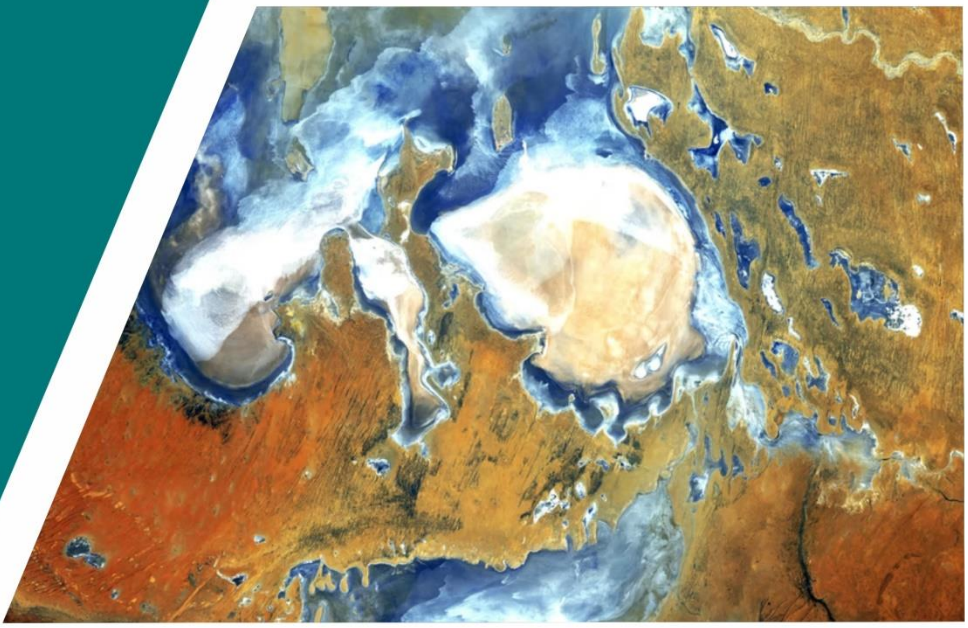
NATIONAL OPEN UNIVERSITY OF NIGERIA

ESM 407



Geographic Information

Module 4

ESM 407 Geographic Information Systems Module 4

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 4 Database Structure

Unit 1 Database Structure

1.0 Introduction

The data used in GIS operations are usually organized in form of databases. A database, in turn, is typically organized according to general data models or structures. Understanding the nature of databases and database management systems is, therefore, crucial to understanding the creation and handling of data in a GIS environment. In this unit, we will be looking at both spatial and non-spatial databases.

2.0 Objectives

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

- explain the concept of database (DB) and database management system (DBMS)
- examine the nature, content, and functions of spatial database systems
- identify non-spatial database structures.

3.0 Main Content

3.1 The Concept of Database

In a GIS environment, the data used for operations is usually organised in database form. Generally speaking, the term 'database' refers to a collection of information about things and their relationship to one another. A database model is a theory of how a database is supposed to look like. Put differently, a database is an integrated collection of data records, files, and other database objects needed by an application. A database could be in analogue or digital format. Here, our concern is with digital database.

Hence, a database is an organized collection of data for one or more purposes, usually in digital form. A database (often abbreviated *DB), is normally* organized in such a way that a computer program can quickly select desired pieces of data. We can therefore think of a database as an electronic filing system.

Traditional databases are organized by *fields*, *records*, and *files*. A field is a single piece of information; a record is one complete set of fields; and a file is a collection of records. For example, a telephone book is similar to a file. It contains a list of records, each of which consists of three fields: name, address, and telephone number.

Data are typically organized relevant aspects of reality (for example, the number of flats in a residential apartment building), in a way that supports processes requiring this information (for example, finding a building with a vacant flat for rental). The term "database" refers both to the way its users view it, and to the logical and physical materialization of its data, content, in files, computer memory, and computer data storage.

Moreover, the term database implies that the data is managed to some level of quality (measured in terms of accuracy, availability, usability, and resilience) and this in turn often implies the use of a general-purpose Database management system (DBMS).

To access information from a database, you need a *Database Management System (DBMS)*. A DBMS is a software package with a collection of computer programs that controls the creation, organization, maintenance, selection and the use of data in a database. The term database is correctly applied to the data and data structures themselves, and is different from the DBMS which is a software system that allows one to store and change the database (i.e., the data), as well as retrieve information from it.

A few examples of commercially available DBMSs include Gemstone, O_2, Versant, Mattise, Codasyl, Sybase, Oracle, DB2, Access, dBase, SQL Server from Microsoft, DB2 from IBM and the Open source DBMS MySQL.

The database concept has evolved since the 1960s to ease increasing difficulties in designing, building, and maintaining complex information systems (typically with many concurrent end-users, and with a large amount of data). It has evolved together with the evolvement of database management systems (DBMSs) which enable the effective handling of databases.

Though the terms database and DBMS define different entities, they are inseparable: A database's properties are determined by its supporting DBMS and vice-versa. The major purpose of a database is to provide the information system (in its broadest sense) that utilizes it with the information the system needs.

3.2 Spatial Database

The GIS technology operates on spatial databases. Spatial databases store information related to objects in space. A spatial database is a collection of spatially referenced data that acts as a model of reality. In other words, a spatial database is a database that is optimized to store and query data that is related to objects in space, including points, lines and polygons.

