals provide more structured interfaces that include specific tools and applications relevant to user's domain interests. Application portals can be tailored to meet specific needs, and the interface designed to provide efficient access to those data and functional services needed. Often application portals store some, if not all, of the data and functional services at the portal site. Application portals provide Web mapping tools to allow users to view and work with the data they find (for example, geo-processing tools such as: route finding, geo-coding, printing, complex query and perhaps even redlining and edit/update functions).

1.4.1.3 Enterprise portals

A third type of spatial portal is emerging as the 'enterprise' spatial portal that integrates spatial data and functionality with business enterprise solutions.

1.4.2 Obtaining data from geoportals

1.4.2.1 Cost or requirements are related to:

- Availability

- Open/free data is downloaded from geoportals free of charge. Whereas some data requires payment in order to have access for downloading.

- Distribution costs
  - These are rise from cost of data bundles for instance and other networking infrastructure that have to be put in place to facilitate transfer data from the supplier. GIS data can be very big; one dataset can be several gigabits.
  - Some organisations do not transfer data through the internet. In such cases costs to obtain data are incurred through postages and freight.
- Registration
  - Many sites require a user to sign up before having access to download data.

1.4.2.2 The methods of obtaining data include:

- Immediate download
- Ordered and sent via disk or tape
- Received from business partners of data catalogue

1.4.3 Geoportals in South Africa

Table 1:1 Examples of Geoportals in South Africa

CSIR Geospatial portal	http://gsdi.geoportal.csir.co.za/
egis	http://egis.environment.gov.za/Discover.aspx?m=23&amid=141
Biodiversity GIS	http://bgis.sanbi.org/
SANParks Data Repository	http://dataknp.sanparks.org/sanparks/
boundary data	http://www.demarcation.org.za/site/?page_id=1836

Other geoportals

- http://www.naturalearthdata.com/
- http://worldclim.org/version2
- http://www.worldpop.org.uk/
- http://download.geofabrik.de/
- https://www.education.gov.za/Programmes/EMIS/EMISDownloads.aspx

1.4.4 Metadata

It is important when searching for secondary data. Metadata can be used to evaluate the usefulness of geospatial data obtained from internet. It refers to Information about data. Metadata files include the following:

- general descriptions about the contents of the file,
- Spatial reference information
- definitions for the various terms used to identify records (rows) and fields (fields)
- the range of values for fields
- the quality or reliability of the data and measurements,
- how the data were collected
- when the data were collected
- who collected the data

1.5 Master input data list (MIDL)

The master input data list (MIDL) is a detailed list of all the data source datasets that need to be entered into GIS system to generate all the information products. A master input data list includes the following four components:

- Data identification details.
- Data volume considerations.
- Data characteristics.
- Data availability and costs.

1.6 REMOTE SENSING AS A SOURCE OF GEOGRAPHICAL DATA

1.6.1 Benefits of Remote Sensing as a source of geographical data

- In comparison to other methods such as field survey remote sensing ensures rapid collection of up to date data over large geographical area. For example a LANDSAT 5 TM scene measures approximately 170km north-south by 183 east-west. Thus data for such a large area can be obtained once through the use of satellite images.
- Also, remote sensing is the best data source to obtain data on regional phenomenon such as geological structure and forests. Remote sensing images are acquired by devices located above the earth surface hence provide an improved synoptic view of the earth surface features.
- Moreover, data is acquired by remotely operated sensors. Remote sensing is the only practical way of obtaining data from inaccessible regions such as the Antarctica and Amazon.
- Most of remote sensing data is acquired through space based satellite platforms. Each satellite as a temporal resolution or revisit period which refers to the number of days it takes before it acquires data at the same location again. Thus data for a certain location is acquired on a regular basis. Many geographical phenomenon exhibits temporal variations. Data acquired over several periods allow comparison of images hence revealing change.
- In addition remote sensing data is the optimum option for global studies as it provides a global data set of uniform quality. Remote sensing provide a uniform data set because images covering the earth surface acquired by the sensor. This will not be possible with field survey for instance.
- Remote sensing images are acquired over broad range of wavelengths allowing detection of certain phenomenon that will not be possible with the human eye. Images are acquired in the near infrared region and such

wavelengths enables detection of unhealthy vegetation for instance which could not be done with the human eye.

- Furthermore, remote sensing is a popular data source for GIS because of systems such as RADAR. It supplies its own electromagnetic energy and operates in the Microwave region of the spectrum. Such systems are advantageous to field surveys for example as they can acquire data in a wide range of conditions. For example RADAR can acquire data at night and can be used to obtain data during severe weather conditions.

1.6.2 Image interpretation and analysis

Figure 1.1 LANDSAT 5 TM scene

Above is a LANDSAT 5 TM scene covering the town of Alice and surrounding villages. The dark spots on the image are water bodies. Therefore, this image can be used to obtain data for water bodies such as area in hectares and/or simply locations of these water bodies. However, in order to take advantage of and obtain data that can be input into a GIS database, we must be able to identify features of interest and measure the required characteristics from the imagery. Data that will be put in a GIS can be extracted by visual interpretation or digital analysis methods from remote sensing images such as the one shown above.

