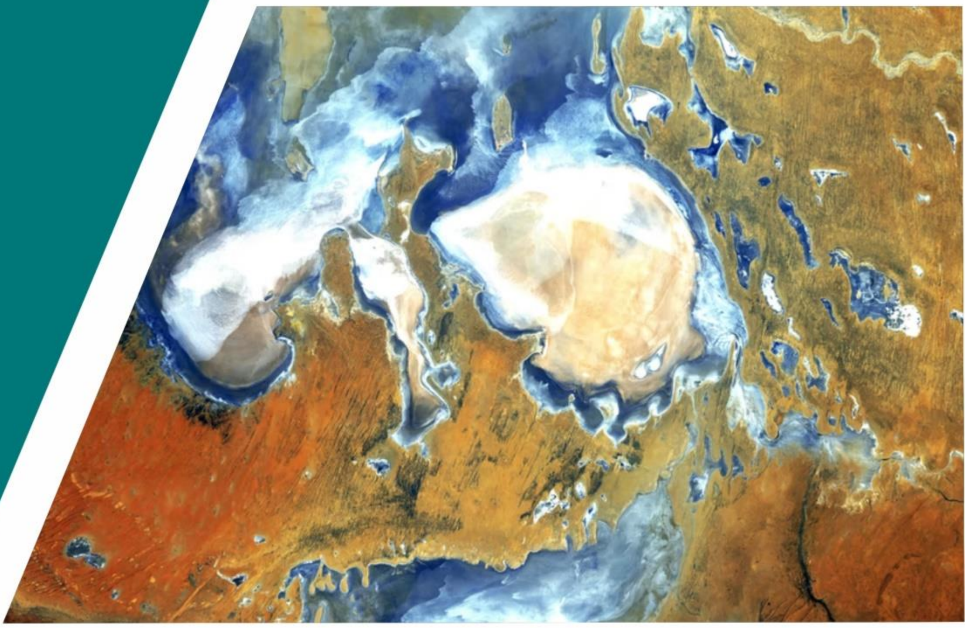
NATIONAL OPEN UNIVERSITY OF NIGERIA

ESM 407



Geographic Information

Module 3

ESM 407 Geographic Information Systems Module 3

Course Developer/Writer
Dr. N. O. Uluocha, University of Lagos

Course Coordinator
Mrs. Cecilia Medupin, National Open University of Nigeria

Programme Leader
Prof. K. T. Obidairo, National Open University of Nigeria

Credits of cover-photo: Henry Ude, National Open University of Nigeria

National Open University of Nigeria - University Village, 91 Cadastral Zone, Nnamdi Azikiwe Express Way, Jabi Abuja, Nigeria

www.nou.edu.ng centralinfo@nou.edu.ng
oer.nou.edu.ng oerunit@noun.edu.ng OER repository
Published in 2021, by the National Open University of Nigeria

How to re-use and attribute this content

Module 3 Functions of GIS

Unit 1 Data Input

1.0 Introduction

Data input is a critical aspect of GIS operations. The quality of the output data is largely dependent on the quality of the underlying database. Creating a GIS database could be quite demanding; careful and rigorous planning and execution is usually required. In this unit, we will focus on such data input operations as geographical referencing, digital data conversion/data capture, data checking/editing, and data integration.

2.0 Objectives

At the end of this unit, you should be able to:

- discuss the concept of geo-referencing
- examine the processes of spatial and attribute data input
- analyse the issues of data checking and editing
- discuss the issue of data integration.

3.0 Main Content

3.1 Projection

Projection is a key component of map making. A projection is a mathematical means of transferring information from a model of the earth, which represents a three-dimensional curved surface, to a two-dimensional medium - paper or a computer screen. Different projections are used for different types of maps because each projection particularly suits specific uses. For example, a projection that accurately represents the shapes of the continents will distort their relative sizes.

Since much of the information in a GIS comes from existing maps, a GIS uses the processing power of the computer to transform digital information, gathered from sources with different projections and/or different coordinate systems, to a common projection and coordinate system.

3.2 Geo-Referencing

The fundamental idea behind the creation of any GIS database is the fact that every object present on the earth can be geo-referenced. Before using any data in a GIS environment, the data should be geo-referenced (Uluocha, 2007). Geo-referencing (also known as geo-rectification, geolocating, geocoding or registering) is the process of assigning spatial location identity to pieces of information. In other words, it is the process of giving a cartographic material such as a digital satellite imagery, aerial photograph, map, or statistical data to a real world coordinate system and map projection. Or, as Kasianchuk and Taggart

(2004) simply put it, “Geo-referencing is the process of establishing a relation between the data displayed in your GIS software and its real-world location.”

Geo-referencing enables us to know exactly where things are positioned on or in relation to the earth’s surface. The geo-referencing process is normally used to relate or link cartographic data to the specific portions on the earth’s surface that they represent or pertain to. Besides, geo-referencing enables one to accurately measure distances, directions, sizes (areas) and shapes of features on cartographic base material. In a GIS environment, unless a piece of data is geo-referenced, it would be impossible to undertake certain spatial analysis operations using the data. Geo-referencing is commonly achieved by using a coordinate system.

There are various spatial referencing systems in use, some of which are rather crude and simple while some are sophisticated and complex. Nevertheless, it has become somewhat customary to classify geo-referencing systems into two broad groups namely, coordinate systems and non-coordinate systems. Examples of coordinate systems are the spherical (geographic) coordinate system and the rectangular coordinate system.

Non-coordinate systems include Postal Addresses and Postal Codes (or ZIP codes in USA), telephone codes, place names, Enumeration Areas (EAs), House Numberings or Street Addresses, etc. Some common geo-coding systems are shown in Table 3.1. It should be noted, though, that some of the non-coordinate systems, e.g. telephone area codes and postal zip codes, exhibit only rudimentary metric properties and do not give information about direction or size of objects (Fabiyi, 2001, p. 62).

For large areas such as states, countries, regions, or continents, the spherical (or geographic) grid coordinate system of latitude and longitude is more useful for geo-referencing. Conversely, the plane rectangular grid coordinate system, which makes use of *x,y* coordinates (or Easting’s and Northing’s) is more suited to geo-referencing small areas like a school compound, census enumeration area, electoral district, village, ward and township.

The spherical coordinate system is composed of a network of infinite number of latitudes and longitudes. The latitudes and longitudes are usually numbered or identified with angular values in degrees, minutes and seconds, e.g. 4°23’14”N, 15°07’25”E. The point of intersection between the Greenwich Meridian and the Equator forms the origin (0) in the spherical coordinate system. In the spherical grid system, the value of latitude is usually given before that of longitude.

