

UNIT 8

DATA INPUT

Structure

- | | | | |
|-----|----------------------------------|------|----------------------------|
| 8.1 | Introduction | 8.8 | Georeferencing |
| | Expected Learning Outcomes | | Geocoding |
| 8.2 | Methods of Data Input | | Geotagging |
| | Raster Data Input | | Rubber Sheeting |
| | Vector Data Input | 8.9 | Level of Input |
| 8.3 | Input Devices | 8.10 | Summary |
| 8.4 | Existing Data and its Conversion | 8.11 | Activity |
| 8.5 | Creating New Data | 8.12 | Terminal Questions |
| | In Situ Data Collection | 8.13 | References |
| 8.6 | Map Digitisation | 8.14 | Further/Suggested Readings |
| 8.7 | Reference Frameworks | 8.15 | Answers |

8.1 INTRODUCTION

Data is the basic requirement to perform any analysis in GIS platform because without data no analysis can be performed. In general data input into a computer system is a process of addition (encoding) of data into a computer-readable form and at the same time writing the database to the digital database. Addition of data into a computer system especially in a GIS environment requires various methods and techniques as the data is characterised as spatial and non-spatial data.

You have read the concepts of GIS, data models and database structures in the previous Block (Block 2), of this Course. In this unit, we will discuss the methods of data input covering raster and vector data input, input devices, existing data and its conversion. In addition we will also discuss reference frameworks, georeferencing, map digitisation and level of input.

Expected Learning Outcomes

After studying this unit, you will be able to:

- ❖ explain methods of data input;
- ❖ describe data input devices;
- ❖ elucidate existing data and its conversion;
- ❖ illustrate creating new data and reference frameworks;
- ❖ discuss georeferencing covering geocoding, geotagging and rubber sheeting; and
- ❖ Know the process of map digitisation and level of input.

8.2 METHODS OF DATA INPUT

Data input is a method of selecting, acquiring and converting data into a digital format that can be stored and analysed in computers. It is also known as *data encoding*. There are number of steps involved in the process of data input in a GIS platform. A sequence of data input process is shown in Fig. 8.1. Database creation is ofcourse the fundamental, expensive and the most time-consuming (i.e. about 80%) part of a GIS project. Data input involves a number of techniques to enter data into a GIS which are collected from various sources. There are many sources of data but following are the commonly used data sources:

- hard copy maps
- aerial photographs
- remotely sensed images
- ground survey with GPS
- reports and publications, and
- existing digital data files.

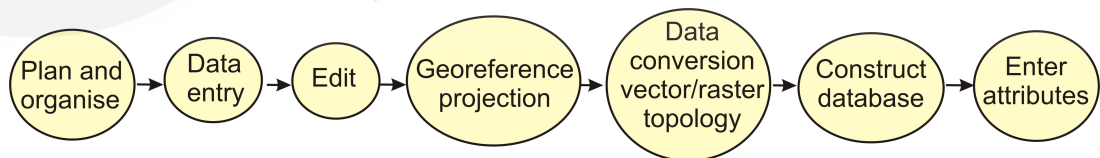


Fig. 8.1: Chain of data input process in GIS. (Source: modified after Davis, 2001)

Data collected both in the analog and digital format need to be encoded to make it compatible with GIS. **Analog data** is normally in paper form like paper maps, statistical tables or hard copy of satellite images. All these data are required to be converted into digital form before using in a GIS project. On the other hand, **digital data** is already in computer readable format e.g., digital map data, digital remote sensing images, data obtained from GPS receivers and other digital databases. Even digital data collected from different sources might not be compatible to a GIS project. Therefore, different GIS software packages are required to convert data from one form to other. It involves a range of methods to convert data into GIS, e.g., maps which come in the digital or

analog format may be entered into GIS by digitising, scanning or by transferring the file directly. On the other hand, aerial photographs may be scanned and put into GIS and satellite images can be downloaded from digital media procured directly from remote sensing. There are different data encoding methods out of which keyboard entry, manual digitisation, automatic digitisation and scanning are most commonly used. Many different types of devices are also used for inputting data into a computer which we will discuss in the subsequent sections.

Geospatial data can be input in the form of vector or raster data and so it is worthwhile to discuss encoding methods for both these data forms. You have already read the basics of raster and vector data in Unit 6 of MGY-101 Course. Now let us discuss methods of raster and vector data input in detail.

8.2.1 Raster Data Input

Aerial photographs and satellite imageries are the examples of raster data in digital format. However, analog maps, aerial photographic film or print out of satellite images can also be converted into digital form by scanners. We will discuss about scanners in detail in the later section of this Unit.

After scanning, the raster image needs to be first corrected for errors caused by scanning. This requires some level of image processing about which we have already read in Unit 17 of MGY-102 Course.

Image distortions are corrected by rubber sheeting, georeferencing, projection, scaling or resizing, despeckling, greyscaling, adjusting brightness or contrast, thresholding, colour separation, subsetting and edge matching. We will discuss georeferencing and other related topics in ensuing sections of this unit.

However, let us get familiar with a few terms related to the processes that improve image quality.

