UNIT 7 DATA INPUT

Structure

7.1 INTRODUCTION

You have been introduced to the concepts of GIS, data, data models and data structures in Block 2, *Fundamentals of Geographic Information System* of MGY-003. Now we will introduce you to the methods of data input. You have already learnt that data is the basic requirement to perform any analysis in GIS platform. In this unit, we will discuss the methods, processes and devices of data input and reference frameworks alongwith a short introduction to linking external database.

Objectives

After studying this unit, you should be able to:

- define data input;
- list sources of data;
- describe data input devices;
- illustrate digitisation process;

- discuss level of input;
- link external databases and metadata; and
- explain reference frameworks and georeferencing.

7.2 WHAT IS DATA INPUT?

The first step in developing the database for a GIS is to acquire the data and to place them into the system.

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. 7.1. Database creation is the fundamental, expensive and time-consuming 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.

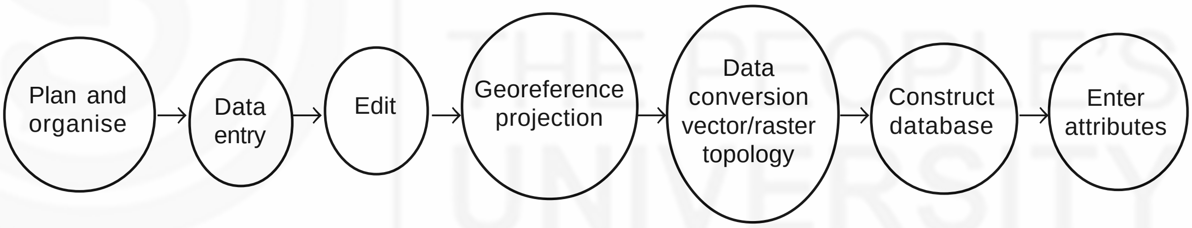


Fig. 7.1: Chain of data input process in GIS (source: modified after Davis, 2001)

Data collected both in the analog and digital format is required to be encoded to make it compatible with GIS. ***Analog data*** is normally in the form of hard copy such as 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 methods of data input 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. Let us discuss these devices in detail.

7.3 METHODS OF GEOSPATIAL DATA INPUT

You have learnt that geospatial data can be in the form of vector or raster data and so it is worthwhile to discuss here data input methods for both these data forms.

You have already learnt the basics of raster and vector data in Unit 5, *Data types and Sources* of Block 2, *Concept of Geospatial Data* of MGY-001.

7.3.1 Methods of Raster Data Input

Aerial photographs and satellite imagery are the examples of raster data in digital format. However, analog maps, aerial photographic film or hard copy print out of satellite images can also be converted into digital form by scanners.

Scanner

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 file in raster format, with 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 (as shown in Fig. 5.8 of MGY-001).

Scanning is the most commonly used method of automatic digitising.

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

In general, the quality of a scanned image can be improved by changing the brightness and contrast levels by using ***gamma correction*** (a method that draws histogram of the image and places points strategically along the histogram to isolate data types) and ***filtering method*** (to remove noise from the scanned document).

Flat-bed Scanners as shown in Fig. 7.2(a) 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 the light reflected back from the map is sensed. Flat-bed scanners are very small and not so accurate.

Drum Scanners as shown in Fig. 7.2(b) differ from flat-bed 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 is scanned either individually and then recombined, depending on the technology used.

Large-format Feed Scanners as shown in Fig. 7.2(c) 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 combinations 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.

(a) (b) (c)

Fig. 7.2: 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)

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 Block 4, *Processing and Classification of Remotely Sensed Images* of MGY-002.

Image distortions are corrected by despeckling, greyscaling, adjusting brightness or contrast, thresholding, etc. We will be discussing georeferencing and related topics later in 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.

7.3.2 Methods of 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 methods of data input for vector data have also improved and a number of input devices are in use for map digitisation. Let us start with vector input devices. Some of the commonly used devices for vector data input are listed below:

- Digitising Table
- Mouse
- Keyboard

Digitising Table

A digitising table as shown in Fig. 7.3 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.

