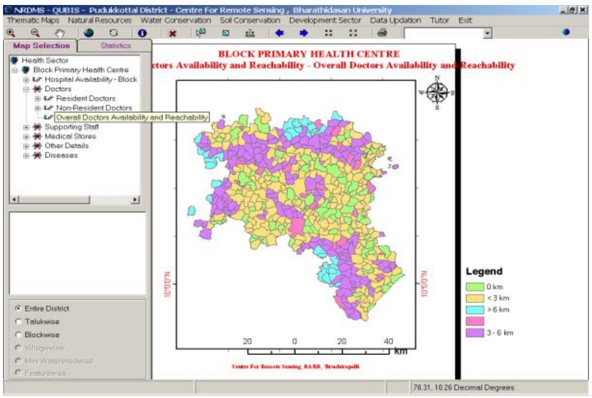
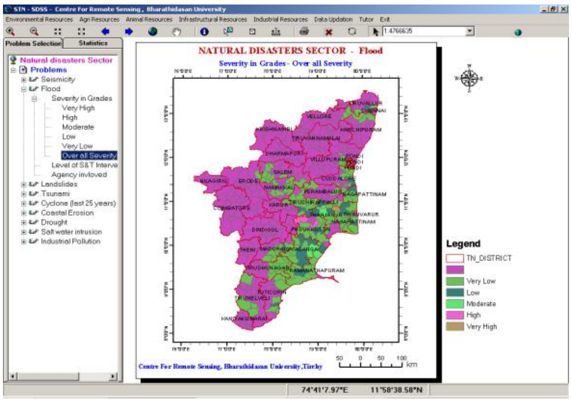
CENTRE FOR REMOTE SENSING
BHARATHIDASAN UNIVERSITY
TIRUCHIRAPPALLI-620023

6 Yr. Int. M.Tech. Geotechnology and Geoinformatics

Reading Material
on

SPATIAL DECISION SUPPORT SYSTEM
Paper code: MTIGT1104





Prepared by

Dr.K.Palanivel
Assistant Professor
Centre for Remote Sensing
Bharathidasan University
Khajamalai Campus
Tiruchirappalli – 620 023.

CENTRE FOR REMOTE SENSING
BHARATHIDASAN UNIVERSITY
TIRUCHIRAPPALLI-620023

6 Yr. Int. M.Tech. Geotechnology and Geoinformatics

September 2015

Paper code: MTIGT1104 – SPATIAL DECISION SUPPORT SYSTEM

Syllabus

1. **Introduction to Decision Support System:** Definition – Concepts – Multicriterian Approach – Usefulness.

2. **Designing of Spatial Database:** Identification of Geographic features – attributes & data layer – Defining the storage parameters for each attribute – ensuring of co-ordinate registration – map projection – Transformation.

3. **Designing of Non spatial Database:** Creation of data table file to hold the attributes – Adding up of description attribute values to table – Different types of sources of data entry – Checking for errors.

4. **Linking of Spatial Database & with Non spatial Database:** Verifying of common item, availability and joining of attribute – table with existing spatial records – spatial display of non spatial data.

5. **Designing & Coding of QUBIS:** Planning for the user requirement – preparation of spatial & Non spatial relational databases – QUBIS Designing – QUBIS Coding (Testing, Error handling, Monitoring, User interface Development.

SPATIAL DECISION SUPPORT SYSTEM (SDSS)

UNIT - 1. **Introduction to Decision Support System:** Definition – Concepts – Multicriterian Approach – Usefulness.

1.1 Definition of SDSS

Decision Making: The process of defining a problem and its environment, identifying alternatives, evaluating alternatives, selecting an alternative and implementing the decision is known as 'Decision Making'.

Decision Support System: An interactive and computer based system that can help decision makers utilize data and models to solve *unstructured or semi structured problems*.

Structured / Routine / Programmed Decision: A type of decision that is repetitive and routine decision situations for which solution techniques are already available is called 'Structured Decision'.

Unstructured Decision: Unstructured decision is one in which the decision maker must provide judgment, evaluation and insight because the decision problem is novel, non-routine and has no agreed-upon procedure for solving it.

Spatial Decision Support System: An interactive computer based system designed to support a user or group of users in achieving a higher effectiveness of decision making while solving a semi structured spatial decision problem

GENERIC CHARACTERISTICS / USEFULNESS OF SDSS / DSS

1. An explicit design to solve ill-structured problems
2. Powerful and easy-to-use user interface
3. Ability to continue analytical models flexibly with data
4. Ability to explore the solution space by building alternatives
5. Capability of supporting a variety of decision making styles and
6. Allowing interactive and recursive problem solving.

DISTINGUISHING CAPABILITIES AND FUNCTIONS OF SDSS

1. Provide mechanisms for the input of spatial data
2. Allow representation of the spatial relations and structures
3. Include analytical techniques of spatial analysis, and
4. Provide output in a variety of spatial forms including maps.

Information and Decision Making

Since the geographic data are raw materials which can not be used directly for decision making, the data are processed by imposing some organization on the measurements to obtain relevant information. Successful decision making depends on the quality and quantity of information available to decision makers.

Broadly, two different types of information are used in the spatial decision making process:

(i) geographical information and

(ii) Information about the decision maker's preferences.

The amount and type of information required depend on the nature of a particular decision problem. Spatial problems typically involve a large number of alternatives that are evaluated on the basis of multiple criteria. Some of the criteria may be qualitative while others may be quantitative.

There may be more than one decision maker or interest group and they may have different preferences with respect to the relative importance of the evaluation criteria and the decision consequences. The spatial decision problem may involve uncertainty and imprecision (fuzziness) that make it difficult to predict the outcome of particular decisions.

1.2 Concepts of SDSS

GIS AND DECISION SUPPORT

The ultimate aim of GIS is to provide support for making spatial decisions. The GIS capabilities for supporting spatial decisions can be analyzed in the context of the decision making process. There are a number of frameworks for analysis of the decision process. Any decision making process can be structured in to three phases:

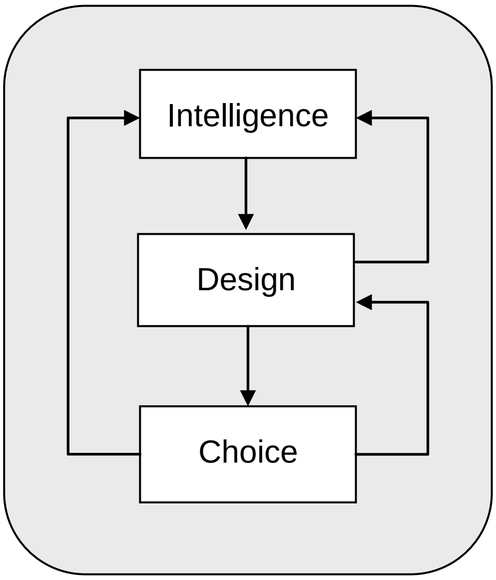
(iii) Intelligence (is there a problem or an opportunity for a change?),

(iv) Design (what are the alternatives?), and

(v) Choice (which alternative is best?).

The three stages of decision making do not necessarily follow a linear path from **intelligence**, to **design**, to **choice** (Figure 1). At any point in the decision making process, it may be necessary to loop back to an earlier phase. For example, one can develop several alternative plans at the design stage but may not be certain whether a specific plan meets the requirements for the decision problem. This requires additional intelligence work. Alternatively, one can be in the process of implementing a decision, only to discover that it is not working, forcing one to repeat the design or choice stage. Each stage of the decision making process requires different types of information. In the context of spatial decision analysis, the critical question is: How and to what extent can GIS provide the support (information) required at each of the three stages of decision making?

Figure 1



COMPONENTS OF SPATIAL INFORMATION SYSTEMS – QUERY BASED INFORMATION RETRIEVAL SYSTEMS (QUBIS)

An automated spatial information system is a toolbox for representing views of the real world via data about locations. Spatial information systems are a technology for processing spatial data. The tools, which may be activated by pushing a button or typing a command, represent processing functions or operations for example, drawing a map, or measuring the distance between two places. The tools work on some or all of the information stored in some systematic way in a database. The information system requires data, software, hardware, 'brainware' and other resources, and exists within some institutional setting as a **resource** to solve problems.

The support for decision making, practical or academic, may consist of virtually instantaneous answers to questions, a **query** type of system. Or the information system may be oriented to producing information, perhaps in the form of maps or tables, for subsequent study. **Products** may be detailed maps for public briefing purposes in a re-zoning case, or may be the complete set of possible routes through a city street system. Whatever the medium, the output is designed for further study, is generally extensive and is not designed for immediate consumption. In contrast, the query mode provides quick responses to highly focused questions usually in the form of small amounts of information.

In providing support for decision making to the user such as a query system and output medium, spatial information should fulfill the following:

1. Provide tools for the creation of digital representations of the spatial phenomena, that is, data acquisition and encoding.
2. Handle and secure these encodings efficiently, by providing tools for editing, updating, managing and storing; for reorganizing or conversion of data from one form of another, and for verifying and validating those data.
3. Foster the easy development of additional insight into theoretical or applied problems, by providing tools for information browsing, querying summarizing and the like: i.e., facilities for analysis, simulation and synthesis.
4. Assist the task of spatial reasoning, by providing for efficient retrieval of data for complex queries.

5. Create people compatible output in varied forms of printed table, plotted map, picture, scientific graph and the like.

The **spatial information system** lies within a large framework that includes physical, informational and guidance systems for phenomena. In a spatial information system context, there is the component of observable phenomena, the component of ‘knowledge’ about those phenomena, and the component of guidance (the planner, forester and politician). Then the information subsystem itself has a physical component (the h/w and data encodings), the documentation component describing the characteristics of the spatial information system, and the guidance component (the person who looks after the database and tools, and solves malfunctions).

Physical Components Spatial Information System

The spatial information system is a software product which has several components and makes connections with other devices in its environment (Figure 2). In structure it includes a database management system (DBMS) for storing and managing data, linked with a graphics management system for cartographic or other visual displays. These two software subsystems are connected in one direction to the computer operating system, and in the other direction, through graphic workstations, to users by means of an interface and a command language interpreter.

The main devices used in a spatial information system are disks for storing various data (alphanumerical, graphic and image), digitizing tablets and scanners to enter graphic data, and plotters and printers to present the results. Similarly, as for all computer systems, it is valuable to have the spatial information system computer linked to a communication network allowing the exchange of data with other people or companies providing or working with geographic or other spatial data.

Disciplines and Fields of Study for Organization of Materials involved in Spatial Information System

1. Disciplines that have developed concepts for dealing with space: Cognitive Science, Geology, Geography, Linguistics, Psychology

2. Fields that develop practical tools and instruments for obtaining or working with spatial data: Cartography, Geodetic Science, Photogrammetry, Remote Sensing, Survey Engineering.
3. Disciplines that provide formalisms and theories fundamental to our working with space and automation: Computer Science, Geometry, Informatics, Artificial Intelligence, Semiology, Statistics.

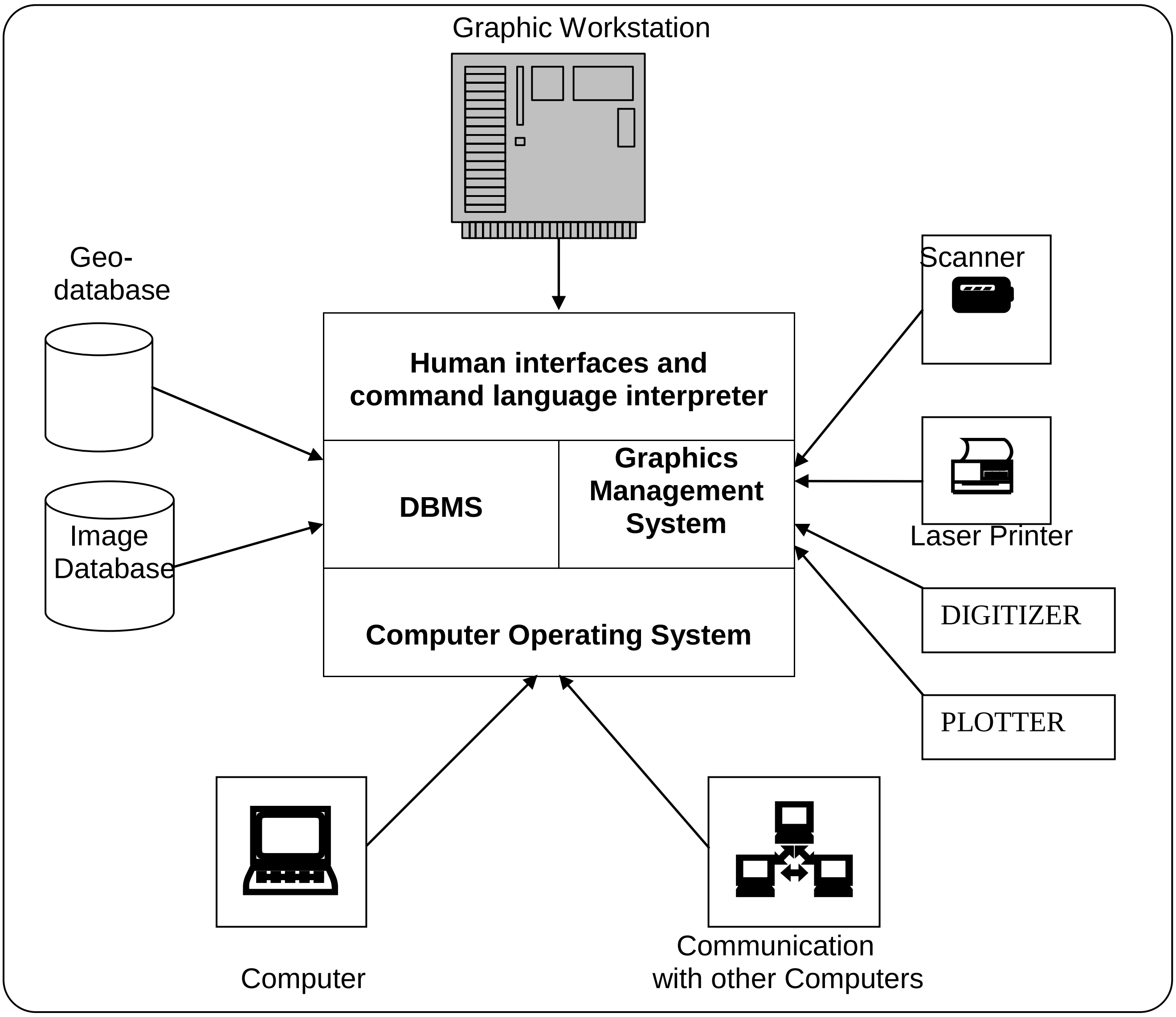


Figure 2: Physical Elements of Spatial Information System

4. Fields making substantial use of automated spatial information systems: Archaeology, Civil Engineering, Forestry, Geotechnics, Landscape Architecture, Urban and Regional Planning.
5. Fields that provide guidance about information: Law, Economics.

The strength and intellectual interest of the spatial information systems subject derives from the cross-fertilization among these many fields. While some scientists use the term geographic information systems to refer to the composite field of study, we prefer the term **geomatics** as an umbrella covering all the fields listed above that are today important for understanding and further developing spatial information systems.

