CHAPTER - III
DATABASE AND METHODOLOGY

3.1 DATA AND DATABASE

Geographical or spatial data are defined as undigested, unorganized, and unevaluated material that can be associated with a location. Geographical data include facts, results of observations, original remote sensing images, basic census figures and statistics all of which are gathered and communicated to the user. Data are of little value in or of themselves. To be useful they must be transformed into information. When data are organized, presented, analyzed, interpreted and considered useful for a particular decision problem, they become information. Accordingly, geographical information is defined as georeferenced data that have been processed into a form that is meaningful and of real or perceived value to decision makers. Decision problems that involve geographical data and information are referred as spatial decision problems. The terms spatial data and geographical data are often used interchangeably (Malczewski, 1999).

While using the tools of Geoinformatics, a database is regarded as an organized collection of data with software to provide to access in different ways. The smallest unit of data has meaning to its users is called data item. It has traditionally being called a field and is now called data element or elementary item. Data items are the molecules of the database. There are atoms and sub-atomic particles composing each molecule (bits and bytes), but they do not convey meaning in their own right and so are of little concern to the users. When the data items are grouped together it forms the records (Martin, 1992). The data item or the field represents the kind of information stored in columns and the rows of the table form the records in a database. It is to be remembered that the database has a data model and data structure and a representation. In database literature, the term data model applies to logical structure of entities and relationships. As the rows in a database represent the records of geographic or any other entities and the columns represent the kinds of information required or generated about the entities for a specific purpose, the database may be organized in the form of a data matrix. The geographical matrix (Berry, 1964) has rows for places and columns for attributes of those places. An
Attribute can be defined as any property that distinguishes a geographical entity. The most essential property of the attributes is that their values vary over geographical space. Therefore, attributes may be referred as variables. Each attributes accounts for a column in the geographical matrix.

Geographical data can be arranged in two distinct tabular forms: a spatial structure data matrix and a spatial interaction data matrix. These two forms are also referred to as a geographical data matrix and a spatial behaviour data matrix respectively. The geographical data matrix approach takes the identity of the objects for granted, and then expects the attributes to be measured for each object. The geographical matrices are the most fundamental techniques for organizing data to be used in spatial decision problems and analysis. Many decision situations require data on both attributes and spatial relationships. Accordingly, a combination of two tabular representations can be used to organize the input data for spatial decision-making.

The organized database in the form of geographical matrix has two fundamental approaches to the analysis of data set, that are filled with all relevant data of entities and their attributes. First, any column of the matrix represents a single spatial distribution pattern, which can be converted to a map format to show geographical distribution of a particular attribute (e.g. land use or population density). Analysis of a single spatial distribution is referred to as single-dimensional analysis. Taking into consideration two or more columns of the matrix, one can study spatial relations and use the data for spatial multi-criteria decision analysis (Malczewski, 1999). Second, each row indicates a character of a particular geographical unit; that is it describes a single geographical entity in terms of a set of attributes. A third dimension can be added to the two-dimensional geographical data matrix to represent the dynamics of a spatial pattern and place characteristics over time, the dynamics can be represented by a number of two-dimensional matrices, showing the data at different points in time (Wilson, 1974; Bennet, 1979). A sequence of interaction matrices can represent changes in the spatial pattern over time (Wilson, 1974; Wilson et al., 1981).

When the rows and columns of the database matrix both refer to the entities of the geographical system, the relationships between the entities can be obtained as the interaction matrix. These relationships may be either measured or assessed levels of
interaction between the two entities. They can also be expressed in terms of distance, time, cost of getting from one location to the other or the degree of connectivity between the two entities. Thus each cell of the matrix is filled by a value or score, which measures the interaction or relationships between the entities. For example, the matrix can represent the volume of movements (flows) of people, goods, services, information or some other entity.

A third dimension can be added to the two-dimensional interaction data matrix. An information system organizes data into a structure, so that simple observations can be converted into more useful information.