Tics are ground-control points for a vector data, which represent known geographic locations and are used to register maps mounted on a digitiser table.

Fig. 7.3: 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 discuss 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 device. Fig. 7.4(a-c) shows different varieties of mouse. Mouse has sensors that respond 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. 7.4(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. 7.4: Mouse and digitisation; (a) roller ball, (b) optical, (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. 7.5). 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 points, lines and polygons can also be encoded with the help of keyboard.

Fig. 7.5: Keyboard

7.3.3 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, which we have already discussed. The features to be digitised can be point, line or polygon (as shown in Fig. 5.8, MGY-001).

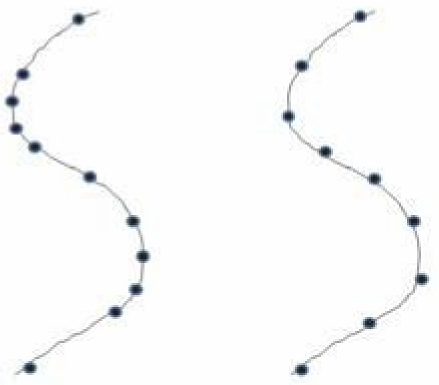
Let us now discuss how to digitise these features.

- **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 starting and ending 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 (Fig. 7.6a), the person carrying out the digitisation task decides the number of points to be placed and the distance between the points. 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. 7.6b). 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.



(a) **(b)**

Fig. 7.6: Two types of digitisation; (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.

7.3.4 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 two 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, and
- avoid extraneous data in a given coverage and make it as specific as possible.

Spend 5 mins

Check Your Progress I

1) List the names of devices used in inputting geospatial data.

....

....

....

....

2) What are the different modes of digitisation?

....

....

....

....

7.4 CAPTURING ATTRIBUTE DATA AND METADATA

You have learnt about methods of spatial data input. Now, let us learn how to capture attribute data, metadata and link external database.

You have been introduced to attribute data and metadata in Unit 4, *Introduction,* of Block 2 *Concept of Geospatial data* of MGY-001.

7.4.1 Capturing Attribute Data

Attribute data are non-spatial properties of a spatial entity. For example, a road may be captured as a line entity and represented as a spatial component in GIS by a certain colour, symbol or a particular location. However, information related to the road i.e. its length, width, type of surface, traffic conditions and estimated number of vehicles per hour, etc. are important and required to be entered into a GIS database. These attribute values can either be stored separately from spatial information or attached to spatial database in the form of object data. Attribute data may come from different sources like paper records, existing databases, spreadsheets, etc. In a GIS database, attributes are managed in tables based on series of simple yet essential relational data concepts. Attribute information in a GIS are typically entered, analysed and reported using a Database Management System (DBMS).