Ideas and tools from many fields are, then, necessary for a full understanding of spatial information systems. As user requirements and expectations increase for such resources and toolboxes, so there are intellectual pressures on the theoreticians and practitioners of geomatics to cooperate in interdisciplinary research studies and development projects to further improve our automation tools.

Our focus is on organizing principles and instruments that help us to understand, evaluate and design spatial information systems – in a sense, a course in application based on many examples, much like an understanding of fine art comes about by looking at and studying many pictures.

While the concepts of data structuring, concepts of spatial relationships, and matching tools of drawing and tabulating existing in the non-automated world have been taken to the automated world, it has not been without cost. So we presume, on the one hand, that the spatial information system in a computerized environment allows some problems to be handled more efficiently and cost effectively. On the other hand, we expect to be able to undertake problems otherwise considered insoluble. Automation allows faster creation and use of large quantities of information, speeds up mathematical or logical operations on data and makes possible tasks that are otherwise unmanageable.

The **spatial information systems,** then, is a computerized environment whereby utility programs performing specific functions are used in an integrated environment, in which the user is shielded from the details of computer processing, to achieve some goal of research, education or decision making. The inherent form of spatial data representation and organization must be designed to support effectively and efficiently the kinds of query and analysis required by many users. The performance of computer systems is a reflection of hardware technology and software engineering, but also reflects the data structures and the quality of algorithms.

Conceptual modeling, is a common ground among users, in part the universe of discourse, in part the universe of modeled entities. Using tools and formalisms the phenomena of interest are identified, their pertinent characteristics described, and how objects relate to each other is mapped out as well as possible. Different views are integrated via a common language and structure. The **logical modeling** stage translates the conceptual organization into something more practical, perhaps simply thought of as putting numerical values into tables of data, but avoiding the details of storage of data on physical media. It is at this level that phenomena are organized into a database as tables of data records and connections to other tables. The internal modeling, which we do not address, treats the organization of data on hardware storage media.

As automation for spatial data handling has proceeded from making graphs and then maps, to automated mapping and thematic cartography, to tools for inventory and management, and then to database resources for decision making or education, so it has become more important to think about the value of the information in the database. The contents, certainly reasonably viewed as a commodity, are presumably valuable for one or more purposes, yet such data alone are not enough. Users provide the added value by interpretation and analysis, and by understanding the nature and quality of the data. It is important to focus on the message, the contents of the envelope, not on the messenger, the postal system. Consequently, the discussion of data structuring in an automated system is put in the context of the semantics of the data.

Conclusion

Automated spatial information systems are:

A data depository

A toolbox

A resource

A technology and

Frames of mind, one or several.

As a software processing routines in a hardware setting they are a new kind of toolbox for practical problem solving. As a new resource compared with paper based map making, they represent a new technology; and through their emphasis on spatial data, they stand, for many people, for a different approach to thinking about problems and knowledge.

NEEDS – PURPOSES AND TYPES OF SPATIAL PROBLEM

Problems to be solved; Tasks to be performed

A sensible approach in designing, evaluating, buying and using spatial information systems is to know what particular, preferably well demarcated problems, scientific or practical, have to be solved. A study of general categories of use, or application domains, provides only a broad framework, leading to a conclusion that these information systems are virtually universal tools. A study of specific questions needing answers allows us to begin to appreciate the data and processing requirements of many users, and especially to reveal the tools required in the automated toolbox.

1.3 MULTICRITERIA SPATIAL DECISION SUPPORT SYSTEMS

DEFINITION

A *spatial decision support system* (SDSS) is an interactive, computer-based system designed to support a user or a group of users in achieving a higher effectiveness of decision making while solving a semi-structured spatial decision problem. It lies at the intersection of two major trends in the spatial sciences: *geographic information sciences* (GIS) and *spatial analysis*. What really makes the difference between a SDSS and a traditional *decision support system* (DSS) is the particular nature of the geographic data considered in different spatial problems and the high level of complexity of these problems. An effective SDSS requires the addition of a range of specific techniques and functionalities, used especially to manage spatial data, to conventional DSS.

According to a SDSS,

(i) it should provide mechanisms for the input of spatial data,

(ii) allow representation of spatial relations and structures,

(iii) include the analytical techniques of spatial analysis, and
(iv) provide output in a variety of spatial forms, including maps.

Multicriteria spatial decision support systems (MC-SDSS) can be viewed as part of the broader fields of SDSS. The specificity of MC-SDSS is that it supports *spatial multicriteria decision making*. Spatial multicriteria decision making refers to the use of *multicriteria analysis* (MCA) to spatial decision problems. MCA is a family of operations research tools that have experienced very successful applications in different domains since the 1960. It has been coupled with geographical information systems (GIS) since the early 1990s for an enhanced decision making.

HISTORICAL BACKGROUND

The concept of SDSS has evolved in parallel with DSSs. The first MC-SDSS have been developed during the late 1980s and early 1990s. Early research on MC-SDSS is especially devoted to the physical integration of the GIS and MCA. These first tools luck interactively and flexibility since GIS and MCA softwares are coupled indirectly, through an intermediate system. Later research concerns the development of MC-SDSS supporting collaborative and participative multicriteria spatial decision making. Web-based MC-SDSS is an active research topic which will be consolidated in the future.

SCIENTIFIC FUNDAMENTALS / CONCEPTS OF SDSS

1.3.1 General structure of SDSS/MC-SDSS

A typical SDSS contains three generic components (Figure 3): a database management system and geographical database, a model-based management system and model base, and a dialogue generation system. The data management subsystem performs all data-related tasks; that is, it stores, maintains, and retrieves data from the database, extracts data from various sources, and so on. It provides access to data as well as all of the control programs necessary to get those data in the form appropriate for a particular decision making problem. The model subsystem contains the library of models and routines to maintain them. It keeps track of all possible models that might be run during the analysis, as well as controls for running the models. The model base management system component provides links between different models so that the output of one model can be the input into another model. The dialogue subsystem contains mechanisms whereby data and information are input to the system and output from the system. These three components constitute the software portion of SDSS. A fourth important component of any decision support system is the user which may be simple users, technical specialists, decision makers and so on.

MC-SDSS can be viewed as a part of a broader field of SDSS. Accordingly, the general structure of a MC-SDSS is the same as that of SDSS. However, the model-based management system is enhanced to support multicriteria spatial modelling and the model base is enriched with different multicriteria analysis techniques.

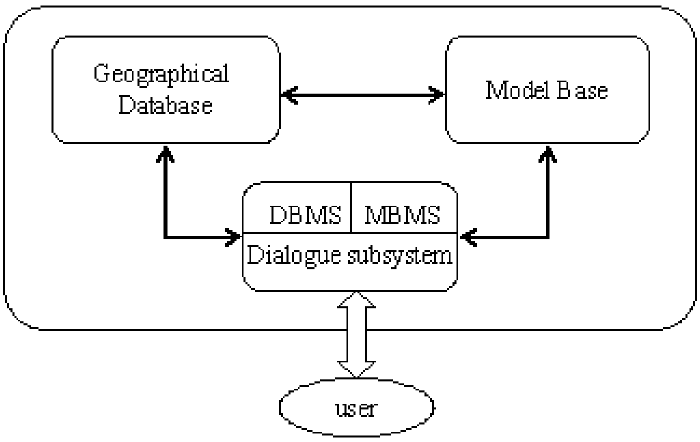


Figure 3: General structure of SDSS

1.3.2 GIS and multicriteria analysis integration modes

The conceptual idea on which most of GIS-based multicriteria analysis is to use the GIS capabilities so as to prepare an adequate platform for using multicriteria methods (Figure 4). The GIS-based multicriteria analysis starts with the problem identification, where the capabilities of the GIS are used to define the set of feasible alternatives and the set of criteria. Then, the *overlay* procedures are used in order to reduce an initially rich set of alternatives into a small number of alternatives which are easily evaluated by using a multicriteria method. Finally, the drawing and presenting capabilities of the GIS are used to present results.

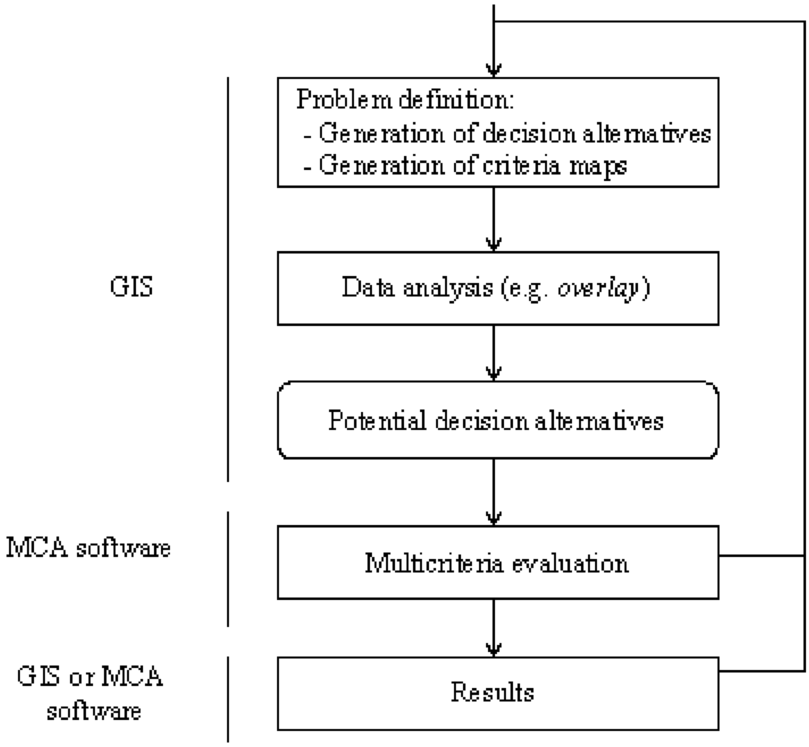


Figure 4: Conceptual schema for GIS and multicriteria analysis integration

Physically, there are four possible modes to integrate GIS and multicriteria analysis tools (Fig. 4):

(i) no integration,

(ii) loose integration,

(iii) tight integration, and

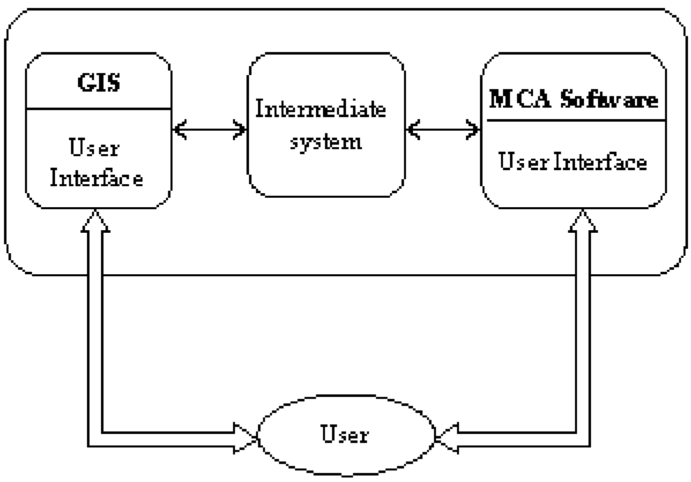
(iv) full integration.

No Integration mode: The first mode corresponds to the situation dominating until late 1980 where the GIS and multicriteria analysis are used independently to deal with spatial problems. The three next modes correspond to increasing levels of complexity and efficiency.

Loose integration mode: The integration of GIS software and a stand-alone multicriteria analysis software is made possible by the use of an intermediate system. The intermediate system permits to reformulate and restructure the data obtained from the overlapping analysis which is performed through the GIS into a form that is convenient to the multicriteria analysis software. The other parameters required for the analysis are introduced directly via the multicriteria analysis software interface. The results of the analysis—totally made in the multicriteria analysis software—may be visualized by using the presentation capabilities of the multicriteria analysis package, or feedback to the GIS part, via the intermediate system, for display and, eventually, for further manipulation. It should be noted that each part has its own database and its own interface, which limited the user-friendliness of the system.

Tight integration mode: In this mode, a particular multicriteria analysis method is directly added to the GIS software. The multicriteria analysis method constitutes an integrated but autonomous part with its own database. The use of the interface of the GIS part alone increases the interactivity of the system. This mode is the first step towards a complete GIS-multicriteria analysis integrated system. Yet, with the autonomy of the multicriteria analysis method, the interactivity remains a problem.

Full integration mode: The third mode yields itself to a complete GIS-multicriteria analysis integrated system that has a unique interface and a unique database. Here, the multicriteria analysis method is activated directly from the GIS interface as any GIS basic function. The GIS database is extended so as to support both the geographical and descriptive data, on the one hand, and the parameters required for the multicriteria evaluation techniques, on the other hand. The common graphical interface enhances the user-friendless of global system.



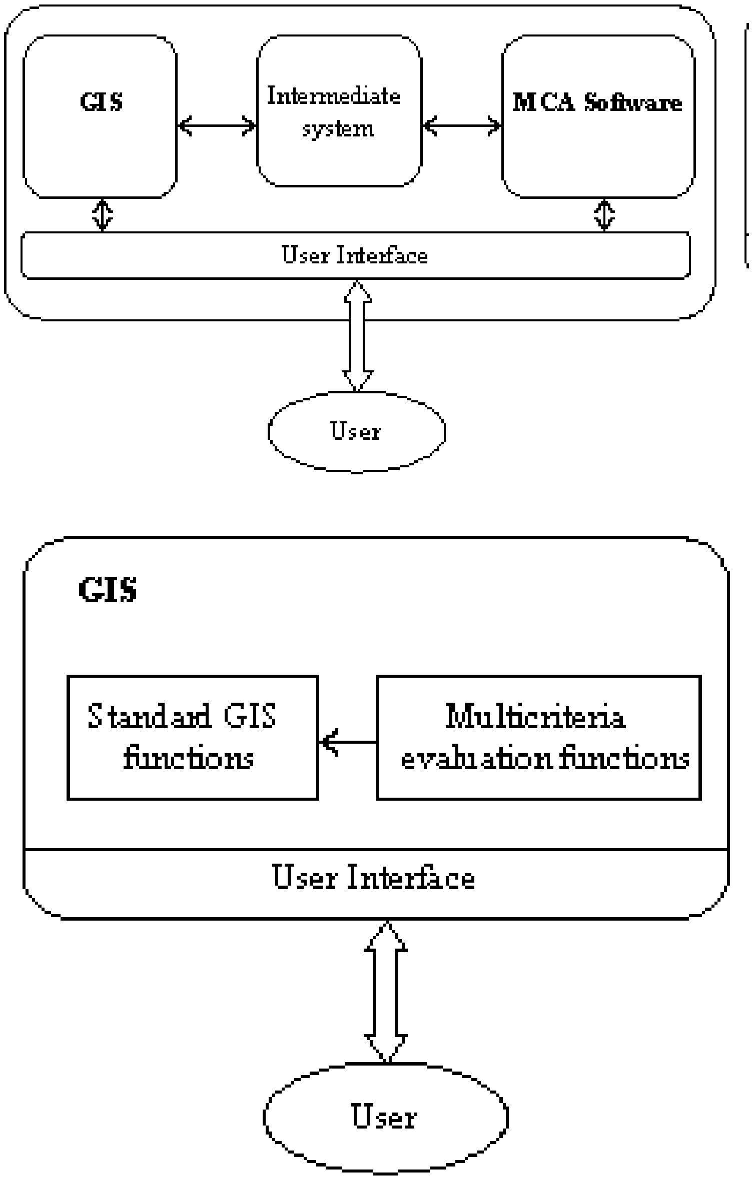


Figure 4: GIS and multicriteria loose (*a*), tight (*b*) and full (*c*) integration modes

1.3.3 GIS and multicriteria analysis interaction directions

We may distinguish five directions of interaction:

(i) no interaction,

(ii) one-direction interaction with the GIS as the main software

(iii) one-direction interaction with multicriteria tool as the main software,

(iv) bi-directional interaction, and

(v) dynamic interaction.