Furthermore, a spatial database system is a database system that offers spatial data types in its data model and query language, and supports spatial data types in its implementation, providing at least spatial indexing and spatial join methods. Spatial database systems offer the underlying database technology for geographic information systems and other applications (Güting, 1994).

As noted by Güting (1994):

- A spatial database system is a database system.
- It offers spatial data types (SDTs) in its data model and query language.
- It supports spatial data types in its implementation, providing at least spatial indexing and efficient algorithms for spatial join.

While typical databases can understand various numeric and character types of data, additional functionality needs to be added for databases to process spatial data types. These are typically called *geometry* or *feature*. Generally, database systems use indexes to quickly look up values.

However, the way that most databases index data is not optimal for spatial queries. Indexes used by non-spatial databases cannot effectively handle features such as how far two points differ and whether points fall within a spatial area of interest. Consequently, spatial databases use a spatial index to speed up database operations.

3.3 Content of Spatial Database

The content of a spatial database depend on its intended purpose, which in turn is dependent on the organisation using it. We can briefly illustrate this with two simple examples namely, transportation and wetland.

Example 1: Transportation

Consider the use of highway data from the different points of view of two agencies namely a natural resources organization and a highway transportation organization. The natural resource organization might only need information on logging routes and the connecting access to secondary or state highways. On the other hand, the transportation organization's main interest is in characterizing highways used by the public. The database might also be used to store detailed highway condition and maintenance information. We would, therefore, expect the transportation organization's need for highway data to be more detailed than would the natural resource organization's need.

Example 2: Wetlands

Let us also consider the need for wetlands data from the different points of view of two agencies namely an ecological organization and a taxing authority. The ecological organization might define wetlands as a natural resource to be preserved and restricted from development.

Thus, that perspective might require considerable detail for describing the wetland's biology and physical resources. On the other hand, the taxing authority might define a wetland to be a "wasteland" and of very little value to society. Thus, that description might require only the boundary of the "wasteland" to be included in the database.

3.4 Basic Characteristics of Spatial Database

A typical spatial database should possess certain characteristics. Such a database should be:

- Contemporaneous - should contain information of the same time period for all its measured variables.
- As detailed as necessary for the intended applications.
- Accurate in terms of the geographical location or positions of features.
- Exactly compatible with other information that may be overlain with it.
- Internally accurate, portraying the nature of phenomena without error - requires clear definitions of phenomena that are included.
- Easily updated on a regular schedule.
- Accessible to whoever needs it and is authorized to use it.

3.5 Functions of Spatial Databases

Spatial databases can perform a wide variety of spatial operations. To perform its functions spatial databases make use of spatial query languages. A spatial query is a special type of database query supported by geodatabases and spatial databases. The spatial queries differ from SQL queries in that they allow for the use of geometry data types such as points, lines and polygons and that they consider the spatial relationship between these geometries.

The following query types and many more are supported by many spatial databases especially the Open Geospatial Consortium (see Wikipedia URL):

Spatial Measurements: Find the distance between points, polygon area, etc.

Spatial Functions: Modify existing features to create new ones, for example by providing a buffer around them, intersecting features, etc.

Spatial Predicates: Allows true/false queries such as "is there a residence located within a mile of the area we are planning to build the landfill?"

Constructor Functions: Creates new features with an SQL query specifying the vertices (points of nodes) which can make up lines. If the first and last vertexes of a line are identical the feature can also be of the type polygon (a closed line).

Observer Functions: Queries which return specific information about a feature such as the location of the center of a circle.

3.6 Non-Spatial Database

As noted earlier, GIS uses raster or vector representations to model location. It is equally important to consider how GIS must also record information about the real-world phenomena positioned at each location and the attributes of these phenomena. That is, the GIS must provide a linkage between spatial and non-spatial data.

The linkage between symbol (map feature) and meaning is established by giving every geographic feature at least one unique means of identification, a name or number usually just called its ID. Non-spatial attributes of the feature are then stored, usually in one or more separate files, under this ID number. In other words, graphic or locational information is linked to specific non-graphic or attribute information in a database.

Non-spatial data (also called *attribute* or *characteristic* data) is that information which is independent of all geometric considerations. It is the set of data that tells more about what geographic features are like. In other words, whereas spatial data gives us information about *where* (location) things are, non-spatial data gives us information on *what* they are like irrespective of their location. For example, a tree is a geographic feature, but the type, height, and age of the tree are non-spatial data because they are independent of the tree's location.

The non-spatial data associated with spatial features can be filed away in several different forms depending on how it needs to be used and accessed. There are numerous types of database models today. These models not only represent how a database looks like but also what kind of operations that can be used to manipulate the data within. Some of the commonly used attribute database models include:

- Hierarchical model
- Network model
- Relational model
- Entity-relationship model
- Object model
- Object/relational database management systems (ORDBMSs)
- Object-oriented database (OODB) model.

In this unit, we will briefly describe the three commonest database models namely hierarchical, network and relational. For more detailed discussion on each of the data models, the interested reader can consult the sources listed in the references section.