7.4.2 Linking External Databases

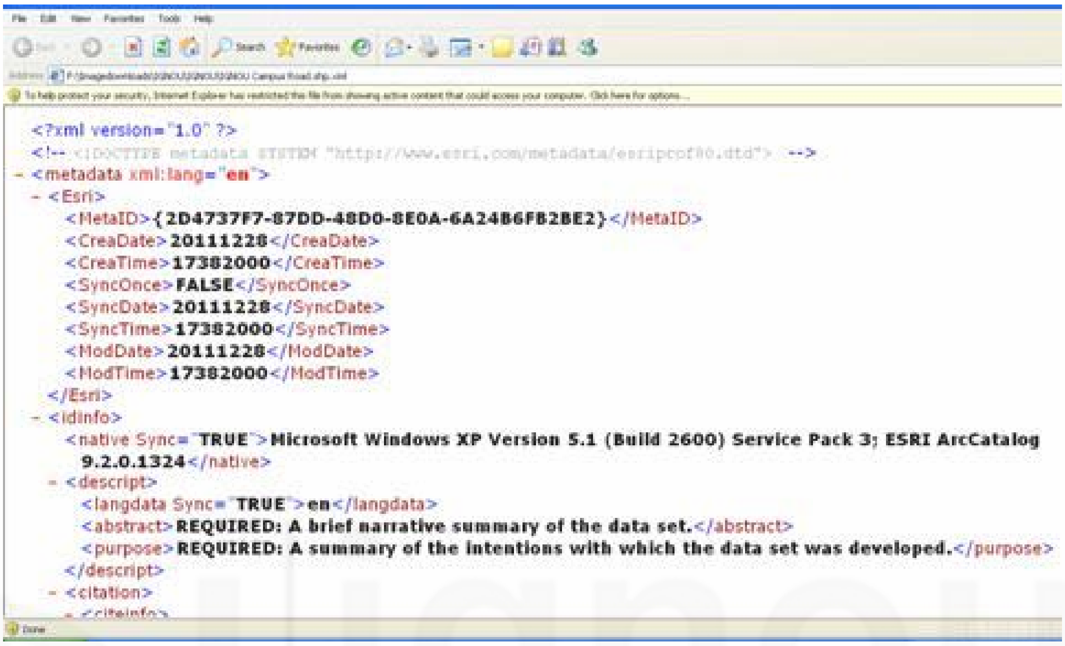
A GIS database needs to be updated time and again as the attributes and the spatial characteristics change in course of time. For this, there has to be an external source of GIS database which can be collected in due course of time and integrated with the master GIS database. This would in fact bring changes in the master database. It is important to note that the new GIS data should be reliable and the format should be compatible with master database as well as GIS software. Now-a-days many governmental agencies, private companies and non-governmental organisations are vendors of GIS database. We should be aware of the quality control procedures followed by the vendor. We should also go through the metadata to ensure that the data source is valid and it is worth entering into the master GIS database. While linking external databases, one has to remove data redundancy or duplicacy as it is time-consuming and affects unit cost of data input.

7.4.3 Metadata

Metadata in simple words is known as data about data. In other words, metadata is a summary document providing content, type, creation, quality and spatial information about a data set. In geoinformatics, it is the information about geospatial data. Metadata provides some basic information as to when the data was created, by whom it was created, why it was created, where it was created, etc. Metadata summarises, indexes, abstracts and describes the quality

HTML (Hyper Text Markup Language) is used for displaying web pages.

and material of the data. Thus, metadata is an integral part of a GIS project and is usually prepared during the data production process. A metadata is usually presented as an XML (Extensible Markup Language) document (Fig 7.7). XML is a markup language similar to HTML. In a metadata XML document, a title may be stored as follows: <title>My Document</title>.



```
<?xml version="1.0" ?>
<!-- <!DOCTYPE metadata SYSTEM "http://www.esri.com/metadata/esriprof80.dtd"> -->
- <metadata xml:lang="en">
  - <Esri>
      <MetaID>{2D4737F7-87DD-48D0-8E0A-6A2486FB2BE2}</MetaID>
      <CreaDate>20111228</CreaDate>
      <CreaTime>17382000</CreaTime>
      <SyncOnce>FALSE</SyncOnce>
      <SyncDate>20111228</SyncDate>
      <SyncTime>17382000</SyncTime>
      <ModDate>20111228</ModDate>
      <ModTime>17382000</ModTime>
    </Esri>
  - <idinfo>
      <native Sync="TRUE">Microsoft Windows XP Version 5.1 (Build 2600) Service Pack 3; ESRI ArcCatalog
        9.2.0.1324</native>
    - <descript>
        <langdata Sync="TRUE">en</langdata>
        <abstract>REQUIRED: A brief narrative summary of the data set.</abstract>
        <purpose>REQUIRED: A summary of the intentions with which the data set was developed.</purpose>
      </descript>
    - <citation>
```

Fig. 7.7: A simple metadata record in XML

a) Metadata Standards

Now we will learn about metadata standards. A metadata standard is a document that contains rules or guidelines for standardising the content of metadata. These standards can be created specifically for use within a particular organisation and also allow to develop tools that can be used to create and modify metadata that meets the standard. In fact, it is the way of verifying the content of metadata in compliance with the standard. Metadata standards contain rules or guidelines for the content, format, creation and update policies, and conformance rules.