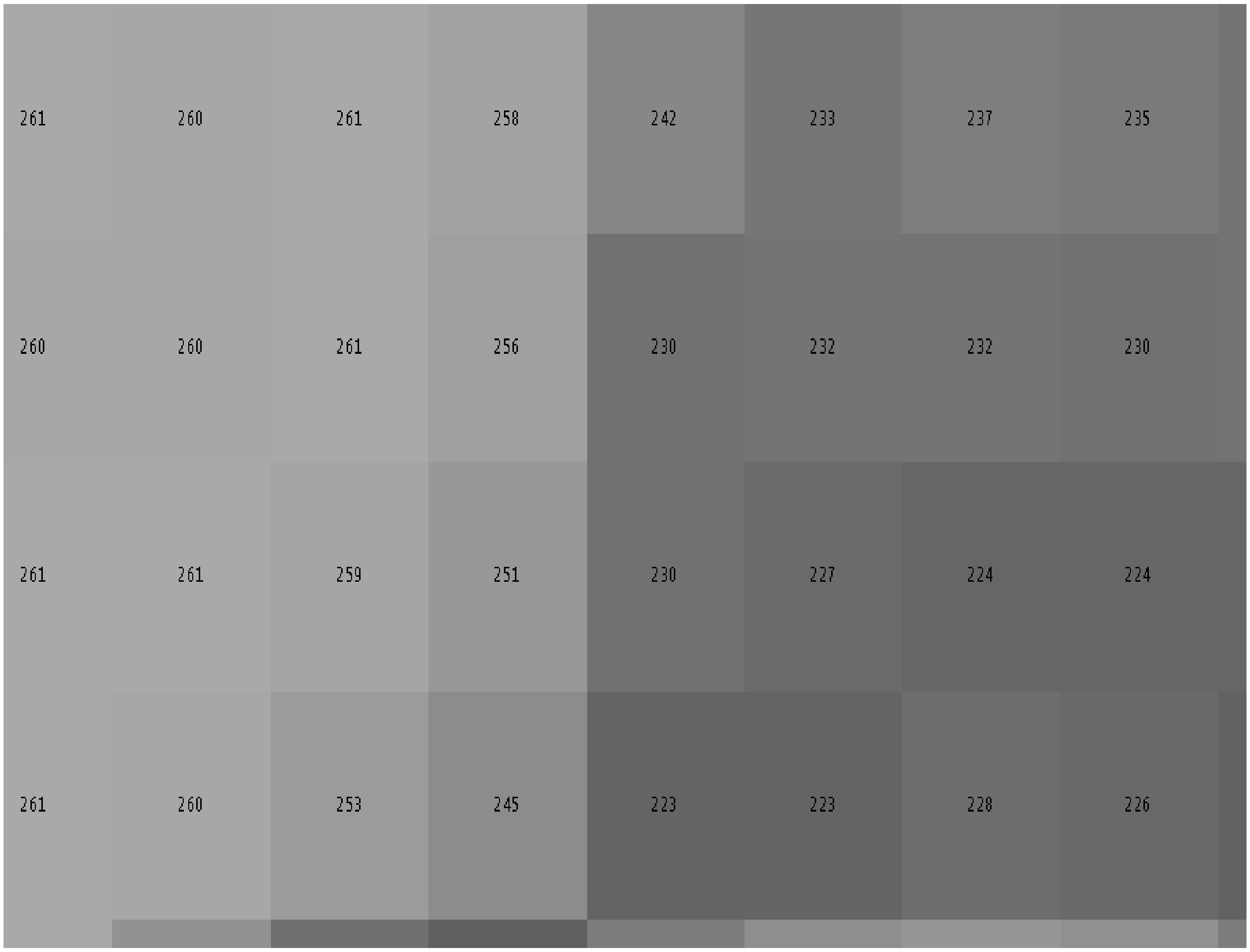


Figure 1.2 LANDSAT 5 TM scene zoomed

The image above is a LANDSAT 5 TM scene that has been zoomed to show cell values. Note that scenes with different cell values have a different shade of grey. Higher cell values are characterised by brighter shades and scenes with low values are the darkest.

1.6.2.1 Visual interpretation

Visual interpretation is performed manually or visually. It involves observing the differences between targets and their backgrounds. Data for various targets is acquired by comparing different targets based on any, or all, of the elements of visual interpretation. These include tone, shape, size, pattern, texture, shadow, and association.

1.6.2.2 Digital processing and analysis

It requires image processing software. Digital processing may be used to enhance data as a prelude to visual interpretation. Digital image processing may involve numerous procedures including the following:

- formatting and correcting of the data,
- digital enhancement to facilitate better visual interpretation
- automated classification of targets and features

Unit 2 DATA INPUT METHODS

2.1 General remarks

Data input methods refers to procedures that create digital spatial data. In other words, it refers to various mechanisms on which digital representation of spatial data such as roads, academic buildings and parking lots on campus are created. The digital representation of this spatial data is then stored in a GIS database. In the previous unit we were exposed to various data sources. In this unit we are going to explore how several data input methods create a digital representation of spatial data from various GIS data sources.

The diagram below helps to understand how digital representations of real word features are created.

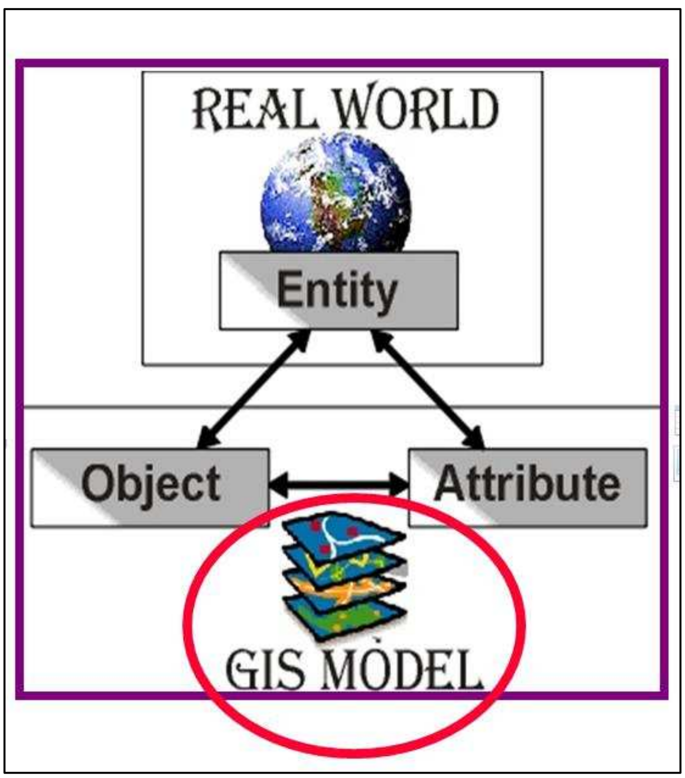


Figure 2.1Data modelling

Importantly, the diagram clearly illustrates that a digital representation is created for a particular earth surface feature that is of interest to a specific GIS project. A GIS project at the University of Fort Hare might create a digital representation of academic buildings. This earth surface feature of interest to a specific GIS project is known as an entity. A GIS database usually consists of several entities. In addition, on the diagram there are arrows pointing from entity label to object and attribute. The digital representation of the selected entity comprises of an object and attribute. Object refers to the graphical representation of an entity in a GIS database. Basic graphic representations of earth surface features are point, line or polygon. Attribute also known as non-spatial data is the characteristics of an entity selected for representation. Academic buildings at Fort Hare University can have attributes such as number of floors, name, number of lecture halls etc. Several entities are usually required to fulfil the needs of a GIS project. The digital representation of these entities as indicated on the diagram form a GIS model.

