UNIT 9  ATTRIBUTE DATA LINKAGES AND  DATA CONVERSION

Structure

9.1 Introduction  
Objectives

9.2 Linking Non-Spatial Data with Spatial Data

9.3 Data Conversion Methods  
GIS Data Conversion from Secondary Sources  
Raster to Vector Conversion  
Vector to Raster Conversion  
Format Conversion  
Coordinate and Projection Conversion  
Thematic Scale Conversion

9.4 Data Integration

9.5 Activity

9.6 Summary

9.7 Unit End Questions

9.8 References

9.9 Further/Suggested Reading

9.10 Answers

9.1 INTRODUCTION

You have been familiarised with spatial and non-spatial data structures, data input and data standards, and topological concepts of GIS in the previous Units 5, 6, 7 and 8 in Blocks 2 and 3 of MGY-003. As you know, there are two types of databases in GIS—spatial and non-spatial (attribute). Therefore, these databases are needed to be integrated together for analysis in GIS environment. You have also studied the spatial data standards in Unit 8 of MGY-003, which is an essential issue for data integration as the uniformity in data standards is required for their integration on the same platform for further visualisation, analysis and obtaining results. The basic component of GIS is the database, where non-spatial database is the description of the spatial database. If there is no non-spatial database, there will be no variation of space except locations. Thus, the non-spatial database is the inevitable part of the GIS database. In this unit you will learn how to link non-spatial (attribute) data with spatial data. We will also discuss data conversion, techniques of data conversion from raster to vector and vector to raster, along with types and methods for data integration.

Objectives

After studying this unit, you should be able to:

• define the data linkages in GIS;
• discuss methods for linking non-spatial data with spatial data;
• describe data conversions methods; and
• explain data integration types and processes.
GIS Database Creation

Spatial data is incomplete and meaningless without its description.

9.2 LINKING NON-SPATIAL DATA WITH SPATIAL DATA

In GIS sense, a data link connects data from different sets (e.g., non-spatial data with spatial data). For example, suppose you want to know what percentage of total food grains production of each state of India is grown for export. Let us assume that you have required data stored in your computer into two separate files in which one file containing data of total food grains production for each state and another is on food grains export. You must combine these two files to solve the problem. Once the files are combined then it becomes simple for computer to process. The production and export situation may be seen spatially on map once these data are linked to the state boundary map. Hence, data linkage is very useful to determine location, conditions, trends, patterns and do modelling.

Linking non-spatial (attribute) data with spatial data serves multiple purposes. Firstly, keeping the non-spatial data separate and joining for analysis and mapping allows the spatial data to be handled easily. Secondly, many tables containing non-spatial database may be joined together making the retrieval faster and maintaining the database redundancy free. Data joining of non-spatial data with spatial ones will be discussed in the following sub-sections. In GIS language, the logical linking of attribute or external data is called ‘relate’ and appending of attribute data is termed as ‘append or join’ (http://webhelp.esri.com). When the data is permanently joined (e.g., with the change in property), the data needs to be updated in the map itself. The temporary join is saved in terms of link only in the project file with various formats or terms used by various software like .apr and .mxd in ArcView old or new version, .wor in MapInfo, etc. In this case, with the change in property, whenever the data in table is updated, the map also gets updated.