The following are some of the suitable standards that are followed while representing metadata:

- **Federal Geographic Data Committee (FGDC)** has developed the content standards for metadata. It lays down specifications for both spatial as well as non-spatial data.

- **International Standards Organisation Technical Committee (ISO/TC 211)** develops standards related to geographic information. Recently ISO has developed an international metadata standard **ISO 19115**. It attempts to satisfy the requirements of all well-known metadata standards.

- **Dublin Core Metadata** is specifically intended to support resource discovery which involves searching, locating and retrieving of information resources on computer-based networks in response to user queries.

b) Metadata Formats

Metadata can be interpreted or valued differently by different users inspite of following a particular standard. Some common metadata formats are listed below:

- unstructured notes or log files
- plain text (ASCII) file
- HTML file format
- summary formats
- formats used for the purpose of indexing or searching but not for viewing
- some new formats being used specific to GIS software which enables to view, create and manage metadata and
- some customisable metadata formats that are becoming popular with XML, etc.

Till 1999, the most common format of metadata used to be ***Outline Format***. Soon ***FAQ*** (Frequently Asked Question) format was popularised as it was easier to read. This format could be easily understood even by non-technical or non-GIS users.

c) Essential Questions for Metadata Creation

We have already discussed in the beginning that metadata provides information about a number of questions related to the data. A general form of metadata has been given in interview approach by U.S. Geological Survey to understand the questions to be answered to create metadata. Some important questions that need to be answered in a metadata are as follows:

- what does the dataset describe?
- who produced the dataset?
- why was the dataset created?
- what was the original source of dataset and techniques of its collection?
- what is the reliability of data?
- what is the availability of data? and
- what is metadata modification?

7.5 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.*

You have already read about maps in Unit 9 *Basic of Mapping* of MGY-001.

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's 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.

The basic geometric transformations of translation, rotation and scale change are essential requirements for computer visualisation and manipulation of map data. Combinations of all these basic transformations are referred to as *affine transformations*.

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

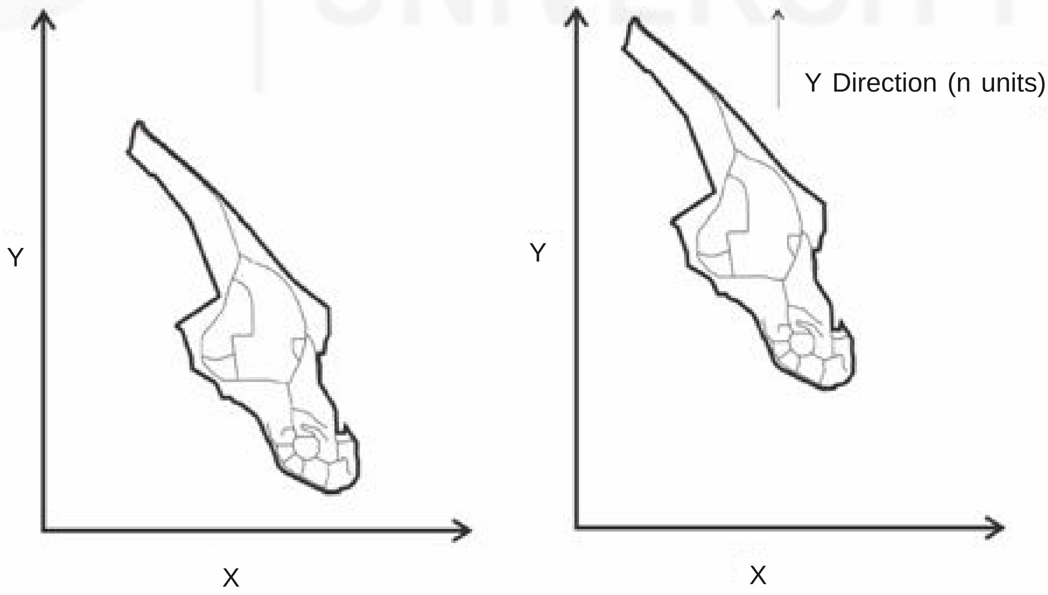
- translation
- scale change and
- rotation

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. 7.8. 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.