Table 3.1: Some Commonly Used Systems of Geo-Referencing

System	Domain of uniqueness	Metric?	Example	Spatial resolution
Place-name	Varies	No	Abuja, Ekenobizi, Ghana	Varies by feature type
Postal address	Global	No, but ordered along streets in most	21, Abayomi Street, Akoka,	Size of one mailbox

		countries	Lagos, Nigeria	
Postal code	Country	No	101017 (University of Lagos, Akoka, Nigeria)	Area occupied by a defined number of mailboxes
Telephone calling area	Country	No	234 (Nigeria)	Varies
Cadastral system	Local authority	Yes	10m x 30m (Dimensions of a land parcel)	Area occupied by a single parcel of land
Public Land Survey System	Western USA only, unique to Prime Meridian	Yes	Sec 5, Township 6E, Range 4N	Defined by level of subdivision
Latitude/ longitude	Global	Yes	6°23'15"N, 10°18'42"E	Infinitely fine
Universal Transverse Mercator	Zones six degrees of longitude wide, and N or S hemisphere	Yes	542500E, 327638N	Infinitely fine
State Plane Coordinates	Unique to state and to zone within state	Yes	55086.34E, 75210.76N	Infinitely fine

Source: Modified from Longley, *et al.* (2001)

3.3 Spatial Data Capture

How can a GIS use the information in a map? If the spatial or cartographic data to be used are not already in digital form, that is, in a form the computer can recognize, various techniques can capture the information. Data capture or conversion is the technical process of entering or putting information into the computer system. Data capture involves identifying the objects on the map, their absolute location on the earth's surface, and their spatial relationships. This process consumes much of the time of GIS practitioners.

Nevertheless, software tools that automatically extract features from satellite images or aerial photographs are now gradually replacing what has traditionally been a time-consuming capture process. There are a variety of methods used to enter spatial data into a GIS where it is stored in a digital format.

Existing spatial data printed on paper or PET film maps can be digitized or scanned to produce digital data. As earlier noted, a digitizer produces vector data as an operator traces points, lines, and polygon (areal) boundaries from a map. Maps can be digitized by hand-tracing with a computer mouse on the screen or on a digitizing tablet to collect the coordinates of features.

Modern GIS technologies use digital information, for which various digitised data creation methods are used. The most common method of data creation is digitisation, where a hard copy map, survey plan or chart is transferred into a digital medium through the use of a computer-aided design (CAD) program, and geo-referencing capabilities. With the wide availability of ortho-rectified imagery (both from satellite and aerial sources), heads-up digitizing is becoming the main avenue through which geographic data is extracted. Heads-up digitizing involves the tracing of geographic data directly on top of the aerial imagery instead of by the traditional method of tracing the geographic form on a separate digitizing tablet (heads-down digitizing).

Digitization or conversion of existing paper based records, plans, maps and charts to digital using any of the three established and tested methods:

- using the digitizing tablet manually
- using the semi-automatic raster chasing method
- using the bundle (fully automatic) method.

Electronic scanners can also convert maps to digits. Scanning a map results in raster data, which could be further processed to produce vector data through a process known as vectorization. Table 3.2 shows the comparative advantages and disadvantages of the manual digitizing process to the automatic scanning technique.

Table 3.2: Digitizing and Scanning Techniques Compared

Manual Digitizing	Automatic Scanning
▪ A time-consuming procedure.	▪ Less time-consuming
▪ The spatial (map) data is recorded in vector format.	▪ The spatial data is recorded in raster (grid cell) format.
▪ Can be used to selectively capture map data (the operator digitises only the required features). This reduces the amount of time spent on cleaning and editing the data.	▪ Automatically captures every feature on the source document (e.g. map, aerial photograph, orthophoto map, satellite imagery). This creates additional editing problem.
▪ The procedure requires a lot of human input (labour-intensive).	▪ Requires less human input.
▪ The captured line work often has a high resolution, hence suitable for map production.	▪ Resolutions of line works are not often high, hence not quite suitable for map production.
▪ The source material to be digitised can easily be geo-referenced.	▪ The process of geo-referencing source material is usually extensive.
▪ Suitable for small mapping projects, which involve very few map sheets.	▪ Suitable for very large mapping and geographical analysis projects requiring the digital conversion of several map sheets, aerial photos or satellite imagery.

Source*:* Uluocha (2007)

Survey data can also be directly entered into a GIS from digital data collection systems on survey instruments using a technique called Coordinate Geometry (COGO). Positions from a Global Navigation Satellite System (GNSS) like Global Positioning System (GPS), another survey tool, can also be directly entered into a GIS. Coordinates from GPS receivers can be uploaded into a GIS. Current trend involves data collection and field mapping being carried out directly with field computers (position from GPS and/or laser rangefinder). New technologies allow for the creation of maps as well as data analysis directly in the field; this makes mapping projects efficient and accurate.

Remotely sensed data also plays an important role in spatial data collection. This consists of sensors attached to a platform such as an aircraft or spacecraft (satellite). Sensors include cameras, digital scanners and LIDAR. Here, satellites use different sensor packages to passively measure the reflectance from parts of the electromagnetic spectrum or radio waves that were sent out from an active sensor such as radar. Remote sensing collects raster data that can be further processed using different bands to identify objects and classes of interest, such as land cover.

The majority of digital spatial data currently comes from photo interpretation of aerial photographs. Soft copy workstations are used to digitise features directly from stereo pairs of digital photographs. These systems allow data to be captured in two and three dimensions, with elevations measured directly from a stereo pair using principles of photogrammetric. Currently, analog aerial photos are scanned before being entered into a soft copy system, but as high quality digital cameras become cheaper this step will be skipped.

Data restructuring can be performed by a GIS to convert data into different formats. Since digital data is collected and stored in various ways, the data sources may not be entirely compatible. Some of the data may be in vector format while some may be in raster format. So a GIS must be able to convert geographic data from one structure to another.

For example, a GIS may be used to convert a satellite image (raster) map to a vector structure by generating lines around all cells with the same classification, while determining the cell spatial relationships, such as adjacency or inclusion; this process is known as raster-to-vector conversion.

3.4 Attribute Data Capture

In addition to collecting and entering spatial data, attribute data is also entered into a GIS. For vector data, this includes additional information about the map objects represented in the system. A typical attribute data consists of statistical facts and figures which are usually presented in tabular form. Hence, the keyboard is the device normally used to put attribute data into the computer. If the data already exists as an electronic file, for example as a spreadsheet, it can be simply downloaded into the GIS.