The basis of data joining between and among various tables or between spatial and non-spatial data is the common identifier known as ‘primary key’ and ‘foreign key’. This identifier is unique in both the files like a unique code which is not repeatable anywhere in the file. Sometimes, with smaller database, names are taken as common identifier but in large database there may be several identical names with different identities. Thus, for each feature unique code is assigned to avoid this kind of confusion. For example, in spatial terms there may be two villages of same names in a block or taluka. If one wants to join the population data of all the villages with the village map, there will be an error in joining the data to the corresponding village of same name. In larger context, the same district names exist in two states in India, for example, Bilaspur exists in Himachal Pradesh and Chhattisgarh states, Hamirpur district exists in Himachal Pradesh and Uttar Pradesh states, Aurangabad district exists in Bihar and Maharashtra states, and Pratapgarh district exists in Uttar Pradesh and Rajasthan states. One example of this can be seen in Fig. 9.1. At lower level, like Block, Panchayat and Villages, there are large numbers of repetitions of names. In this case, data of one spatial unit (district, block, panchayat or village) may be linked to another one having same name. For instance, data of Aurangabad district of Bihar may go to the Aurangabad district of Maharashtra and vice-versa after joining, if there is no unique identity for each district in the spatial data (map) as well as non-spatial data (attribute or table). For this reason, unique identity is required for data joining. Example of different spatial units is given in Fig. 9.2.
Fig. 9.1: Map of India showing districts with same name “Bilaspur” in Himachal Pradesh and Chhattisgarh with different spatial locations
Fig 9.2: Non-spatial and spatial data joining based on unique identities

Unique ID in attribute table
This way, various small attribute tables containing specific details of the spatial features are joined with spatial features using the unique but common identities, where a common field like DIST_ID in above example is essential in all attribute tables for data joining as an identifier. This relates the various tables or attribute table (non-spatial data) with spatial data to the exact feature.

Check Your Progress I

1) What are non-spatial (attribute) data linkages? Describe the difference between data relate and join.

Data conversion is the transfer or translation of data from one form to other. The data conversion includes the conversion of spatial information into GIS format. The conversion may be from tabular information, textual information, survey data, drawing or sketch, remote sensing imagery, aerial photographs, GPS data, etc. (United Nations, 2000; Montgomery and Schuch 1993; Buckley, 1997). It also includes the conversion from analogue to digital, digital to analogue, shape to coverage or autocad drawing to shape, raster to vector and vector to raster. It may also be conversion from one format of GIS to other. You will learn it in detail in the following sub-sections.

There are various methods of data conversion in GIS starting from a very conventional method to a latest sophisticated and automated system. Data conversion may be classified into six categories:

- GIS data conversion from secondary sources
- raster to vector conversion (vectorisation)
- vector to raster conversion (rasterisation)
- format conversion
- coordinate and projection conversion
- thematic/conceptual/scale conversion

Let us discuss all these methods one by one.

9.3.1 GIS Data Conversion from Secondary Sources

There are sources which capture the data (spatial and attribute) but the data are not in the GIS format. This kind of data needs to be converted into GIS format for further use in GIS platform. It involves the following:

a) Conversion From Field Inventory

Some of the spatial data are still collected through conventional methods like plain table survey or chain/tape survey and recorded in the Field Measurement Books (FMBs) known by different terminologies in various places. The data conversion of information stored in the FMBs is converted into GIS database based on the distance and bearings. The example may be seen in Fig. 9.3.
GIS Database Creation

Fig. 9.3: Field measurement book and field drawing (source: National Land Records Modernisation Programme, Govt. of India)

The coordinate geometries (locations, distance and direction of a straight line, and radius, angle, arc length and direction of curved line) of geographical features recorded by a surveyor are translated into the computer in creating new features as GIS layers. In this process, point, line and polygon features are captured in the survey book (FMB) and thereafter GIS database. Thus, it is the conversion of conventional survey data into the GIS environment.

b) Data Conversion from Global Positioning System

In this case, the data conversion is done from transferring and manipulating the data captured by Global Positioning System (GPS)/ Differential Global Positioning System (DGPS) or Electronic Total Station Survey (ETS) into GIS environment following certain error correction methods. You should note here that the instruments like GPS/DGPS and ETS are the automated versions of manual survey and therefore data is directly captured in the storage devices in electronic form, that reduce time, effort and cost in translating these data into GIS format as these instruments have inbuilt capability to capture XY locations (latitude and longitude, distance, direction, etc.) along with height of the location. An illustration of ETS survey and reading is given in Fig. 9.4.
The data obtained from these instruments are converted in the GIS formats with the help of relevant software.

c) **Keyboard Entry**

The keyboard entry is mainly done while placing the coordinates, distance and bearings captured in the field survey. It is also helpful to enter the coordinate values, properties or identifiers.