2.2 Digitising

Digitising is the process where features on a map or image are converted into digital format for use by a GIS. There are three primary methods for digitizing spatial information:

- Manual Methods include:
    - Tablet Digitising
    - On screen digitising /Heads-up Digitising
- An Automated Method includes:
    - Scanning

To ensure that maps are digitised most efficiently and accurately, 4 steps need to be followed:

- Use good base maps

    - The quality of the map manuscripts from which you digitise directly affects the accuracy of your digital data. You should always get the most reliable, most current maps possible.
- Define your procedures
    - Determine how the maps will be digitised.
    - Establish a standard sequence of procedures, for example, you may want to digitise all the line coverages before the point coverages, etc.
    - Establish a sequence for digitising features and map sheets so you can track which portions of the database have already been digitised.
    - Establish standard naming conventions.
    - Establish schedules and shifts - digitising is tiring work, so scheduling digitising helps ensure accuracy, e.g. a schedule of 2 hours on, 2 hours off, might be appropriate.
- Prepare your maps
    - Map preparation helps minimize problems at the digitising and editing phases.
    - Overall, the goal is to minimize the number of times the person who digitises and edits will have to stop work.
- Digitise your maps

2.2.1 Tablet Digitising

Tablet digitising is done by placing a paper map on a digitising tablet and entering all the elements of the map into the data base by means of a sensitive digitizing puck. A digitizing tablet is a hardened surface with a fine electrical wire grid under the surface. The diagram below shows a digitising tablet with a map prepared for digitising.

Figure 2.2 Digitising tablet

A digitizing puck is an electrical device with cross hairs and multiple buttons to perform data entry operations.

Figure 2.3 Digitising puck

An operator enters data by placing the digitising puck over the points on the map attached to the digitising board and pressing different buttons on the puck, thus indicating the type of each point. A point can be either an individual element, or a part of a larger element such as a line or a polygon.

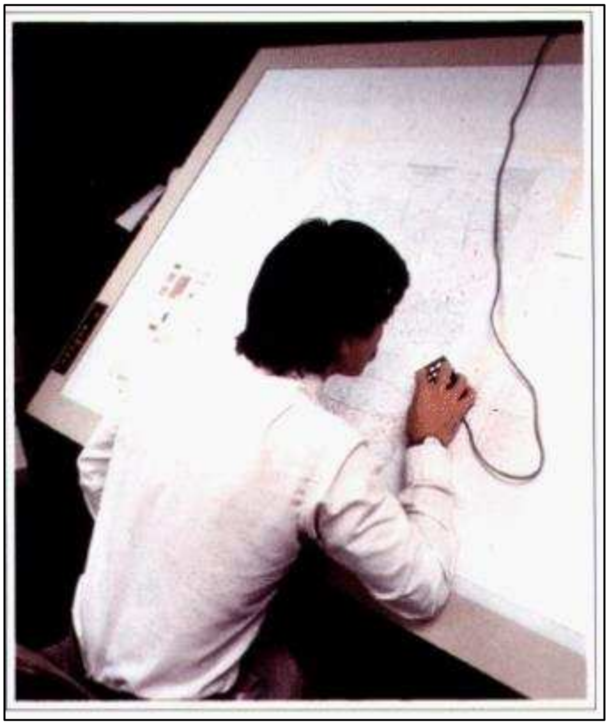


Figure 2.4 Digitising in action

Lines and polygons are defined by a set of points (vertices) entered by an operator and chains/segments. Therefore, the accuracy of the data depends of the accuracy of the location of the points. In the case with lines and polygons, the more points entered, the smoother the curves of the line will appear.

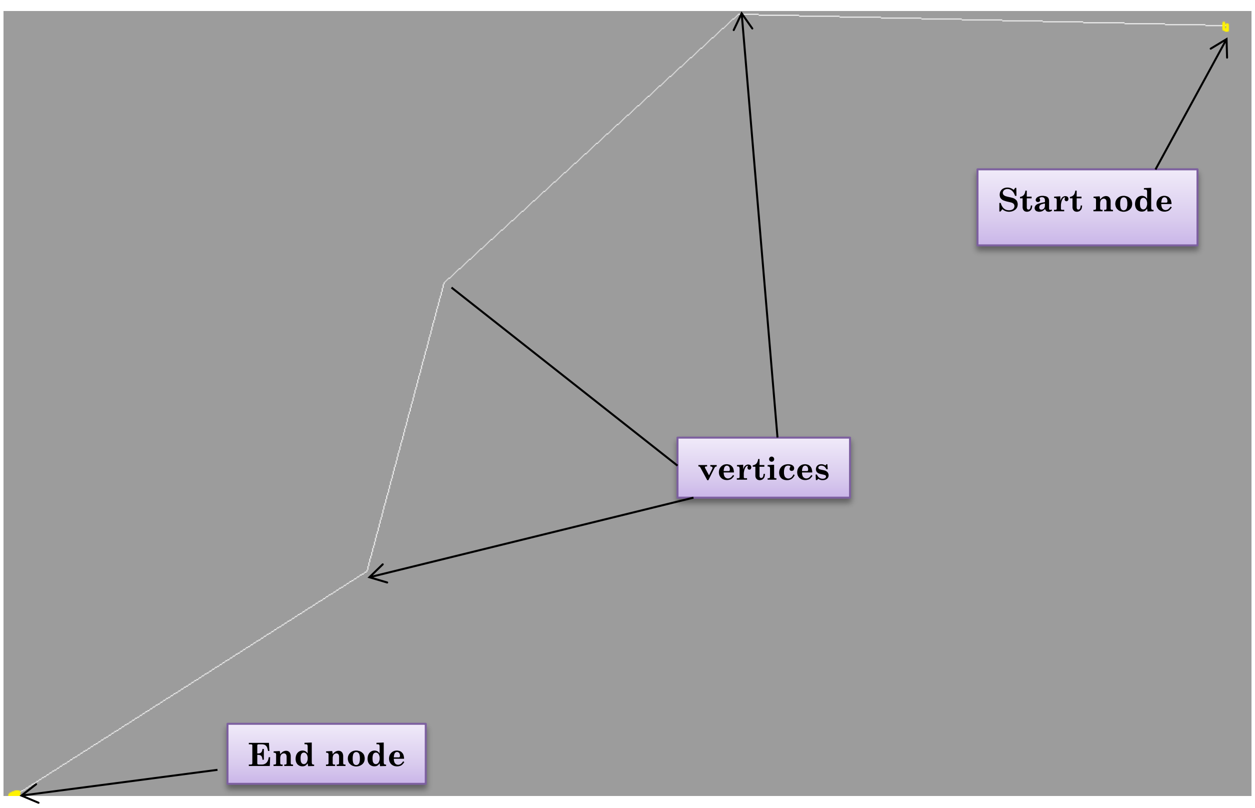
2.2.2 On screen digitising

On screen digitising is often called heads-up digitising. It uses the computer mouse and digitising software to manually trace the features on a source documents displayed on the computer screen. The source documents may be scanned images or digital aerial and satellite images. Most GIS software systems support on-screen digitising. On-screen digitising generally follows the steps below:

- Scan the source document into a digital image if paper map or aerial photo used.
- Add the image into a GIS
- Georeference the image
- Trace features on the image using the mouse
- Edit features
- Add attribute data

2.2.3 Data Elements Created

- Digitising point features is simply the process of creating a single point feature with an x,y coordinate. In most cases to create a point you first left click and then right click to make the point permanent.
- Digitising lines involves creating a line feature that consists of a node (start and end) and a series of vertices which indicate a change of direction along that line. Straight features require fewer vertices; curved/complex features require more vertices.