(a) (b)

Fig. 7.8: Schematics showing the process of translation; (a) before translation, and (b) after translation. Note 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. 7.9. 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.

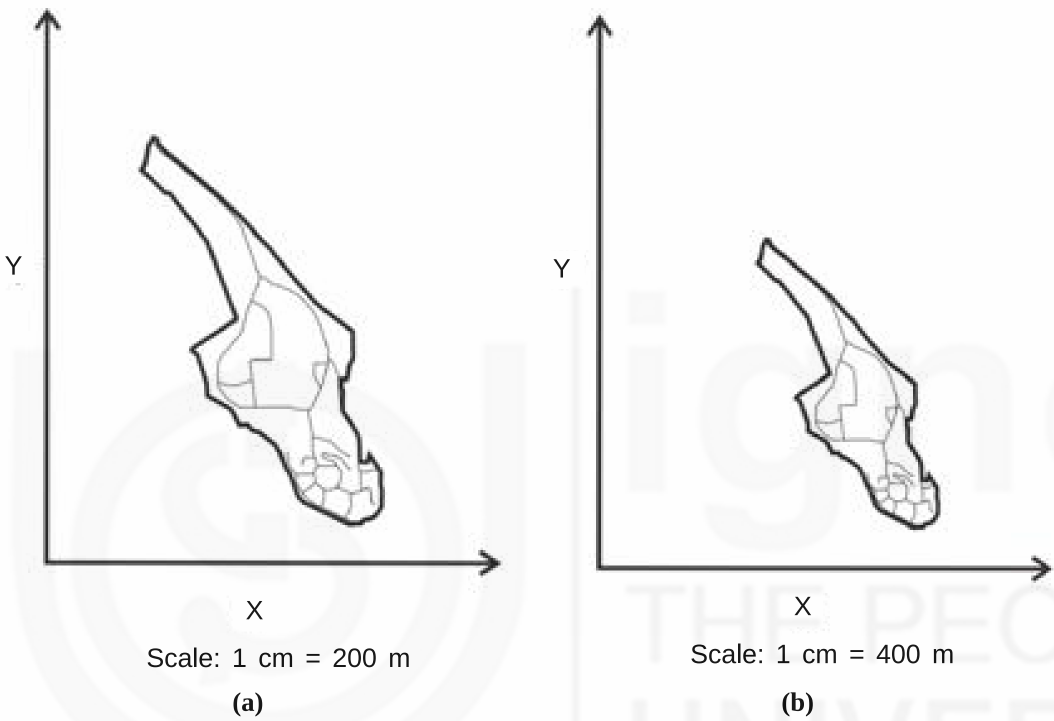


Fig. 7.9: Schematics showing the process of scale change; (a) original object and (b) after scale change. Note that the same object has become smaller after scale change

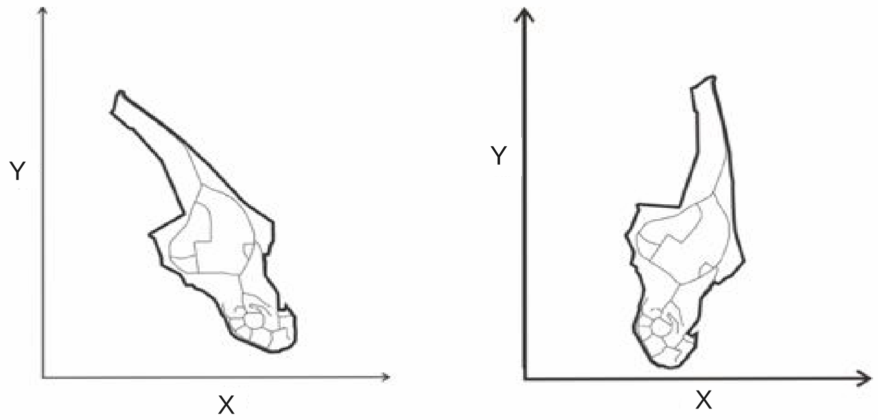
Rotation: It sets the x, y and z rotation angles around a fixed point (the origin) for objects contained within the track as you can see in Fig. 7.10. Rotation around the x-axis is the ***roll,*** or ***yaw angle***. Rotation around the y-axis is the ***inclination,*** or ***pitch angle***. Rotation 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 (è) and then adding that value to the original Y- coordinate multiplied by the sine of the theta (sin è). 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 è.

$$X' = X \cos è + Y \sin è,$$

and

$$Y' = – X \sin è + Y \sin è$$

where, è is the angular displacement needed.