A GIS serves the distinctive needs when geographic descriptions play a central role in the observations. The geographic measurements require choices that can be organized as measurement frameworks. Differences in the measurement frameworks best explain the technical choices of representation for geographic information; measurements and representation, in turn strongly influence the operations that can be performed with the information (Chrisman, 1997). Finally transformation can convert from one measurement framework to another. In the observation, measurement and transformation processes, the data of various kinds are generated. Accordingly the data in a GIS framework are very simple to very complex.

### 3.2 DATABASES USED IN THE STUDY

In a very broad sense, the two kinds of databases are used in this study. The first is the spatial database or map database, which may also be referred to as geographical database. The other kind of database is on the attribute information of the spatial data.

As per the objectives of the study outlined in the foregoing chapter, the spatial digital databases have been developed for various types of map features or map units. Some of the features are developed interactively during the process of spatial operations on various occasions. Spatial or the map database contains the geometric attributes of respective features. These are generally developed in digital vector format as Arc Info coverage and at a later stage converted into Arc View shape files or other formats for further analysis. The spatial database that have been created for the following types of map units or the features are: -
1. Polygon features of administrative units - 156 villages and 20 Gram Panchayats and the block
2. Polygon features for various land use classes, various physical units of geology, geomorphology, physiography, relative relief, watersheds, soils, wetlands and relevant features.
3. Linear features for boundaries of villages, Gram Panchayats, block, contours, roads, railways and drainage network
4. Point features of villages, spot heights, sample sites and important places
5. Digital image database created from the satellite data.

The important attribute databases to be integrated in spatial database of polygon, line and point features stated above and these are created from various sources (subsequently to be described in details), which can be grouped into several aspects. Some of the very important and relevant aspects in management of resources and infrastructure in Kathiatoli Development Block are:

1. Database on physical characteristics – such as altitude, relief, slopes aspects, geological and geomorphological, climates, soils, natural vegetation and forests, drainage networks, watersheds, ground water levels and relevant derived variables for various spatial features that are outlined above.
2. Database on demographic characteristics and some derived parameters for certain type of analyses such as density or dependency ratio, man land ratio and many others, which are to be integrated on the point and polygon map features in carrying out further spatial analysis and modelling.
3. Database on socio-economic and cultural characteristics of population and their derivatives
4. Database on amenities and infrastructure facilities available in various locations
5. Database on the classified or analyzed digital data, either for map features or for the extracted features from satellite images
It is important to note that in digital environment a large volume of some temporary or intermediate data set including their spatial and attribute databases are created during the process of spatial analyses, modelling and integration of various themes. Many of them are converted into permanent spatial database along with their attributes for further analyses.

3.3 DATA SOURCES

Generally, precise maps at village level meeting all cartographic standards are not available. This is a great problem in developing spatial database for GIS. A wide variety of data sources have been adopted to generate spatial data and their attributes. Specifically, mention may be made of the following sources:

1. Office of the Deputy Commissioner, Nagaon District, Assam
2. Office of the Sub-Deputy Collector (SDC), Kampur and Nagaon Revenue Circle, Nagaon District, Assam
3. Office of the Senior Block Development Officer, Rengbeng, Kathiatoli, Nagaon District, Assam
4. Office of the Sub-Divisional Officer, P. W. D., Nagaon, Assam
5. Directorate of Census Operations, Guwahati, Assam
6. Directorate of Panchayat and Rural Development, Guwahati, Assam
7. Survey of India, Dehradun
8. Geological Survey of India, Kolkata
11. Field Observation

The planning cell of the district administration is maintaining the records of the villages at block level in a map on 1:63,360 scale. This map is based on the topographic maps prepared by Survey of India, Dehradun. Similarly, the cadastral maps of the revenue villages at the scale of 16 inches to a mile are available with the local revenue authorities. These maps are used as reference for the villages of the study area. The census authority has maintained a map for the villages of the block that has been based on
the map prepared by the Directorate of Land Revenue, Assam. This map is also available on the scale of 1:63,360. The list of the villages under Kathiatoli Development Block is maintained by various sources for different purposes, which do not match each other. So, to avoid the anomaly on the list of villages, the available records from the District Rural Development Agency, Nagaon; Office of the Senior Block Development Officer and Office of the Deputy Commissioner, Nagaon District and Office of the Sub-Deputy Collector (SDC), Kampur and Nagaon are consulted to finalize the villages and GPs. Accordingly, the village and GP level maps are compiled and spatial database for these as polygon features are developed.