- Digitising Polygons involves creating a set of connected lines. Polygons have a single start-end node and vertices in between. Similar to lines polygons have vertices at every location where its boundaries change direction.

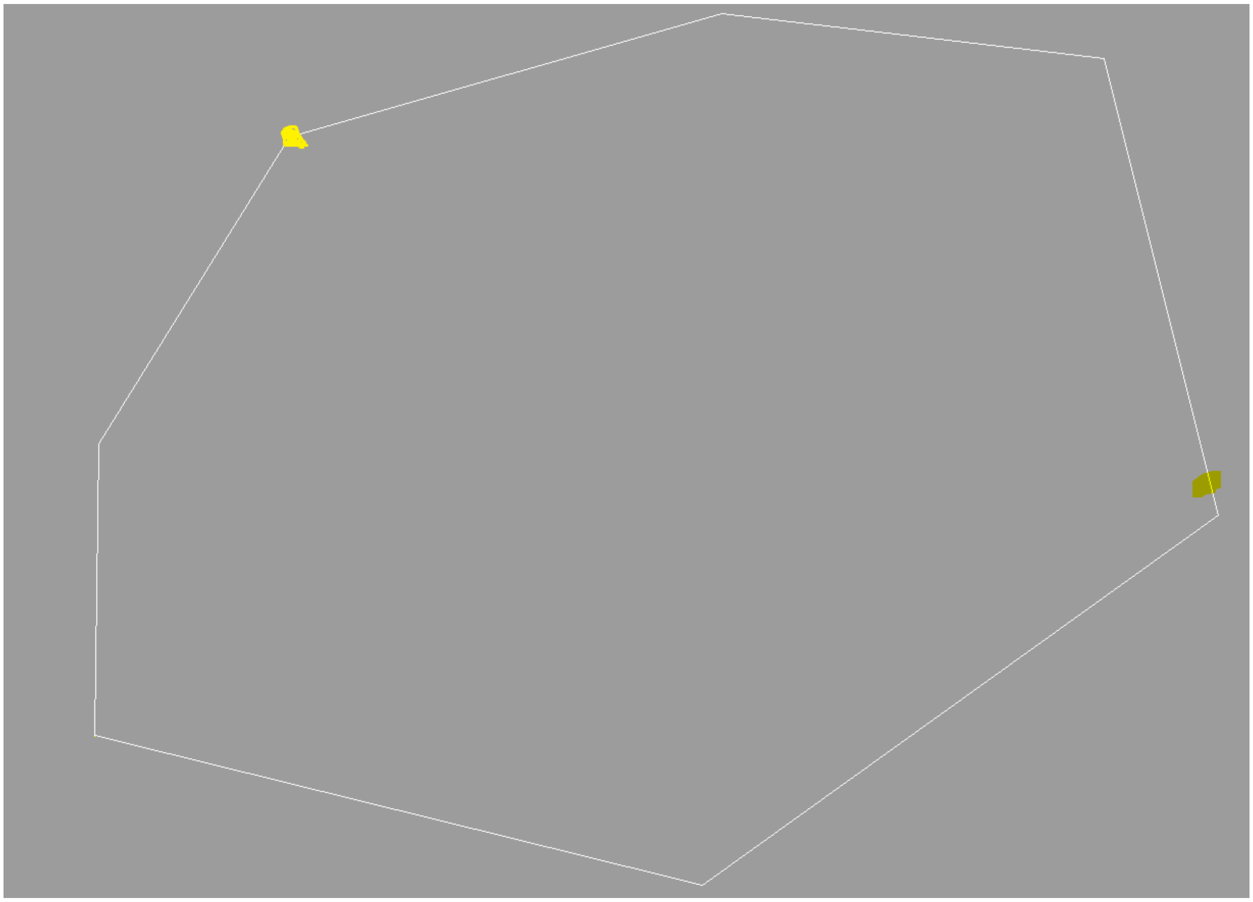
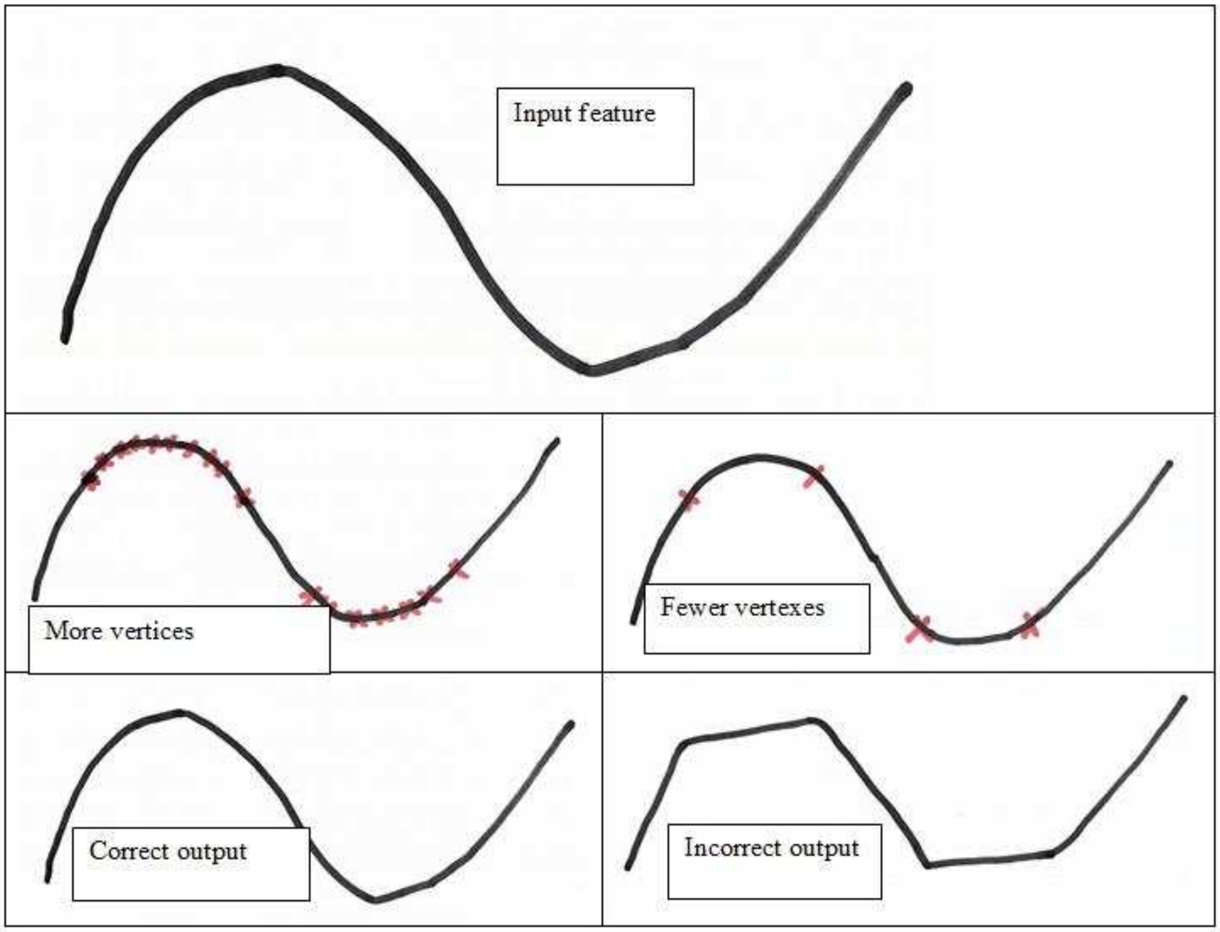