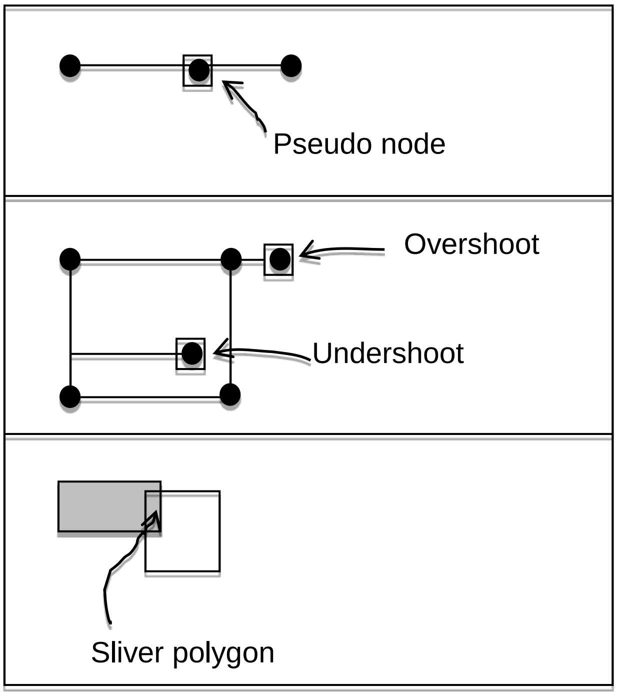
3.5 Checking and Editing

The data capture process is never error-free. Hence, after capturing the geographical data or keying in the statistical (attribute) data into a GIS, the data usually requires checking and editing, to identify and remove any errors, or further processing. For vector data it must be made "topologically correct" before it can be used for some advanced analysis.

For example, in a road network, lines must connect with nodes at an intersection. Errors such as undershoots and overshoots must also be removed. For scanned maps, blemishes on the source map may need to be removed from the resulting raster. For example, a fleck of dirt might connect two lines that should not be connected.

The possible errors that could occur in a digitised map include (Figure 3.1):

- Pseudo nodes (unwanted nodes)
- Overshoots and undershoots (unwanted dangling arcs/nodes)
- Sliver polygons (unwanted overlapping polygons).



Φιγ. 3.1: Χομμον Ερρορσ ιν Διγιτισεδ Μαπ Δατα

The possible attribute data entry errors include the following:

- spelling errors
- entering of wrong digits (numerical figures)
- wrong field naming (e.g. designating a field as "character" instead of "numeric", and vice versa)
- too long or too short field width (size)
- omission of some data items
- inclusion of unwanted data items.

3.6 Data Integration

A GIS makes it possible to link, or integrate, information that is difficult to associate through any other means. In other words, GIS is effectively used to relate information from different sources.

Thus, a GIS can use combinations of mapped variables to build and analyse new variables (Figure 3.2). For example, using GIS technology, it is possible to combine agricultural records with hydrographic data to determine which streams will carry certain levels of fertilizer runoff. Agricultural records can indicate how much pesticide has been applied to a parcel of land. By locating these parcels and intersecting them with streams, the GIS can be used to predict the amount of nutrient runoff in each stream. Then as streams converge, the total loads can be calculated downstream where the stream enters a lake.

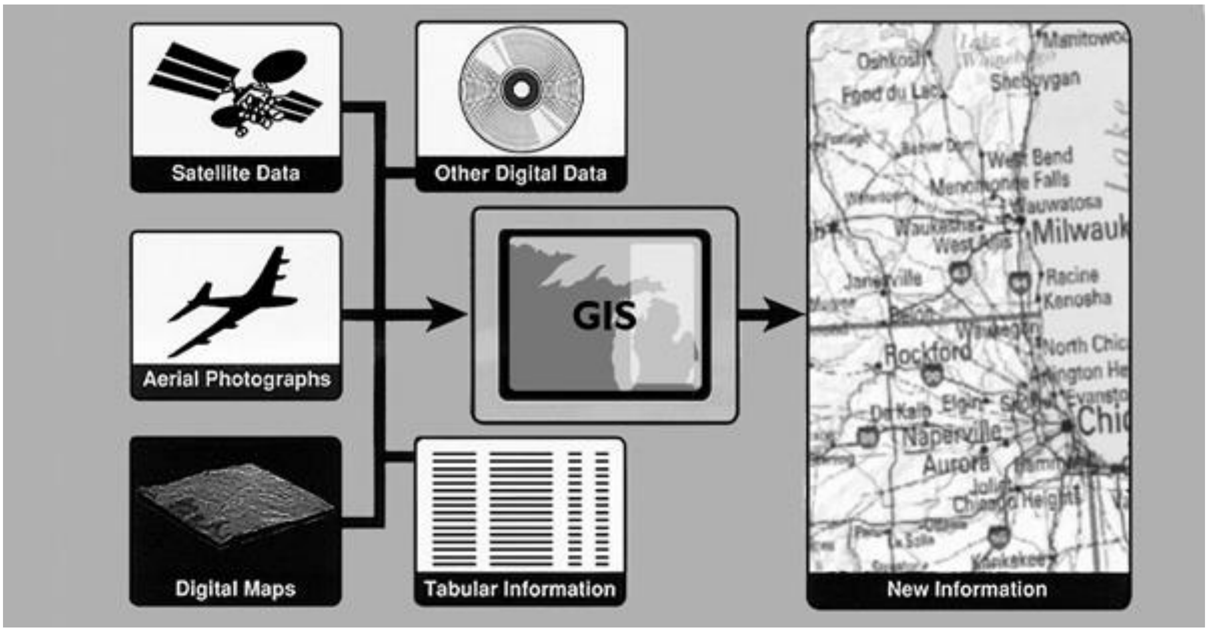


Fig. 3.2: Data Integration: the linking of information in different forms through a GIS

The power of a GIS comes from the ability to relate different information in a spatial context and to reach a conclusion about this relationship. Most of the information we have about our world contains a location reference, placing that information at some point on the globe.

For instance, when rainfall information is collected, it is important to know where the rainfall is located. This is done by using a location reference system, such as longitude and latitude, and perhaps elevation. Comparing the rainfall information with other information, such as the location of marshes across the landscape, may show that certain marshes receive little rainfall. This fact may indicate that these marshes are likely to dry up, and this inference can help us make the most appropriate decisions about how humans should interact with the marsh. A GIS, therefore, can reveal important new information that leads to better decision-making.