Hierarchical Database Model

The hierarchical data model organizes data in a tree structure (Figure 4.1). There is a hierarchy of parent and child data segments. In other words, to create links between record types, the hierarchical model uses Parent-Child Relationships. For example, an organization might store information about an employee, such as name, employee number, department, salary.

The organization might also store information about an employee's children, such as name and date of birth. The employee and children data forms a hierarchy, where the employee data represents the parent segment and the children data represents the child segment. If an employee has three children, then there would be three child segments associated with one employee segment. In a hierarchical database the parent-child relationship is one-to-many.

In a hierarchical database model data is stored in more than one type of record. One field is usually recognized as key to all records, but data in one record does not have to be repeated in another. This system allows records with similar attributes to be associated together. The records are linked to each other by a key field in a hierarchy of files. Each record, except for the master record, has a higher level record file linked by a key field "pointer". In other words, one record may lead to another and so on in a relatively descending pattern.

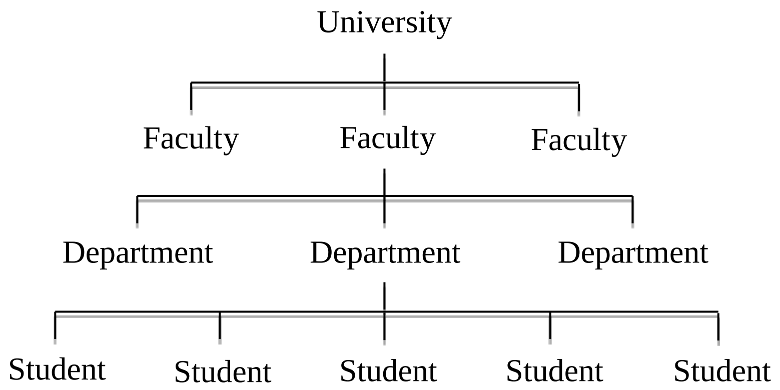


Fig. 4.1: Hierarchical Data Structure

Network Model

The network model uses records and sets, which are its two basic structures, when organizing data. Record contains the fields, while the set is the one defining the relationship existing between the records. Unlike what obtains in the hierarchical data structure, data in network structure are modeled with more than one parent per child (Figure 4.2). So, the network model permits the modeling of many-to-many relationships in data. The basic data modeling construct in the network model is the set construct. A set consists of an owner record type, a set name, and a member record type. A member record type can have that role in more than one set; hence the multi-parent concept is supported. An owner record type can also be a member or owner in another set. The data model is a simple network, and link and intersection record types (called junction records by IDMS) may exist, as well as sets between them.

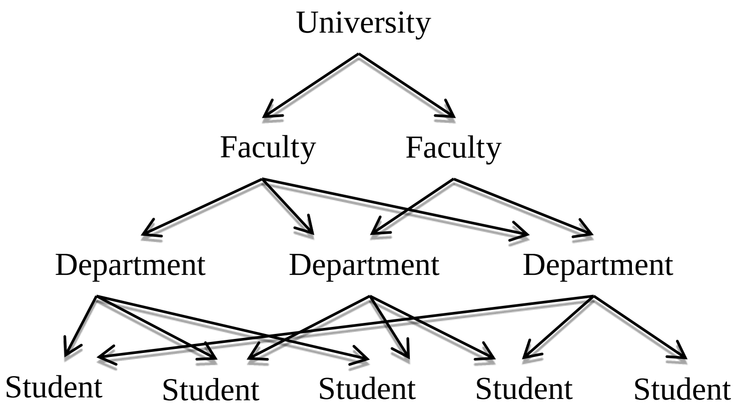


Fig. 4.2: Network Data Structure

Relational Database Model

A relational database allows the definition of data structures, storage and retrieval operations and integrity constraints. In such a database, the data and relations between them are organized in tables (Table 4.1). A table is a collection of records and each record in a table contains the same fields. A common link of data is used to join or associate records. The link is not hierarchical. A "matrices of tables" is used to store the information.

As long as the tables have a common link they may be combined by the user to form new inquires and data output. This is the most flexible system and is particularly suited to SQL (structured query language). Queries are not limited by a hierarchy of files, but instead are based on relationships from one type of record to another that the user establishes.

Perhaps, this is the simplest method, where each geographic feature is matched to one row of data. All records in this kind of database have the same number of "fields". Individual records have different data in each field with one field serving as a key to locate a particular record. For example, as a person your national ID card number may be the key field in a record of your name, address, phone number, sex, ethnicity, place of birth, date of birth, religion, and so on.

Similarly, for a plot of land there could be hundreds of fields associated with the record (plot), such as owner, address, size, usage, etc. Because of its flexibility this system is the most popular database model for GIS (Foote and Huebner, 2000). The relational database structure is further discussed in Unit 3 of this module.

Table 4.1: Relational Data Structure

Bld_ID	Owner	City	Str_No	Str_Name	Type	No_Floors	...
1	Ben Ike	Yaba	6	Ayodele	Duplex	2	...
2	Segun Ajayi	Yaba	13	Ekanem	Bungalow		...
3	Ayo Oni	Yaba	2	Okoro	Bungalow		...
4	Idris Musa	Yaba	7	Lawanson	Duplex	3	...
5	Uche Ojeh	Yaba	21	Obayan	Bungalow		...
...	...	...	...	...	...	...	...