Figure 2:5 Polygons

2.2.4 Digitising curved lines

The diagram below illustrates how curved lines are digitised to preserve their properties. On a curved line change of direction is very abrupt and this can be preserved by inserting many vertices close together to capture the changes in direction. On the diagram it is shown that if a curved line is digitised with a few vertices it will not represent the earth feature accurately as it will not show the curvature of a river for an example. In addition zooming in may be required to digitise a curved section accurately.



2.2.5 Advantages of manual digitising

- The ability to correct errors or distortions in the original maps at the time of data capture.
- Highly reliable human recognition of map objects.
- The ability to interpret ambiguous or incomplete information and select the relevant required information at the time of data capture.
- Can be performed on inexpensive equipment;
- Requires little training

2.2.6 Disadvantages of manual digitising

- Manual digitising labour intensive and therefore very time-consuming and costly.
- The quality of results is highly dependent on the operator experience.

- The results may be inconsistent due to varying operator conditions, stress, and fatigue.

2.3 Scanning

Scanning is also known as automated digitising. Hardy copy documents are converted into digital images with a scanner. Scanning produces a digital image of the map by recording amount of light reflected by the surface of the map. Common types of scanners are flatbed scanners and rotating drum scanners. If good source documents are available, scanning can be an efficient time saving mode of data input.

2.3.1 Flat-bed scanner

A flatbed scanner is usually composed of:

- a glass pane,
- bright light which illuminates the pane,
- and a moving of photodetectors

On a flat-bed scanner the map is placed on a flat scanning stage. They are mainly used for small format maps and aerial photographs.

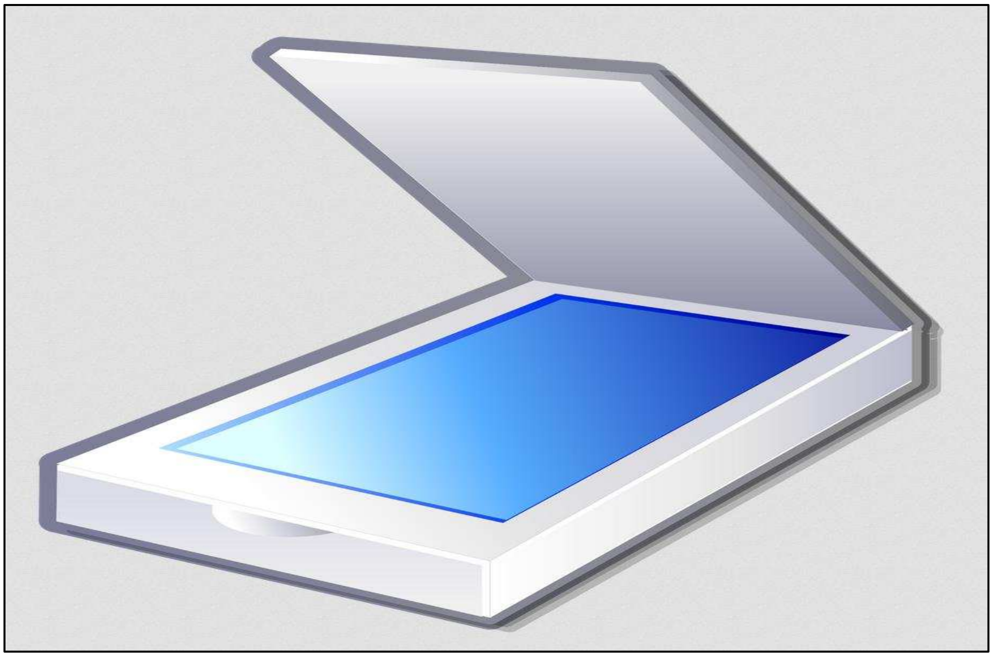


Figure 2.6 flat-bed scanner

A flatbed scanning involves an array of photodetectors which extract data from several rows of the raster simultaneously. The detectors move across the map in a swath. When all the columns have been scanned, the detector moves to a new swath of rows.

2.3.2 Drum Scanner

The map is mounted on a cylindrical drum which rotates while the detector moves horizontally across the map. As the drum rotates about its axis, a scanner head containing a light source and photodetector reads the reflectivity of the target graphic band digitizing this signal. It creates a single column of pixels from the graphic at time. The scanner head moves along the axis of the drum to create the next column of pixels, and so on through the entire scan.

Figure 2.7 Drum Scanner

2.3.3 Requirements for scanning

- Documents must be clean (no smudges or extra markings)
- Lines should be at least 0.1 mm wide
- Complex line work provides greater chance of error in scanning
- Text may be accidently scanned as line features
- Contour lines cannot be broken with text
- Automatic feature recognition is not easy (e.g. contour lines vs. road symbols) diagram
- Special symbols (e.g. marsh symbols) must be recognized and dealt with

2.3.4 Scanning – limitations

- Requires expensive equipment,
- Involves expert personnel,
- Usually entails considerable editing

- Needs clean maps with well-defined lines
- Produces large quantities of data

2.4 Input of existing digital data

It is becoming increasingly popular. In some cases, geoportals may have a variety of spatial data that can fill some of the needs of a project. If the data was created using different software, a translation must occur to exchange the data and put it into a suitable format. Several ad hoc standards for data exchange have been established in the market place. GIS vendors have developed and provide data exchange/conversion software to go from their format to those considered common in the market place. Data transfer relies on the exchange of data in mostly proprietary file formats, using the import/export functions. Integrating these heterogeneous data sources requires considerable knowledge of GIS data-integration methods to produce a complete and seamless digital database.