4.0 Conclusion

Data input is a major function of GIS; it is one of the most critical and cumbersome aspects of the system. Creating a robust, reliable and comprehensive database is crucial to the success of GIS operations. The quality of the outcome obtainable from GIS cannot be different from the quality of the underlying database – it is still a case of "garbage in, garbage out". Thus, in any GIS project the data input operation should be properly planned and meticulously executed.

5.0 Summary

Data is the raw material that GIS processes to produce the much desired information. But before this data can be processed it has to be entered into the system through a process known as data input. The spatial data, which could be in form of a map, aerial photo or satellite image, has to be captured into the system.

If the data source is in analogue form (hardcopy) it can be digitised directly into the system or scanned before digitising it on-screen. The relevant attribute or non-spatial data are usually entered into the system via the computer keyboard. After inputting the data it has to be cleaned up by carefully checking for and correcting any possible errors. The edited data is then stored in any suitable electronic storage medium, for further use.

6.0 Self-Assessment Exercise

1. Critically examine the processes of spatial and attribute data input.
2. Analyse the issue of data integration.
3. Discuss the concept of geo-referencing.

7.0 References/Further Reading

Fabiyi, S. (2001). *Geographic Information Systems: Techniques and Methods*. Ibadan: Research Support Services.

Kasianchuk, P. & Taggart, M. (2004). *Introduction to Arc GIS I*. Redlands: Environmental Systems Research Institute (ESRI).

Longley *et al* (2001). *Geographic Information Systems and Science*. New York: John Wiley & Sons Inc.

Uluocha, N. O. (2007). *Elements of Geographic Information Systems*. Lagos: Sam Iroanusi Publications.

Unit 2 Data Storage

1.0 Introduction

Perhaps data maintenance in a GIS environment starts with having in place a good data storage system. Once a database has been created it needs to be properly stored for safe-keeping and easy access. There are various digital data storage devices available today. In this unit, we will take a quick look not only at the available devices but also the peculiar requirements and qualities of devices for the storage and handling of geospatial data.

2.0 Objectives

At the end of this unit, you should be able to:

- identify the various electronic data storage devices used in GIS
- discuss the qualities of good storage devices.

3.0 Main Content

3.1 Data Storage

It is not just enough to digitally compile data; once compiled, the digital map files and the related attributes data files in the GIS should be stored on magnetic or other digital media. In a GIS environment, data storage is based on a *Generic Data Model* that is used to *convert map data into a digital form*. As already identified, the two most common types of spatial data models are *Raster* and *Vector*. Both types are used to simplify the data shown on a map into a more basic form that can be easily and efficiently stored in the computer. On the other hand, the tabular *Relational* data model is commonly used to store attribute data.

It is instructive to note that the particular model – raster or vector - used to store spatial data matters a whole lot. As we have already discussed, comparatively speaking, each of the models has its own merits and demerits. Hence, in deciding on which model to choose for data storage, the intended application of the database and the expected output (end product) should be taken into consideration. Moreover, it must be borne in mind that certain operations are more efficiently executed using one type of data model than the other.

3.2 Storage Devices

Various data storage devices exist for GIS configurations. These devices are commercially available in varying physical dimensions and storage densities. Conventionally, the magnetic tape and the optical disk are the two types of storage devices used in GIS. However, not too long ago the Zip Drive, compact disk (CD) and flash drive joined the family of computer storage media. Diskettes, which were once in vogue, are hardly in use nowadays.

Usually, large storage capacities are required for GIS applications. This should be so because GIS databases apart from being traditionally large often include graphic data, which normally make high demand on computer storage space (Uluocha, 2007). Presently, optical disks with a capacity of several terabytes exist. Such very high storage capacity media can conveniently be used to handle large geographic databases. It should also be noted that for efficient and effective GIS operations, a storage device with an efficient read/write mechanism, hence a fast input/output (I/O) rate, is most desirable.

Ordinarily, owing to the large volume of a typical geographic database coupled with the graphic component, it took a while to retrieve and view data in a GIS environment. However "with more efficient read/write mechanisms, higher capacity I/O channels, and intelligent disk controlling devices" (Croswell and Stephen, 1988), it is now a lot faster to retrieve, view, query and manipulate geographical databases.

3.3 Qualities of a Good Storage Device

Owing to the fact that geospatial data are characteristically voluminous coupled with the peculiar nature of some GIS operations, a storage device that meets certain qualities is usually desirable. Fast access rate to data, which allows for real-time processing. For a storage device to be considered good enough for data storage in a GIS environment, it should have the following qualities:

- Very high storage density, which can conveniently support the often large volume of geographic databases.
- Efficient read/write mechanism
- Fast input/output (I/O) rate
- Cost effective
- Durable
- Less prone to virus attack.

4.0 Conclusion

Once the data have been digitally compiled, digital map files and the associated attributes data files in the GIS are stored on magnetic or other digital media. For smooth operations and to obtain good results, appropriate data model and robust devices must be chosen, for data storage.

5.0 Summary

Data storage is an important aspect of every GIS operation. Both the spatial and non-spatial data captured must be properly stored. In storing the data files, an appropriate data storage model should be used; for spatial data this could either be raster model or vector model, while the relational structure is mostly used for non-spatial data. There are many electronic data storage devices. However, given the oft voluminous nature of the data used in GIS, a storage device that is robust, efficient, durable, resilient, and with very high storage density is most preferable.

6.0 Self-Assessment Exercise

Discuss the qualities of good storage devices.

7.0 References/Further Reading

Croswell, P. L. & Clark, S. R. (1988). *Trends in Automated Mapping and Geographic nformation System Hardware, Photogrammetric Engineering and Remote Sensing*. Vol. 54, No. 11, pp. 1571-1576.

Uluocha, N. O. (2007). *Elements of Geographic Information Systems*. Lagos: Sam Iroanusi Publications.

Unit 3 Data Manipulation and Analysis

1.0 Introduction

Spatial analysis is one of the major functions a GIS performs. There is a vast range of spatial analysis techniques that have been developed over the past half century. The subject of spatial analysis is a rapidly changing field, and GIS packages are increasingly including analytical tools as standard built-in facilities or as optional toolsets, add-ins or 'analysts'.

In many instances such facilities are provided by the original software suppliers (commercial vendors or collaborative noncommercial development teams), whilst in other cases facilities have been developed and are provided by third parties. Furthermore, many products offer software development kits (SDKs), programming languages and language support, scripting facilities and/or special interfaces for developing one's own analytical tools or variants. In this unit, we will look at some of the spatial analytical operations that can be carried out using GIS.