One-direction interaction provides a mechanism for importing/exporting information via a single flow that originates either in the GIS or multicriteria software. This type of interaction can be based on GIS or multicriteria as the principle software.

In the **bi-directional interaction** approach the flow of data/information can originate and end in the GIS and multicriteria modules.

Dynamic integration allows for a flexible moving of information back and forth between the GIS and multicriteria modules according to the user's needs.

1.3.4 Design of a MC-SDSS

Different frameworks for designing MC-SDSS have been proposed in the literature. A part differences in GIS capabilities and multicriteria techniques, most of these frameworks contain the major components introduced earlier. This framework is conceived of in such a way that it supports GIS-MCA integration and is also open to incorporate any other OR/MS tool into the GIS (Figure 5).

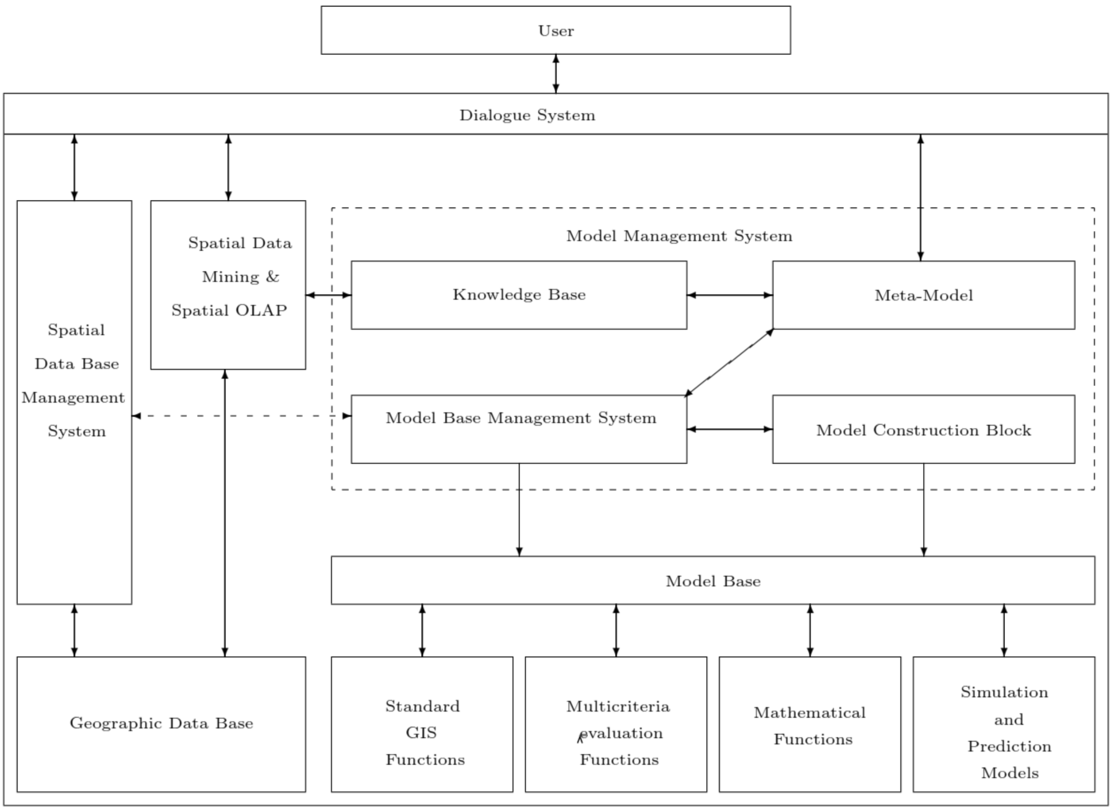


Figure 5: A design of a multicriteria SDSS

1.3.4.1 Spatial data base management system

The spatial data base management system is an extension of the conventional database base management system. It is used specially to manage spatial data.

1.3.4.2 Geographic database

The geographic data base is an extended GIS database. It constitutes the repository for both (i) the spatial and descriptive data, and (ii) the parameters required for the different OR/MS tools.

1.3.4.3 Model base

The model base is the repository of different analytical models and functions. Among these functions, there are surely the basic GIS ones (e.g. statistical analysis, overlaying, spatial interaction analysis, network analysis, etc.). The model base contains also other OR/MS models. Perhaps the most important ones are those of MCA. Nevertheless, the system is opened to include any other OR/MS tool (e.g. mathematical models, simulation and prediction models, etc.), or any other ad hoc model developed by the model construction block.

1.3.4.4 Model management system

The role of this component is to manage the different analysis models and functions. The model management system which contains four elements: the meta-model, the model base management system, the model construction block and the knowledge base.

1.3.4.5 Meta-model

This element is normally an Expert System used by the decision maker to explore the model base. This exploration enables the decision maker to perform a “what-if” analysis and/or to apply different analytical functions. The meta-model uses a base of rules and a base of facts incorporated into the knowledge base. The notion of meta-model is of great importance in the sense that it makes the system open for the addition of any OR/MS analysis tool. This requires the addition of the characteristics of the analytical tool to the base of rules, and, of course, the addition of this model to the model base.

1.3.4.6 Knowledge base

Knowledge base is the repository for different pieces of knowledge used by the meta-model to explore the model base. Practically, the knowledge base is divided into a ***base of facts*** and a ***base of rules***. The base of facts contains the facts generated from the model base. It also contains other information concerning the uses of different models, the number and the problems to which each model is applied, etc. The base of rules contains different rules of decision which are obtained from different experts, or automatically derived, by the system, from past experiences. This base may, for instance, contains: *If the problem under study is the concern of many parties having different objective functions then the more appropriate tool is that of MCA*.

1.3.4.7 Model base management system

The role of the model base management system is to manage, execute and integrate different models that have been previously selected by the decision maker through the use of the Meta-Model.

1.3.4.8 Model construction block

This component gives the user the possibility to develop different ad hoc analysis models for some specific problems. The developed ad hoc model is directly added to the model base and its characteristics are introduced into the base of rules of the Knowledge Base.

1.3.4.9 Spatial data mining and spatial on line analytical processing

Data Mining and On Line Analytical Processing (OLAP) have been used successfully to extract relevant knowledge from huge traditional databases. Recently, several authors have been interested in the extension of these tools in order to deal with huge and complex spatial databases. In particular, underlines that *spatial data mining* is a very demanding field that refers to the extraction of implicit knowledge and spatial relationships which are not explicitly stored in geographical databases. The same author adds that *spatial OLAP* technology uses multidimensional views of aggregated, pre-packaged and structured spatial data to give quick access to information. Incorporating spatial data mining and spatial OLAP into the MC-SDSS will undoubtedly ameliorate the quality of data and, consequently, add value to the decision-making process.

1.3.4.10 Dialogue system

The dialogue system represents the interface and the equipment used to achieve the dialogue between the user and the MC-SDSS. It permits the decision maker to enter his/her queries and to retrieve the results.

1.4 KEY APPLICATIONS OF MC-SDSS - USEFULNESS

MC-SDSS have been used in a wide range of practical applications of spatial multicriteria decision making problems including nuclear waste disposal facility location, solid waste management, land-use planning, corridor location problem, water resource management, habit site development, health care resource allocation, land suitability analysis. In the rest of this section we provide a brief description some SDSS.

• Open Spatial Decision Making (OSDM) is an Internet-based MC-SDSS designed to support the selection of suitable sites for radioactive waste disposal by the public in Great Britain. An important characteristic of OSDM is that it does not require prior knowledge of GIS or MCA.

•Spatial Group Choice (SGC) is a GIS-based decision support system for collaborative spatial decision support making. The system has been used successfully for habit site selection in the Duwamish Waterway and area and for health care resource allocation.

•IDRISI/Decision Support is a built-in decision support module for performing multicriteria decision analysis. This system has been applied in different real-world applications. The case study described in illustrates the use of the system for analyzing land suitable for a housing project in Mexico.

•DOCLOC has been designed for aiding for aiding health practitioners in the selection of practices in the state of Idaho. One limitation to this system is the use of the loose coupling strategy.

•Collaborative Planning Support System (CPSS) provides an example of a system employing multi-objective fuzzy decision analysis. It is a multicriteria collaborative spatial decision support system for sustainable water resource management.

FUTURE DIRECTIONS

Use full integration modes

The first limitation concerning MC-SDSS is relative to the integration mode adopted. In fact, most of the proposed works use loose or tight integration modes. One possible solution to permit a full integration is to identify a restricted set of *multicriteria evaluation functions* and their incorporation in the GIS. These functions represent elementary operations required to implement the major part of multicriteria methods. This integration strategy avoids the necessity of programming the different multicriteria methods. In addition, it permits a full integration since the multicriteria evaluation functions are generic and can easily be incorporated in the available commercial GIS.

Incorporation of many multicriteria methods

It is well established that each multicriteria method has its advantages and disadvantages. This means that a given method may be useful in some problems but not in others. One intuitive solution to this problem is to incorporate as many as possible multicriteria methods in the MC-SDSS. However, this idea has several limitations: (i) the obtained system is not flexible enough, (ii) it requires a considerable effort for programming the different methods, and (iii) there is no way to develop “personalized” methods. The integration strategy proposed in the previous paragraph permits to handle this limitation. In fact, the multicriteria evaluation functions are defined in a generic way and can be used to implement different existing multicriteria methods or even to create *ad hoc* methods adapted to the problem under consideration.

Formal methodology to select the multicriteria method to apply

Disposing of a large number of multicriteria methods in the MC-SDSS permits to extend and reinforce the analytical potentiality of the GIS. However, a new problem appears: how to choose the method to use in a given problem? There are generally three possible solutions to the multicriteria method selection problem: (i) the use of a classification tree (ii) the use of a multicriteria method, and (iii) the use of an Expert System or a decision support system. We think that the last solution is more appropriate in a perspective of GIS and multicriteria analysis integration. The development of a rule-based system needs (i) the characterization of the spatial decision problems, the multicriteria methods and the decision maker(s) (ii) the identification and quantification of knowledge about multicriteria methods, and (iii) the establishing of a corresponding between the elements enumerated in (i). The result is a collection of rules. These last ones are then used, by the inference system, as a basis for selecting the most appropriate method.

Choice of the standardization / weighting techniques

Among the problems that are not sufficiently treated in GIS-based multicriteria systems is the selection of the standardization and the weighting techniques. There are many different standardization/weighing techniques that can be used in MC-SDSS. It is important to note that different standardization/weighting techniques may lead to different results. The development of a formal framework for aiding the decision maker during the selection of the standardization/weighting technique—similar to the one proposed for the selection of the multicriteria method—is a good initiative.

Developing multicriteria spatial modelling environment

The use of multicriteria analysis in the GIS is complicated by the lack of an appropriate multicriteria spatial modelling environment. A possible solution is to develop a script-like programming language supporting the different multicriteria evaluation functions. **Decision Map Algebra** (DMA), proposed in and inspired from Tomlin's map algebra seems to be a good start point.

Web-based multicriteria spatial decision making

Web-based MC-SDSS is a recent and active research topic. This is particularity important since it permits to share geographical information and facilitates multicriteria collaborative spatial decision making.

***

UNITS 2 & 3

UNIT-2 Designing of Spatial Database: Identification of Geographic features – attributes & data layer – Defining the storage parameters for each attribute – ensuring of co-ordinate registration – map projection – Transformation.

UNIT-3 Designing of Non spatial Database: Creation of data table file to hold the attributes – Adding up of description attribute values to table – Different types of sources of data entry – Checking for errors.

DESIGNING THE DIGITAL SPATIAL AND ASPATIAL DATABASE

The first step in developing your digital database is to determine what the contents of the database will be. By spending some time designing your database before actually automating it, you will ensure that when you are ready to perform the analysis and create the final products, all of the coverage features and attributes you need will be there. Having to modify a database in the middle of an analysis or mapping procedure is time consuming and costly. A well designed database can also ensure that the data will be usable for future projects.

You should note that there are many factors influencing the design of the database. Database design consists of three major steps: identify the geographic features, attributes and data layers required; define the storage parameters of each attribute; and ensure coordinate registration. Keep in mind that the available source data, whether map manuscripts or digital data, plays an important part in the design process. You should therefore do some preliminary research to determine what source data is available for your study area.

Identifying data layers and attributes: The first step of database design is to determine what data to include in the database. This is a three-step process:

- Identify geographic features and their attributes
- Organize the data layers
- Identify coverages to be automated.

Identify geographic features and their attributes: First you will want to identify the geographic features needed in the database and the attributes associated with each. This will be determined directly by the analysis you wish to perform and the map products you will create. There may be several required attributes for each feature, based on the criteria for the analysis and subsequent maps. For example, consider the following criteria:

- Identify soils suitable for development
- Use land use codes to select brushland for development
- Estimate a purchase cost, based on area and the land use class

Based on these criteria, you would need soils polygons as one geographic feature with suitability rating as an attribute, and land use polygons as another feature with land use code and per-unit cost as attributes. Your list of features and attributes would look like this:

Geographic feature	Feature class	Feature attributes
Soils Land use	Polygons Polygons	Suitability Land use code Cost per hectare

If the map of the results is also to include land use polygons labeled by land use type, locations of roads drawn with line symbols to indicate their surfacing, and major streams, you would need to expand your list of geographic features and attributes.

Geographic feature	Feature class	Feature attributes
Soils Land use Roads Streams	Polygons Polygons Lines Lines	Suitability Land use code Cost per hectare Land use type Road code Stream class

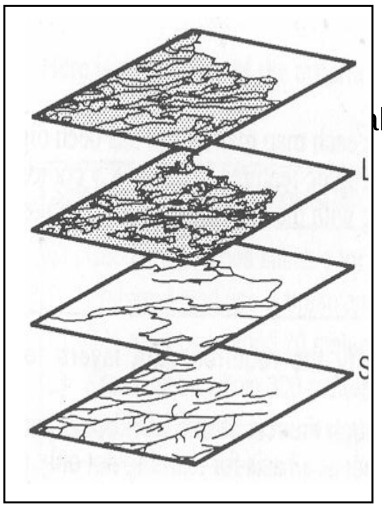
Organize the data layers: Once you have identified the necessary features and their attributes, you can begin organizing the geographic features into layers of data.

A number of factors influence layer organization in a geographic database, and they differ with each application. Two of the most' common considerations for organizing layers include feature types (point, line or polygon) and thematic grouping of features.