- **Despeckling:** It removes speckles or stray pixels that appear in an image when we scan a dirty or wrinkled image.
- **Greyscaling:** It converts a colour image into greyscale which is obtained by adding the values of red, green and blue (RGB) channels in an image and dividing the value by three.
- **Brightness and Contrast:** It can be adjusted in a colour or greyscale image. Increasing the contrast enhances the distinction between dark and light areas whereas increasing the brightness lightens the image so as to enhance even the shadow areas.
- **Thresholding:** It segregates the image grey values into two distinct values (that is 0 for black and 255 for white) by a threshold value.
- **Colour Separation:** It means that colour channels can be separated to bring out individual colours of an image.
- **Subsetting:** It is a process that breaks a larger image into a smaller one so as to use only the area of interest needed for the project.
- **Edge Matching:** It intends to create a seamless join across two or more maps as there might be some minor mismatches between adjacent map sheets.

8.2.2 Vector Data Input

Vector data is usually captured or digitised from a hardcopy print out or a digital raster image. With the advancement of technology data encoding methods for vector data have also improved and a number of input devices are in use for map digitisation. We will discuss about input devices in detail in the later section of this Unit.

8.3 INPUT DEVICES

You have read about the methods of raster and vector data input in the previous section. Now let us discuss the input devices used for raster and vector data input.

Aerial photographs and satellite imageries are the examples of raster data in digital format. However, analog maps, aerial photographic film or print out of satellite images can also be converted into digital form by scanners. Vector data is usually captured or digitised from a hardcopy printout or a digital raster image. With the advancement of technology data encoding methods for vector data have also improved and a number of input devices are in use for map digitisation. Let us discuss about the various devices used for raster and vector data.

Devices used for Raster Data Input

Scanners are used for converting analog maps, aerial photographic film and satellite imagery prints into digital format. Let us discuss about the scanners in detail.

Scanners

Scanners are used to convert images from analog maps or photographs into digital image data in raster format, which is then converted to vector format through digital tracing or digitisation. This process of tracing is also called vectorisation. Scanning converts the map into binary scanned file in raster format, each pixel having a value of either '1' representing the map feature or '0' representing the background (Chang, 2010). Map features are shown as raster lines i.e. a series of connected pixels on the scanned file.

A scanner has a light source, a background for source document and a lens. There are three different types of scanners:

- Flat-bed scanners
- Drum scanners
- Large-format feed scanners

Flat Bed Scanners as shown in Fig. 8.2a have a flat surface on which the map is placed. It has a mat or hinged cover which is kept on the top of map. An optical train emitting light is then passed over the map and light reflected back from the map is sensed. Flat-bed scanners are very small and not so accurate.

Drum Scanners as shown in Fig. 8.2b differ from flatbed scanners as they employ a rotating cylinder. A map is fixed onto the surface of this cylinder which is then set to rotate at a uniform velocity. They also use optical detection of reflected light to sense map elements. Drum scanners are too slow and

expensive. Both drum scanners and flat bed scanners can give monochromatic or colour output. For obtaining colour output each of the three primary colours are scanned either individually and then recombined, depending on the technology used.

Large-format Feed Scanners as shown in Fig. 8.2c are most suitable of all scanners mentioned above as they are very accurate and inexpensive. It uses contact image sensor technology: red, green and blue light-emitting diodes that produce white light. Contact Image Sensors (CIS) use sensors and mirrors in combination with a cold cathode ray fluorescent lamp to scan documents. However, large format scanner would not be ideal to scan artful compositions of objects with more depth than a thick piece of paper.

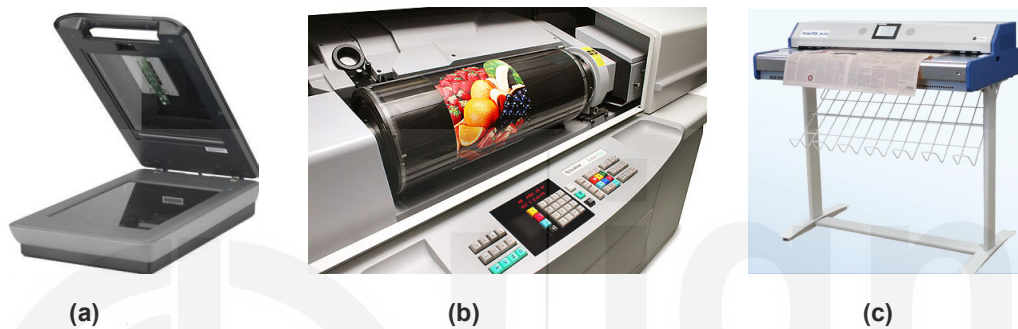


Fig. 8.2: Different types of scanners: a) flat bed; b) rotating drum; and c) large-format feed scanners. [Sources: www.techfuels.com/scanner/1579-flatbed-scanners.html (a); <http://postcardprinters.us/drum-scanner> (b); and www.chinasystem.com.hk/en/main.htm (c)]

Devices used for Vector Data Input

Some of the commonly used vector input devices are:

- Digitising Table
- Mouse
- Keyboard

Digitising Table

A digitising table as shown in Fig. 8.3 is the most common device used for **hard copy digitisation**. It has an in-built electronic mesh, which can sense the position of a cursor. Cursor is a mouse like device also called **puck** which contains a cross hair encased in a glass or transparent plastic that allows the operator to place the cursor accurately over the map elements. Cursor also has buttons to indicate the start and end of line or polygon or define left and right polygons. A map document is fixed to the center of digitising table with a sticky tape. Digitising table uses a local rectilinear coordinate system. Map and the digitiser must be registered so that vector data can be captured in real-world coordinates. This is achieved by digitising a series of four or more control points also called reference points or tics along the four corners so that it is well spread out and then their real-world values are entered. Digitiser control software calculates the transformation matrix and then automatically applies this to any future coordinates that are captured.