Keyboard entry of database for GIS in terms of attribute data conversion is also an important method. In this case, some of the Optical Character Recognition (OCR) software help in automatic conversion with full verification. This is also helpful in data conversion in terms of computations.

### 9.3.2 Raster to Vector Conversion

Raster to vector data conversion or vectorisation converts the data from various formats into GIS or discrete vector formats, where every feature is a distinct feature selectable, editable and updatable. These functions are not possible with other than GIS formats. The conversion formats are briefly discussed in the following sub-sections.

a) **Scanning of Paper Maps and Drawings**: Scanning is a process of converting the paper maps or drawings into electronic form which can be saved in computer storage devices like hard disk, floppy disk, CD or DVD, magnetic tape, pen drive, etc.

b) **Map Digitising (Vectorisation)**: Map digitising is a conversion process where either GIS database is created from the paper maps or satellite imagery or aerial photographs. In this process, the conversion is done from paper maps/drawings and scanned digital copy of the maps or drawings. In other words, it is a process of tracing of maps digitally. The computer accessories which are used for digitisation have been discussed in Unit 7 of MGY-003. The digitisation is done in two ways, one is table top and another is heads-up or on-screen. In table top digitisation, the paper map is placed on the digitising table or pad and various features are traced with the digitising mouse. It is a manual process and becomes tedious at times.
The heads-up or on-screen digitisation involves the digitisation over the scanned paper map, imagery or drawing on the computer screen using simple mouse. The example of this is shown with the digitisation of scanned cadastral map in Fig. 9.5. It could also be a drawing, paper map or satellite imagery on screen. In this method, there are three options – manual, semi-automatic and automatic.

The manual digitising is the same as the table top digitising with a difference that one is done at digitiser table or pad and other is done on computer screen. In the semi-automatic mode, the software asks the operator to pick the lines or colours, direction of a line from a node and to trace the features. In the automatic mode, the software automatically traces or digitises the features over the scanned map or drawing. The scanned map or drawing file needs to be in binary format in the automatic mode of digitisation.

The advantage of automatic digitising is that it is very fast and done within few seconds after making the scanned binary image noise free. The disadvantage is that it does not maintain the accuracy as that of the manual or semi-automatic digitising.

c) Conversion from Remote Sensing Imagery: In this conversion method, the sources of GIS data are satellite imageries or aerial photographs of various resolutions. It is called image interpretation in case of GIS database creation from satellite imagery, and photogrammetry in case of database conversion from aerial photograph. Here also for coarse resolution or very small scale maps, automated database creation is done in case of satellite imagery by image classification. Depending on the photo quality of very high resolution satellite data like any other aerial photographs, the features of satellite data are traced out manually. The example of data conversion from satellite imagery is shown in Fig. 9.6 with the digitisation on very high resolution satellite imagery of the IGNOU main campus obtained from Google Earth.
9.3.3 Vector to Raster Data Conversion

Vector to Raster conversion or rasterisation is done for specific applications. In this process, the entire vector data (point, line or polygon) is converted into an array of cells with corresponding attribute data attached to it (Fig. 9.7). The data formats may be BMP, DAT, GIF, IMG, JPEG, JPEG 2000, PNG, TIFF, etc.

![Fig. 9.7: Vector to raster data conversion; (a) vector data and (b) raster data](source: www.google.com/earth)

9.3.4 Format Conversion

It is an important data conversion function where data is converted from one format to other as per the requirement in terms of compatibility, software availability and application. For example, the vector formats like AutoCad Map drawing file (.dwg) is converted into Arcview shape file (.shp) for the usage with number of other software and applications which are compatible with shape file. It can be easily integrated in the basic programmed software. Similarly, data from one raster format to others are converted based on the compatibility needs. For instance, data of BIL (Band Interleaved by Line) format is converted into .img format to work with ERDAS Imagine or ArcGIS software. A data with .img format may be converted to GRID format for spatial analysis with raster data. Similarly, data from .img format to .jpg format for easy opening and usage, which can be opened with any commonly available image viewing software.