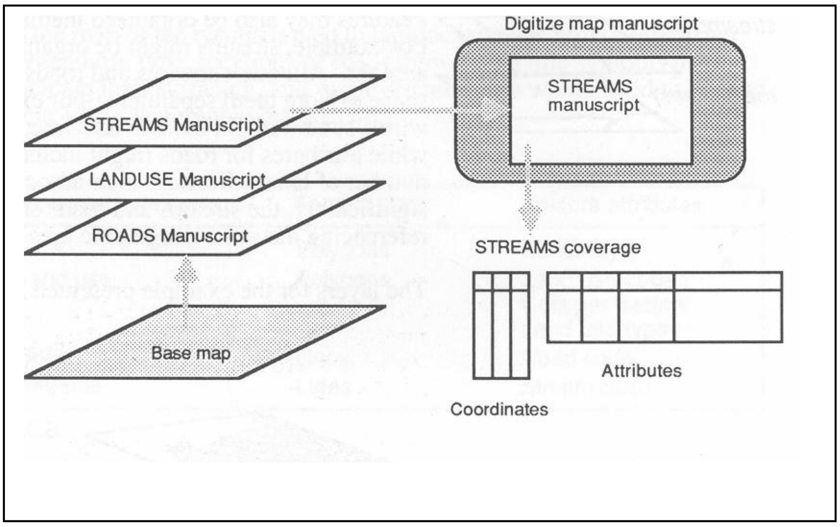
Wells (Points), Roads (lines): Typically, layers are organized so that points, lines, and polygons are stored in separate layers. For example, well sites represented by points might be stored in one layer, while roads represented by lines are organized in another layer.

Streams (lines), Roads (lines): Features may also be organized thematically by what they represent. For example, streams might be organized in one layer and roads in another. Although streams and roads are both line features, it makes sense to store them separately. For example, the attributes associated with a stream might include its name, stream class and rate of flow, while attributes for roads might include a name, surface type, and number of lanes. Because their associated attributes differ significantly, the streams and roads should be stored in separate layers referencing the same geographic area.

The layers for the example presented above would look like this:

Layer	Feature type	Feature class	Attributes
	SOILS	Polygons	Soil class bility
	LANDUSE	Polygons	Land use code Cost per hectare Land use type
	ROADS	Lines	Road code
	STREAMS	Lines	Stream class

Identify coverages to be automated: The process of identifying the geographic features and their attributes, and organizing this information into layers determines what coverages your digital geographic database will contain. In some cases, the data layers will be available on separate maps (for example, a map showing only parcel boundaries) or will already be in digital format on the ~ computer. In other cases, you will have to automate layers from a single base map. In these instances, it is often easier to create separate map manuscripts for each layer, since the amount of information on the base map may make data capture more difficult. This is often done by tracing the necessary features onto MYLAR- or some other durable transparent material.



Once each map manuscript has been digitized, you'll have the needed geographic features stored as x, y coordinates in the digital database, along with the attributes in the coverage feature attribute table.

Identify the required data layers for the project

Through the rest of this workbook, you'll be working on a small GIS project as a basis for learning not only the functionality of ARC/INFO, but the flow of a GIS project as well.

Here then is an introduction to your GIS project. A local university is planning to construct a small lab and office building to perform research and extension projects in aquaculture. They've narrowed the area down to a coastal farming area near several small towns and now need to select a specific site that meets the following requirements:

- At least 2,000 square meters in size
- Must overlay soils suitable for construction of buildings
- To reduce costs from clearing land, the site should not be forested. And since a regional agricultural preservation plan prohibits conversion of farmland in this area, the current land use must be brushland.
- A local ordinance designed to prevent rampant development allows new construction only within 300 meters of existing sewer lines.
- A recent national water quality act requires that no construction occur within 20 meters of streams.

You've been given the task of assembling a list of potential sites, with estimated purchase cost. You'll also need to create a map showing the location of the potential sites, along with improved and semi-improved roads, and a report listing each site with its area and estimated purchase cost.

Here is a summary of the criteria for the project:

Selection criteria for proposed lab site

- Choose soil types suitable for development
- Preferred land use is brushland
- Site must be beyond 20 meters of existing streams
- Site must be within 300 meters of existing sewer lines
- Site must contain an area of at least 2,000 square meters.

In addition, the final map should show locations of roads.

Each criterion can be translated into the need for a specific data layer with specific feature attributes (with the exception of the site area requirement - this criterion will be derived from a new layer resulting from the spatial join of the other layers). Based on the list of criteria given above, complete the following table. For example, the first criterion is to find sites with soils suitable for development. Thus, one of the layers will be SOIl..s, a polygon coverage. Fill in the remaining lines. This illustrates the process of identifying the data layers you'll need for the project. (See Appendix D for the exercise answers.)

Criterion		Layer Feature	Type Feature Class
Suitable soils	-	SOILS	POLYGONS
Land use is brushland	-		
More than 20 m from streams	-		
Within 300 m of sewers	-		
Map showing roads	-		

Defining each attribute: Once the attributes needed for each layer in the database have been determined, we have to decide the specific parameters for each attribute and the types of values to be stored. Deciding now makes the task of building the data files to hold the attribute values much easier (as you will see in Lesson 6).

Coding: You saw in the last lesson how attributes are stored in the computer as numbers and characters. You'll need to decide which attributes to store as numbers and which, as characters. For example, if you wish to label streets with their names, you must store the name as it should appear (e.g., N Main ST). Similarly, for attributes representing a numeric value, such as street width or length, the actual value must be stored (e.g., 321.375).

Some attributes described by a character string are better represented as a code in the computer. If the attribute describes a class, it might be easier and more efficient to store a code for the class rather than a description. For example, you could store the character string 'Urban/Developed' as either a character code (UD) or a numeric code (100) that you define. Coding attributes makes it easier to select and draw features of a particular class. For example, the code could be used to look up a symbol number in another table, so that each feature having that code will be drawn using the same symbol.

Attribute table

AREA ID	LUCODE	
4322	10	200
3901	11	300
5200	12	300
1698	13	100
2004	14	200

Lookup table

LUCODE	SYMBOL	LABEL
100	16	Urban
200	45	Agriculture
300	24	Brushland

Also, attributes having many repeating values are best represented using code values to reduce the database size.

Numeric values representing a range may also be easily stored as a code. For example, coverage of polygons representing slope classes of 0-10%, 11-30%, 31-45%, and over 45% may be more easily represented in the computer by codes 1, 2, 3 and 4, respectively.

SLOPE	CODE
0-10	1
11-30	2
31-45	3
45+	4

When designing your database, remember that features can always be grouped into fewer classes at a later date. It is much more difficult to separate a few classes into more classes if they have not been coded that way initially.

Allocating storage: In addition to deciding how each attribute will be stored, you must also decide on the amount of storage required for each. For example, how many characters are needed to store the street name (determined by the longest street name). For numeric items, determine the number of required digits and decimal places. Of major importance is the overhead of storing large amounts of data. Less space used for each attribute will result in smaller data files, less disk storage used on your computer, and faster processing times.

Instead of requiring that all descriptive data be stored in a single feature attribute table, temporary relates between tabular files. Tables can be related if they share a common item. Records in one file will be temporarily connected to records in the related file sharing the same values for the common item. For example, the lookup table LANDUSE.LUT can be related to LNDA TO5.P AT via the common item in each named LUCODE.

LNDAT05.PAT

AREA	PERIMETER	LNDAT05	LNDAT05_ID	LUCODE
12.505	42.933	1	59	400
2.514	13.252	2	60	200
2.004	7.244	3	61	400
0.727	5.104	4	62	200
7.482	18.752	5	63	300
10.298	29.731	6	64	500

LANDUSE.LUT

LUCODE	TYPE
100	Urban
200	Agriculture
300	Brushland
400	Forest
500	Water

The values for LUCODE are used to associate records between the two tables. Relates such as this allow you to minimize redundancy and data storage requirements.

Using the JOIN operation instead of JOINITEM offers several advantages for designing and implementing your tabular database operations.

Relates reduce redundancy in data storage for items with many repeating values. For example, a descriptor, symbol number, and text label might be associated with each value for LUCODE. Each of these values can be stored once for each land use type in a related table instead of repeating these values for each land use polygon.

Such data reduction results in more efficient software performance.

Different departments can refer to the same coverage, yet have their own descriptive data available in a separate table which can be maintained independently of other departments.

Many applications operate only on the descriptive data -not on the locational data (e.g., tax assessor files or permit tracking by parcel). The ability to store attribute data separately from the coverage allows you to utilize tabular data for other purposes independent of the coverage information.

The JOIN operation in ARC/INFO is available in TABLES, ARCEDIT, and ARCPLOT.

Building a data dictionary: The data dictionary is a list maintaining, for each coverage, the names of the attributes and a description of the attribute values (including a description of each code, if necessary). Creating a data dictionary for your database will be invaluable as a reference during the project, as well as in transferring information to others.

Here is a sample data dictionary in table form:

Layer Feature	**Class**	**Attributes**	**Value**	**Description**
SOILS	Polygons	SOILCODE	Abbreviation for soil type	
		SUIT	0	Unsuitable
			1	Poor suitability
			2	Moderate suitability
			3	Good suitability
LANDUSE	Polygons	LUCODE	100	Urban
			200	Agriculture
			300	Brushland
			400	Forest
			500	Water
			600	Wetlands
			700	Barren
		Cost per hectare (COSTHA)	Actual monetary value	
STREAMS	Lines	STRMCODE	1	Major stream
			2	Minor stream
SEWERS	Lines	DIAMETER	Actual diameter is stored	
		SYMBOL	1	60 cm pipe
			77	45 cm pipe
ROADS	Lines	RDCODE	1	Improved
			2	Semi-improved

Coordinate registration: Once you've identified all of the coverages required for your database, you should ensure that the coverages register correctly to each other. As we've seen, in most cases, your database will consist of a number of coverages representing various features for the same geographic area. When combining the data from one coverage with another (for example, when overlaying two coverages to create a new one), the coincident data must match exactly. If the coordinate registration is close, but not exact, you'll experience offset problems: sliver polygons when performing overlays; ragged edges when generating maps; and inaccurate measurements when compiling data for reports.

For example, the LANDUSE and SOIS layers in the figure below represent the same geographic area, but each uses a different set of tics (coordinate registration points). Thus even minor errors in registration, introduced when the maps were automated, will result in offset between the layers.

Advanced topic -Building an application system

In this lesson, we have described the steps involved in performing a single GIS project, including the development of a database specific to the project. In some cases, there is an economy of scale to be gained by developing an application system. This is a system consisting of a comprehensive database that can meet the needs of a variety of users within a specific geographic area. The database can be used not only for special projects, but for ongoing tasks.

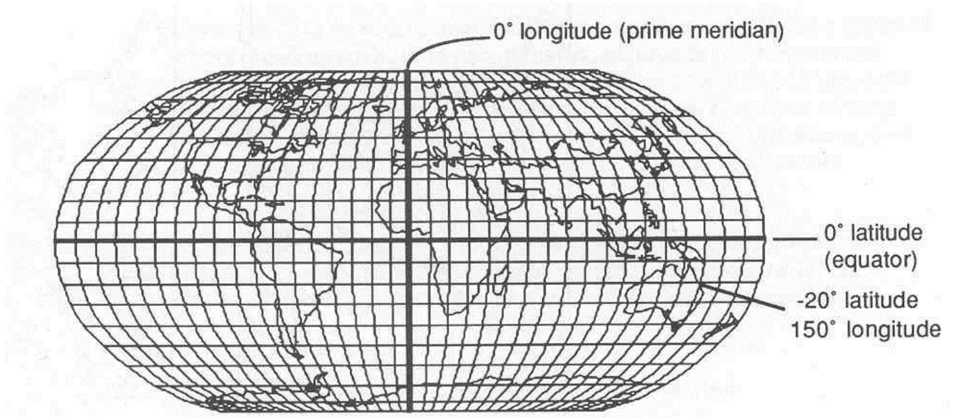
Although activities differ between agencies within a geographic area, they share a common jurisdiction. Problems surface in a manual system when data needed to execute these tasks is located within a single department, making access difficult. Conversely, if each department maintains its own data it can mean duplicate effort and inconsistent files. A properly planned and implemented application system can provide separate users with different views of the same database and eliminate redundancy and inconsistency.

The first step in designing such a system is to develop a thorough user-needs assessment. This may consist of interviews with potential users of the system to ensure that the data they require for their tasks is included. Once the needs have been determined, the design phase can proceed. Finally, a pilot project will identify any missing elements in the database design and allow for correction before major time and resource commitments are made to automating the entire database.

MAP PROJECTION IN GIS

Most maps display coordinate data by conforming to a recognized systems global coordinate system; for example, the Universal Transverse Mercator (UTM), the Albers Conic Equal-Area, and the Polar Stereographic coordinate systems. These are examples of map projections used to represent spherical geographic features on a flat surface. A projection is used to ensure a known relationship between locations on a map and their true locations on the earth.

Latitude-longitude: The most familiar locational reference system is that of latitude and longitude. This system can be used to identify the locations of points anywhere on the earth's surface.



This Robinson projection illustrates latitude and longitude reference locations.

Latitude and longitude are angles measured from the earth's center to a point on the earth's surface. Latitude is measured North and South, while longitude is measured East and West. A graticule or referenced grid of latitude and longitude lines can be superimposed on the earth's surface to geographically reference various locations. Longitude lines, sometimes called meridians, begin and end at the North and South poles. Latitude lines, sometimes called parallels, encircle the globe with parallel rings.

Latitude and longitude are traditionally measured in degrees, minutes, and seconds (DMS). For latitude, O^{o} is at the equator, 90^{o} is at the North Pole, and -90^{o} is at the South Pole. For longitude, 0^{o}, the Prime meridian, starts at the North Pole, passes through Greenwich, England, and ends at the South Pole. Longitude is measured positively, up to 180^o, when traveling east from Greenwich and measured negatively, up to -180^o, traveling west from Greenwich. For example, Australia, which is south of the equator and east of Greenwich, has positive longitudes and negative latitudes.

Latitude-longitude, however, is a geographic reference system - not a two-dimensional (planar) coordinate system. In the figure above, notice how meridians converge at the poles but separate or diverge as they get closer to the equator. Thus the length of one degree of longitude will vary depending upon the latitude at which it's measured. For example, one degree of longitude at the equator is 111 kilometers (69 miles) in length, but the length of one degree of longitude at the North Pole or South Pole converges to O. Since the measurement units aren't associated with. a standard length, they can't be used as an accurate measure of distance. And since this reference system measures angles from the center of the earth, rather than distances on the earth's surface, it's not a planar coordinate system.