(a) (b)

Fig. 7.10: Schematics showing the process of rotation; (a) object before rotation and (b) same object after rotation by angle è

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

Spend 5 mins

Check Your Progress II

1) What is meant by scale change?

....

....

....

....

....

....

....

....

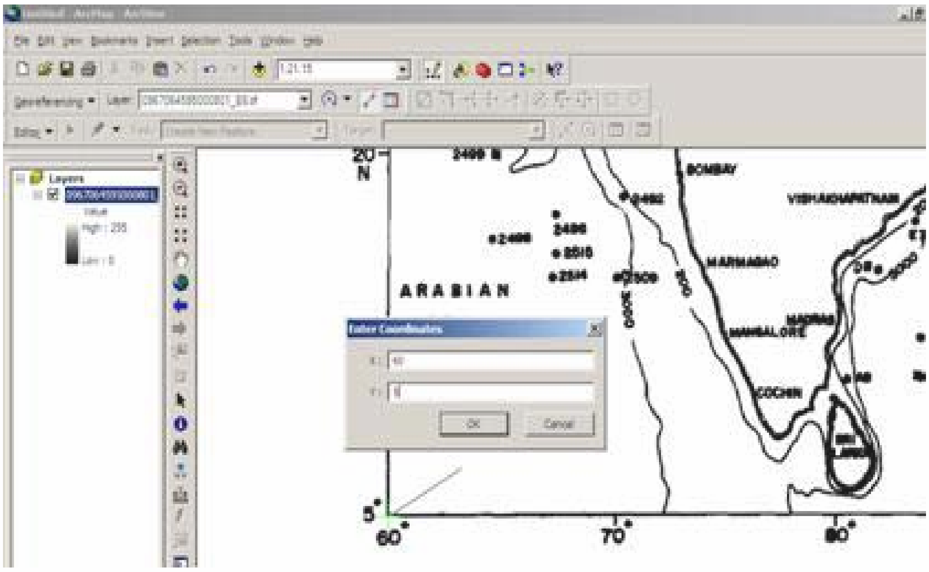
....

....

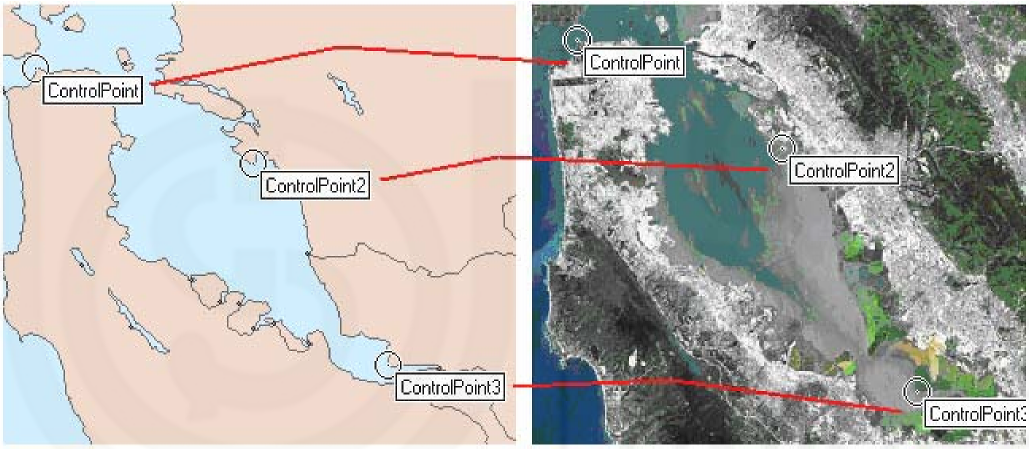
7.6 GEOREFERENCING

Georegistration is another term which is used alternatively with georeferencing.