2.4.1 Problems associated existing digital data.

- Software developers are reluctant to publish the exact file formats their systems can handle. Therefore, import routines are sometimes unstable and frequently lose some of the information contained in the original data files.
- Another set of problems are encountered due to lack of metadata or incomplete metadata. As a result, it is difficult to assess the quality of the digital information for instance. Also missing information about the geographic reference framework might make it impossible to convert data from the external data set's coordinate system to the one used by the organization.
- Challenges to using existing data to construct GIS data base may occur due to differences in the data itself. These differences can arise due to variations in definitions and coding schemes. The use of different cartographic reference systems. Incompatible spatial scales and varying

accuracy standards, which may result in features that should match across two databases being displaced.

2.5 Keyboard entry

Keyboard entry is mainly used to input coordinates of a number of point features into a GIS. It involves reading or obtaining coordinates of point features from maps or other source documents. The coordinates are typed into a file in a tabular format and reading the file into a GIS.

2.6 Attribute data entry

All geographic objects have attributes of one type or another. Attribute data is stored in tables. The following steps are followed to ensure that attribute data is entered into tables.

- It starts with table creation. This involves setting the name and how the table will relate to other tables.
- Each table has columns or fields in which values of the chosen characteristics for representation are entered. A table field is created by specifying the name, data format and length. Field data format can be integer, text, float, etc. one field should be specified as a primary key.
- After creating the fields values are entered by selecting a table cell and typing in through the keyboard.
- In a GIS database attribute (non-spatial) data should be linked to map elements. Most software prompts you to enter attribute data immediately after finishing digitising a point, line or polygon so that they are linked.
- If the database comprises of tables that are related, it will be necessary to define how the tables are related. This is done by linking similar fields in both tables.

Unit 3 DATA QUALITY

3.1 Basic definitions

Accuracy - The degree to which information on a map or in a digital database matches true or accepted values. Consider horizontal and vertical accuracy with respect to geographic position, as well as attribute, conceptual, and logical accuracy.

Precision - The level of measurement and exactness of description in a GIS database. High precision does not indicate high accuracy nor does high accuracy imply high precision.

Data quality refers to the relative accuracy and precision of a particular GIS dataset.

Error encompasses both the imprecision of data and its inaccuracies.

3.2 Data source errors

3.2.1 Field data

The skills and motivation of the surveyor influences the errors in spatial data collected through a field survey. The skills are an important consideration for factors such as use of measuring instruments and on identification of phenomenon of interest. Error can be introduced for example when a geologist fails to identify certain rock types in the field. Also even if the data collect is well knowledgeable of the geographical phenomenon under study errors can occur because they are less motivated or just not industrious. As a result, data may not be collected for attributes that are easy to measure or the number of places visited may not be adequate.

In addition, errors in data collected from field survey can be caused by mis-calibrated data collection instruments. Data collection instruments should be calibrated to ensure a standard scale of recordings. As an example, when using more than one GPS to collected coordinates of control points that will be used to georeference an aerial photograph. Errors can occur if some GPS are set to record coordinates of a location in decimal degrees and others in degree minutes and seconds format. Also, despite the settings of the GPS location data collected will be inaccurate if the GPS is not functioning properly.

3.2.2 Remote sensing

Raw digital images usually contain geometric distortions so significant that they cannot be used as maps. The source of these distortions range from variations in the altitude, attitude, and velocity of the sensor platform, to factors such as panoramic distortion, earth curvature, atmospheric refraction, relief displacement, and nonlinearities in the sweep of a sensor's Instant Field of View [IFOV].

Atmospheric constituents affect the electromagnetic radiation that is recorded by sensors due to scattering and absorption. These processes add or diminish the true ground-leaving radiance as result the values on the satellite image may not be true representation of the ground features.

Each sensor has detectors that are responsible for recording electromagnetic radiation reflected by earth surface features. Remote sensing data can have error due to malfunction of these detectors.

The errors in remote sensing data are corrected various image pre-processing procedures. Geo-rectification procedures correct geometric errors and atmospheric correction and radiometric correction procedures are meant to correct for defects due atmospheric effects and detector malfunctioning respectively.

3.2.3 Maps

Maps are common secondary data sources. On paper maps errors can be caused by markings and scribblings done during its previous use. Folding and shrinking of paper maps also lead errors. Scanned paper maps may also have geometric distortions. Also, errors may be present in the map data. These could come from the source data or were introduced during map creation. During map creation data errors may result from inaccurate interpretation of the source and bias of the map maker.

3.3 Data input errors

This part examines the errors that occur during digitising. Refer to input of existing data for errors that result from conversion of existing data.

3.3.1 Scanning maps or images

Scanned maps and images can have errors that may arise due to properties of the scanner and skills of the operating personnel. Size of a scanner may have negative effects. A map may be scanned more than once were it is folded to match the size of a scanner. It will be difficult to georeference and join the map parts consequently disrupt the actually spread of features on the ground. Depending on the resolution of the scanner some lines may be cut and possibility of colour changes. Considerable knowledge of scanning and high levels of concentration is required during scanning. A map should be placed on the scanner in a way that maintains its orientation. Slanting maps will result in geometric distortions. As mentioned earlier, complex maps are difficult to scanner as it will be difficult to identify features on a scanned map.