4.0 Conclusion

Every GIS has a database management system (DBMS), which is used to create and handle the datasets used for various operations.

5.0 Summary

Data items used in GIS operations are usually arranged and stored in database format. A database, which is normally held in digital form, is an organized collection of data records, files, and other database objects needed by an application, for one or more purposes. We have the spatial database as well as the attribute (non-spatial) database. A spatial database is a collection of geographically-referenced data that acts as a model of reality. On the other hand, an attribute database is used to hold descriptive information about the geographical features represented in a spatial database.

6.0 Self-Assessment Exercise

1. Discuss the concept:

a. Database (DB)

b. Database management system (DBMS)

2. Identify non-spatial database structure.

7.0 References/Further Reading

Date, C. J. (1994). *An Introduction to Database Systems*. New York: Addison-Wesley Publishing Co.

Foote, K. E. & Huebner, D. J. (2000). *Database Concepts*. The Geographer's Craft Project, Department of Geography, The University of Colorado at Boulder.

http://www.colorado.edu/geography/gcraft/notes/datacon/datacon_f.html (Accessed 22/6/11).

Güting, R. H. (1994). An Introduction to Spatial Database Systems**.** *The International Journal on Very Large Data Bases*, Vol. 3, No. 4, October 1994.

Yeung, A. K. W. & Hall, G. B. (2007). *Spatial Database Systems: Design, Implementation and Project Management*. New York: Springer Science+Business Media.

http://www.webopedia.com/TERM/D/database.html

http://en.wikipedia.org/wiki/Spatial_database

Unit 2 Spatial Data Model

1.0 Introduction

Real world features are represented in GIS using a spatial data model. Simply put, a data model is an abstract structure that provides the means to effectively describe specific data structures needed to model an application. GIS data represents real objects (such as roads, land use, elevation, trees, waterways, etc.). Real objects can be divided into two abstractions: discrete objects (e.g., a house, a roundabout, or a borehole) and continuous fields (such as rainfall amount, vegetation, or elevations).

Traditionally, there are two broad methods or models used to store spatial data in a GIS: *raster images* and *vector*. In this unit, we will closely look at each of the models.

2.0 Objectives

At the end of this unit, you should be able to:
- identify and discuss the two spatial data models commonly used in GIS: vector and raster
- compare and contrast the models
- examine the conversion from one model format to another.

3.0 Main Content

3.1 Vector Structure

In a GIS, geographical features are often expressed as vectors, by considering those features as geometrical shapes. Different geographical features are expressed by different types of geometry.

Vector digital data have been captured as points, lines (a series of point coordinates), or areas (shapes bounded by lines) (Figures 4.4a and b). An example of data typically held in a vector file would be the property boundaries for a particular housing subdivision.

Points

Zero-dimensional points are used for geographical features that can best be expressed by a single point reference — in other words, by simple location. Examples include wells, peaks, features of interest, and trailheads. Points convey the least amount of information of these file types. Points can also be used to represent areas when displayed at a small scale. For example, cities on a map of the world might be represented by points rather than polygons. No measurements are possible with point features.

Lines or Polylines

One-dimensional lines or polylines are used for linear features such as rivers, roads, railroads, trails, and topographic lines. Again, as with point features, linear features displayed at a small scale will be represented as linear features rather than as a polygon. Line features can measure distance.

Polygons

Two-dimensional polygons are used for geographical features that cover a particular area of the earth's surface. Such features may include lakes, park boundaries, buildings, city boundaries, or land uses. Polygons convey the most amount of information of the file types. Polygon features can measure perimeter and area.

Each of the above geometries is linked to a row in a database that describes their attributes. For example, a database that describes lakes may contain a lake's depth, water quality, pollution level. This information can be used to make a map to describe a particular attribute of the dataset. For example, lakes could be coloured depending on level of pollution. Different geometries can also be compared. For example, the GIS could be used to identify all wells (point geometry) that are within one kilometer of a lake (polygon geometry) that has a high level of pollution.