Fig. 8.3: Digitising table and puck. (Source: <http://proceedings.esri.com/library/userconf/proc01/professional/papers/pap894/p894.html>)

Digitising tables can range from small paper sized formats to large formats. Large sized tables also have adjustable stands to alter the elevation of the digitiser as per convenience of the operator. Modern digitisers provide good resolution of about 0.001" and an accuracy of about 0.003". A good digitiser should have the properties of stability, repeatability, linearity, resolution and skew (DeMers, 2009). Let us read about these properties in detail.

- **Stability:** It deals with the tendency of the digitiser not to change as its temperature rises.
- **Repeatability:** It is a measure of the precision of the digitiser. Suppose, if the operator is able to place the cursor at the same location twice and the difference between the first and the second readings is of the order of 0.001", the digitiser is said to have good repeatability.
- **Linearity:** It is a measure of the ability of the digitiser to be within a specified distance of the correct value as the cursor is moving over large distances.
- **Resolution:** It deals with the ability of the digitiser to handle even smaller units of measures with precision.
- **Skew:** It represents squareness of results on a digitising table, which deteriorates as the table becomes old and withered along the edges, thus reducing the ability to digitise the entire table.

Mouse

Mouse is the simplest and the most accurate type of digitiser. Fig. 8.4(a-c) shows different varieties of mouse. Mouse has sensors that responds to the motion of a rubber ball found inside it. This type of digitisation is popular now-a-days and is used for **on-screen digitisation or heads-up digitisation** as shown in Fig. 8.4(d). In this, the analog map after being converted into digital raster data is input into the GIS software, which displays it on the computer screen or monitor. Then it is digitised with the help of a mouse or digitiser. It provides greater accuracy in digitisation due to inbuilt facilities of zoom, pan, etc. in the GIS software package. The roller ball mouse has gradually been phased out over the past ten years, replaced by the optical mouse and more recently by the laser mouse.



Fig. 8.4: Mouse and digitisation: a) Rollar ball; b) Optical mouse; c) Laser mouse; and d) On-screen digitisation with the help of mouse.

Keyboard

Keyboard is the simplest device to input data into a computer (Fig. 8.5). The process is also known as **key coding**. This technique is mostly used to input attribute data into a GIS. These attributes can be linked to map features in a spatial database using identification codes. The coordinates of special features like point, line and polygons can also be encoded with the help of keyboard.



Fig. 8.5: Keyboard.

8.4 EXISTING DATA AND ITS CONVERSION

You have read about the input devices used for raster and vector data input in the previous section. Now let us discuss about existing data and its conversion.

There are different types of spatial datasets that are created and distributed in many formats. Some of these formats are point, line, polygon, raster, vector and image, etc. These formats are created by considering the different computer related software and operating systems. For example, Computer operating systems like Windows and Unix, and GIS related software like GRASS, Idrisi, ArcGIS, etc. Hence, there are many numbers of currently

available spatial datasets that are accessible in different format. To convert these datasets to a GIS related compatible format several methods and tools are being used. Presently most of the GIS packages are being developed with an inbuilt functionality to import and export these widely shared common formats.

There are different types of spatial dataset formats that are directly or indirectly used in GIS environment. These are i) Generic data formats ii) Published data formats iii) GIS-specific data formats.

- i) **Generic data formats:** In a GIS system generic spatial data such as vector polygons, vector lines, raster polygons, raster lines are stored in several formats. For example, the vector polygons are stored as a continuous chain of points and marked by a unique ID number. In GIS certain formats are suitable for analysis and some formats are for displaying only. Moreover, at times data are accessible in only one format and should be converted to the required format to suit the requirement of the given GIS project.
- ii) **Published data formats:** These are spatial dataset formats which are published by the government institutions and accessed for their convenience. Most of the available GIS packages are facilitated with a built-in functionality of these formats to import and export data. Some of these commonly used formats are Digital Elevation Model (DEM), Spatial Data Transfer Standard (SDTS), Digital Line Graph (DLG), Topologically Integrated Geographic Encoding and Referencing (TIGER), Drawing Exchange Format (DXF), etc.
- iii) **GIS-specific data formats:** These are data formats that are specifically supported by some of the commonly used GIS packages. For example, in ArcGIS a proprietary export/import format is supported where data can be easily exchanged between different operating systems such as Unix and windows.

8.5 CREATING NEW DATA

In the previous section you have read about the existing data and its conversion. In this section we will discuss about creating new data.

For any GIS project we require not only existing data as input, but also some sort of data that need to be created (as new input data) for the requirement of the GIS project. We have already discussed various modes of input data acquired through different devices. We can create new input data by in situ data collection methods. Let us discuss about in situ data collection in detail.