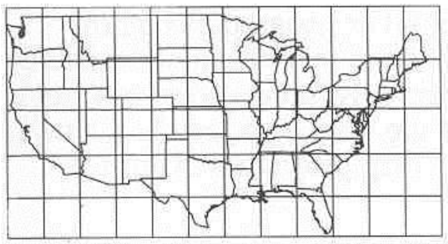
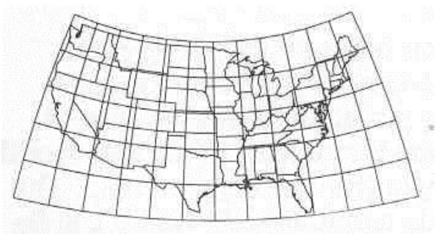
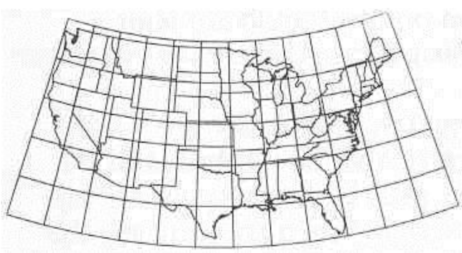
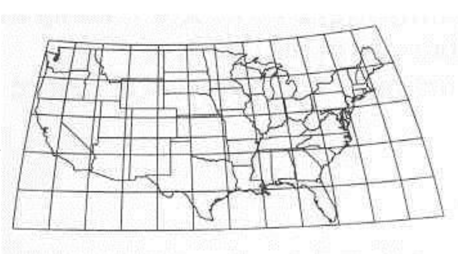
Planar coordinate Systems: Planar coordinate systems (often called Cartesian coordinate systems) have several properties which make them useful for representing real world coordinates on maps:

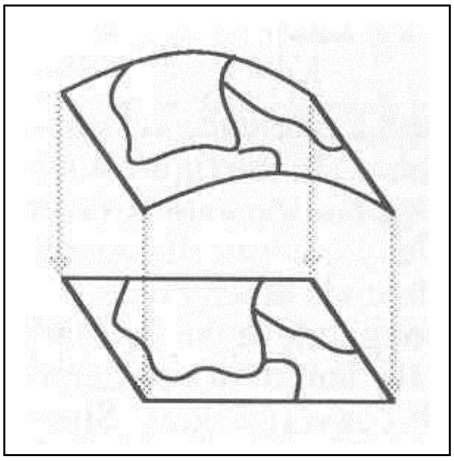
- There are two dimensions: x measures distance in a horizontal direction, and y measures distance in a vertical direction
- Measures of length, angle, and area are constant across the two
- Various mathematical formula exist to ‘project’ the earth’s s 1>0 spherical surface onto a flat, two-dimensional surface

Geographic Information Systems, like flat maps, use various planar 11 coordinate systems to map the earth's surface. Each coordinate, system used is based on a particular map projection.

Map projections: Since the earth is a spheroid, a mathematical conversion must be used to create a flat map sheet from the spherical surface. This mathematical conversion is commonly referred to as a *map projection.*

Some commonly used map projections In the U.S. and North America

Projection	Characteristics	Maps
Mercator 	Shape: Preserved Area: Definite distortion Distance: Distortion increases toward poles Direction: Some distortion	• Navigational maps • Time zone maps
Universal Transverse Mercator (UTM) *UTM is not used for projecting very large areas.*	Shape: Preserved Area: Distortion increases away from Central Meridian Distance: Distortion similar to area Direction: Some distortion	• USGS 1 :250,000 scale quadrangle map series (i.e., 1^o latitude by 2^o longitude)
Transverse Mercator 	Shape: Preserved Area: Definite distortion Distance: Some distortion Direction: Some distortion	• $7^{1/2}$' and 15' quadrangle maps for 22 states within the U.S. • Maps of North America
Albers Equal-Area Conic 	Shape: Some distortion Area: Preserved Distance: Some distortion Direction: Some distortion	• United States maps • Some section maps of the U. S.
Lambert Conformal Conic 	Shape: Preserved Area: Some distortion Distance: Some distortion Direction: Some distortion	• $7^{1/2}$' and 15' quadrangle maps for 32 states within the U.S. • State base maps

State Plane Coordinate System 	Not really a map projection, it is a definition of a set of coordinate systems which vary from state to state. East-West State Plane Zones use the Lambert Conformal Conic projection while North-South State Plane Zones use Transverse Mercator. Each zone has a set of parameters that define the projection for that zone. All $7^{1/2}$' and 15' quadrangle maps for the U.S use State Plane Zones.

Since every base map is stored in a particular projection, you should determine the projection of your base map
before automating your map sheet.

Understanding the uses of map projections and latitude-longitude reference points is very important as you begin to use ARC/INFO to develop your database. For the moment, please note the following:

- Any representation of the earth's surface in two dimensions always involves distortion of some parameter, be it shape, area, distance, or direction.
- Different projections produce different distortions.
- The characteristics of each projection make them useful for some applications and not useful for others.

Though a full understanding of map projections is not needed to ensure that the coordinate values in your database are measured in a real-world coordinate system, it's an important issue when dealing with large amounts of map data whose projection sources may vary.

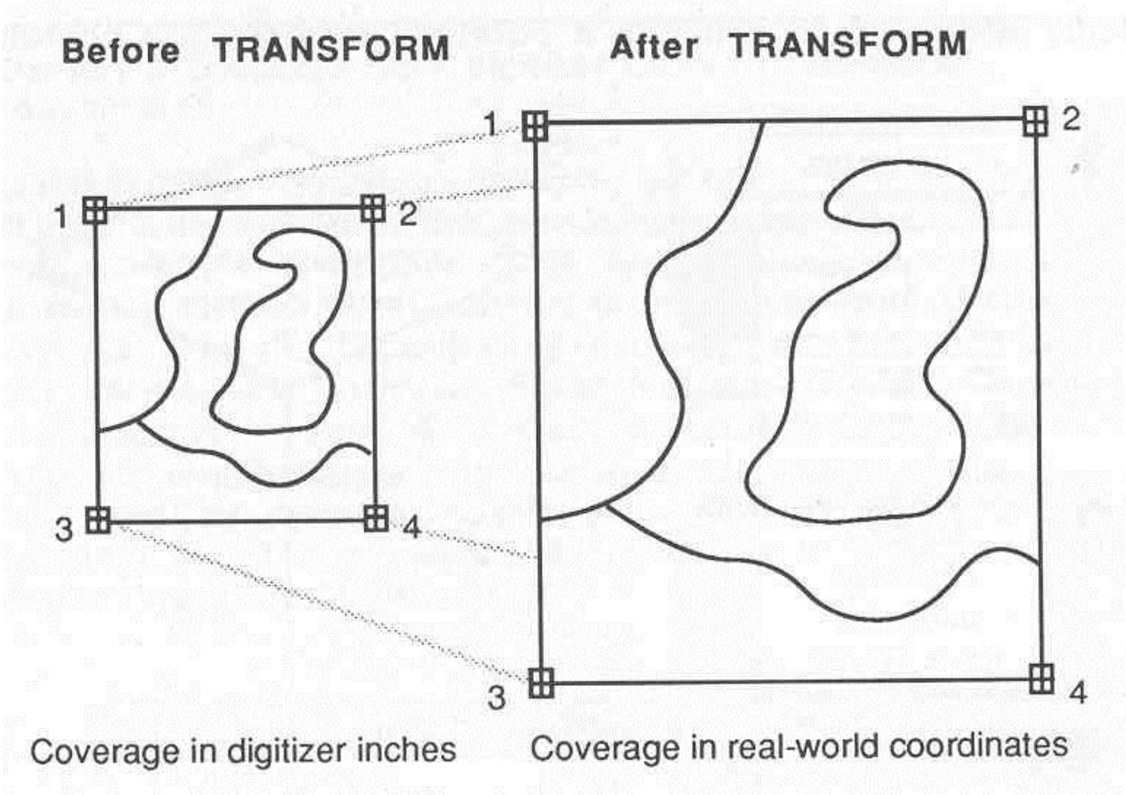
Tic features: Always establish tic points (representing known real-world locations) for a study area before the digitizing stage. These common tic locations should then be recorded on each map sheet to provide a common locational reference for each coverage. Once a map is digitized, the tic points become registration or geographic control

points for that coverage allowing all coverage features to be registered to a common coordinate system. Additionally, other coverages, such as adjacent coverages or other layers of the same area, can be spatially related using the same tics and the same map projection to ensure geographic control.

Transformation: When a map is digitized, the x- and y-coordinates are held initially in digitizer measurements. To make this information meaningful and also to impose a scale factor, it is necessary to convert these measurements to the real-world coordinate system in the same projection as the original map. This process is known as *transformation.*

Points representing tics can be located on maps by identifying the x- and y-coordinates from a known real-world coordinate system. These are usually given in both the map projection units, usually in meters or feet, and degrees of latitude and longitude. Often real-world coordinates for these tics are taken straight from an original map. But if these are not available, then the latitude and longitude values can be used to identify geographic registration points, which are later converted to coordinates in the same projection used by the base map.

Typically, a tic table is developed manually, before any processing takes place. This table contains a list of all the tic numbers and the known locational reference for each tic, read from the original map. By recording each tic location in the projected map coordinates, the transformation process converts the existing digitizer measurements to real-world coordinates. If the tic locations are in latitude and longitude, they must be converted into the base map projection before performing the transformation.



1 2

The table on the following page outlines the process used to create a coverage in real-world coordinates. You've already completed the tasks listed in Step 1 (in Lesson 3) and Step 2 (in Lessons 4 and 5). Now it's time to complete the remaining steps.

UNIT- 4

UNIT-4 Linking of Spatial Database & with Non spatial Database: Verifying of common item, availability and joining of attribute – table with existing spatial records – spatial display of non spatial data.

Data Processing in GIS

INTRODUCTION

GIS programs link attribute data files to digital maps. The previous Chapter focused on the map side of this equation. Let's focus now on the attribute data files. Like the previous chapter, this chapter examines several key concepts and covers the preprocessing of your GIS data, but it specifically focuses on attributes, data files, and the editing of your attribute data. The concepts focus on attribute data and principles of raster and vector database management. Understanding these concepts will help you to effectively edit and manage your attribute data. The bulk of the chapter focuses on various preprocessing routines including adding and deleting fields, deleting records, joining data files, selecting and sorting records, calculating attributes, and geocoding. The chapter ends with a short discussion regarding attribute verification.

ATTRIBUTE DATA

As described in the previous chapter, spatial data occupies geographic space. It has a specific location that is tied to one of the world's geographic referencing systems (like latitude and longitude). Besides spatial data, GIS files contain non-spatial attributes that describe the spatial features. This section focuses on these non-spatial attributes.

Related to the discussion of "measurements of scale" in Chapter 2, your attributes can be classified as either qualitative or quantitative and actual or derived. Quantitative data focus on numbers and frequencies rather than on subjectivity, meaning, and experience. They are easy to analyze statistically, and their values are often the result of field work and laboratory experiments. Maps exhibiting quantitative data depict differences in magnitude among features.

Qualitative data, by contrast, often provide deeper description and meaning. Maps displaying qualitative data show differences in kind or type. You might subjectively judge whether a quantity is low, medium, or high. You might also classify detailed land uses into broader categories of residential, commercial, and industrial. The statistical options are narrowed too due to the subjectivity of the data and the categorization of data into classes.

Data can also be defined by whether they represent some intrinsic characteristic of the feature being measured (absolute), or whether they are in a sense “created” (derived). Absolute data consists of both the quantitative and qualitative data just described, but it represents phenomena that are measured (like election data or the amount of water stored), the ranking and rating of attributes (even though this process can be subjective), and personal, subjective accounts gained from questionnaires and surveys.

Derived attributes either do not occur naturally, or they cannot be directly gathered; they are the result of statistical manipulation that produces the data. An example is average July temperatures, which is the calculated result of averaging many actual temperature values. Derived data may result from averaging actual values like these, or they represent the relationships between already gathered attribute data, which take three forms: ratio, proportion, and percentage.

- *Ratio* attributes are derived when the value of one attribute is divided by the value of another. Population density is a good example. The total number of people within a particular region is divided by the region’s area. Both the population and area attributes may be “actual” values, but the calculated population density attribute is derived.
- *Proportion* compares the value of one attribute to the total value of all related attributes. The proportion of all African-Americans to the total population is derived by dividing the number of African-Americans (actual data) by the total number of people (also actual).
- Many people think of proportions as *percentages*; they are similar, but percentages multiply proportions by one hundred.

PRINCIPLES OF DATABASE MANAGEMENT – VECTOR

Let’s turn our discussion from characteristics of data to how these values are organized within a data file. Data files are the basic “database” for many programs including spreadsheets, statistic programs, and GIS. Within a GIS, there is a data file for each particular type of geographic feature (e.g. streets, street lights, buildings, and parcels of land). They are the database's version of your features. The data files are automatically created when feature layers are defined in your GIS. You place into them the attributes related to the features.

Data files, often called "tables," arrange attributes within a matrix of fields and records. Fields form the columns of a data file (see Figure 4.1), and they contain the values for each specific attribute you are collecting. For example, parcels might include attributes such as area, land use, and Assessor's Parcel Number (APN). In this example, you would have at least three fields: one called area, another titled land use, and one labeled APN.

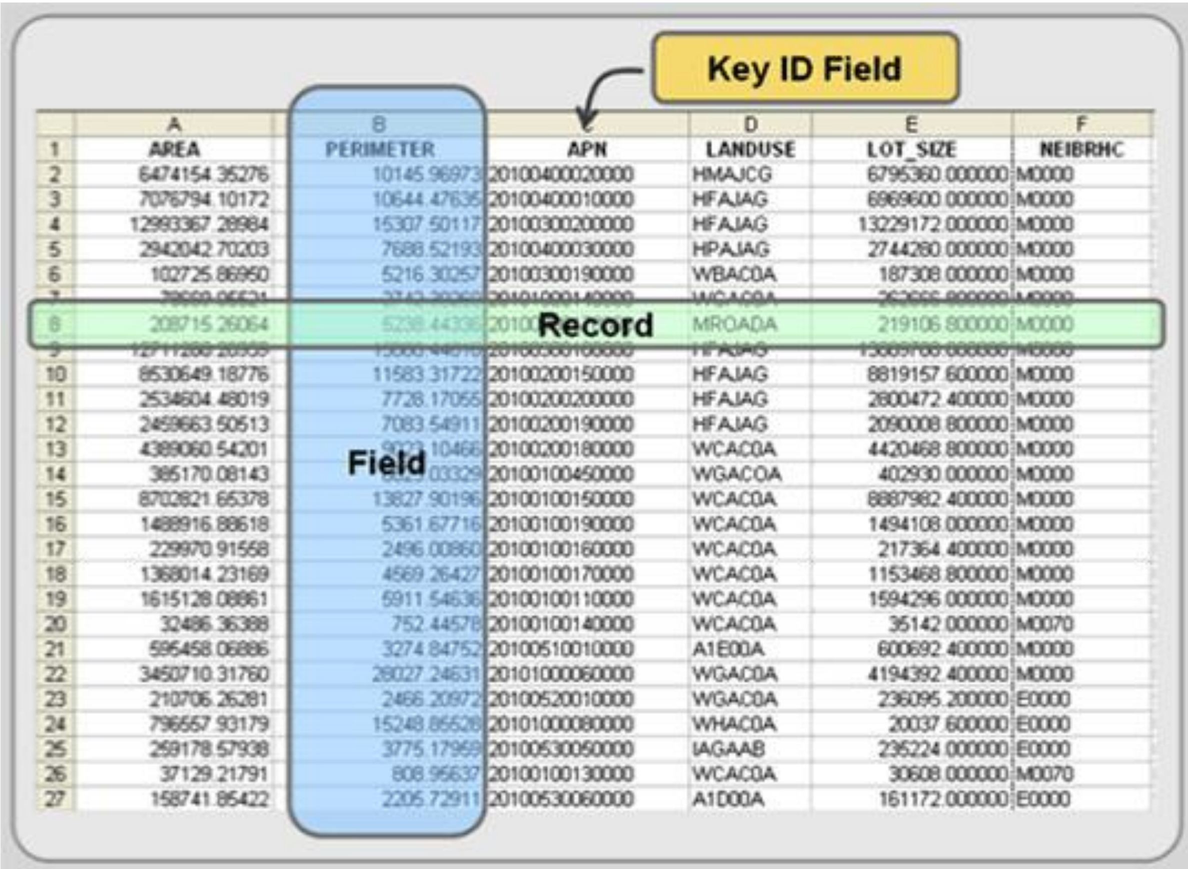


Figure 4.1: Key parts of a data file.