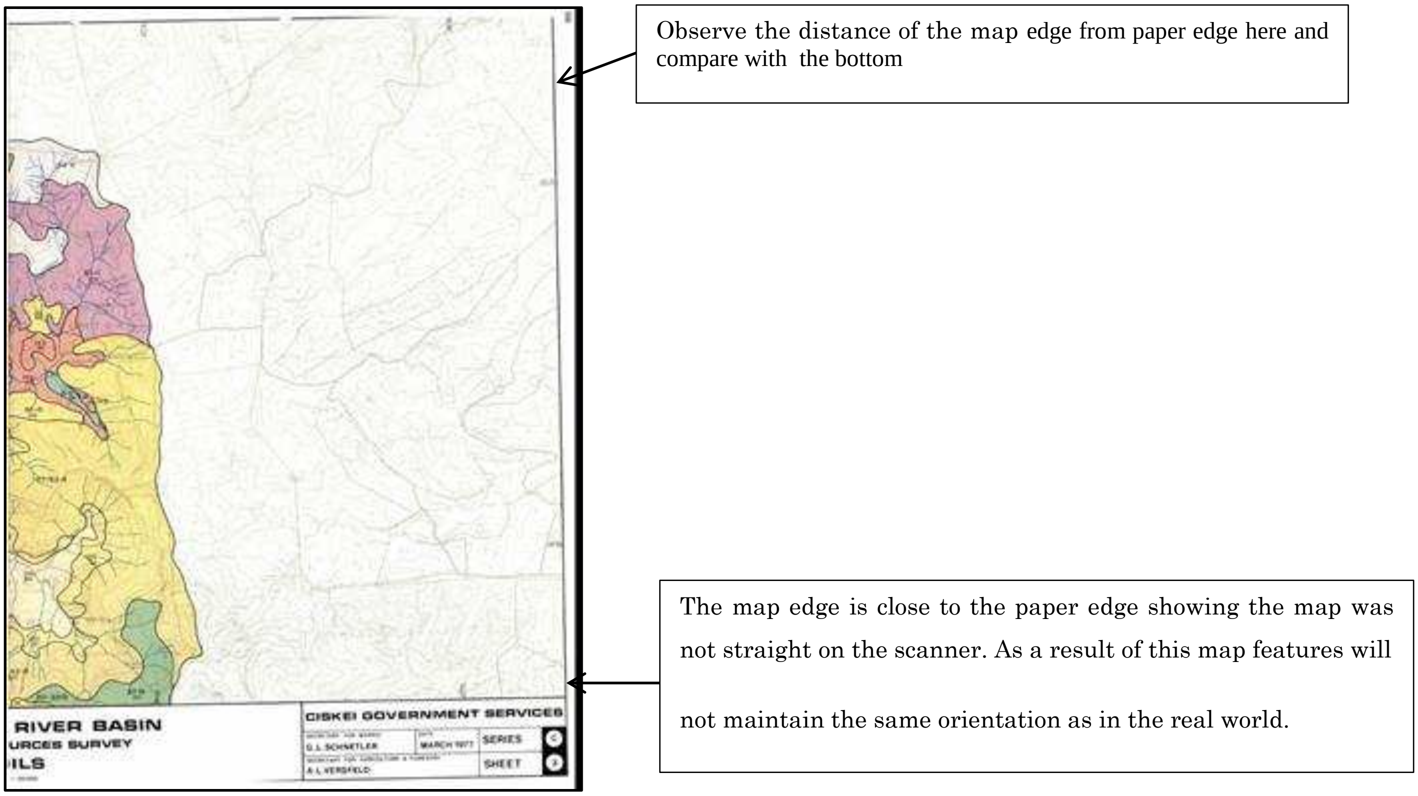


Figure 3.1Scanning errors

3.3.2 Georeferencing

In most cases inaccuracies during georeferencing are a result of human errors. In order to accurately georeference an image or map the operator should correctly identify control points. If the control points are not match all the features on the map will be wrongly placed. In addition, these control points should be spread across the area. Especially, if the coordinates of control points are obtained through a field survey, the surveyor should be motivated to visit many sites that are a distance apart. If the control points are only located on the central parts of the map the output will shrink.

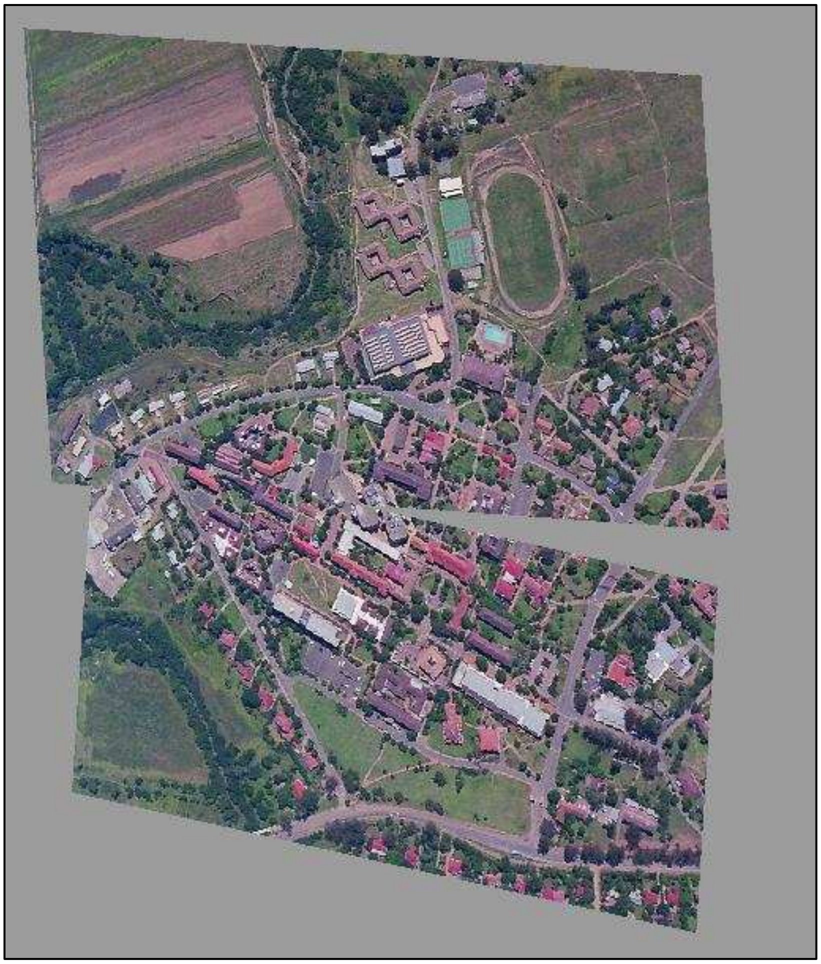


Figure 3.2 Georeferencing errors

The diagram above illustrates how positional errors occur to wrongly placed coordinates of control points. In the real the space covered by the two images is conjoined thus an error occurs since the digital images are not joining. Moreover, it is important to understand coordinate system parameters and how they are set on particular software. Wrong coordinate system will result in positional errors. Some software has limitations that may lead to errors as they only take decimal coordinates. The operator should take note of these and make necessary conversion before inserting coordinates of control points.

3.3.3 Digitising

3.3.3.1 Pseudo Nodes

These are unintentional nodes that occur at along a line or a polygon. They can be due to misplaced point or to push wrong button during digitising. GIS software has capability to detect pseudo nodes. To correct pseudo notes, first determine whether they are indeed errors. This can be achieved by comparing with the data sources. Incorrect pseudo nodes can be selected manually and deleting them. GIS software can also do it automatically.

