CHAPTER - VII

Geographical Information System
Geographical Information System

General Statement

To integrate the multidisciplinary information on the geology, geomorphology, groundwater and soil into a land system database GIS technology has been adopted in the present program. The term GIS stands for Geographical Information system. The GIS is considered a tool for effective and efficient storage and manipulation of spatial and non-spatial data. It is defined as a potential set of tools for collecting, storing, retrieving at will, transforming and display spatial data from real world (Burrough, 1987). The word "geographic" implies that the location of data items are known, or can be calculated in terms of geographic co-ordinates, "information" implies that the data are organized to yield useful knowledge and word "system" implies that it is made up from several linked components with different functions.

The history of GIS dates back to prehistory when early man charted the territory around him and kept mental notes of spatial locations of the objects of his need - food, water and shelter. The advent of the cartography, development of number theory and emergence of computers contributed to the growth of GIS. The advent of computer changed the concept of documentation of spatial information into dynamic system. The GIS developed through integration of capabilities of the computer technology for data base management, graphic representation and spatial analysis.
The early GIS focused on processing attribute data and their geographic analysis, the graphic capabilities were rudimentary. In 1970s the GIS function in pilot and demonstration mode. The GIS uses in 1980s were in Application mode. The linkage of data bases to graphics was a major development (Anon, 1988). The development of the interface devices and software's facilitated online linkages between data bases to facilitate the concept of distributed data bases networking. The development of metadata has operationalized the GIS (Medyckyj-Scott et al., 1996). In India GIS concept has remained in R&D mode, NRDMS, NRIS (NNRMS) and IRDP have attempted to integrate the data bases in spatial information system for development (Fig. 19).

A brief introduction to components of GIS, Database structure and Data structure in GIS have been discussed. The present program has attempted GIS for Land System Studies in parts of Yamuna Basin. The MGE (Modular GIS Environment) Software was used.

The Components of GIS

The components of Geographical Information System are Computer hardware, software and a proper organizational set-up, required for optimal functioning of the system.
Computer Hardware

The desired computer hardware shall include a computer, printer, digitizer, plotter, etc. A high speed computer is desired with color monitor, it should have large disk storage space so as to store the data and images. The inter computer communication can also be made through modem. A digitizer is used to convert the analogue map into a digital map. Plotter is used to present results of the data processing (Fig. 20).

GIS software modules

The software package for a GIS consists of five basic technical modules. (Fig. 21). These basic modules are sub-systems for the following operations:

a. Data input and verification
b. Data storage and database management
c. Data output and presentation
d. Data transformation
e. Interaction with the user

a. Data input and Verification

Covers all aspects of transforming data captured from existing maps, field observations and sensors (including aerial photography, satellites and recording instruments) into a computer compatible digital form. A wide range
of interactive terminal, digitizer, scanners and the devices necessary for recording data already written on magnetic media such as tapes, drums and disks are used as input devices. Figure 22 shows the main input components.

b. **Data storage and database management**

It concerns the way in which the data about the position, linkages (topology) and attributes of geographical elements such as points, lines and areas are structured and organized both in respect to the way they must be handled in the computer and how they are perceived by the users of the systems. The computer program used to organize the database is known as Database Management System (DBMS).

c. **Data output and presentation**

It concerns the way the data are displayed and the results of analyses are reported to the users. Data may be presented as maps, tables and figures (graphs and charts) as ephemeral image on a visual display unit, hard copy outputs drawn on printer or plotter or information recorded on magnetic media in digital form. (Fig. 23).

d. **Data transformation**
Embraces two classes of operation, namely (i) Transformation needs to remove errors from the data sets and (ii) the methods that can be applied to the data in order to achieve answers to the queries asked for GIS. Transformation can operate on spatial and non-spatial aspect of the data, either separately or in combination.

e. Interaction with the user

The last module in the list for geographical information systems mentioned above, that of interaction with the user - query input - is absolutely essential for acceptance and use of any information system. Certainly this is an aspect which deserves much attention. The widespread introduction of PC's and programs that are operated by commands chosen from a menu (a list), or that are initiated by a response to be requested in (an English - like) command language Standard Query Language (SQL) of verbs, nouns and modifiers has broken down the barriers.