9.3.5 Coordinate and Projection Conversion

It is one of the most important conversion of GIS database for data integration into single platform where all the database need to be converted into same coordinate and projection system. Data created with different projection systems need to be converted into a single projection system. For example, if one has to
work on UTM WGS84 projection system and the data collected are in polyconic or other projection systems, the prerequisite is to bring the data of other projection systems into the UTM WGS84 for seamless integration.

### 9.3.6 Thematic Scale Conversion
Thematic or conceptual conversions are done to eliminate the conceptual anomaly in the two sets of database. The best examples are different land use or soil classes in two datasets prepared based on different scales, which are made identical by aggregating or disaggregating classes for integration and analysis.

#### Check Your Progress II

1) What is data conversion in GIS? Differentiate between vectorisation and rasterisation.

---

### 9.4 DATA INTEGRATION
Data integration as a process of making different data sets compatible with each other, so that they can reasonably be displayed on the same map and their relationships can sensibly be analysed (Flowerdew, 1991). Data integration helps to bring all the data together at the same platform to answer variety of questions related to spatial data. In Unit 8, you have studied that GIS data standardisation is needed for the seamless integration of various information layers or themes to get the real picture of any situation or place to make better decisions based on the spatial relationships, connections and patterns, and attribute properties. This is not possible without integration of various data sets.

The data creation standards are still not uniform all across the globe. As a consequence, data is created in different standards like coverage or boundaries, time period, data structure, different formats, types, accuracy levels of various kinds, etc. as per the needs. It is created from various sources and by using various methods.

In data integration, data from different sources or of different standards are standardised to make compatible to place one over other with matching locations and boundaries for spatial analysis and mapping. The major conversion in this case is done for coordinate and projection system and scale. The other standards are conceptual or logical. For example, the other standards like the land use data of one part of the region is done for three level classification and the other part is level four. In this case, both the parts need to be brought into the same level by either expanding or reducing one class for uniformity to integrate and work.

The following types of integration may be done:

**i) Horizontal Integration:** It is a process to merge the data of all the adjacent areas seamlessly. This is done when the area of interest is digitised in two or
several contiguous spatial sub-units. For example, topographical map-wise
digitisation of resources for a district or state are required to be mosaiced finally
as a single unit. In this process, the adjacent sheets are brought together and
matched for point to point locations doing edge matching with various techniques.
For example, Fig. 9.8 shows horizontal integration in a simple way where each
state digitised separately in a mosaiced form to make the map of India.

![Fig. 9.8: Horizontal integration of adjacent sheets/maps to prepare one single map](image)

**ii) Vertical Integration:** It is a process to keep all the databases of
common area one over other with matching all points or locations in all the
data layers (Fig. 9.9). It is superimposition on computer like keeping various
maps on tracing sheets/transparencies of same area layer-by-layer.

![Fig. 9.9: Vertical integration of various thematic layers (source: http://webhelp.esri.com)](image)

The other vertical integration is the conceptual or thematic adjustment for
identical entities in all the layers if the data are of different time period. For
example, making identical land use classes or soil classes for all the years for
integration and studies or keeping the same administrative boundaries of an area
for temporal studies as the datasets are often created at different scales and themes for various requirements over the time.

You should note that the above horizontal and vertical data integration is done for the attribute data also.

iii) **Spatial and Non-Spatial Data Integration:** It is also referred to data linkages, which has already been dealt with in sub-section 9.2 of this Unit. Attribute data attachment with spatial data is also a data integration, which makes data compatible as per the spatial units where the number of features will be equal to the number of rows in any table.