8.5.1 In Situ Data Collection

An in-situ data are collected on the ground by involvement of humans or by the use of special data collection instruments. For example, the involvement of people in data collection are the individuals who visit door to door for collection of socio, economic and demographic data. On the other hand the use of different instruments such as rain gauge, thermometer, anemometer to collect rainfall, temperature and wind parameters respectively are the classical examples of data collected by instruments.

The primary characteristic of in situ data sets while considering the spatial data is the documentation of each measurement in geographic space known as *ground truthing*. In ground truthing survey the measurement of the data collected will be the same as the actual location on earth. Interestingly, the most commonly used device to locate the data collected in the field is Global Navigation Satellite System (GNSS) instrument that is used for navigation, position accuracy and for time services on a comprehensive global or regional scale. The device helps to map the obtained ground survey data in a GIS environment. There are many GNSS launched by many countries such as GPS, GAGAN, GLONASS, etc. which you have already read in Unit 2 of this Course. However, let us discuss by taking the example of Global Positioning Systems (GPS).

Depending on the particular equipment utilised and the techniques used, GPS are capable of recording position to a high level of accuracy. The system was designed and is maintained by the US Department of Defense (DoD) as an accurate, all-weather navigation system. Though designed as a military system, it is available with certain restrictions to civilians. It is used to determine the exact x, y, and z location on both land and sea surface. Though there are limitations in using the GPS, it has many advantages and used for various purposes.

SAQ I

- a) What is analog Data?
- b) List the names of devices used for input of geospatial data.
- c) Define published data formats.
- d) What is the primary characteristic of in situ data sets in relation to the spatial data?

8.6 MAP DIGITISATION

You have read about creating new data in the previous section. In this section we will discuss about map digitisation.

It is the process of converting features on a paper map or simply tracing a paper map into digital format. Manual digitising is the most common type of encoding spatial features. Manual digitisation is of two types: on-screen digitisation and hard copy map digitisation. The features to be digitised can be point, line or polygon. Let us now discuss one by one how to digitise these features.

Digitising point feature: A point has a zero dimension and each point is just clicked once to record its location. Thus, point features are recorded as single digitised points. For example, spot heights, location of railway station, bus station, telephone exchange, etc. A unique code or identifier is added to each point feature so that the attribute information may be attached to it.

Digitising line feature: Line features such as roads or streams are digitised as a series of points which are connected together with straight line segments with the help of software. Lines are also referred to as arcs and its start and end

points are called nodes. A unique code is also added to a line feature to which attributes can be added.

Digitising polygon feature: Area or polygon features are digitised as the series of points linked together by line segments in the same way as line features. In a polygon feature, the start and end points are joined or closed to form a complete area. A centroid is created for each polygon with the help of GIS software. A unique identifier can also be added to polygon centroid to which attributes can be added. For example, a polygon representing a district can have attributes like district name, male and female population, sex ratio, literacy rate, etc.

While digitising line features one can follow either *point mode* or *stream mode*. In point mode digitisation, the person carrying out the digitisation task decides the number of points to be placed and the distance between the points (Fig. 8.6a). If the line is curved or bent at places then more points are required and the points should be close to each other. If the line is more or less straight then fewer points can be placed. In stream mode digitisation, points or vertices are added at a preset interval, also known as **stream tolerance** (Fig. 8.6b). Hence, stream digitisation partially automates the digitisation process. It instructs the digitiser control software to automatically collect vertices every time a preset distance or time threshold is crossed.

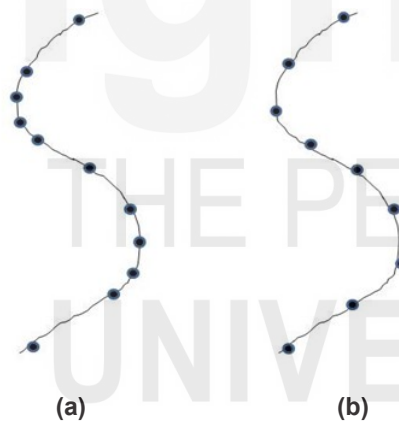


Fig. 8.6: Two types of digitization: a) point mode; and b) stream mode.

There are a number of digitisation errors related to vector data. This requires quality checking and cleaning of vector data. After cleaning the vector data, topology is created. Topology is the relationship between entities (point, line and polygon) in a vector dataset. These relationships are related to connectivity, adjacency and containment between entities which we will discuss detail in next unit (Unit 9) of this course.

8.7 REFERENCE FRAMEWORKS

You have read about map digitisation in the previous section. In this section we will discuss about reference frameworks.

A map is the representation of earth's pattern as a whole or a part of it, on a plane surface, with conventional signs, drawn to a scale and projection and every point on it corresponds to actual terrestrial position. Thus, maps are representations of three-dimensional reference globe projected onto a flat

surface. You know that the process of transforming three-dimensional surface of globe or a part of it on a two-dimensional or flat surface is called map projection.

Digitising helps us to reduce sophisticated information of map projection to a set of Cartesian coordinates in case of digitiser. Prior to digitisation, we need to provide information regarding the type of projection used, information about grid system, zones of origin, etc. to GIS software. This helps us in transforming the map to its original projection after it is input. GIS produces a number of transformations so as to project from Cartesian coordinates on the digitiser to a two-dimensional map projection coordinates and then through a process called **inverse map projection** to three-dimensional latitude and longitude coordinates. This process eventually needs to be repeated to produce cartesian coordinates for output devices.