2.0 Objectives

At the end of this unit, you should be able to:

- identify some of the major data manipulation and analysis operations carried out in GIS
- discuss in detail some of the geographical analysis procedures.

3.0 Main Content

3.1 Data Manipulation/Analysis Operation

Although the data input is, in general, the most time consuming part, it is for data analysis that GIS is used. What distinguish the GIS system from other information system are its spatial analysis functions. The heart of GIS is the analytical capabilities of the system. The analysis functions use the spatial and non-spatial attributes in the database to answer questions about the real world.

Geographic analysis facilitates the study of real-world processes by developing and applying models. Such models illuminate the underlying trends in geographic data and thus make new information available. The objective of geographical analysis is to transform data into useful information to satisfy the requirements or objectives of decision-makers at all levels in terms of detail. Results of geographical analysis are mostly communicated with the help of maps, and/or graphs (charts).

GIS offers the user several data manipulation and analysis options. The facilities available in GIS for data processing functions are known as "*Toolkits*." A toolkit is a set of generic functions that GIS user can utilize to manipulate and analyze geographic and attribute data. Toolkits provide processing functions such as data retrieval, query, measuring area and perimeter, overlaying maps, performing map algebra, and reclassifying map data. Data manipulation tools include:

- coordinate change (for changing from one geographical coordinate system to another)
- projections (for changing from one map projection to another)
- rescaling (for changing map scale)
- edge matching (or rubber sheeting), which allows a GIS to reconcile irregularities between map layers or adjacent map sheets called Tiles.

Similarly, GIS is usually equipped with a number of analytical tools for conducting various kind of geographical analysis. Among the broad range of major geographical analysis procedures in GIS are:

- Database query
- Map overlay
- Proximity analysis
- Network analysis
- Digital Terrain Modeling (DTM)
- Statistical and Tabular Analysis.

3.2 Some Geographical Analysis Procedures

This subsection looks briefly at some of the geographical analysis functions carried out using GIS.

3.2.1 Slope and Aspect

Slope, aspect and surface curvature in terrain analysis are all derived from neighbourhood operations using elevation values of a cell's adjacent neighbours. There are various techniques for calculating slope and aspect. Slope is a function of resolution, and the spatial resolution used to calculate slope and aspect should always be specified.

3.2.2 Data Modelling

A GIS, however, can be used to depict two - and three-dimensional characteristics of the earth's surface, subsurface, and atmosphere from information points. For example, a GIS can quickly generate a map with isopleth or contour lines that indicate differing amounts of rainfall. Such a map can be thought of as a rainfall contour map.

Many sophisticated methods can estimate the characteristics of surfaces from a limited number of point measurements. A two-dimensional contour map created from the surface modelling of rainfall point measurements may be overlaid and analyzed with any other map in a GIS covering the same area. For example, with a GIS one can easily relate wetlands maps to rainfall amounts recorded at different points such as airports, television stations, and high schools.

Additionally, from a series of three-dimensional points, or digital elevation model, isopleth lines representing elevation contours can be generated, along with slope analysis, shaded relief, and other elevation products. Watersheds can be easily defined and delineated for any given reach, by computing all of the areas contiguous and uphill from any given point of interest.

3.2.3 Topological Modelling

A GIS can recognize and analyze the spatial relationships that exist within digitally stored spatial data. These topological relationships allow complex spatial modelling and analysis to be performed. Topological relationships between geometric entities traditionally include adjacency (what adjoins what), containment (what encloses what), and proximity (how close something is to something else).

3.2.4 Network Analysis

The GIS can be used to undertake various network analyses. With the GIS, for instance, one can study the network density, network characteristics, network behaviour, and network function. The flow of materials and energy in a network can be modelled using GIS.

Similarly, the potential impacts of a given network can equally be examined using GIS. For instance, if all the factories near a wetland were accidentally to release chemicals into the river at the same time, how long would it take for a damaging amount of pollutant to enter a recipient wetland reserve? A GIS can simulate the routing of materials along a linear network. Values such as slope, speed limit, or pipe diameter can be incorporated into network modelling to represent the flow of the phenomenon more accurately.

Network modelling is commonly employed in transportation planning, hydrology modelling, and infrastructure (utility) modelling.

3.2.5 Hydrological Modelling

GIS hydrological models can provide a spatial element that other hydrological models lack, with the analysis of variables such as slope, aspect and watershed or catchment area. Terrain analysis is fundamental to hydrology, since water always flows down a slope. As basic terrain analysis of a Digital Elevation Model (DEM) involves calculation of slope and aspect, DEMs are very useful for hydrological analysis. Slope and aspect can then be used to determine direction of surface runoff, and hence flow accumulation for the formation of streams, rivers and lakes. Areas of divergent flow can also give a clear indication of the boundaries of a catchment.

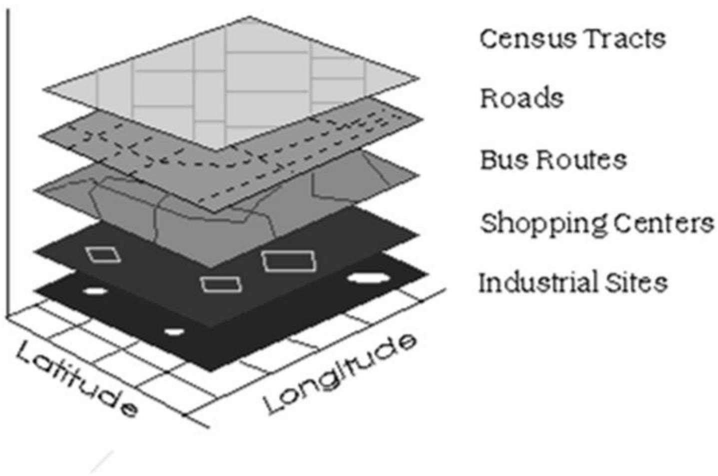
Once a flow direction and accumulation matrix has been created, queries can be performed that show contributing or dispersal areas at a certain point. More detail can be added to the model, such as terrain roughness, vegetation types and soil types, which can influence infiltration and evapotranspiration rates, and hence influencing surface flow. These extra layers of detail ensure a more accurate model.

3.2.6 Cartographic Modelling

The term "cartographic modelling" refers to a process where several thematic layers of the same area are produced, processed, and analyzed to obtain a composite map. The map overlay (or simply overlay) method is generally used to achieve this.