Georeferencing 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. 7.11a). 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. 7.11b & c).



(a)



(b) (c)

Fig. 7.11: 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)

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. 7.11(a). 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.

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).

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. 7.10(b). 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.

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

7.6.1 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. The Table 7.1 and associated figure is shown in Fig. 7.12.

Table 7.1: 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

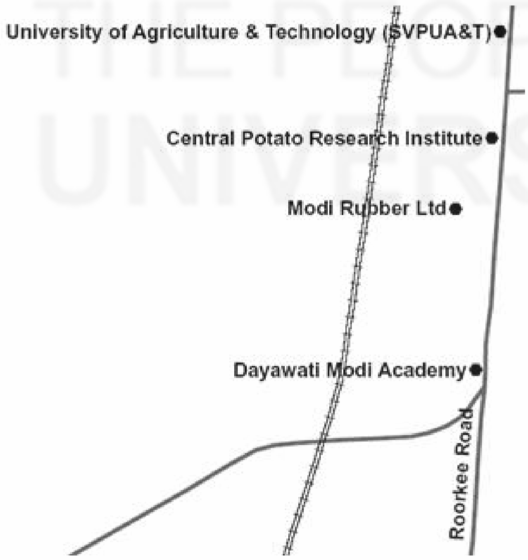


Fig. 7.12: Geocoding with the help of coordinate information of places given in table 7.1

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

7.6.2 Geotagging

Geotagging enables the user to find location of a particular media. 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. 7.13).

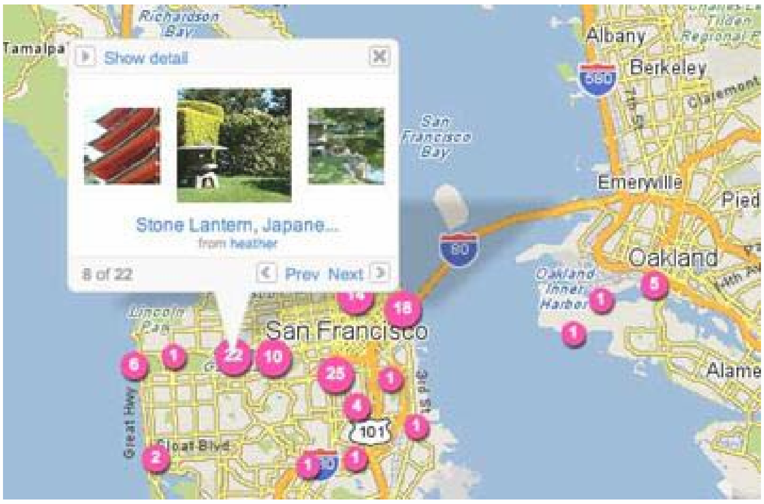


Fig. 7.13: Geotagging (source: www.web-strategist.com/blog/category/geo-tagging/)

7.6.3 Rubbersheeting

Rubbersheeting is a non-uniform adjustment of a dataset based on the movement of known control points to new locations. 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. 7.14). 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.