Remember from Chapter 2 that each of these fields has a specific "data format" that defines the type and length of the value that can be directly entered into the data file. Frequently attributes are coded as one of the following, but there are many data formats and the specific name of the data format often changes from one software program to another. Broad data format categories include:

Integer	Numeric values consisting of whole numbers. No decimals.
Real	Numbers consisting of integers with decimals.
Byte	Numeric values ranging from 0 to 255.
Character	Alphanumeric values.

Figure 4.2: Data format categories.

A single record, a row in the data file, represents the database's version of a single feature, including all of its specific attribute values (see Figure 4.1). A few of these attributes may be system variables that the GIS needs for data integrity reasons and to link the data file to the feature's spatial files. In addition, some GIS programs automatically generate length calculations for line features and both area and perimeter calculations for polygon features. Each data file should have a key identifier field that uniquely identifies each feature (i.e. each record). The remaining attributes are up to you and the purpose of your study.

Data files are a collection of related records. If you have 25 street lights within your GIS, you will have 25 street light records in its attribute file. As briefly described above, a largely empty data file is created when a new layer is defined within a GIS program. It is your job to add fields and attribute values to the data file. These descriptive attributes can be entered by hand or imported from external sources. It is likely that you will enter some attributes by hand (and it can be time consuming and tedious), but many—if not most—of the attributes you seek will be imported or "joined" from separate, non-GIS data files. This is because many non-spatial data files predate your need for their incorporation into a GIS, but it is deeper than that. Data manipulation within GIS is clumsy, and since most GIS users are familiar with data management programs like Excel and Access, they prefer working with these programs and then exporting their data and "joining" the external data file to the GIS data file. The joining process is described later in this chapter.

These external data files are coded in one of many "file formats". Some file formats are specific to a particular software program while others are somewhat universal. Even those using a program's proprietary format can export the data file into one of many formats that most GIS programs can read. Some of the file formats that can be read by most GIS programs include:

dBase This industry standard format is read by just about every GIS program. Many GIS programs use this format internally rather than creating their own.

Excel and **Access** – Microsoft's file formats for Excel and Access can be read by many GIS programs. If your GIS program does not read these formats, open the data file in Excel or Access and export it into a format that your system reads.

ASCII (American Standard Code for Information Interchange) – Since most computers use ASCII to represent text, it is possible to transfer data from one computer to another in this format. It is also read and written by most GIS programs, but it is rarely used as the primary GIS file format (with the exception of some raster-based GIS programs). Some government data sets are contained in this file format. Text files come in several different "delimited" forms, and all may include numeric or alphanumeric content (see "Joining Data Files" later in this chapter).

Data files contain a matrix of fields and records for each feature layer. A database is a collection of several related data files (like parcels, street lights, and buildings). In other words, databases contain data files for related layers. Accessing these data files are done through either the GIS software or increasingly from external database management systems (DBMS) that are linked to the GIS. DBMS are specialized programs that organize, manipulate, and report non-spatial data and help you store your data more efficiently. They are particularly valuable when working with large data sets because you can select a subset of your records and fields to work with. The entire attribute file does not have to be used. Examples of external DBMS programs include Access, Oracle, Ingres, SQLServer, INFORMIX, and to a lesser degree Excel, which can serve as an elementary database program. Regardless of whether you are accessing the data files within the GIS software or from an external DBMS, all databases have standard operations which include sorting and selecting records, deleting records and fields, and editing fields and attributes.

Different databases have different structures or ways to organize data. The hierarchical and network data models are two examples, but they are rarely used for GIS (and so will be skipped in this section). For vector systems, the relational database model is the most common data model arguably because they are more flexible, the table structure is easy to understand and program, and outside of GIS, data files are commonly held in relational databases.

Linking or joining data files is the relational database model's strength. Key identifiers, found in multiple data files, are used to link records from one data file to another. In other words, you cross reference multiple data files using common attributes and attach (or join) these external data files to your internal GIS data file. This link takes the selected fields in the data file you wish to join and relates them to the appropriate records in the GIS data file. This requires that each data file have at least one common field to perform a join. There are different names for the key identifier including key and primary key. This process is highlighted later in this chapter.

Many, however, think that the relational database model does not adequately represent spatial data. For some, records in a relational data file are too discrete; they do not properly depict the continuous and multi-dimensional nature of the features they are representing. We use relational data models because they are simple and convenient, but we artificially bend geographic features to conform to existing database standards that were created for non-spatial data.

This has led to the development of object-oriented data structures, which are seen as a more sophisticated database model. The database discards many of the foundational concepts that we have applied throughout this book. Features are defined differently; object-oriented features blur the line between points, lines, and polygons. Also, instead of having multiple files for each GIS layer, the geography and attribute data are integrated into a single file. This allows for simultaneous geographic and attribute editing and quicker processing. The more sophisticated model, however, is a more complex model, and that may have slowed its spread even though "object-oriented" databases were one of the hottest topics in GIS in the 1990s. It may still be the touted successor of the relational model, but it seems that the relational model, despite its drawbacks, has significant pluses—including its ease of use—that will help it dominate at least into the near future.

PRINCIPLES OF DATABASE MANAGEMENT – RASTER

As described in Chapter 1, the raster data model aligns the Earth's surface into a grid of columns and rows. Cells, or pixels, the building blocks of the raster data model, form at the intersection of the columns and rows, and each cell contains a single attribute value, representing the condition of a specific portion of the Earth's surface. That means that a single raster layer only contains the values for one specific attribute across space. That last point is important because raster layers fill space. Their attributes occur everywhere in the study area; there are no blank spaces. Empty areas get a "0" value, but every pixel gets a value. If you need more than one attribute, you construct multiple layers, each containing a single specific attribute for the same area. Conceptually, it is a simple model. As in Figure 4.3, your study area is divided into cells, and each cell of each layer has a single attribute that represents that area.

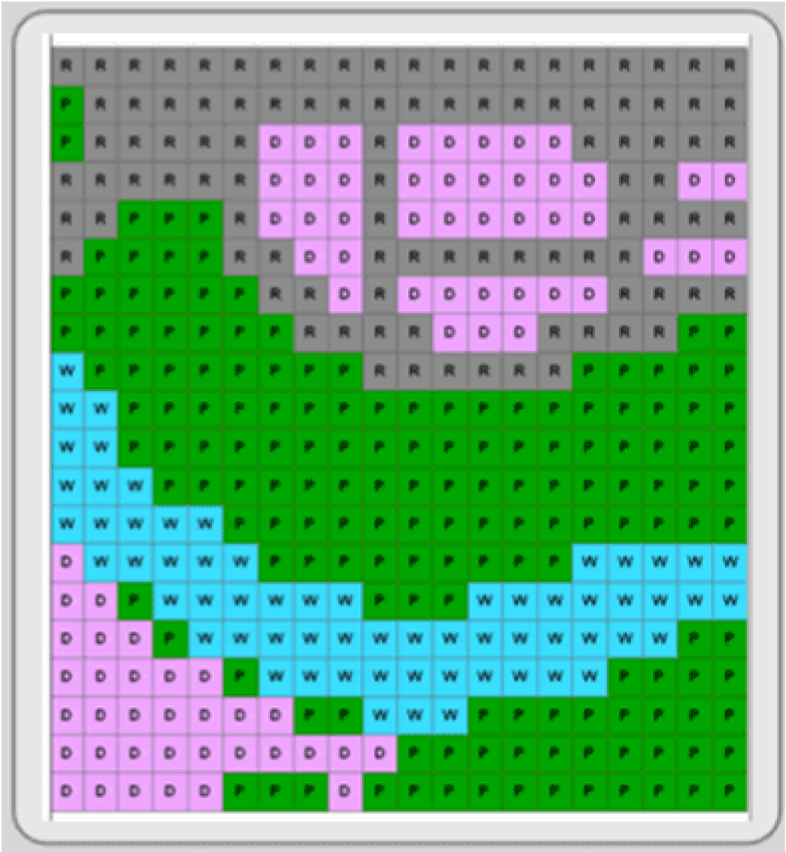


Figure 4.3: Raster image. Image by Mike Tuck.

There are many ways—some more complex than others—the raster data model may be stored. The two general categories are regular and irregular. The regular structure is conceptually simple, and includes two types: full raster encoding and run-length encoding. Full raster encoding creates a data file that records the attribute value for every pixel. It’s as though you read an image’s pixels like a book, starting in the upper left corner and reading from left to right and downward row by row. The data file looks a bit different. It records each pixel’s attribute value on a separate line, so if you had an image with 640,000 pixels, your data file would have 640,000 lines, making it a very long data file. Figure 4.4 is a simplified example.

Figure 4.4: Full raster encoding. This figure is the beginning—just the first three rows—of the data file for the image in Figure 4.3. Color is added to highlight the different attribute values.

Run-length encoding is more efficient than full raster encoding. Since the same values often occur in runs across several cells, run-length encoding enters the attribute values as pairs: the first number is the run length and the second number is the cell's value. This substantially reduces file size especially if contiguous pixels have the same value. Contrast Figure 4.5 with Figure 4.4.

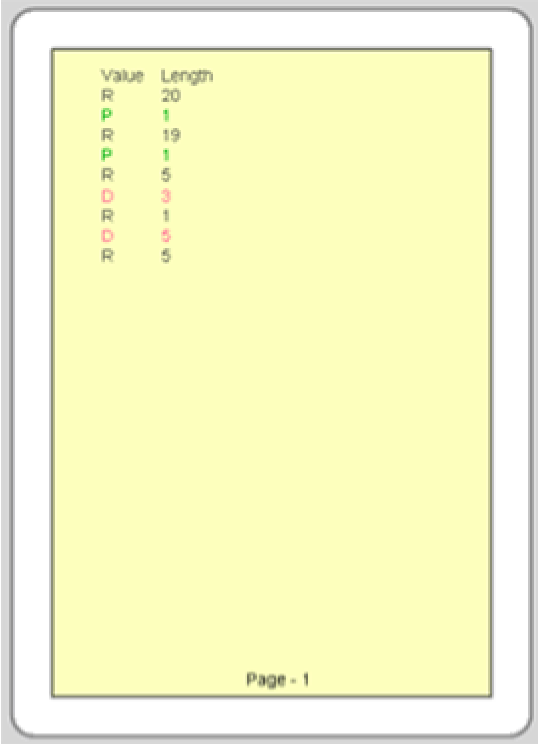


Figure 4.5: Run-length encoding. This figure also depicts the first three rows of Figure 4.3. Compare run-length encoding with full raster encoding (Figure 4.4). Color is added to highlight the different attribute values.

Irregular raster data structures, like quadtree and others, are more complex, proprietary, and beyond the scope of this e-text. They usually make file size smaller and provide ways to store raster data for quick retrieval.

ATTRIBUTE PREPROCESSING AND EDITING

When you add feature layers, containing both spatial and attribute data, to an active workspace, the attribute data file might not be immediately visible. Opening and editing the attribute files are easy processes, but they are specific to individual programs. Once the attribute table is open, you can enter data by typing attribute values directly into the data file or loading and joining external data files to it. Other processes like editing attributes, adding or deleting fields, deleting records, querying attributes (record selection), calculating fields, and geocoding are completed through the data file interface.

Adding and deleting fields

As described above, fields define feature attributes. Most GIS programs provide a way for you to add or delete fields from within your open data file. The GIS program will instruct you to define a new field. You will give it a name and select from options that determine the data format of the values that will be placed into the field. Deleting a field usually involves selecting the field and deleting it.

Deleting records

You can delete a single record or a group of records in a data file by first selecting them and then deleting them. Since records are the database representation of features, when you delete records in the attribute file, you are also permanently discarding their spatial representation. The entire feature, graphic and record, is deleted.

Generally, you can not add a record through the data file interface because it must also be represented spatially. See Chapter 3 for how to add a feature. Its record is automatically created when the graphic feature is added to the workspace.

Joining Data Files

Once a GIS layer is created, its attribute file can be linked ("joined") to external data files. Joining is one of the most frequently performed data file processes because it brings together feature attributes that are contained in multiple digital data files. To perform a join, a unique matching field, the key identifier, must be observed in both data files. As stated in Chapter 3, the key identifier could be something like a social security number or an assessor parcel number. It is a field that gives the feature a unique identification. Once linked, the join can be temporary or made permanent.

The external files that you load into the GIS to perform a join are typically in file formats such as dBase, ASCII, Microsoft Excel, or Microsoft Access. The precise steps involved in joining together two files are software specific, but it usually involves:

1. loading the external file that you wish to join to the GIS attribute file,
2. selecting the external file and the GIS attribute file that you wish to join,
3. selecting the field (containing the key identifier) in each file, and when joined,
4. making sure that the join was successful.

In the example in Figure 4.6, the parcel layer exists, but it does not include assessed value. It does contain a field named APN (Assessor's Parcel Number) whose values are unique to each record and which could be used to join other data files. A spreadsheet file, with assessed value, also exists, and it must be loaded into the GIS either in its native format (if accepted) or exported from the spreadsheet program to a format that the GIS can read. The spreadsheet has a field named APN_NUM, which, after a visual check, has the same values as those under APN in the parcel layer, and it can be used to perform the join.

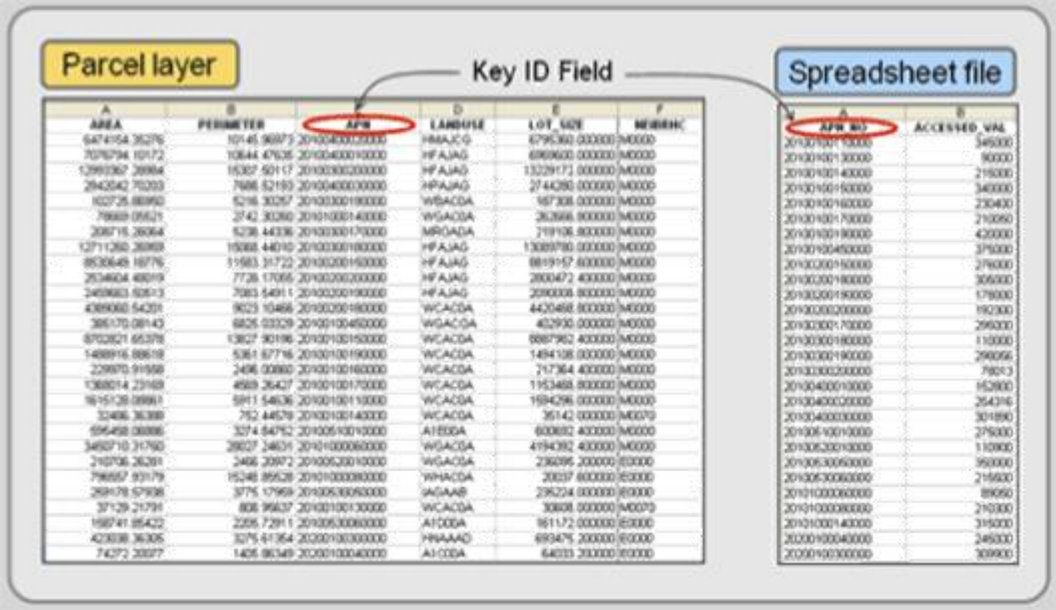


Figure 4.6: Joining two attribute files together requires that the two files each have a common key identifier.