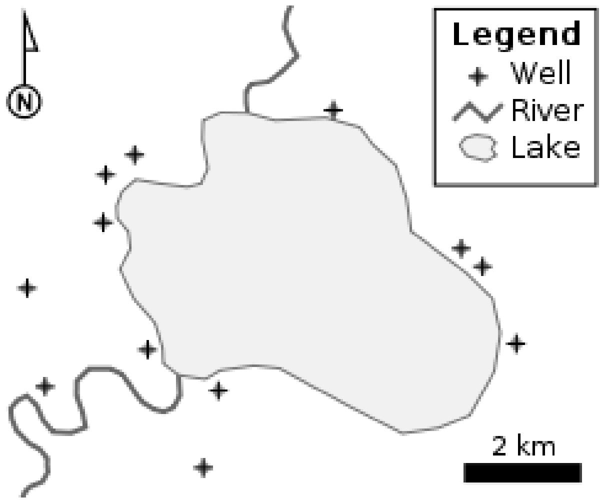


Fig. 4.4a: A Series of Point Coordinates

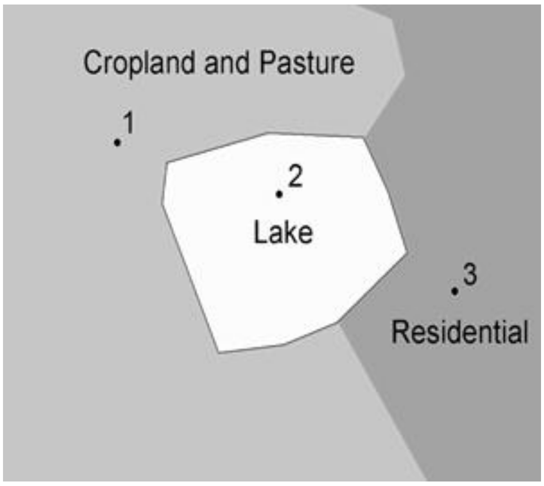


Fig. 4.4b: Shapes Bounded by Lines

3.2 Raster Structure

A raster data type is, in essence, any type of digital image represented by reducible and enlargeable grids (also known as cells or pixels). Figure 4.5 illustrates the representation of point, linear and areal geographical features in a raster format. Anyone who is familiar with digital photography will recognize the raster graphics pixel as the smallest individual grid unit building block of an image, usually not readily identified as an artifact shape until an image is produced on a very large scale.

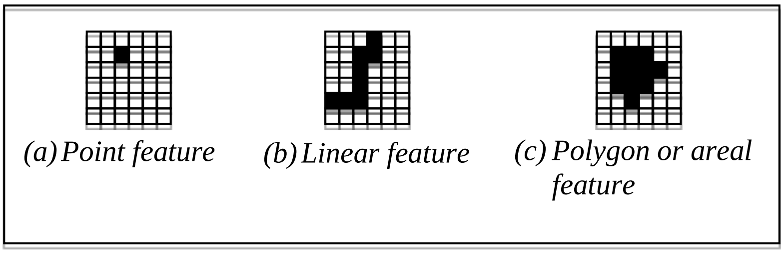
A combination of the pixels making up an image colour formation scheme will compose details of an image, as is distinct from the commonly used points, lines, and polygon area location symbols of scalable vector graphics as the basis of the vector model of area attribute rendering.

While a digital image is concerned with its output blending together its grid based details as an identifiable representation of reality, in a photograph or art image transferred into a computer, the raster data type will reflect a digitized abstraction of reality dealt with by grid populating tones or objects, quantities, co-joined or open boundaries, and map relief schemas. Aerial photos are commonly used form of raster data, with one primary purpose in mind: to display a detailed image on a map area, or for the purposes of rendering its identifiable objects by digitization.

Additional raster data sets used by a GIS will contain information regarding elevation, a digital elevation model, or reflectance of a particular wavelength of light, Landsat, or other electromagnetic spectrum indicators.

Raster data type consists of rows and columns of cells, with each cell storing a single value (Figure 4.6). Raster data can be images (raster images) with each pixel (or cell) containing a colour value. Additional values recorded for each cell may be a discrete value, such as land use, a continuous value, such as temperature, or a null value if no data is available.

While a raster cell stores a single value, it can be extended by using raster bands to represent RGB (red, green, blue) colours, colourmaps (a mapping between a thematic code and RGB value), or an extended attribute table with one row for each unique cell value. The resolution of the raster data set is its cell width in ground units.



Φιγ. 4.5: Ρεπρεσεντατιον οφ Γεογραπηιχαλ Φεατυρεσ ιν Ραστερ Φορματ

Raster data is stored in various formats; from a standard file-based structure of TIF, JPEG, etc. to binary large object (BLOB) data stored directly in a relational database management system (RDBMS) similar to other vector-based feature classes. Database storage, when properly indexed, typically allows for quicker retrieval of the raster data but can require storage of millions of significantly sized records.

Typically, raster data files consist of rows of uniform cells coded according to data values. An example is land cover classification (Figure 4.6). Raster files can be manipulated quickly by the computer, but they are often less detailed and may be less visually appealing than vector data files, which can approximate the appearance of more traditional hand-drafted maps.

1	1	1	1	1	1	1	3	3	3
1	1	1	1	1	1	1	3	3	3
1	1	1	1	1	1	3	3	3	3
1	1	1	2	2	2	2	3	3	3
1	1	1	2	2	2	2	3	3	3
1	1	1	2	2	2	2	3	3	3
1	1	1	1	2	2	2	3	3	3
1	1	1	1	1	1	3	3	3	3
1	1	1	1	1	1	1	3	3	3
1	1	1	1	1	1	1	1	3	3

Fig. 4.6: Raster Data Type (Land cover Classification)

3.3 Comparison of Raster to Vector Mode

There are some important advantages and disadvantages to using a raster or vector data model to represent reality:

- Raster datasets record a value for all points in the area covered which may require more storage space than representing data in a vector format that can store data only where needed.
- Raster data allows easy implementation of overlay operations, which are more difficult with vector data.