GIS software has to perform a number of graphical manipulations that results into different types of projections. It is important to mention here **three primary processes** that often occur simultaneously (DeMers, 2009). They are as follows:

- translation
- rotation
- scale change

Translation: It sets the distance to move the objects contained in the track, in the x, y and z directions. In translation, objects slide to the new position without being rotated. This is done by adding or subtracting the coordinate values necessary for X and Y coordinates of the object as shown in Fig. 8.7. The new X-coordinate say X' would be equal to the original X-coordinate plus some value T_x . Similarly, the new Y coordinate Y' for each graphic object would be equal to Y coordinate plus some value T_y .

$$X' = X + T_x$$

$$Y' = Y + T_y$$

where, the values of T_x and T_y can either be positive or negative.

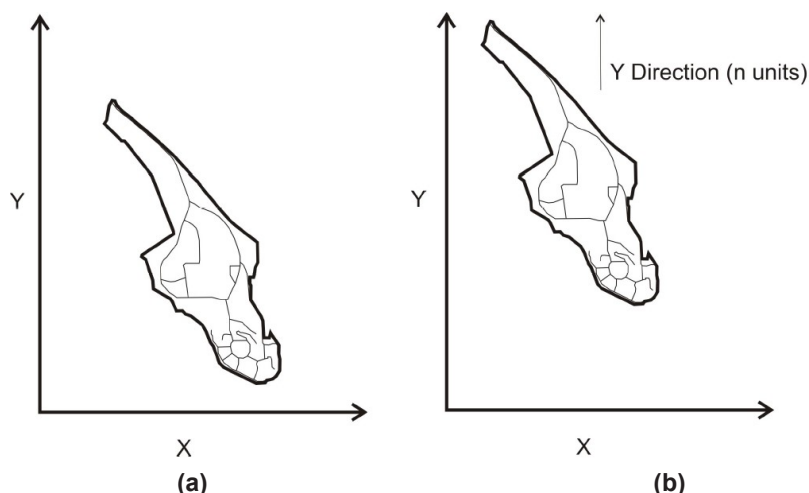


Fig. 8.7: Translation: a) before translation; and b) displacement of the same object in Y-direction after translation.

Scale Change: It sets the x, y, and z scale factors to make objects in the track larger or smaller as shown in Fig. 8.8. This is also useful in comparing differently scaled maps. Even the output can be represented in different scales. This is done by multiplying the overall X-coordinate extent by a scale factor s_x , and each set of Y-coordinates by a scale factor s_y .

$$X' = X s_x$$

$$Y' = Y s_y$$

where, s_x and s_y represent the amount or percentage of scale change.

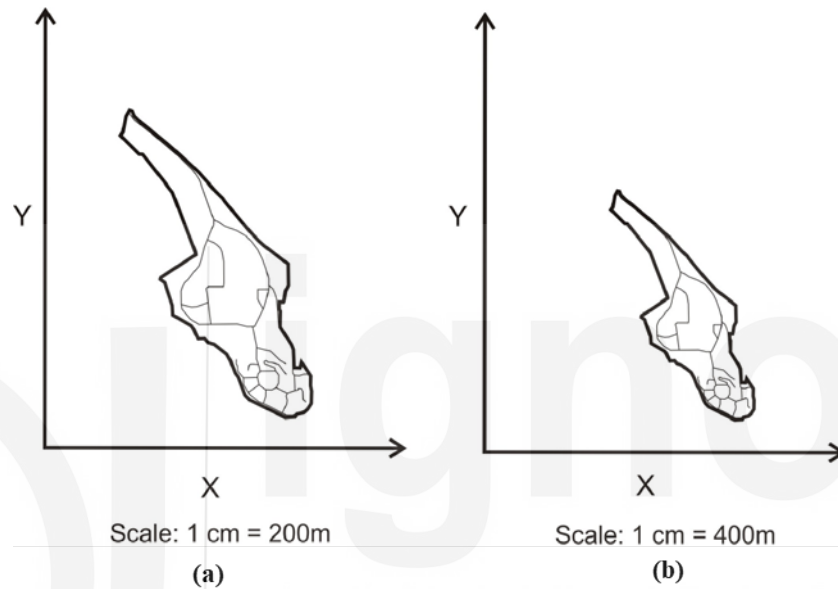


Fig. 8.8: Scale change; a) original object; and b) same object has become smaller after scale change.

Rotation: It sets the x, y and z rotation angles around a fixed point (the origin) for objects contained within the track as you can see in Fig. 8.9. Rotation around the x-axis is the **roll, or yaw angle**. Rotation around the y-axis is the **inclination, or pitch angle**. Rotation around the z-axis is the **azimuth, or heading angle**. It is used frequently during the process of projection and inverse projection and uses the basic trigonometry. For X-coordinate locations, the new location X' would be found by multiplying the original X location by the cosine of the new angle (θ) and then adding that value to the original Y-coordinate multiplied by the sine of the theta ($\sin\theta$). The new Y-coordinate location Y' is found by multiplying the negative of the original X value by the sine of the angle and again adding that to the product of Y coordinate and $\sin\theta$.

$$X' = X \cos\theta + Y \sin\theta,$$

and

$$Y' = -X \sin\theta + Y \cos\theta$$

where, θ is the angular displacement needed.