Map overlay (Figure 3.3) involves the combination of several vector spatial datasets (points, lines or polygons) to create a new output vector dataset, visually similar to stacking several maps of the same region. These overlays are similar to mathematical *Venn diagram* overlays.

A *union* overlay combines the geographic features and attribute tables of both inputs into a single new output. An *intersect* overlay defines the area where both inputs overlap and retains a set of attribute fields for each. A *symmetric difference* overlay defines an output area that includes the total area of both inputs except for the overlapping area.



Φιγ. 3.3: Τηε Χονχεπτ οφ Μαπ Οϖερλαψ

3.2.7 Geo-Statistics

Geostatistics is a point-pattern analysis that produces field predictions from sample data points. It is a way of looking at the statistical properties of those special data. It is different from general applications of statistics because it employs the use of graph theory and matrix algebra to reduce the number of parameters in the data.

When phenomena are measured, the observation methods dictate the accuracy of any subsequent analysis. Due to the nature of the data (e.g. traffic patterns in an urban environment; weather patterns across Nigeria), a constant or dynamic degree of precision is always lost in the measurement. This loss of precision is determined from the scale and distribution of the data collection.

Usually the larger the sample size the more accurate will the result of the analysis be. To determine the statistical relevance of the analysis, an average is determined so that points (gradients) outside of any immediate measurement can be included to determine their predicted behaviour. This is due to the limitations of the applied statistics and data collection methods, and interpolation is required to predict the behaviour of particles, points, and locations that are not directly measurable.

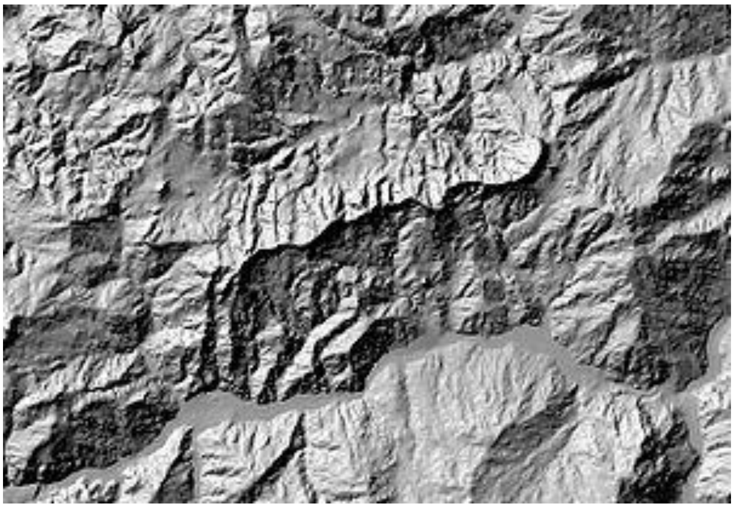


Fig. 3.4: Hillshade Model derived from a Digital Elevation Model (DEM)

Interpolation is the process by which a surface is created (Figure 3.4), usually a raster dataset, through the input of data collected at a number of sample points. There are several forms of interpolation, each of which treats the data differently, depending on the properties of the data set.

In comparing interpolation methods, the first consideration should be whether or not the source data will change (exact or approximate). Next is whether the method is subjective, a human interpretation, or objective. Then there is the nature of transitions between points: are they abrupt or gradual? Finally, there is whether a method is global (it uses the entire data set to form the model), or local where an algorithm is repeated for a small section of terrain.

Digital Elevation Models (DEM), Triangulated Irregular Networks (TIN), edge finding algorithms, Thiessen polygons, Fourier analysis, (weighted) moving averages, inverse distance weighting, kriging, spline, and trend surface analysis are all mathematical methods to produce interpolative data.

3.2.8 Address Geo-Coding

Geo-coding is interpolating spatial locations (X,Y coordinates) from street addresses (i.e. street names and house numbering), or any other spatially referenced data such as Postcodes or ZIP Codes, parcel lots and address locations. A reference theme is required to geo-code individual addresses, such as a road centerline file with address ranges. The individual address locations have historically been interpolated, or estimated, by examining address ranges along a road or street segment. These are usually provided in the form of a table or database.

The GIS will then place a dot approximately where that address belongs along the segment of centerline. For example, an address point of 50 will be at the midpoint of a line segment that starts with address 1 and ends with address 100. Geo-coding can also be applied against actual parcel data, typically from municipal tax maps (cadastral maps). In this case, the result of the geo-coding will be an actually positioned space as opposed to an interpolated point. This approach is being increasingly used to provide more precise location information.

There are several potentially dangerous caveats that are often overlooked when using interpolation. Various algorithms are used to help with address matching when the spellings of addresses differ. Address information that a particular entity or organization has data on, such as the post office, may not entirely match the reference theme. There could be variations in street name spelling, community name, etc.

Consequently, the user generally has the ability to make matching criteria more stringent, or to relax those parameters so that more addresses will be mapped. Care must be taken to review the results so as not to map addresses incorrectly due to overzealous matching parameters.

4.0 Conclusion

What distinguish the GIS system from other information system are its spatial analysis functions. As a matter of fact, the heart of GIS is the analytical capabilities of the system; it is for data analysis that GIS is used.

5.0 Summary

One data are stored in a GIS, many manipulation options are available to users. The analysis functions use the spatial and non-spatial attributes in the database to answer questions about the real world. The objective of geographic analysis is to transform data into useful information to satisfy the requirements or objectives of decision-makers at all levels in terms of detail.

6.0 Self-Assessment Exercise

Identify and discuss some of the major data manipulation and analysis operations carried out in GIS.

7.0 References/Further Reading

GIS Primer, *http://gis.nic.in/gisprimer/analysis3.html* (Retrieved on 27/7/11).

http://en.wikipedia.org/wiki/File:Dem.jpg

Unit 4 Data Output

1.0 Introduction

A critical component of a GIS is its ability to produce graphics on the screen or on paper to convey the results of analysis to the people who make decisions about resources. Wall maps, Internet-ready maps, interactive maps, and other graphics can be generated, allowing the decision-makers to visualize and thereby understand the results of analysis or simulations of potential events.

2.0 Objectives

At the end of this unit, you should be able to:

- examine the issue of data display in a GIS environment
- discuss document and printing formatting
- analyze final data output.