Data, Data Base and Database Structures

Data is the basic input, from which information system is developed. The data are organized into records and files, which constitute the database. Database is a collection of data that can be shared by different users. It is a group of records and files that are organized to minimize the redundancy.
Database structures consists of data in many files, in order to access data from one or more files easily, it is necessary to have some kind of database structure. The commonly used data models in GIS applications are Hierarchical, Network and Relational data structures (Fig. 24). The development of multi-disciplinary approach in GIS has necessitated development of metadata (Medyckyj-Scott et al., 1996) which provides data about the data for GIS applications.

Hierarchical Data Structure

The data in this structure have one to many relation as soil series within a soil family, hierarchical method provide quick and convenient means of data access. This data structure assumes that each part of the hierarchy can be reached using a discriminating criterion (Key) that fully describes the structure. (Fig. 25 a)

The advantage of hierarchical system is that they are easy to understand, update and expand. Data access via the keys is easy. This structure is very good for data retrieval if the structure of all possible queries can be known beforehand. This is commonly the case in bibliographic, bank or airline retrieval systems.

The disadvantages of this database structure are that large index files are to be maintained and certain attributes are to be repeated, leading to data redundancy thus increasing the storage and access costs.
Network Database Structures

In this database structure unlike the hierarchical system where travel within the database is restricted to the paths up and down the taxonomic pathways, the linkages are direct and many, m:n linkages. This database structure is used when a rapid linkage is required especially in data structure for graphic features where adjacent items in the map or figure need to be linked together even if the actual data about their co-ordinates may be written in different part of database. Each entity set with its attribute is considered to be a node in the network. (Fig. 25 b).

The network data base for simple structure is easy to develop but is far more difficult for complex structures. The main disadvantage of this structure is that this does not allow links between records of the same type. The Network database structure is more flexible than hierarchical database structure for handling complex spatial relationships.

Relational Database Structures

This is the most popular and flexible structure for GIS. In this type of Database Management Structure, data are organized in two dimensional tables. These table are easy to develop and understand. This structure is called as relational structure as each table represents a relation. The Columns are homogeneous, all items are of the same field. Each column is assigned a
distinct name. Each row has a distinct identifier. Duplicate rows are not allowed. Both the columns and rows can be viewed in any sequence at any time without affecting the information content of the table.

Since different users see different data and different relationship between them, it is necessary to extract subsets of the table columns for some users and to join tables together for others to form large tables. The mathematics provides the basis for extracting some columns from the tables and for joining various columns. This capability is normally not available in hierarchical or network database structures. Each entity set is represented by relations of rows and columns in the form of tables, while each row or tuple in the table represents the data for an individual entity. Each column holds data on one of the attributes of the entire set. Since the relationships between entities are directly represented as tables, there is no need for linkages or pointers between the data records.

The important feature of relational databases are:

- **Primary key** An attribute in the data set of a row which is unique and serve as address is termed Key. When there are more than one keys then one is designated as primary key and others are secondary keys.

- **Relational Joins** It is the mechanism of linking data in different tables. In this the value in column or columns in one table are matched with values in columns in second table and second table can be matched with corresponding columns in third table and so on until the necessary data from the requisite number of tables have been retrieved. Matching is
frequently based on primary key in one table linked to column in the second which is termed as a foreign key.

- **Normal Forms** Certain amount of redundancy in data is implicit as the join mechanism matches the column values between tables. Without careful design unnecessary redundancy may be introduced into the database. A relational model with different values in every row and column position, is said to be normalized. There are two models for a relational database to be normalized:
  a. All the tables should have rows and columns which have no repetition of data.
  b. Every column must be fully dependent on the primary key.

The advantages of relational database system are:

i. Rigorous design methodology based on sound mathematical formulations.

ii. All the other database structures can be reduced to a set of relational tables, which is the most general form of data representation.

iii. Ease of use, implementation, modification and flexibility compared to other types of data models and powerful query language facilities, makes it widely acceptable for GIS applications.