- Vector data can be displayed as vector graphics used on traditional maps, whereas raster data will appear as an image that may have a blocky appearance for object boundaries (depending on the resolution of the raster file).
- Vector data can be easier to register, scale, and re-project, which can simplify combining vector layers from different sources.
- Vector data is more compatible with relational database environments, where they can be part of a relational table as a normal column and processed using a multitude of operators.
- Vector file sizes are usually smaller than raster data, which can be 10 to 100 times larger than vector data (depending on resolution).
- Vector data is simpler to update and maintain, whereas a raster image will have to be completely reproduced. (Example: a new road is added).
- Vector data allows much more analysis capability, especially for "networks" such as roads, power, rail, telecommunications, etc. (Examples: Best route, largest port, airfields connected to two-lane highways). Raster data will not have all the characteristics of the features it displays.
- Raster files can be manipulated quickly by the computer, but they are often less detailed and may be less visually appealing than vector data files, which can approximate the appearance of more traditional hand-drafted maps.

3.4 Data Model Conversion

From the foregoing it is obvious that digital spatial data are collected and stored in different ways; vector or raster. The two data models are not entirely compatible. Therefore, a GIS must be able to convert data from one structure to another. Data restructuring or conversion can be performed by a GIS to convert data between different formats.

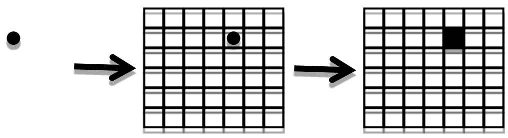
For example, a GIS can be used to convert a satellite image map (raster data) to a vector structure by generating lines around all cells with the same classification, while determining the spatial relationships of the cell, such as adjacency or inclusion (Figure 4.7a and 4.7b).

Vector to Raster Conversion

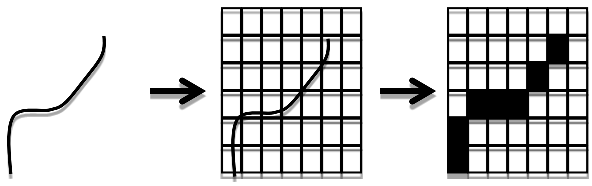
Converting vector data to raster involves using grid cells to define or represent the location of a point, line or polygon (area) feature held in vector format. To rasterise a vector map an artificial matrix of cells is first imposed on the map. With the fine mesh of grid lines now covering the map features it becomes easy to identify which pixels (cells) lie on or within the boundaries of the features. With the aid of a suitable rasterisation algorithm the cells accommodating the particular spatial object(s) of interest can be identified and assigned a value or shaded in a chosen pattern to represent that object type. Figures 4.8a, b and c illustrate the rasterization of point, linear and areal features respectively.

Fig. 4.7a: Views of the Satellite Image Map

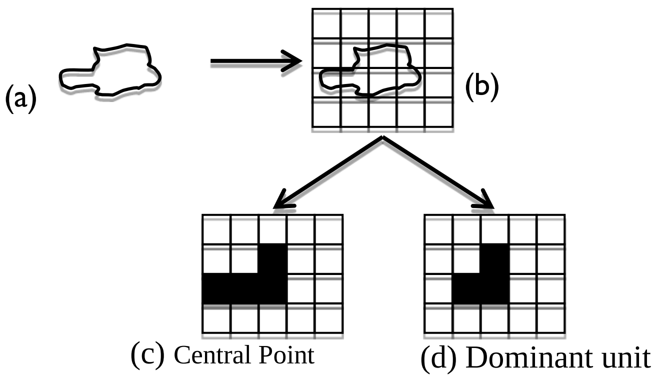
Fig. 4.7b: Magnified Views of the Same GIS Data File Converted into Vector Format



Φιγ. 4.8α: Ποιντ Φεατυρε Ραστερισατιον (Υλυοχηα, 2007)



Φιγ. 4.8β: Ραστερισινγ α Λινε



Φιγ. 4.8χ: Πολψγον Ραστερισατιον

4.0 Conclusion

There are two main spatial data models used in storing geographically-referenced data in GIS. These are the vector model and the raster model. The selection of a particular data model, vector or raster, is dependent on the source and type of data, as well as the intended use of the data. Whereas certain analytical procedures require raster data others are better suited to vector data.

5.0 Summary

Traditionally, geospatial data has been stored and presented in the form of a map. However, various types of spatial data models have been developed for storing geographic data digitally. The two models mostly used are raster and vector. The raster model makes use of a matrix of cells to represent and store data while vector uses the more familiar cartographic symbols (points, lines, and polygons). Each of the spatial data models has its own merits and demerits. It is possible to convert from one data model to the other.

6.0 Self-Assessment Exercise

1. Identify and discuss the two spatial data models commonly used in GIS.
2. Compare and contrast the two models.

7.0 References/Further Reading

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

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

Unit 3 Attribute Data Model

1.0 Introduction

Additional non-spatial (also known as aspatial or *attribute*) data can also be stored along with the spatial data represented by the coordinates of a vector geometry or the position of a raster cell. Software is currently being developed to support spatial and non-spatial decision-making, with the solutions to spatial problems being integrated with solutions to non-spatial problems. The end result with these Flexible Spatial Decision-Making Support Systems (FSDSS) is expected to be that non-experts will be able to use GIS, along with spatial criteria, and simply integrate their non-spatial criteria to view solutions to multi-criteria problems. This system is intended to assist decision-making.