These are the basic three types of graphical manipulations which help in all necessary transformations.

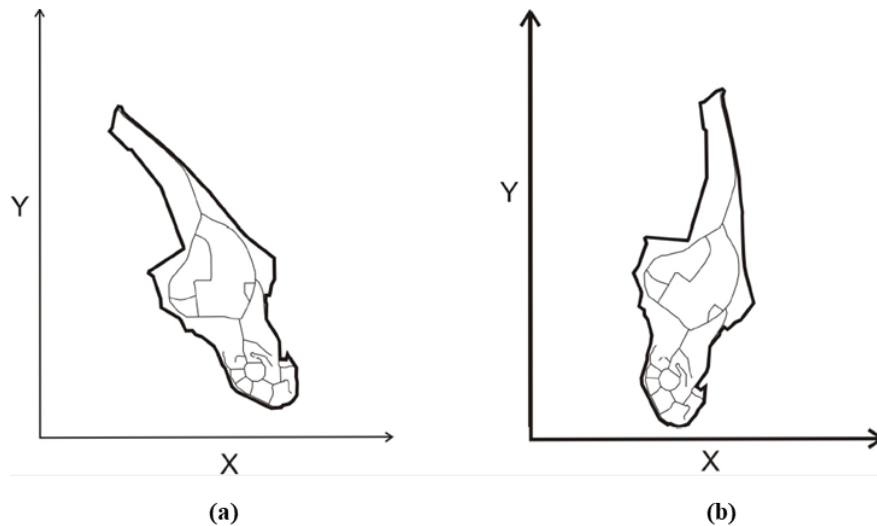


Fig. 8.9: Rotation: a) object before rotation; and b) same object after rotation by angle θ .

8.8 GEOREFERENCING

You have read about map reference frameworks in the previous section. In this section we will discuss about georeferencing.

Georeferencing also known as georegistration is the process of locating any entity in real-world coordinates. Raster data is commonly obtained by scanning maps, topographical maps or collecting aerial photographs and satellite images. Scanned maps normally do not contain spatial reference information. However, sometimes they might have coordinate information marked along it with the help of which it can be georeferenced (Fig 8.10a). If there is no coordinate information, we can georeference it by adjusting the map or image to the geographic location of a “known good” reference image or map. The image or map being used as a reference is called **reference** component and the image or map being adjusted is called **target** component (Fig.8.10b & c).

Georeferencing involves the following steps:

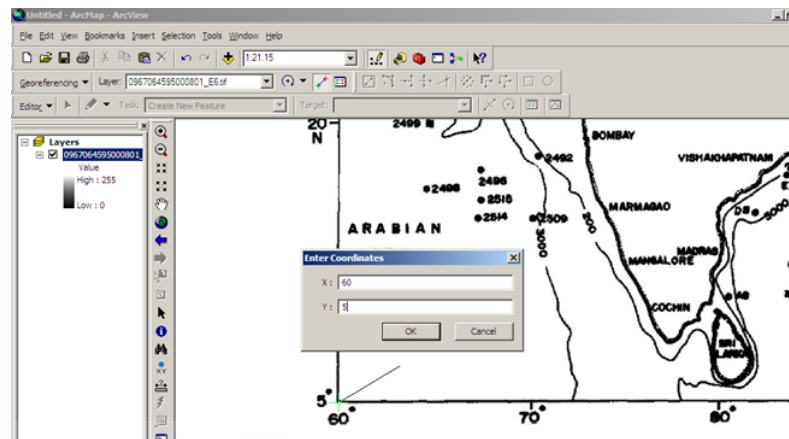
Step 1: Raster dataset needs to be opened up in GIS software;

Step 2: Add control points in the raster dataset that links it to known positions in map coordinates as shown in Fig. 8.10a. Control points are locations that are more or less permanent features on the earth and can be accurately identified on the raster dataset and in real world coordinates e.g., intersections of road, street corners, etc. They are also called GCPs (Ground Control Points). The control points are used to build a polynomial transformation that will convert the raster dataset from its existing location to the spatially correct location. We should now save the control points; and

Step 3: Raster dataset should be transformed permanently or rectified.

Transformation uses a set of control points and transformation equations to register a digitised image, a satellite image or an aerial photograph. Map to map or image to image transformation uses a set of control points to establish a mathematical model that relates the map coordinates of one system to another. An example of this has been shown in Fig. 8.10b. The Root Mean Square Error (RMSE) is a quantitative measure that can determine the quality of geometric

transformation. It measures the displacement between the actual and estimated locations of control points. **Resampling** fills in each pixel of the transformed image with a value that is derived from original image. The georeferenced image can be saved in different formats like IMG, TIFF, BMP, GIF, JPEG, etc.



(a)



(b)

(c)

Fig. 8.10: Georeferenced images: a) georeferencing by adding control points; b) and c) georeferencing of map/image by using a reference image (b) and target image (c). (Source: www.georeference.org/doc/georegistration.html)

Now let us get acquainted with some commonly used terms such as geocoding, geotagging and rubbersheeting, etc.