The disadvantages of relational database system are:

In a multi-user environment requirement for processing resources is increased and queries involving multiple relational joins may give slower response than desirable. However indexing can solve the problem of slow response. Availability of faster computers have solved this problem to a large extent.
Relational Database Management System

A RDBMS provides a set of more or less flexible and sophisticated tools for handling information and perform the following functions:

a. Define a database
b. Query a database
c. Add, edit and delete data
d. Modify the structure of the database
e. Secure data from public access
f. Communicate within networks
g. Export and import data

Data Structure in GIS

There are two fundamental ways of representing the data. There are Raster representation and Vector representation.

Raster representation

In this type of representation of the geographical data, a set of cells located by the co-ordinates are used, each cell is independently addressed with value of an attribute.
The Raster Data Structure

The simplest raster data structure consists of an array of grid cells. Each grid cell is referenced by a row and column number and contains a number representing the type or value of attribute being mapped. In raster structure a point is represented as a single grid cell; a line by a number of neighboring cells strung out in given direction and an area by an agglomeration of neighboring cells (Fig. 26). Since each cell in a two dimensional array can only hold one number, different geographical attributes must be represented by separate sets of Cartesian arrays, known as overlays. In its simplest form, the overlay concept is realized in raster data structures by stacking two dimensional arrays, thus resulting in three dimensional structure.

The advantage of raster data structures are:

a. Data structure is simple.

b. The overlay combination of mapped data with remotely sensed data is easy.

c. Various kinds of spatial analyses are easy.

d. Simulation is easy because each spatial unit has same size and shape.

e. The technology is cheap.

Disadvantage of raster data structures are:

a. Volume of graphic data is enormous.

b. The use of large cells to reduce data volumes means that
enomenologically recognisable structures can be lost and this may lead to loss in information.
c. The output maps are burden with in-built cartographic error and boundaries are hatched.
d. Network linkages are difficult to establish.
e. Projection transformation are time consuming.

Vector Representation

The three main geographical entities are points, lines and areas. The cells grid in raster representation are replaced by the points in vector representation. Lines and areas are set of interconnected co-ordinates that can be linked to given attributes.

The Vector Data Structure provides representation of an object as exactly as possible. The co-ordinate space is assumed to be continuous, not quantized as with raster space, allowing all positions, lengths and dimensions to be defined precisely.

The Point entities can be considered to embrace all geographical and graphical entities that are positioned by single XY co-ordinate pair. Besides the XY co-ordinates, other attribute data must be stored to indicate what kind of point it is and the information associated with it.
The Line entities can be defined as linear features built up of straight line segments made up of two or more co-ordinates. The simplest line required the storage of two XY co-ordinates, an origin point and end point. An arc or chain or string is a set of n number of XY co-ordinates pair describing a continuous complex line. The chains can be stored with data records indicating the type of display line symbols to be used.

The Area entities or polygons can be represented in various ways in a vector database. The simplest way of representing a polygon is in form of extension of the simple chain. The names or symbols used to define each polygon are entered as a set of simple text entities.

The limitations of vector representation are:

a. Line between adjacent polygon must be digitized and stored twice.
b. There is no neighbourhood information.
c. Islands are impossible except as purely graphical constructions.
d. There is no easy way to check if the topology of the boundary is correct or burdened with dead ends or topologically inaccessible loops/weird polygons etc.

The vector data structure is useful for archiving phenomenologically structured data and for network analyses.

The vector data structure is used when high quality results in terms of line drawings, etc. are desired. The combination of both vector and raster are used for plotting high quality lines in combination with efficient area filling in
color, the lines are held in vector format and the attributes in compact raster structures.

The raster and vector methods of data representation are distinctly different approaches for modelling geographical information. The raster and vector data are valid methods for representing spatial data and both structures are interconvertible. The conversion from vector to raster is simplest and there are many well known algorithms, performed automatically in many display screens by inbuilt microprocessors. The raster to vector conversion is also possible but is much complex operation and is complicated by need to reduce the number of co-ordinates in the resulting lines by weeding.