In unit 1, we identified and briefly discussed some of the major attribute (non-spatial) database models used in GIS. In this unit, we aim at taking a further look at the Relational Database Model, which is the most commonly used attribute database model in GIS. The nature, qualities and errors commonly associated with the use of the relational structure will be discussed.

2.0 Objectives

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

- discuss the major attribute data model used in GIS
- identify the errors commonly made during attribute database creation.

3.0 Main Content

3.1 Relational Database Model

In a GIS environment, attribute data are commonly stored in a tabular form. As earlier mentioned in Unit 1, there are various data structures or models that could be used for attribute data files, including hierarchical structure, network structure and relational structure. However, the relational model is the most commonly used type.

A *relational database structure* is simply a two-dimensional table made up of rows (also known as *tuples* or records) and columns (also known as *domains*, fields, or attributes). Each row contains a single record representing an entity or object, while a column contains an attribute or characteristic of the entity.

In vector data, the additional data contains attributes of the feature. For example, as illustrated in Figure 4.9, a building polygon may also have an identifier (ID) value and other information about it such as the name of the owner, the type of building, street address, number of floors (if storey building), number of flats, colour, age (year built), etc. In raster data the cell value can store attribute information, but it can also be used as an identifier that can relate to records in another table.

In a GIS-based relational attribute table, the fields could be of different types. The field type affects the way GIS will recognize and process the data contained in the field. For instance, if the values (statistical figures) in a field are meant for some arithmetic or mathematical calculations, GIS will not be able to use the figures for such operations if the field was not created as a 'numeric' field. Hence, it becomes quite imperative to accurately define the type of each field in an attribute database during the creation of the database.

The fields in an attribute database can be defined as:

- Alphanumeric
- Numeric
- Date
- Logic

Table 4.9: Sample of a Relational Table

ID	Address	Owner	Type	No_Floors	No_Flats	Occupier	....	....
001	5, Ulo Close, Eluigwe	Chief B.C Dede	Duplex	2	4	Owner		
....								
....								

Presently, the Relational Database Management System (DBMS) is the most widely used commercial data management tool in GIS implementation and application. The relational DBMS is attractive to GIS users for a number of reasons, including its:

- *Simplicity* in organization and data modeling
- *Flexibility* - data can be manipulated in an ad hoc manner by joining tables
- *Efficiency* of storage – proper design of data tables can reduce redundancy
- *Queries* do not need to take into account the internal organization of data.

3.2 Data Editing

The accuracy and precision of the attribute data to be used for GIS operations should be ensured. Hence, after creating the database effort should be made to check through the data for possible errors. Any errors detected should be promptly rectified. The common errors which one can possibly make during the creation of an attribute data file include:

- Wrong spelling
- Improper definition of a field

- Omission of some vital details
- Wrong entry (e.g. keying in 17,018 instead of 17,108).

4.0 Conclusion

In GIS a separate data model is used to store and maintain attribute data. This data model may exist internally within the GIS software, or may be reflected in external commercial Database Management Software such as the DBMS. A variety of different data models exist for the storage and management of attribute data. However, the most widely used attribute data model in GIS is the *relational database model.*

5.0 Summary

Like the spatial data, the non-spatial or attribute datasets used in GIS are usually designed, created and maintained using a certain database model. There are different data structures or models that can be used for attribute data files, such as hierarchical structure, network structure and relational structure.

However, the relational database model is the most widely accepted for managing the attributes of geographic data. The relational database organizes data in *tables*. Each table, is identified by a unique table name, and is organized by *rows* (records) and *columns* (fields). Each column within a table also has a unique name. Columns store the values for a specific attribute, e.g. tree age, tree height, etc. Rows represent one record in the table. In a GIS, each row is usually linked to a separate spatial feature, e.g. a building.

Accordingly, each row would be comprised of several columns, each column containing a specific value for that geographic feature. Good practice demands that after creating a database and keying in the data items it should be edited, to identify and correct any existing errors.

6.0 Self-Assessment Exercise

Identify the errors commonly made during attribute database creation.

7.0 References/Further Reading

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

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

Unit 4 Data Quality

1.0 Introduction

Data quality has to do with usefulness of a set of data vis-a-vis the intended use. It is a measure of the level of fitness-for-use of the data. It is often necessary to ascertain the quality of data before using it for a GIS operation. One should always bear in mind the popular slogan: *garbage in, garbage out*. The quality of the data used in a GIS project goes a long way to determining the success or otherwise, of the project. In this unit, we will focus on examining the factors for evaluating or measuring the quality of geospatial as well as attribute data.

2.0 Objectives

At the end of this unit, you should able to:
- discuss the parameters for assessing spatial data
- highlight the qualities of attribute data.

Main Content

3.1 Data Quality

The quality of data is usually determined using certain criteria. The criteria used to conduct quality assessment on both spatial and attribute data are basically the same. Such criteria include *lineage, logical consistency, completeness, accuracy/precision, currency/timeliness, and scale (resolution)*. Each of the data quality assessment parameters is briefly discussed as follows.

3.1.1 Lineage

This gives account of the origin or source of the data, the date of collecting the data, and the methods adopted in data collection/database creation. Knowing the source of data helps one in determining whether or not the data is reliable.