3.0 Main Content

3.1 Data Display

Data display is a form of data output – softcopy output. To work interactively with the computer system the data has to be displayed. The VDU (visual display unit) also known as monitor or screen, is usually the medium of data display. Both the graphic (map and chart) and textual (attribute) data can be displayed. The attribute data is usually displayed in tabular format. The spatial data is commonly displayed in map form. Most GIS operations involve a lot of graphics; consequently, a high-resolution VDU with a powerful GUI (graphical user interface) is often desirable.

There are various graphic display techniques. The VDU can be used to present spatial (map) data as a planimetric (2-D) or altrimetric (3-D) displayed, depending on the nature of the data. Traditional maps are abstractions of the real world, a sampling of important elements portrayed on a sheet of paper with symbols to represent physical objects. People who use maps must interpret these symbols. Topographic maps show the shape of land surface with contour lines or with shaded relief.

Today, graphic display techniques such as shading based on altitude in a GIS can make relationships among map elements visible, heightening one's ability to extract and analyze information. For example, two types of data could be combined in a GIS to produce a perspective view of a portion of an area. This can be done by using the digital elevation model, consisting of surface elevations and a satellite image for the same coordinate points, pixel by pixel, as the elevation information.

3.2 Document Formatting

Before the hardcopy of a document is printed (or plotted) some form of formatting may be required. Formatting involves preparing and presenting the document in the desired final output form. This is a form of customising the document.

Thus, document formatting may involve defining certain specifications or settings relating to the document to be produced. Formatting applies to both the graphics and texts. It equally involves defining printing or plotting options. The graphics, texts and printing formatting exercises are discussed below.

Graphics (map) formatting: This relates to the modifications made on the map to enhance its aesthetics and communication efficiency. The formatting exercise may include:

- modifying feature colour (gray tone (black & white), or coloured)
- modifying symbols
- inserting neatlines (borderlines) around the map
- choosing North arrow symbol
- adding/modifying graticules (lines of latitude and longitude), legend and scale bar
- adding inset map
- adding/editing labels (map lettering)
- proper positioning of map elements (e.g. map area, scale bar, legend box, north arrow, title, source, disclaimer, copyright, etc.), to achieve balance
- map embellishment.

Text formatting: This has to do with certain modifications that could be applied to the text and font. This may involve specifying font:

- type (e.g. Arial, Times New Roman, Tahoma, etc.)
- colour
- size (e.g. 8, 12, 36, 72 point size)
- style (e.g. normal, italics, bold)
- underlining
- effects (e.g. shadow, panels, balloons)
- orientation (horizontal, vertical, diagonal) etc.

Paper/Printing Formatting:

- selection of paper size (Letter, A4, A3, A2, A1, A0)
- indicating paper orientation (portrait, landscape)
- setting page margins

- inserting page number
- inserting date and other special remarks, symbols/logos, watermarks
- defining printing options (e.g. gray scale, colour, number of copies, etc.)
- print quality (usually specified as number of dots per inch (dpi) e.g. 300dpi, 600dpi, 1200dpi)
- specifying page range to print (e.g. all, current page, pages – to -)
- selection of printer or plotter type.

3.3 Data Output

Cartographic data output is quite crucial in GIS operations. Visualization or cartographic display of spatial data in a GIS environment is a key component of GIS analytical operations. Cartography is the design and production of maps, or visual representation and communication of spatial data. The vast majority of modern cartography is done with the help of computers, usually using a GIS. Most GIS software gives the user substantial control over the appearance of the data.

Cartographic work serves two major functions: First, it produces graphics on the screen or on paper that convey the results of analysis to the people who make decisions about resources. Web Map Servers facilitate distribution of generated maps through Web browsers using various implementations of Web-based application programming interfaces (AJAX, Java, Flash, etc.).

Second, other database information can be generated for further analysis or use. An example would be a list of all building addresses within one mile (1.6 km) radius of a toxic spill.

The hardcopy (paper) data output can be obtained using a printer or a plotter. Printers are used for producing relatively small size papers. Currently, the largest paper size a wide carriage printer can print on is A3. To obtain hardcopy outputs on paper sizes larger than A3 a plotter is commonly used.

4.0 Conclusion

Data manipulation and analysis operations often yield some output. This output could be in soft or hard copy version. The data output subsystem of GIS offers tools that can be used to format and package an output as may be desired.

5.0 Summary

The data output function of GIS involves displaying the result(s) of data processing and obtaining an analogue version of it. Usually, to obtain a softcopy the data is visualised on the screen of the workstation. Permanent or hardcopies are produced on paper (or some other printing media) using a printer or a plotter. The output whether softcopy or hardcopy could be a draft or final version. But before obtaining the output, the document should be properly formatted.

6.0 Self-Assessment Exercise

1. Discuss document and printing formatting.
2. Critically examine the issues of data display in a GIS environment.

7.0 References/Further Reading

Clarke, K. C. (1995). *Getting Started with Geographic Information Systems*. (3rd ed.). New Jersey: Prentice Hall.

Jones, C. (1997). *Geographical Information Systems and Computer Cartography*. Essex: Addison Wesley Longman Limited.

Kraak, M. & Ormeling, F. (1996). *Cartographic Visualisation of Spatial Data*. Essex: Addison Wesley Longman Limited.

Longley *et al.* (2001). *Geographic Information Systems and Science*. New York: John Wiley & Sons Inc.

Unit 5 Data Updating

1.0 Introduction

There is always the need to keep the data items in a GIS database as current as possible. Both the spatial and attribute databases should be periodically updated, as the need arises. Hence in this unit, our focus is on issues relating to the update of data entries in a GIS environment. It should be noted that in GIS there are two broad categories of data we deal with namely *graphics* (point, line, polygon) and *text* (alphanumeric and numeric). In other words, GIS data updating involves revising both spatial and non-spatial (attribute) data.

2.0 Objectives

At the end of this unit, you should be able to:

- define the concept of data updating
- discuss the process of spatial data updating
- explain the process of attribute (non-spatial) data updating.

3.0 Main Content

3.1 Concept of Data Updating

Basically, data updating is the process of bringing an already existing piece of data up to date. This process may demand deleting an obsolete data item and replacing it with a more recent one. It may also involve simply adding a freshly acquired data to an existing body of related data, without necessarily deleting any existing data.

It should be noted that the decision to update a database as well as the format and extent of the update depends on the intended application. For instance, if a database is to be used for some historical studies or time-series analysis, both the old and contemporary data items alike will be very much needed.

If, for example, a land-use/land-cover study of a place is to be carried out for, say, between 1960 and 2010, then maps, aerial photographs, and/or satellite images of different time periods in-between the two time segments will be needed. This, of course, is in spite of the fact that, for instance, a 1960 topographic map of an area will be considered obsolete when compared to a 2010 edition of the same area.