The geological information has been extracted from the map prepared by Geological Survey of India. This map was drawn at the scale 1:250,000 for the Kopili basin. The spatial database has been developed for the basin and the study area is taken out from the basin area, as a result the information is highly generalized.

Similarly, the information on soils are given on the basis of the soil map prepared by the National Bureau of Soil Survey and Land Use Planning, Regional Office, Jorhat and its national head quarter at Nagpur. This map is available for the districts of Assam at 1:250,000. Hence, the details in this respect are also found in generalized form.

One of the most important data sources that has been used to develop some database layers such land use, roads, contours, drainage and location of important places is the topographical sheets of Survey of India at 1:50,000 scale. To have full coverage of the study area, the required topographical sheet numbers are – 83 B/11, 83 B/12, 83 B/15, 83 B/16, 83 C/09 and 83 C/13.

In preparing resource maps for natural resources of the area, IRS and LANDSAT TM digital data have been acquired from the National Remote Sensing Agency, Hyderabad for various dates / time periods. The digital data products used in this study are-

1. IRS-1A LISS-I Full Scene Digital Data in Bands-1, 2, 3 and 4 for Path 015 and Row 049 corresponding to the topographical sheets (83 B/11, 83 B/12, 83 B/15, 83 B/16, 83 C/09 and 83 C/13) at scale 1:50,000. The date of acquisition was 26th January, 1989.
2. LANDSAT – TM Full Scene Digital Data in Bands - 2, 3, 4 and 5 for Path 136 and Row 042 corresponding to the study area. The date of acquisition was 27th January, 1994.

3. IRS-1C LISS-III Full Scene Digital Data in Bands-1, 2, 3 and 4 for Path 111 and Row 053 corresponding to the study area. The date of acquisition was 10th March 1997.

4. IRS-1B LIS-II Full Scene Digital Data in Bands –1, 2, 3 and 4 and Geocoded Photographic Data in Bands – 2, 3 and 4 for Path 015 and Row 049 at the Scale 1:50,000. The date of acquisition was 4th January, 1998.

5. IRS-1D LISS-III Full Scene Digital Data in Bands - 2, 3, 4 and 5 for Path 112 and Row 053 corresponding to the study area. The date of acquisition was 4th January, 2000.

6. IRS-1C LISS-III Full Scene Digital Data in Bands-1, 2, 3 and 4 for Path 111 and Row 053 corresponding to the study area. The date of acquisition is 27th November, 2002.

7. IRS-1D LISS-III Full Scene Digital Data in Bands - 2, 3, 4 and 5 for Path 112 and Row 053 corresponding to the study area. The date of acquisition is 2nd February, 2003.

3.4 **TOOLS AND TECHNOLOGY**

While dealing with various kinds of data in digital environment, computer hardware and software become very essential. There is unbelievable progress and development in hardware and software in 1990s. Varieties of computer platforms are now-a-days available ranging from a personal computer to super computer. Similarly, there are increasing number of GIS and Digital Image Processing software in the market with various capabilities and functionalities. These are available in very low cost personal budget to very high cost institutional infrastructure budget. It is worth mentioning here that during the period of this study, there was a sea change in configuration of hardware and software systems that had taken place. The new sets of tools with minimum
configurations of various items that are used presently appear in the brackets along with the earlier ones listed below. Some of the tools like digitizing systems are obsolete today.