Once the spreadsheet file is loaded, you begin the joining process by specifying the two files (the layer's table and the spreadsheet file) and the two field names that the join will be made on. APN and APN_NUM are the key identifiers of these two files (see Figure 4.7), and even though the field names are not identical, the GIS will be able to join these two files together provided that the values under the two field names match.

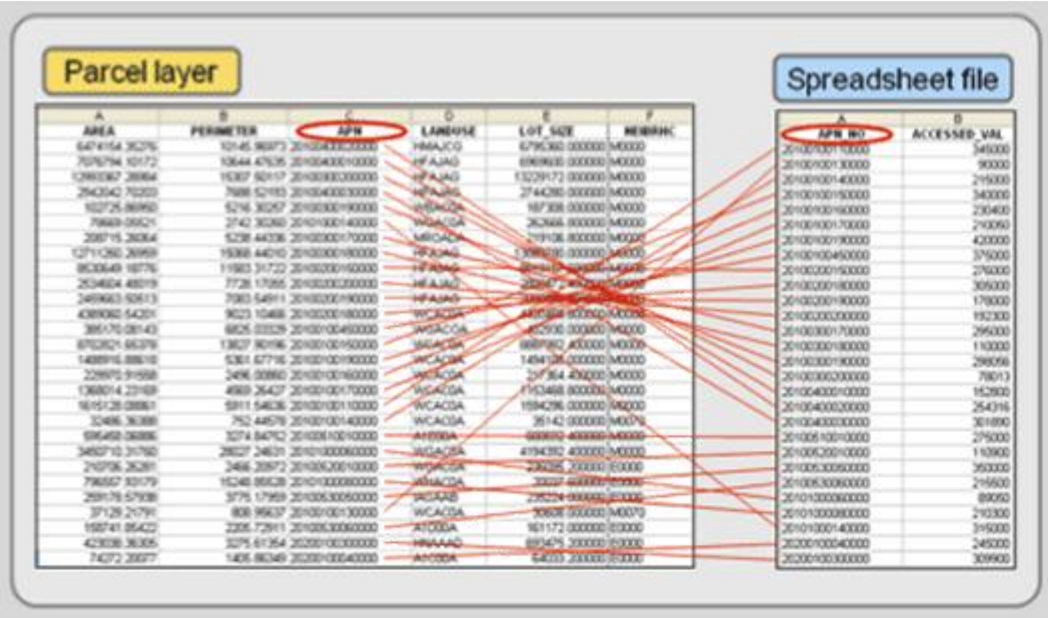


Figure 4.7: Matching key identifiers.

If the match is successful, your two files will be joined together into a single file (see Figure 4.8).

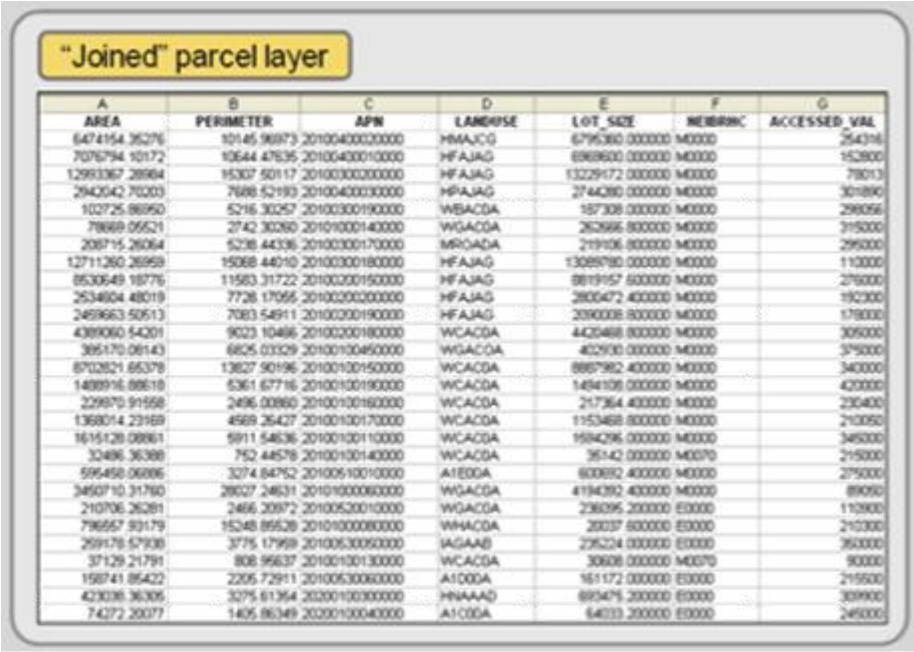


Figure 4.8: A joined file with accessed values a one of the attributes.

Perhaps the most time consuming tasks are the first and fourth steps. Loading an external data file should be easy —and frequently it is—but sometimes the imported data file may be misformatted or unreadable. If it is, return to the host program (your spreadsheet or DBMS programs) and save it in a different format. The probability of your GIS program being able to read the external data file usually improves as you go from more sophisticated file formats (like Excel and Access) to dBase to ASCII (basic formats). Many data files are coded in ASCII because of its almost universal compatibility with computers and software programs, but it does have its complications—it comes in several forms. Below are four of the most used variants of ASCII based on what delimits the file's fields.

Whitespace delimited ASCII files differentiates fields by the use of one or more spaces. Since spaces separate fields, fields that have no value must be represented by a non-blank code and character attributes cannot contain spaces between words (underscores can be used to separate words). You can open ASCII files in any word processer or text editor. A whitespace-delimited ASCII data file with five records might look something like the following:

M1 Betsy_Burns Yes 38.5 0.85

P1 Dan_Arreola No 45.7 0.99

M2 Frank_Aldrich Yes 32.8 0.55

P2 Fritz_Steiner No – -

P3 Ruth_Yabes No 37.72 -

Spacequote delimited ASCII is a variant of whitespace delimitation, but the attributes containing multiple words are enclosed in double quotes, and consequently, they can contain embedded spaces between words. The spacequote delimited ASCII file may look like the following in a text editor:

M1 “Betsy Burns” Yes 38.5 0.85

P1″Dan Arreola” No 45.7 0.99

M2 “Frank Aldrich” Yes 32.8 0.55

P2 “Fritz Steiner” No – -

P3 “Ruth Yabes” No 37.72 -

Tab delimited files separate fields by the use of a single tab. Two tabs in a row signify a blank field. Values within an attribute field cannot contain embedded tabs. A tab delimited ASCII file would look like the following in a text editor.

M1   Betsy Burns   Yes   38.5   0.85

P1   Dan Arreola   No   45.7   0.99

M2 Frank Aldrich Yes   32.8   0.55

P2   Fritz Steiner   No

P3   Ruth Yabes   No   37.72

Comma delimited, also known as comma-quote delimited and CSV, separate fields by commas. Character fields may be enclosed in double quotes, and need to be if they contain an embedded comma. Two commas in a row signify that the field is blank. Usually whitespace is not allowed before or after fields (although this may be tolerated in the CSV form). The comma-delimited ASCII file might look like the following in a text editor:

M1,”Betsy Burns”,Yes,38.5,0.85

P1,”Dan Arreola”,No,45.7,0.99

M2,”Frank Aldrich”,Yes,32.8,0.55

P2,”Fritz Steiner”,No,,

P3,”Ruth Yabes”,No,37.72,

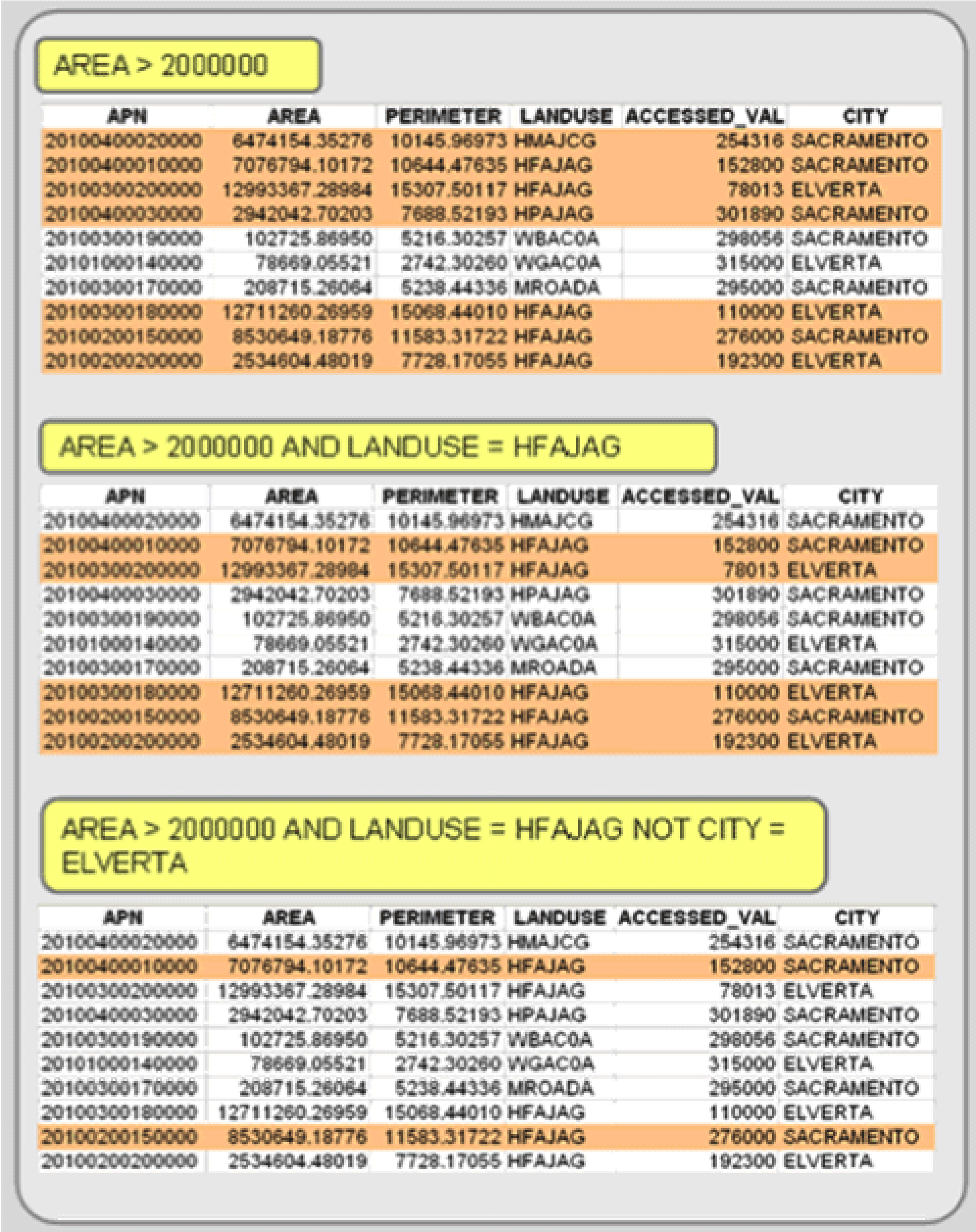
Sorting records

Sorting temporarily rearranges your data file records, so you can view, select, update, or print them in the new sorted sequence. Although the specifics vary by program, you generally choose the field (or fields) you want to sort by. The first sort field arranges, usually in ascending or descending order, the records based on the field's contents. For example, a class roster might be sorted alphabetically by last name. Some systems allow you to choose a second sort field (or more), which arranges records (in ascending or descending order) when two or more records have the same first field value. In the example above, if your alphabetical list has four students with the last name Smith, those four records could be rearranged in alphabetical order based on their first name.

Record selection/Attribute Query (Boolean Selection)

Selecting specific records is one of the most common database functions. Often called attribute query, it consists of highlighting a subset of the records based on a specific criteria. In other words, you create an expression—a formula—that queries all the records in the data file and the GIS highlights—both in the data file and on the map display—only those features that fit the criteria.

Most GIS programs use a Standard Query Language (SQL) interface to conduct attribute queries. If one is using an external relational DBMS program (like Access or Oracle), SQL makes the call to the external database and isolates only the necessary records that you will use. SQL uses set algebra, Boolean algebra, and arithmetic operators (=, -, *, /) for attribute queries. Set Algebra includes the use of less than (<), greater than (>), equal to (=), and not equal to (<>) operations. You can create an expression like that found below (see Figure 4.9) to isolate only those records that fit your criteria. You can extend or constrain the selected features by using Boolean algebra, which uses the conditions OR (extend), AND (constrain), and NOT to further select or isolate records. Each record is queried and added to the set if it meets the criteria.



APN	AREA	PERIMETER	LANDUSE	ACCESSED_VAL	CITY
20100400020000	6474154.35276	10145.96973	HMAJCG	254316	SACRAMENTO
20100400010000	7076794.10172	10644.47635	HFAJAG	152800	SACRAMENTO
20100300200000	12993367.28984	15307.50117	HFAJAG	78013	ELVERTA
20100400030000	2942042.70203	7688.52193	HPAJAG	301890	SACRAMENTO
20100300190000	102725.86950	5216.30257	WBACOA	298056	SACRAMENTO
20101000140000	78669.05521	2742.30260	WGACOA	315000	ELVERTA
20100300170000	208715.26064	5238.44336	MROADA	295000	SACRAMENTO
20100300180000	12711260.26959	15068.44010	HFAJAG	110000	ELVERTA
20100200150000	8530649.18776	11583.31722	HFAJAG	276000	SACRAMENTO
20100200200000	2534604.48019	7728.17055	HFAJAG	192300	ELVERTA

APN	AREA	PERIMETER	LANDUSE	ACCESSED_VAL	CITY
20100400020000	6474154.35276	10145.96973	HMAJCG	254316	SACRAMENTO
20100400010000	7076794.10172	10644.47635	HFAJAG	152800	SACRAMENTO
20100300200000	12993367.28984	15307.50117	HFAJAG	78013	ELVERTA
20100400030000	2942042.70203	7688.52193	HPAJAG	301890	SACRAMENTO
20100300190000	102725.86950	5216.30257	WBACOA	298056	SACRAMENTO
20101000140000	78669.05521	2742.30260	WGACOA	315000	ELVERTA
20100300170000	208715.26064	5238.44336	MROADA	295000	SACRAMENTO
20100300180000	12711260.26959	15068.44010	HFAJAG	110000	ELVERTA
20100200150000	8530649.18776	11583.31722	HFAJAG	276000	SACRAMENTO
20100200200000	2534604.48019	7728.17055	HFAJAG	192300	ELVERTA

APN	AREA	PERIMETER	LANDUSE	ACCESSED_VAL	CITY
20100400020000	6474154.35276	10145.96973	HMAJCG	254316	SACRAMENTO
20100400010000	7076794.10172	10644.47635	HFAJAG	152800	SACRAMENTO
20100300200000	12993367.28984	15307.50117	HFAJAG	78013	ELVERTA
20100400030000	2942042.70203	7688.52193	HPAJAG	301890	SACRAMENTO
20100300190000	102725.86950	5216.30257	WBACOA	298056	SACRAMENTO
20101000140000	78669.05521	2742.30260	WGACOA	315000	ELVERTA
20100300170000	208715.26064	5238.44336	MROADA	295000	SACRAMENTO
20100300180000	12711260.26959	15068.44010	HFAJAG	110000	ELVERTA
20100200150000	8530649.18776	11583.31722	HFAJAG	276000	SACRAMENTO
20100200200000	2534604.48019	7728.17055	HFAJAG	192300	ELVERTA

Figure 4.9: Select records based on their attributes by using SQL expressions.