8.8.1 Geocoding

It is the process of converting the street addresses to latitude and longitude or to some universal coordinate system (Longely et al., 2011). So if a database has addresses mentioned in the form of house number, street name, zip code, etc. it can be mapped and entered into GIS. This process is also called **address interpolation**, which starts with a textual description of a location and translates that into the x,y coordinate that can be plotted on a map.

Another method of geocoding is generating points from a table in which at least two fields contain latitudinal and longitudinal coordinates (Table 8.1). The table and associated figure is shown in Fig. 8.11.

Table 8.1: X and Y coordinates of places.

FID	Shape	Id	Place_name	Latitude (Y)	Longitude (X)
0	Point	A1	Modi Rubber Ltd	29.062	77.705
1	Point	A2	Dayawati Modi Academy	29.05	77.707
2	Point	A3	Central Potato Research Institute	29.068	77.708
3	Point	A4	University of Agriculture and Technology (SVPUA&T)	29.075	77.709

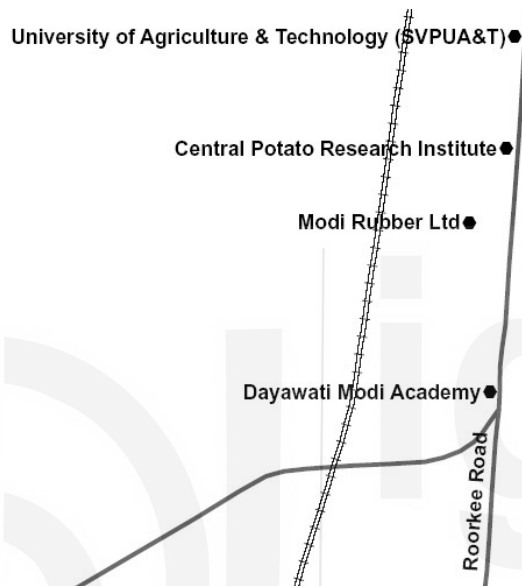


Fig. 8.11: Geocoding with the help of coordinate information of places given in Table 8.1.

Reverse geocoding can also be done by finding the street address with the help of associated geographical coordinates.

8.8.2 Geotagging

It is the process of adding location specific information or geographical identifiers to photographs, films, videos, websites, etc. This is also called **geospatial metadata** and usually consists of coordinate information in the form of latitude-longitude along with information related to altitude, bearing, distance, place names, etc. (Fig. 8.12).

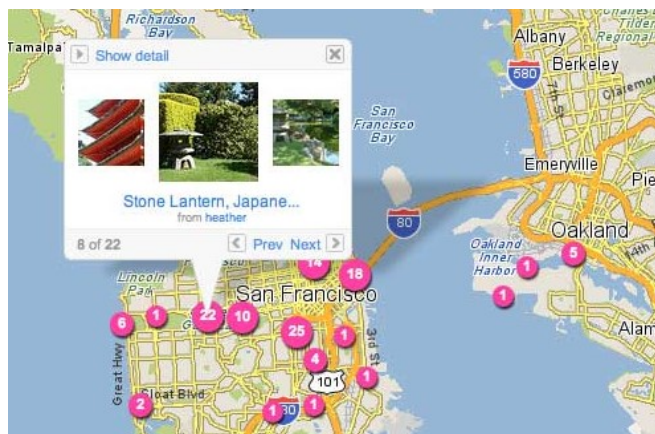


Fig. 8.12: Geotagging. (Source: www.web-strategist.com/blog/category/geo-tagging/)

8.8.3 Rubber Sheeting

It is a process that corrects flaws in source map or vector drawing through geometric adjustment of coordinates with the help of a more accurate target layer (Fig. 8.13). Errors may be due to imperfect image registration, scanning, inaccurate flight alignment or camera inaccuracies in case of aerial photographs, etc. Thus, inaccurate data can be stretched or rubbersheeted over the accurate data using control points and place marks common to both data sets.

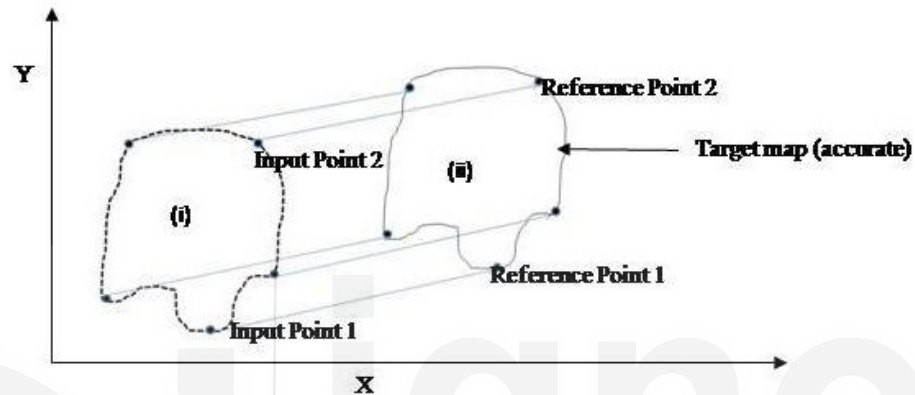


Fig. 8.13: Illustration of rubber sheeting; (i) vector drawing which is more accurately adjusted to the target drawing (ii).