**Hardware**

**Personal Computer System**

- Pentium-II at 233 MHz (Pentium-IV at 2.4 – 3.0 GHz)
- 4.3 Giga Bytes (GB) Hard Disk Space (40 - 80 GB Hard Disk)
- 128 MB RAM (256 - 512 MB RAM)
- Digitizing System - Summagrid V (High Resolution Colour Scanning System)
- Ink Jet Colour Printers, with printing speed 1 page per minute (ppm) (Various Laser Jet Printers, with printing speed 14 ppm or more / Very High Speed Plotters)

**Software**

- Windows 95 / 98 (Windows 2000)
- dBase-III (Visual dBase for Windows)
- Data Automation Kit (PC Arc Info Version 4.5 or More)
- Arc View G. I. S. Version 3.2 (Version 8.3 or More)
  - With the application modules – ArcMap, ArcCatalog, Spatial Analyst, 3-D Analyst, Network Analyst, Arc Editor, Geostatistical Analyst and many more.

In addition to the GIS software that have been used, the following new user friendly and versatile products capable of Digital Image Processing and GIS are now available and the important functionalities and modules of these have been intensively used in the work:

**Geomatica Version 9.0**

- The main application modules in this programme are - Geomatica Prime, OrthoEngine and Fly!

**ERDAS Imagine Version 8.4**

- The main application modules in this programme are – Viewer, Import, DataPrep, Composer, Interpreter, Catalog, Classifier, Modeler, Vector, Radar, VirtualGIS and OrthoBase.
3.5 METHODOLOGY

In this section, an attempt has been made to describe general methodology for data creation, modelling and their transformations. The methodology for fulfilling specific objectives using various sets of created data will be described in relevant section of the study. The major components of GIS and digital image processing – data and data sources, input and output devices in terms of hardware and software have already been outlined very briefly. The desired results are highly subjected to the limitations of various components of the system and the working knowledge and experience of the user. Similarly, the methodology will also vary depending on the kinds data available and / or the kinds of data to be used to get the best results as per the objectives.

The data models that have to be prepared for various operations in GIS and Digital Image Processing need very careful attention, because they are very complex in nature. One has to deal with the graphic elements of various types besides the numerical, textual or logical elements coherently involved in each graphic element. Moreover there are hierarchies of these graphic elements. It is to be remembered that the computation process is solely dependent on the machine hardware and the logic of applications with spatial reference. In a computer database, location is based on analytical geometry. For practical purposes, the spatial component of geographic information is represented in the form of co-ordinates, which are ordered measurements inside a spatial reference system. These measurements may be angles on an ellipsoid or orthogonal distances on a projection plane. There are different ways to encode equivalent measurements, even on a plane. Each GIS package will implement a certain range of alternative spatial reference systems as specific representations of the mathematical possibilities.

The two predominant models of representation of space in GIS are - vector and raster. Based on analytical geometry principles a vector model builds a complex representation from the primitive objects for the dimensions: points, lines and areas. These primitives have a nested dependencies: areas are described by closed boundary lines and location of a line can be approximated by string of line segments connecting a series of points following a coordinate system at the base. The other major family of representation models is called the raster. The raster model divides the region into rectangular / square building blocks generally referred to as grid cells or pixels that are
filled with measured attribute values. Many hardware systems like remote sensing sensors, printers or the cathode ray tube displays all contributed to the popularity of raster structure. A spatial data structure that organizes a hierarchical structure of square cells through iterative division into four daughter cells is called a quadtree.