Software Used for Digital Image Processing and GIS

In the present study to overcome the deficiencies of vector based GIS (ArclInfo) and raster based GIS (ILWIS), MGE (Intergraph) has been adopted.

The Modular GIS Environment (MGE) software has both the vector and raster capabilities. Besides, it is having Digital Image Processing (DIP) capabilities.
The modules of MGE software are:
MGE Basic Nucleus
MGE Base Mapper
MGE Basic Administrator
MGE ASCII Loader
MGE Grid Analyst
MGE Terrain Analyst
MGE Map Finisher
MGE Image Translator
MGE Base Imager
MGE Advance Imager

The MGE Basic Nucleus

It provides a single, consistent entry point for accessing MGE project data, various GIS software routines and other application products. It offers project management, co-ordinate system operations, data query and access and multiple configuration options as an efficient, affordable GIS baseline.

The MGE Basic Nucleus is foundation of all applications in the MGE Software applications and ensure compliance and integration of all MGE applications.
The MGE Base Mapper

It provides tools for capturing and managing GIS project data in Intergraph's Modular GIS Environment. The MGE Base Mapper includes both interactive and batch tools for data entry, cleaning and manipulation. The module allows users to locate and edit attribute information through interactive query on graphics elements representing features and provide easy to use utilities for feature creation and symbol modification.

The MGE Basic Administrator

This module provides project and database management tools that link the Intergraph Modular GIS Environment to a relational database. It provides project definition and maintenance tools for defining database schema's, map categories, features and attribute tables.

The MGE Grid Analyst

MGE Grid Analyst provide capabilities for creation, query, analysis and display of topologically structured geographical data. It is a powerful spatial analysis tool for building topological modules and querying the relationship between geographic features. The MGE Grid Analyst provide facility for building the topology automatically from map features and attributes then perform sophisticated query on the results. It can create spatial queries, create buffer zones around features, remove common boundaries between features
and create thematic displays of the results of many queries of multiple topological files.

The MGE Terrain Analyst

The MGE Terrain Analyst is used to manage three-dimensional surface features. It uses both TIN and Grid data formats to produce vector and raster outputs. Map projection and co-ordinate systems are built into every surface module. It gives a complete 3-D Is analysis solution. Terrain Analyst has an intuitive, easy to use graphical interface as well as a complementary set of command line function for creating and manipulating digital terrain model data on disk.

MGE ASCII Loader

The MGE ASCII Loader is a tool for loading and unloading ASCII data into MGE project format. Applications include loading survey data, GPS or maritime point data, digital terrain data, etc. The MGE ASCII Loader accepts ASCII files containing co-ordinates for vector geometry and associated attribute information.
The MGE Map Finisher

It provides features based map composition and symbolization for screen display and color plots. The MGE Map Finisher provides symbolization and data re-organization capabilities so that user defined features representation may be generated for plotting. To create hardcopy plots Intergraph’s I/Plot software may be used to plot to any number of devices including thermal, inkjet, electrostatic and pen plotters.

The MGE Image Translator

The MGE Image Translator provide automatic access to the most commonly used satellite and aerial photograph image formats which are commercially available. It provide an option for reading and writing generically formatted data, as well as utilities for renaming and copying tapes. All the functions are supported by user friendly Window interface. It provides image input and translation of SPOT, SPOTVIEW, AVHRR, ERS, EOSAT and USGS data. Besides these, it can access all the data in BIL, BIP, BSQ and TIFF formats. Image Translator is a beneficial product for GIS, remote sensing image mapping and photogrammetric specialists having a requirement to convert raw data from different agencies selling the data into widely accepted formats.
The MGE Base Imager

This is a basic level desktop image processing tool providing facilities of image display, enhancement and manipulation. MGE Base Imager is suitable to meet the demands of intensive image mapping and cartographic production environments. It provides sophisticated gray scale image processing and color image display for a variety of image oriented applications.

The MGE Advanced Imager

The MGE Advanced Imager is a fully functional multispectral Image Processing environment. It is having the capabilities for training, classification and post classification routines besides full set of enhancement tools. MGE Advanced Imager includes a set of advanced functions for gray scale image processing; including Fourier transformation and filtering, noise reduction, filters and texture processing for radar and multi-spectral imagery and complex logical operations. All the image enhancement, analysis and display features are having user friendly windows interface and are supported with associated command line options for better processing.