3.1.2 Logical Consistency

This is a measure of the degree of conformity of internal data structures to specified data modeling rules. This measure of data accuracy and quality is normally applied to spatial data; it is used to establish the authenticity (or otherwise) of the data structure created for a data set. Hence, the logical consistency search is carried out to identify any spatial or topological errors in a spatial data structure such as incorrect line intersections, duplication of lines or boundaries, gaps in lines (discontinuity), and so on.

3.1.3 Completeness

Measures the extent to which the data cover the population of items of interest. For instance, if there are 15,000 properties in a Local Government Area (LGA) and the database holds records for only 9,000, the database is obviously incomplete.

3.1.4 Positional Accuracy and Precision

This index is used to determine the difference between measured (or observed or computed) location values (coordinates and altitude) and their true values. In other words, positional accuracy and precision is an index that gives one an idea of how close a recorded positional value is to its true or generally accepted value.

3.1.5 Attribute Accuracy and Precision

This is the extent to which recorded attribute values correspond to their true or real world or generally accepted values.

3.1.6 Currency

This is a measure of the obsoleteness or up-to-datedness of data. By considering or assessing data currency, one would be able to know whether or not the data are recent and timely (up-to-date) in relation to the intended application. (It should be noted that, that a piece of data is out of date does not mean it is no longer useful. Such data are often essential to time-series or temporal analysis).

3.1.7 Scale and Spatial Resolution

This deals with the degree of graphic representation of details about real world features. Various GIS applications require mapped data at various scales and hence spatial resolutions. Thus effort should be made to ascertain if the available maps, air-photos, or satellite imagery (paper or digital) are at a scale suitable for the task at hand.

For instance, a 1:50,000 topographical map may not be ideal for undertaking detailed site analysis and landscaping for the purpose of constructing a petrol station; a larger scale map, say 1:500, would be needed for such a purpose.

4.0 Conclusion

Data quality refers to the appropriateness of a set of data for an intended use. Data that is appropriate for use with one application may not be fit for use with another. The quality of any data set must be ensured before using it in GIS.

5.0 Summary

Data quality is an important aspect of any piece of data to be used in GIS. Quality here can simply be defined as the fitness for use for a specific data set. It is fully dependant on certain parameters such as the source, scale, accuracy, currency, completeness and extent of the data set.

6.0 Self-Assessment Exercise

Highlight the qualities of attribute data.

7.0 References/Further Reading

Faiz, S. & Boursier, P. (1996). "Geographic Data Quality: from Assessment to Exploitation." *Cartographica*, Vol. 33, No.1, pp. 33-40.

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

Unit 5 Sources of Data

1.0 Introduction

Data is critical to the successful execution of any GIS project. The data must not only be available, but also accessible. Moreover, the available data must be relevant, reliable and usable. Sourcing and obtaining useful data is often a major task. Various sources of data exist. However, one must be careful to properly assess the reliability of any source before making use of it. In this unit, we will try to identify the various sources and techniques of spatial and attribute data acquisition for GIS operations.

2.0 Objectives

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

- identify sources of spatial data
- identify sources of attribute data.

3.0 Main Content

3.1 Sources of Data

Generally, the data for GIS projects can be obtained from primary and/or secondary sources. Primary data are basically the set of data collected originally by the user. On other hand, secondary data is an already existing data.

In Nigeria, there are a number of government agencies as well as private enterprises that collect or retail spatial and statistical data that could be used for GIS projects. Some of such sources of data include:

- Federal Surveys Department (now Office of the Surveyor-General of the Federation (OSGF)).
- Various State Survey Departments.
- National Population Commission (NPC).
- Nigerian Meteorological Agency (NIMET).
- Abuja Geographical Information System (AGIS).
- Nigerian Geological Surveys.
- National Space Research and Development Agency (NASRDA).
- Statistical Department of various Ministries.
- Various Private Geoinformation Service Providers.
- Research Institutes.

3.2 Sources of Spatial Data

Geospatial data can be generated from the following sources/techniques:

- Land or terrestrial surveys
- Photogrammetric surveys
- Satellite imagery
- Aerial photographs (see Figure 4.10)
- Global Positioning System (GPS) Reading
- Existing maps (analogue or digital)
- Existing digital boundary files, which usually contain geometric description of administrative units
- Geophysical data files
- Digital environmental data files
- Digital Elevation Model (DEM)
- Gazetteer of Geographical Names (Place-names)
- Postcode Directory
- Google Earth

Fig. 4.10: A Gray-Tone Aerial Photograph

3.3 Sources of Non-Spatial (Attribute) Data

- Socio-economic statistical records
- Questionnaire surveys
- Census (demographic and housing) surveys
- Market (customer) survey.

4.0 Conclusion

To a large extent, the source of data determines the quality of that data. Given that there are several possible sources of both spatial and attribute data for GIS operations, effort should be made to acquire data only from authentic and reliable sources.

5.0 Summary

The spatial and non-spatial data used in GIS are usually acquired from different sources and using different techniques. The data could be obtained first-hand as primary data by the user, or it can be obtained as secondary data from already existing sources. Maps, aerial photographs, satellite imagery, surveys, and official statistical records are some of the major sources of GIS data.

6.0 Self-Assessment Exercise

In a tabular form, identify sources of both spatial and attribute data.

Reference/Further Reading

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