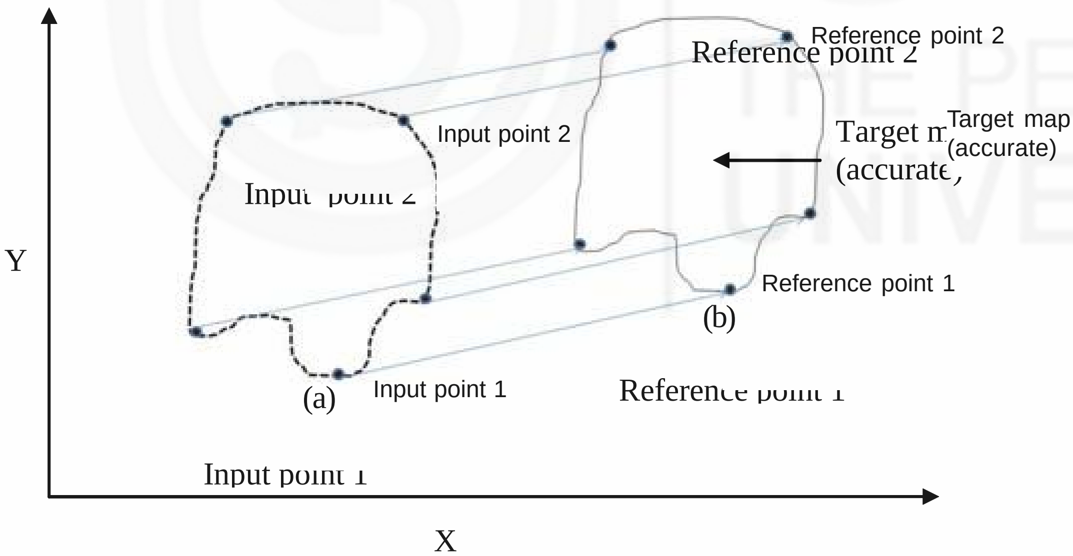


Fig. 7.14: Illustration of rubbersheeting; (a) vector drawing which is more accurately adjusted to the target drawing (b)

7.7 ACTIVITY

1) Scan a paper map which has coordinate information at the corners.

2) Georeference the map using a suitable coordinate and projection system with the help of GIS software. Then input the rectified or georeferenced map into the GIS software and digitise it (on-screen digitisation). Digitise different features like roads, rivers, lakes, forests, etc. and input them as different layers in the GIS database. Now add attributes to your digitised features so as to represent their names, location and other attributes.

3) Now you have learnt the basics of data inputting. After finishing this you can scan topographical maps of a place and continue doing the same process as mentioned above. If the topographical maps are in continuity then digitised features can also be edge-matched once the georeferenced toposheets open up side-by-side.

Note: For scanning of topographical maps prior permission from the concerned authorities is needed.

7.8 SUMMARY

In this unit, you have learnt that:

- Data input is a method of getting data into a computer.
- There are different types of devices like digitising table, mouse, scanner and keyboard which are used for data input.
- Map digitisation is the process of converting features on a paper map or simply tracing a paper map into digital format with the help of digitising table or on-screen digitisation by displaying raster maps on a computer screen with the help of a GIS software. Digitised features can be in the form of points, lines or polygons.
- Metadata is the data about data and is prepared mostly in the form of XML during the time of data production. It forms the most integral part of a GIS and reveals basic information about data as to when it was created, by whom it was created, why it was created, where it was created, etc.
- A metadata standard is a document that contains rules or guidelines for standardising the content of metadata.
- Before digitisation, it is important to georeference map using a suitable projection system. Georeferencing is the process of locating any entity in real-world coordinates. Some other processes related to georeferencing are geocoding (adding coordinates to street locations), geotagging (adding geographical information to photographs, films, etc.) and rubbersheeting (geometric adjustment of coordinates with the help of a more accurate map). GIS software has to perform a number of graphical manipulations that results into different types of projections. They are translation, scale change and rotation.

Spend 30 mins

7.9 UNIT END QUESTIONS

1) What do you understand by map digitisation?
2) Briefly describe different methods of raster and vector data input.
3) Give a brief discussion of metadata in GIS.
4) Explain the concept of georeferencing.

7.10 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/

The data from all the above web pages was retrieved between 10th March, 2012 and 20th March, 2012.

7.11 FURTHER/SUGGESTED READING

- Anjireddy, M. (2008): *Textbook of Remote Sensing and Geographical Information Systems*, BS Publications, Hyderabad.

7.12 ANSWERS

Check Your Progress I

1) The names of different devices used for inputting geographical data are, mouse, digitising table and puck/cursor, scanners, keyboard, etc.

2) The different modes of digitisation are point mode and stream mode digitisation.

Check Your Progress II

1) 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.

Unit End Questions

1) Refer to section 7.3.

2) Refer to section 7.3.3

3) Refer to section 7.4.3

4) Refer to section 7.6.