Whatever the reason or nature of any data updating exercise, it requires undertaking some form of data upgrade or modification. The whole idea is to keep the database as timely, relevant, useful and accessible as possible.

The general procedure for updating data in a database can be summarised thus:

- locate and retrieve file containing database to be updated
- open and display the database

- make required modifications to data (e.g. adding new data items, deleting obsolete data items, etc.) using the appropriate software tools
- validate data (attempting to update a database can result in some errors, hence after modifying the database, edit it by checking for possible any errors; correct any identified errors)
- save the updated database.

Reasons for data updating:

- to make the data up-to-date
- to improve on the accuracy of the data
- to enhance the use value and hence of the data
- to increase data accessibility (for authorised users).

3.2 Spatial Data Updating

Spatial data updating largely involves updating the vector map database; the essence is to improve on the quality of the database. In other words, a major goal of a map updating program is to ensure that revised graphics meet the accuracy specifications for the intended use. Updating becomes necessary if certain errors are noticed in the data. The updating process can also be used to capture and reflect recent changes that have taken place in the area covered by the map. Owing to the dynamics of land-use/land-cover change, it often becomes necessary to update maps held in a GIS in order to better fit the current landscape.

Updating spatial or cartographic data involves improving the geometric and horizontal (positional) accuracy of features represented on the old map. Usually, current data are more accurate than older map bases and can therefore be used to improve the accuracy of the map data.

Map or spatial data updating is mostly done using data from primary and secondary data sources, including the following:

- Existing maps and atlases
- Aerial photographs
- Satellite imagery
- Orthophotos
- Stereophotographs
- Gazetteer of place-names
- GPS data
- Direct field measurements
- In-situ personal field observations (ground truthing).
- Geographically-referenced databases from federal, state and local government agencies.

In updating a digital map (vectorised or scanned map), aerial photograph, orthophoto (orthorectified), or satellite image could be used as a back-drop raster image. The vector map to be updated is displayed as an overlay on the raster image. GIS image-warping methods can be used to adjust an old map base to newer, more accurate control. This is done to ensure that the map to be updated aligns or registers perfectly well with the scanned image.

For proper registration, the digital image and the map to be updated should be consistent with each other; this can be achieved by making them to be of the same projection and scale. The image rubber sheeting facility available in most commercial GIS has greatly improved the accuracy of some maps and improved the consistency of clusters of maps. Once the vector map and the raster image are properly aligned, new objects can be traced (on-screen) from the image and added to existing objects on the map.

On the other hand, obsolete, unwanted, or no-longer-existing objects on the map could be deleted. Similarly, in updating/editing a GIS map lines can, for example, be moved, rotated, intersected, or joined. More so, points of lines can be added, deleted, or moved. GPS readings can also be downloaded and used in updating map data in a GIS environment.

It should be noted that in using remote sensing imagery (aerial photographs, orthophotos and satellite images) to update cartographic data some sort of quality control should be conducted. The quality of the imagery can be checked visually on a workstation. In doing this, the radiometry, geometry, histogram, brightness and contrast properties of the image should be properly checked.

3.3 Attribute Data Updating

As noted earlier, the spatial data in a GIS works in conjunction with the associated attribute data. Hence, just as it often becomes necessary to update the spatial data, it is equally imperative to update the attribute database. As a matter of fact, certain data items in an attribute database are automatically updated as certain properties of map features in a spatial database are modified. This is so because in reality attribute data updating involves adding, changing or deleting attributes of map objects.

The attribute database is commonly held in tabular form. One important operation in maintaining tables is the ability to update the data contained in the table. Most commercial GIS have features or commands that allow one to perform various data updating functions. The rows (records) as well as the columns (fields) of tables can be updated. You can update all the rows in a table or a selection of rows. Also, you can update a single column or a group of columns. Some of the commands used in modifying an attribute table are INSERT, UPDATE, or DELETE.

Some of the attribute table modifications that can be carried out in a GIS environment include:

- Add a temporary column or update an existing column with data from another table
- Update a table
- Place graphic information into visible columns
- Add a new column

- Modify data in a cell (entering new values that will replace the current ones)
- Edit/Update data in an existing column
- Delete column
- Change column name
- Delete data in a column
- Change column type (e.g. from alphanumeric to numeric)
- Increase/Reduce column size.

The process of updating an attribute data may involve bringing data from one table into another. In doing this, you can either add a temporary column or you can update an existing column, depending on the flexibility and requirements of the GIS software in use. Depending on the magnitude of the modification to be made, the attribute data may be updated directly in the GIS environment, or the table could be created in another GIS-compatible database management system such as Access, Oracle or Microsoft Excel and later imported into GIS.

Many contemporary GIS software offer tremendous flexibility with how information is imported and what information is transferred. However, in using other software to create a new tabular data that will be used to update an existing GIS attribute database, care must be taken to ensure that the format of the new database is consistent with the format of the already existing.

4.0 Conclusion

The credibility of GIS hinges so much on the quality and currency of the database. The need for constant updating of GIS data, therefore, cannot be overemphasised. Having current and useable data is essential to maximising the potential of GIS. Hence, a well thought out plan for periodic data upgrading should form a key component of every GIS implementation programme.

5.0 Summary

The ultimate goal of adopting and operating the GIS technology is to provide reliable, useful and timely information for decision making. The success or otherwise of any GIS scheme hinges largely on the quality and dependability of its database. To keep the data 'fit-for-use' at any time there is need to regularly update the data. Data updating involves upgrading the quality of an existing piece of data in order to bring it up to date. Doing this may take the form of adding fresh items to the database and/or deleting certain obsolete ones.

6.0 Self-Assessment Exercise

1. Define the concept "Data Updating".
2. Analyze the process of attribute data updating.

7.0 References/Further Reading

Cheng *et al.* (2008). A Study on Multi-agent Spatial Database Updates Mechanism based on Wiki Idea. In: L. Liu, X. Li, K. Liu, X. Zhang & A. Chen (Eds). *Geo-Simulation and Virtual GIS*

Environments, Proceedings of the SPIE Geoinformatics 2008 and Joint Conference on GIS and Built Environment. Volume 7143, pp. 71431K-71431K-9.

Scheu, M., Effenberg, W. & Williamson, I. (n.d.). *Incremental Update and Upgrade of Spatial Data. Retrieved from:* citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.6.5286&rep=rep1&type=pdf