Microsoft Windows NT

It the Operating System used to run the MGE softwares. Windows NT is a 32-bit operating system, which offers multitasking, multiprocessing and
multithreading features. The architecture of Windows NT is designed to offer the most reliable desktop environment. Windows NT is designed to offer the best networking support and applications compatibility while protecting data with full security in heterogeneous environment. Windows NT offers long term vision, it is highly responsive, scaleable and portable.

**Microstation 95**

Microstation 95 is a very powerful CAD software. The Microstation is a common Cad engine for total integration among CAD, GIS and Image Processing software's. It provide graphic foundation for all MGE application softwares. It adds further to the existing capabilities of any CAD software, it is faster, gives superior production tools and allow to customize working environment to match the CAD requirement of user’s decipline.

**The Sybase SQL Server**

The sybase Standard Query Language is used as a GIS database management tool. It creates geographical features with intelligent database linkages. SQL database tool helps to update and maintain existing data and incorporate new data, while ensuring data security and integrity. SQL query perform multi-theme spatial analysis and query. Query results are viewed as a report from the database or as a graphic display with user-specified symbolization.
The work flow in MGE Software is:
Creating directory
Creating Schema
Locating and Activating Schema
Creating Project
Defining MCSO working units and Registration of maps
Digitization of maps
Feature building
Map Cleaning
Graphic processing
Attribute Processing
Non-spatial data entry
Building a topological file
Creating query set
Query display
Overlay analysis, etc.

The procedure adopted during the present study is described as under:

Creating directory The directory is to be created while starting Microstation, open the directory menu in the Microstation menu by clicking the mouse button on it and enter the name for new directory in the directory field.

Create Schema For creating Schema the command to be followed in sequence are MGE > Utilities > RIS Schema Manager > Schema definition >
Create Schema > Schema Name > Get Client Address > Define project directory > OS Type > Database name > DBMS location > OS Username > Apply. Schema Created > Close.

Locating and Activating Schema Show Client Location > Locate Client > Click local > Apply > Client located successfully > Close.

Show Schema File Location > Locate Schema File > Enter file name > Apply > Schema File located successfully > Cancel.

Creating Project MGE > File > New Project
Enter working directory press TAB

Applications > associate applications which are required for the project > OK.

Defining MCSO working units and Registration of map

Microstation > Applications > MCSO > File > Co-ordinate System > Primary > Define Projection System > Define Parameters - Latitude of origin; Longitude of origin > Define the Geodetic datum

MCSO > File > Working Units > Map scale > OK.

Microstation > Main Palette choose Place Active Point and Change elements Attribute (Color and Weight).
Microstation > Applications > MCSO > Keyin > Primary > Geographic Enter latitude and longitude (co-ordinates) of tic points.

Microstation > Settings > Tablet > Partition define the limit of the map > OK. Microstation > Settings > Digitizing > Set-up > Use digitizer cursor first button > Click at tic point 1 > Tic at corresponding point on the screen and similarly all four tic points will be referenced.

Then the digitizer cursor will be bought in the area bounded by 4 tic points and button number 2 of the digitizer will be clicked. The level of error will be displayed and if it is < 1.0 then accept. Close Digitizing tablet window.
Hierarchical

Network

Relational

Database Structures in GIS

Fig. 24
Geographical Information Systems are the result of linking parallel developments in many separate data processing disciplines.

The major hardware components of a Geographical Information System include:
- Digitizer
- Disk Drive
- CPU
- Plotter
- Tape Drive
- Visual Display Unit

The main software components of a Geographical Information System are:
- Data Input
- Query Input
- Geographic Database
- Display and Reporting
- Transformation

Fig. 19
Fig. 20
Fig. 21
Fig 22. Data Input

Fig. 23. Data Output
Fig. 26. Elements of a map. A. Point, B. Line and C. An area

Fig. 25a An example of Hierarchical data Structure

Fig. 25b An example of Relational Data Structure