The geographic vector data model (map data) have been created using Data Automation Kit / PC Arc Info Version 4.5 and Summagraphic-V digitizing system. The vectors that have been created are further processed and transformed into geographic co-ordinate system from the digitizer co-ordinates taking the spheroid as the map model and Everest as the datum. Finally, these vectors data are converted into Universal Transverse Mercator (UTM) projection system with co-ordinates in metres, the World Geodetic System, 1984 (most commonly referred to as WGS84) as reference map model and the datum for such transformations and data modeling processes. The satellite images are geometrically corrected with reference to the Survey of India topographic sheets after Georeferencing in geographic co-ordinates. Like vector data, the raster maps are also transformed into UTM with the same datum. The vector data, which are converted into rasters for various types of analysis and modeling, are also treated in the same manner. Visual dBase for Windows and Database Management Systems (DBMS) of Arc View Version 3.0a and Arc GIS Version 8.3 are used extensively in building of attribute databases. Quantitative assessment and graphical representations have been made and analyses have been made for various themes individually or in combinations. Geoprocessing tools and overlay tools in the GIS package have been used in integration of themes. The principles of Relational Database Management System are followed in integrating the information. The cartographic representation of collected information, integration and analysis of databases have been done by Arc View GIS, Arc View 3-D Analyst and Arc View Spatial Analyst. In some cases, the functionalities of ILWIS 2.2, Geomatica 9.0 and ERDAS Imagine 8.4 have been used. All the available tools will fruitfully be utilized to derive meaningful layer of thematic information and generating spatial models by integrating the databases.
Fig. 3-1: Schematic Diagram on Database and Methodology
The standard methodology of data collection, classification, integration and interpretation has been adopted to build the database on various aspects. In a very broad manner, the schematic representation and interrelationships among various components of database and methodology is shown in Fig. 3-1. Some of the components shown in the schematic diagram represent more than one aspect such as data and database also indicate the data sources. Similarly, the Database Management System (DBMS) indicates the methods, tools and generation of information from various data. All these processes/stages of methodology have been performed in three distinct phases, viz.-

(1) Pre-field work
(2) Field work
(3) Post field work

3.5.1 Pre-Field work

This work primarily consists of the selection of the study area, collection of study materials, collection of necessary data and maps, preparation of manuscript maps and development of digital databases. The list of performed activities are summarized below:

(a) Collection study materials and development of skills on tools and technology.

(b) Identification of villages and Gram Panchayats of Kathiatoli Development Block, Nagaon District, Assam.

(c) Preparation of base map for the Kathiatoli Development Block using Survey of India topographic sheet at 1:50,000 scale and the map of the villages prepared by the revenue authority/census authority.

(d) Preparation of manuscript maps for the various themes such as revenue villages, Gram Panchayat, drainage, roads, contours, significant features and land use units.

(e) Acquisition of satellite data in photographic and digital format.

(f) Visual interpretation of satellite data and topographical sheets of Survey of India in preparation of tentative (preliminary) thematic maps on – geomorphology, geology, hydrology, soil, drainage, slope/relief, land use etc. of the study area.
(g) Collection of socio-economic and cultural data from various sources – official records at different levels, published reports and materials.
(h) Selection of sample points / areas for ground verification of the attributes.
(i) Preparation of attribute database at village and Gram Panchayat level from the available data from Census of India and other sources.
(j) Digitization of manuscript maps and creation of spatial database.
(k) Digital Image Processing of satellite data to create themes.
(l) Map data modelling and transformations.

3.5.2 Field work

This work is essentially targeted for verification of some information that have been collected through secondary data, generation of some primary information specially on social, economic, cultural and demographic variables and it is intended for collection of ground truths to ascertain some features identified on satellite data including the GPS based collection of ground control points. The major tasks accomplished in this section are:

(a) Verification of data on interpreted details on Survey of India maps, satellite data, and available official records through fieldwork.
(b) Collection of data on resources, amenities and infrastructures of all the villages of Kathiatoli Development Block.
(c) Collection of data on demographic and socio-economic parameters through standard questionnaire from nearly 10% households.
(d) Verification of block boundary and road network at some selected sites.
(e) Collection of geographic co-ordinates at various road junctions using Global Positioning System for rectification of satellite data.

3.5.3 Post - Field Work

The most crucial portion of the study is the post-field work session. It consists of finalization and authentication of data from primary and secondary sources, development, processing, integration and analysis of digital databases; creation of thematic, query and criteria based maps including spatial modelling. Intermediate data layers have been
created in this stage for various kinds of analysis and synthesis to achieve targeted objectives of study.

(a) Generation of thematic layers on data interpreted and verified by fieldwork.
(b) Creation of spatial database from maps on various themes and attribute database.
(c) Integration and Management of themes using GIS functionality and tools.
(d) Query of information layers with specific criteria and generation of composite map output.
(e) Overlay of need specific thematic layers using GIS to evolve action plans / strategies / models.