While doing integration there are some conversion principles that should be taken into account which are considered step-wise as listed below:

- the conversion of data from analogue to digital form
- all the data are converted into the same data format like .shp of arcview before integrating
- coverage or boundaries and locations are matched for the integration, if the data are of non-matching pairs
- temporal matching is also taken care of before integration. For example, a data is of 1991 and other data of 2001, both need to be of same day and month based on their nature and applications
- standardisation of projection and reference system, which is prerequisite for all datasets to be integrated together. If the various data layers of same area are prepared in different projections and reference systems, the seamless integration is not possible due to different transformation models, and
- scale and accuracy of various data layers are also required to be standardised for integration.

Similarly, the standardisation is also needed in the attribute data of any administrative or spatial unit with the change of the administrative unit or boundary of the spatial unit. For example, if a district is divided into two districts, the attribute data like population also needs to be divided accordingly before integration.

**Check Your Progress III**

1) What do you mean by data integration? Explain data integration with suitable examples.
9.5 ACTIVITY

1) You have already studied about data linking concepts and methods; therefore try to create a small example with few geographical features associated with non-spatial data and try to link them in both modes - permanent and temporary. Also try for one to one, one to many, many to one and many to many data joining.

2) Try for various spatial data conversion methods mentioned above like GIS data conversion from secondary sources, data conversion of field survey data (from field inventory book or field measurement book assigning coordinates and bearing; and keyboard entry, scanning of paper maps and drawings, raster to vector conversion (vectorisation), map digitising (vectorisation) from remote sensing imagery, vector to raster data conversion (rasterisation), format conversion and coordinate and projection conversion from one to other.

3) Try to integrate the data acquired from various sources or various layers prepared by you with various standards using GIS software and make a seamless dataset for analysis and mapping.

9.6 SUMMARY

In this unit, you have learnt about:

- Linking of non-spatial data with spatial data and types of data linkage.
- The process and methods of data conversion such as GIS data conversion from secondary sources, raster to vector conversion (vectorisation), vector to raster conversion (rasterisation), format conversion, coordinate and projection conversion and thematic/conceptual/scale conversion.
- Spatial data integration, which is an important part of GIS, without which full functionality and capability of GIS cannot be explored. The data integration in brief is bringing all the data from various sources and periods together at the same platform for viewing with spatial relations, analysis and mapping.

9.7 UNIT END QUESTIONS

1) What is data linkage? Discuss the types of data linkages.

2) What do you mean by data conversion? Explain various kinds of data conversion.

3) What is data integration? Why is it important in GIS? Explain the types and processes of data integration.

9.8 REFERENCES

GIS Database Creation


- www.google.com/earth


Data from above-mentioned websites was retrieved between 29th July 2011 and 5th July 2012.

9.9 FURTHER/SUGGESTED READING


3. www.webhelp.esri.com

9.10 ANSWERS

Check Your Progress I

1) Non-spatial data linkage is relating or joining related attribute tables with maps. The difference between data relate and join is the temporary joining of attribute data is called relate and permanent joining of attribute data is termed as append or join.

Check Your Progress II

1) Data conversion is a process of transformation of spatial and non-spatial data from one form to other. There are data conversions from drawing and textual/tabular form, paper maps, satellite imagery and aerial photographs, etc. The other conversions are vectorisation and rasterisation of GIS database. Vectorisation is a process where spatial data is converted into vector GIS format from various forms. Rasterisation is a process to convert vector GIS data into array of cells representing the spatial features.
Check Your Progress III

1) Data integration is a process of bringing all the data sets from various sources into one platform. The examples are bringing the spatial data of geology, geomorphology, land use/land cover, soil, drainage, water bodies, mines and minerals, forest cover, watershed, ground water prospects map, cadastral boundaries, administrative boundaries, transport network, health infrastructure, education infrastructure, development/backwardness maps, planning maps, etc. together for micro-level or cadastral level planning and various applications.

Unit End Questions

1) Refer to section 9.2.

2) Refer to section 9.3

3) Refer to section 9.4