8.9 LEVEL OF INPUT

GIS database should have relevant, authentic and optimum amount of information. It is important to follow certain rules while inputting data into a GIS project. Too much data input could confuse the user, while too little data could leave certain problems unanswered. DeMers, 2009 has listed out some rules or guidelines for inputting data into a GIS database which are given below:

- determine the real need for building a GIS database
- limit the level of input to coverage or layers that are used
- define the goals that specify the objectives of the project
- avoid exotic sources when conventional sources are available
- use the best and most accurate data needed for the project
- decide the level of accuracy of the available data
- import relevant features from the maps one by one as different layers into GIS
- avoid extraneous data in a given coverage and make it as specific as possible

SAQ II

- a) List the different modes of digitisation.
- b) What are the three basic types of graphical manipulations required for a GIS software to perform that results in different types of projections?

- c) What is meant by scale change?
- d) What is rubber sheeting?

8.10 SUMMARY

Let us summarise what you have studied in this unit.

- Data input is a method of selecting, acquiring and converting data into a digital format that can be stored and analysed in a computer system and known as data encoding. Geospatial data can be input in the form of vector or raster data.
- Analog maps, aerial photographic film and satellite imageries prints are mostly the raster data that can be converted into digital form using scanners. Vector data can be digitised using input devices such as digitising table, mouse and keyboard.
- Different types of spatial datasets like point, line, polygon, raster, vector and image are created and distributed in many formats. These formats are created by considering the different computer related software and operating systems.
- In-situ data are collected on the ground by involvement of humans or by the use of special data collection instruments. The primary characteristic of in situ data sets while considering the spatial data is the documentation of each measurement in geographic space known as ground truthing.
- Map digitisation is the process of converting features on a paper map or simply tracing a paper map into digital format. Manual digitising is most common type of encoding spatial features and it can be on-screen digitisation or hard copy map digitisation.
- GIS software performs a number of graphical manipulations that results in different types of projections. The three main processes that are involved in graphical manipulations are translation, rotation and scale change.
- Georeferencing also known as georegistration is the process of locating any entity in real-world coordinates. The image or map being used as a reference is called reference component and the image or map being adjusted is called target component.
- GIS database should have relevant, authentic and optimum amount of information. It is quite essential to follow certain rules while inputting data into a GIS project.

8.11 ACTIVITY

- Scan a paper map which has coordinate information at the corners. Then, georeference the scanned map using GIS software in a suitable coordinate and projection system.

8.12 TERMINAL QUESTIONS

1. Explain the different methods of data input in GIS.
2. Discuss the spatial data formats used in GIS.

3. Explain the process of map digitisation.
4. Explain the concept of georeferencing.

8.13 REFERENCES

- Chang, K.-T. (2010) Introduction to Geographic Information Systems. Tata McGraw-Hill, New Delhi.
- Davis, B.E. (2001) GIS: A Visual Approach. Onword Press, Canada.
- DeMers, M.N. (2009) Fundamentals of Geographic Information Systems. John Wiley and Sons Inc., New York.
- Longely, P.A. Goodchild, M.F. Maguire, D.V. Rhine, D.W. (2011) Geographic Information Systems and Science. John Wiley and Sons, New York.
- <http://postcardprinters.us/drum-scannerstem.com>
- <http://proceedings.esri.com/library/userconf/proc01/professional/papers/pap894/p894.htm>
- www.chinasystem.com.hk/en/main.htm
- www.georeference.org/doc/georegistration.htm
- www.techfuels.com/scanner/1579-flatbed-scanners.html
- www.web-strategist.com/blog/category/geo-tagging/
- <file:///C:/Users/hp/Downloads/eScholarship%20UC%20item%209q94d63j.pdf>

The data from all the above web pages was retrieved between 10th March, 2012 and 20th March, 2012; 15 November, 2022 and 21 January, 2023.

8.14 FURTHER/SUGGESTED READINGS

- Anjireddy, M. (2008) Textbook of Remote Sensing and Geographical Information Systems. BS Publications, Hyderabad.
- Bolstad, P. (2008) GIS Fundamentals: A Text Book on Geographic Information System. Eider Press, Minnesota.

8.15 ANSWERS

SAQ I

- a) Analog data is the data that is generally represented in paper form like paper maps, statistical tables or hard copy of satellite images.
- b) The names of different devices used for input of geographical data are, mouse, digitising table and puck/cursor, scanners, keyboard, etc.
- c) Published data formats are spatial dataset formats which are published by the government institutions and accessed for their convenience.
- d) The primary characteristic of in situ data sets while considering the spatial data is the documentation of each measurement in geographic space known as ground truthing.

SAQ II

- a) The different modes of digitisation are point mode and stream mode digitisation.
- b) The three basic types of graphical manipulations required for a GIS software to perform that results into different types of projections are i) translation ii) rotation iii) scale change.
- c) Scale change sets the x, y and z scale factors to make objects in the track larger or smaller. This is done by multiplying the overall X- coordinate extent by a scale factor s_x , and each set of Y coordinates by a Y scale factor s_y .
- d) Rubber sheeting is a process that corrects flaws in source map or vector drawing through geometric adjustment of coordinates with the help of a more accurate target layer.

Terminal Questions

1. Please refer to section 8.2.
2. Please refer to section 8.4.
3. Please refer to section 8.6.
4. Please refer to section 8.8.



ignou
THE PEOPLE'S
UNIVERSITY