3.3.3.2 Dangling Nodes

The dangling nodes are defined as a single node connected to a single line entity. They result from three possible mistakes:

- failure to close a polygon,
- failure to connect the node to the object it was supposed to be connected to (undershoot)
- going beyond the entity you were supposed to connect to (overshoot)

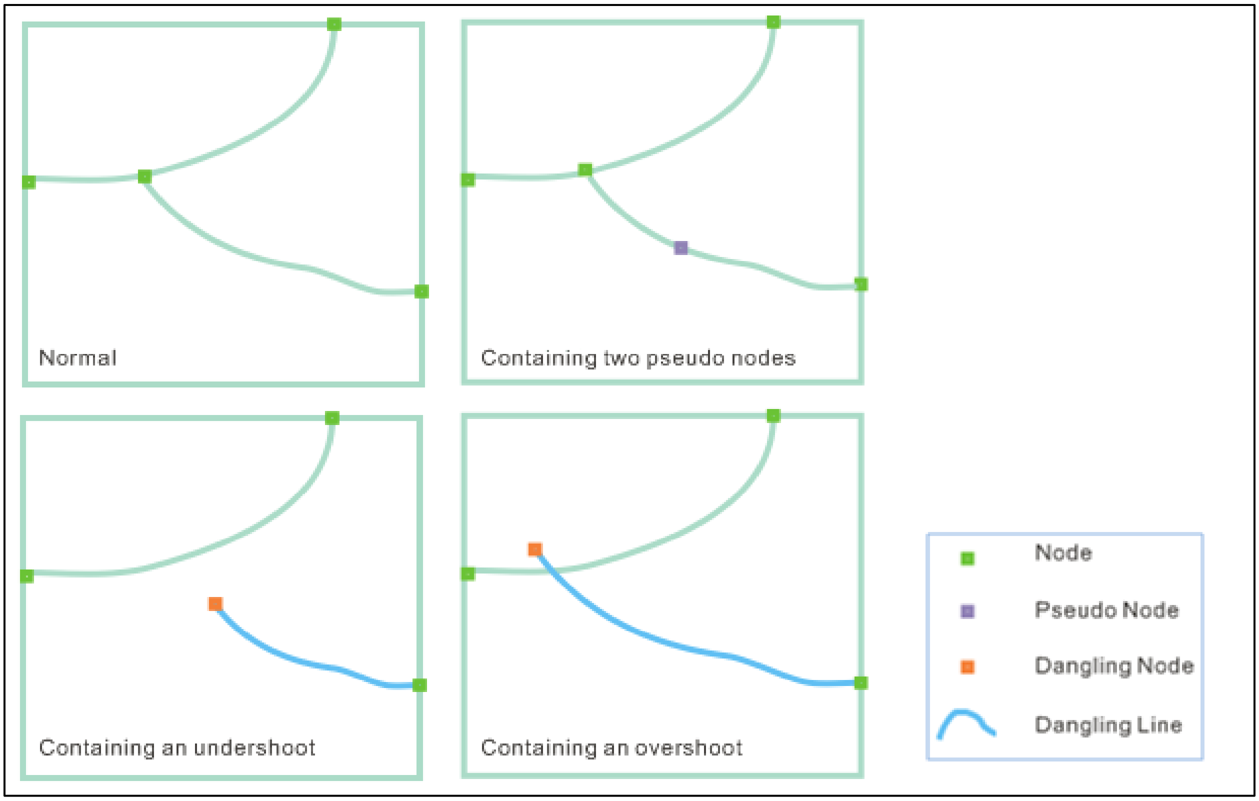


Figure 3.3 Digitising errors

If the dangling node is an open polygon, the software alerts you by providing the number of complete polygons in the dataset. If the number is different from expected the errors can then be corrected manually. In the case of an open polygon, you merely move one of the nodes to connect with the other. For undershoots, the node is identified and is moved or 'snapped' to the object to which it should have been connected. Overshoot errors are corrected by identifying the intended line intersection point and 'clipping' the line so that it connects where it is supposed to.

3.3.3.3 Sliver Polygons

This type of digitising error most commonly occurs between two adjacent polygons. Small polygons are created when the same line is digitised twice. Sliver polygons can also occur as a result of overlay operations. Detection of sliver polygons is very difficult. Most often, the only way to find them, is to move through your image, searching for suspect polygon boundaries, and then zooming in to see the slivers.

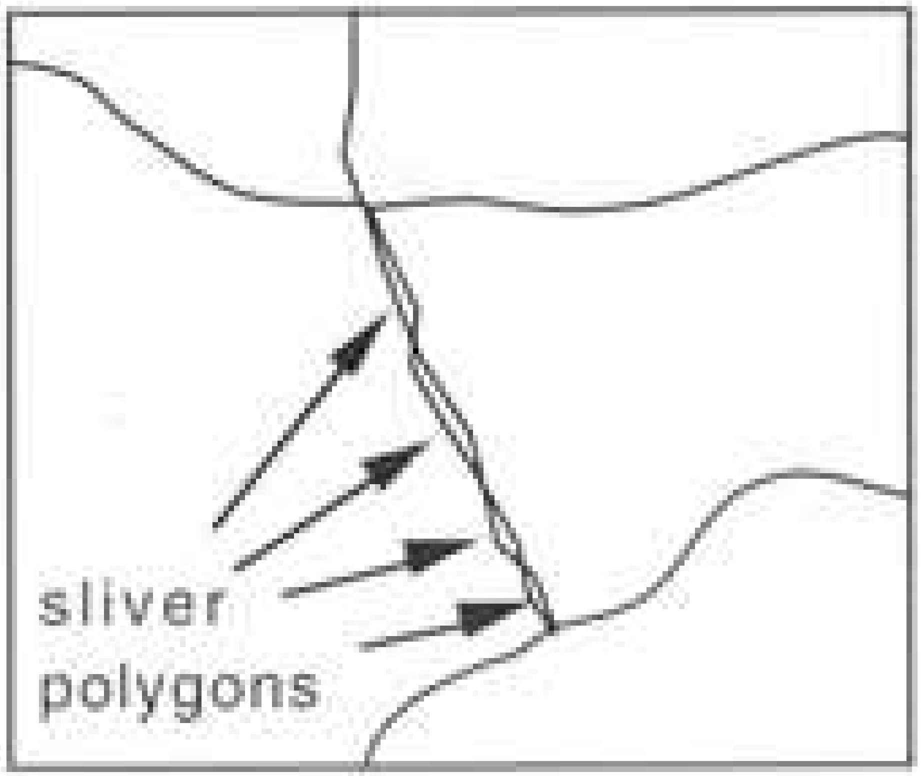


Figure 3.4 Sliver polygons

3.3.4 Attribute Errors

Attribute data errors are more difficult to identify than locational errors as they are not apparent until later on in the data processing analysis. Simple data entry errors such as missing or duplicate data records may become evident when linking spatial and attribute data. Attribute data errors may include:

- Incorrect assignment of features unique identifiers.
- Missing data records or too many records.
- Missing attribute
- Incorrect attribute value

Attribute data error may result from:

- Observation or measurement errors.

- Data entry errors
- Out dated data

Attribute data editing in GIS main includes the following on attributes associated with features and their values:

- Adding
- Deleting
- Updating