Once the records are selected, you can work with just those records. This is helpful for viewing, sorting, editing, calculating fields, generating statistics, using the selected features to select features in another GIS layer, creating a new layer with only the selected features, and isolating specific records to perform analysis functions on (like buffering selected features).

In addition, spatial queries, selecting features based on their geographic location (see Chapter 5), can be combined with attribute queries for more sophisticated queries. There is more on attribute and spatial queries in Chapter 5.

Calculate Attributes

Within an open data file, you can create new attributes by using values in existing fields, mathematical expressions, and text functions (see Figure 4.10). Mathematical operations allow you to add, subtract, multiply, and divide existing fields or values to create new, derived attributes. Text functions allow you to populate fields with data, copy values from one field to another, concatenate fields (and or values), truncate attributes, and convert text to different formats. Before calculating the new field, however, you need to create a new attribute field, which includes defining its field name and its data properties). Calculations can be performed on a single record, several selected records, or on every record in the data file. The calculate function can also be used to copy data from one field to another.

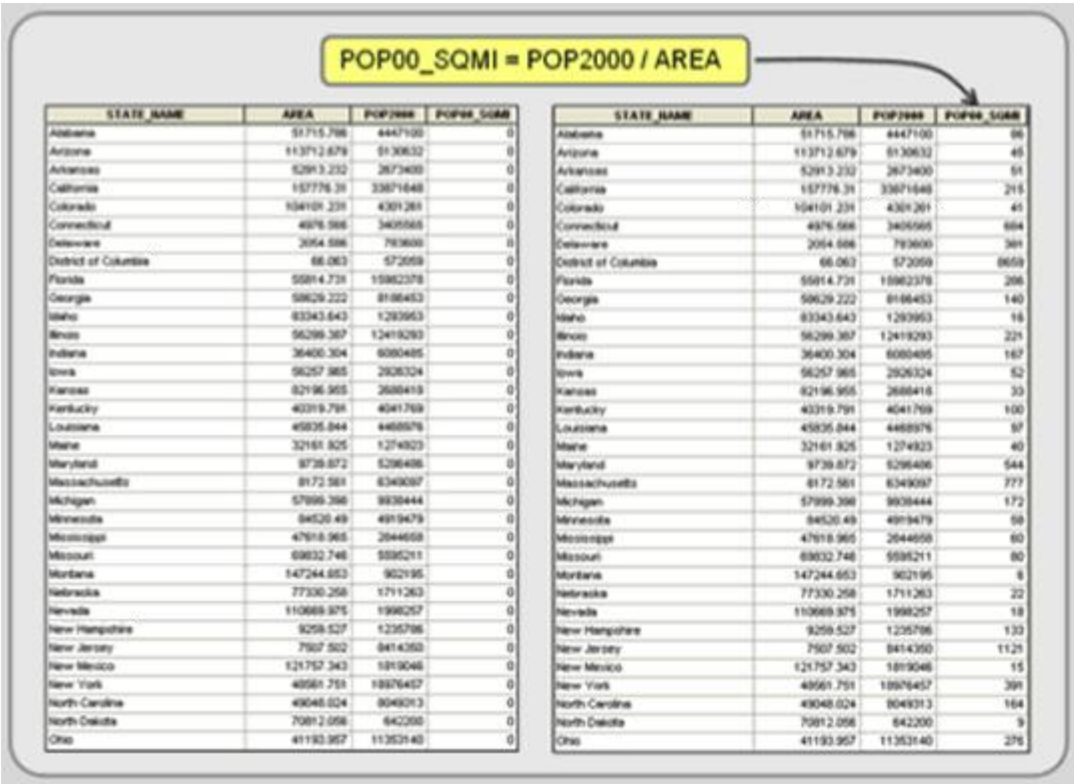


Figure 4.10: Calculating fields. In this example, population density is calculated by dividing population by area. First, the field must be added. Then, you calculate the results directly into the new field.

Geocoding

There is a way to create geographic data directly from attribute data. The process, called geocoding, assigns geographic locations to features directly from attribute fields that contain locational information within a data file. This is a popular way to create GIS feature layers; you create or obtain a spreadsheet or data file with location information, open the attribute table in your GIS, and direct the system toward the appropriate attribute fields. There are two types of geocoding: coordinate locations and address matching.

Spatial features can be created from data files containing fields with x,y coordinate values. The coordinates need to be separated into two separate fields: one for the x coordinate and one for the y coordinate. The process is straightforward; you direct the GIS to the data file's appropriate x,y fields, and it creates a spatial layer of point features from the coordinates. One possible complication is that the data file's coordinates are different than the coordinate system you are using. This requires that you open the file in a temporary workspace registered to the data file's coordinate system and then convert the new spatial layer to the desired coordinate system.

Address matching is another type of geocoding. It matches records in two data files—one containing a list of addresses and the other having street network attributes—to create a new layer (see Figure 4.11). In other words, it creates a layer of point features alongside street segments when addresses in the two data files match. It essentially looks up the address in the first record of the external data file and tries to find a match along the street network layer. If multiple possibilities exist, the routine will present them for user input. After the first record is matched or not, it moves to the second record and tries again. The resultant file is assigned the street network's coordinate system.

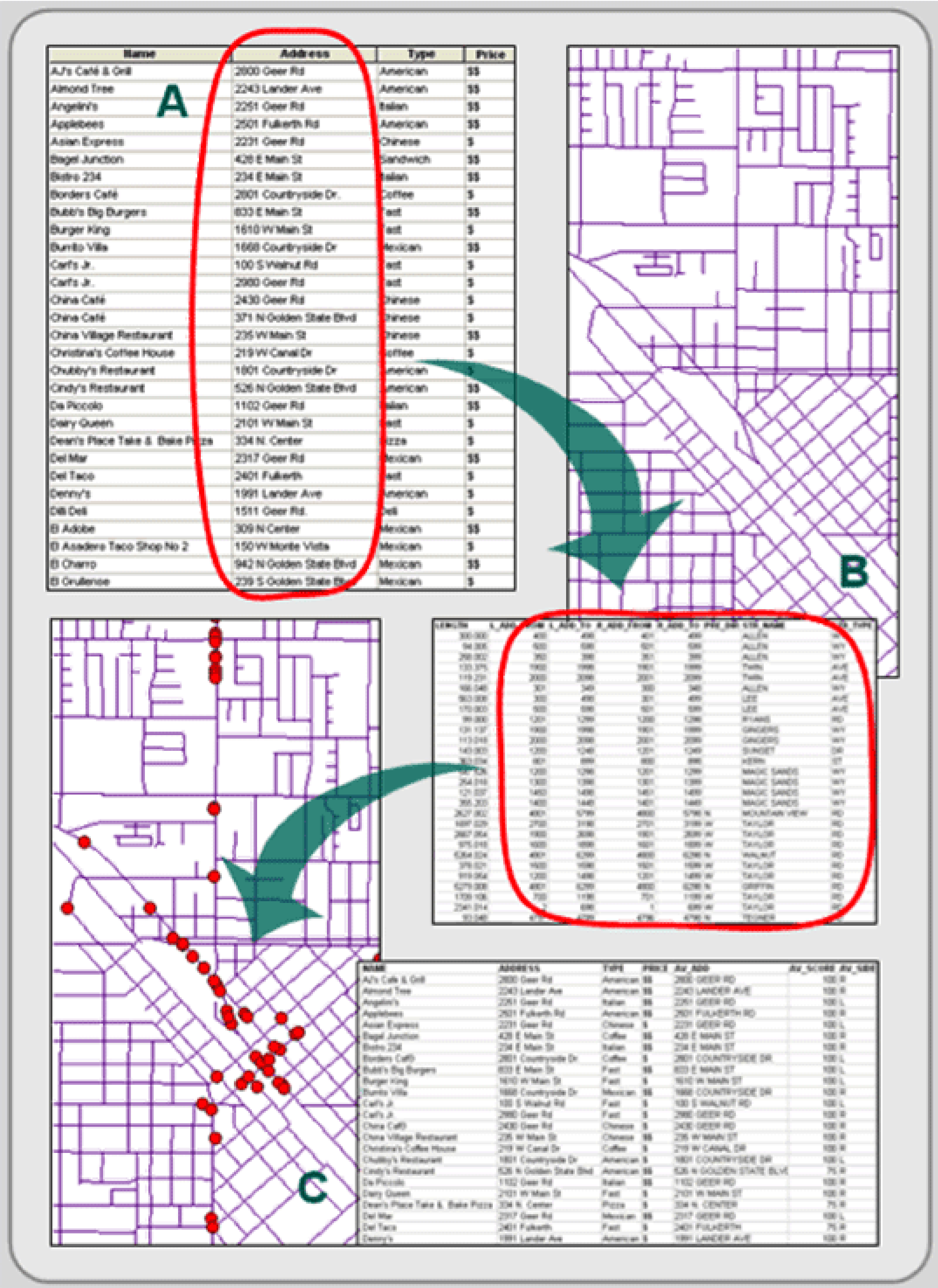


Figure 4.11: Address matching. The addresses in an external data file (A) are compared to a street network's (B) attribute fields, and if a match is made, the record in the external data file gets a point on the map (C).

Both the street network layer and the external data file need address data (street name, street type, and an address range for start and end of each line segment), and perhaps even more information like city, state, and Zip code attributes to make your address information unique (multiple cities will likely contain streets with the same name). The process works well if the addresses in both the external data file and the street network layer are accurate and complete, but address matching is a time consuming process.

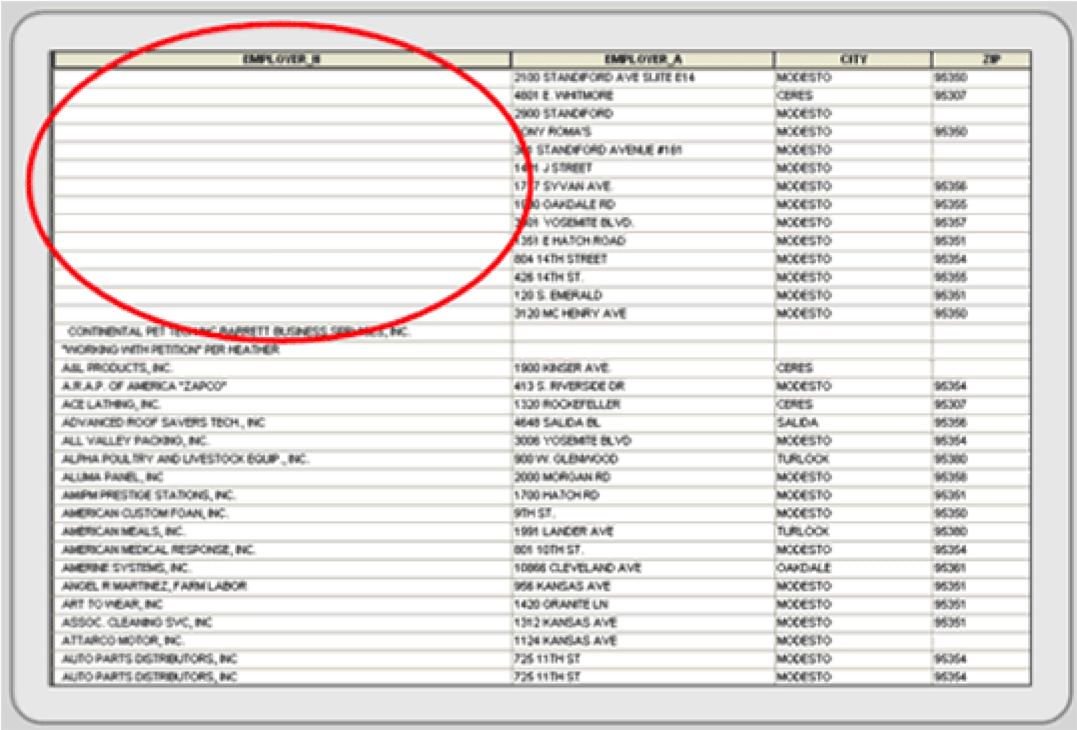
Data Export

Exporting your GIS layers, including their geographic and attribute data files, are covered in Chapter 6. Most GIS programs can export your layer's attribute file in a number of formats including dBase and ASCII. The exported files can then be used in database, statistic, and spreadsheet programs for additional analysis.

ATTRIBUTE VERIFICATION

This section looks at verifying the accuracy of attributes. The verification process looks for both missing attributes and incorrect attribute values. Unlike geographic verification, there are no attribute verification procedures built within the software to verify their accuracy.

Instead, the layer's data file can be displayed and sorted by each attribute in ascending order to identify missing attributes (see Figure 4.12). Map features missing a value for a particular field are revealed at the top of the table. Selecting those features from the data file and highlighting them on the screen can be a handy way to reference those features that have missing attributes. The selected features can then be investigated and updated. You can also sort the attributes alphabetically and glance down the field looking for spelling mistakes.



EMPLOYER_B	EMPLOYER_A	CITY	ZIP
	2100 STANDIFORD AVE SUITE E14	MODESTO	95350
	4801 E. WHITMORE	CERES	95307
	2900 STANDIFORD	MODESTO	
	[illegible]ONY ROMA'S	MODESTO	95350
	[illegible] STANDIFORD AVENUE #181	MODESTO	
	14[illegible]1 J STREET	MODESTO	
	17[illegible]7 SYVAN AVE.	MODESTO	95356
	1[illegible]0 OAKDALE RD	MODESTO	95355
	[illegible]01 YOSEMITE BLVD.	MODESTO	95357
	[illegible]351 E HATCH ROAD	MODESTO	95351
	804 14TH STREET	MODESTO	95354
	426 14TH ST.	MODESTO	95355
	120 S. EMERALD	MODESTO	95351
	3120 MC HENRY AVE	MODESTO	95350
CONTINENTAL PET TE[illegible] INC BARRETT BUSINESS SER[illegible], INC.			
"WORKING WITH PETITION" PER HEATHER			
A&L PRODUCTS, INC.	1900 KINSER AVE.	CERES	
A.R.A.P. OF AMERICA "ZAPCO"	413 S. RIVERSIDE DR	MODESTO	95354
ACE LATHING, INC.	1320 ROCKEFELLER	CERES	95307
ADVANCED ROOF SAVERS TECH., INC	4648 SALIDA BL	SALIDA	95358
ALL VALLEY PACKING, INC.	3006 YOSEMITE BLVD	MODESTO	95354
ALPHA POULTRY AND LIVESTOCK EQUIP., INC.	900 W. GLENWOOD	TURLOCK	95380
ALUMA PANEL, INC	2000 MORGAN RD	MODESTO	95358
AMPM PRESTIGE STATIONS, INC.	1700 HATCH RD	MODESTO	95351
AMERICAN CUSTOM FOAM, INC.	9TH ST.	MODESTO	95350
AMERICAN MEALS, INC.	1991 LANDER AVE	TURLOCK	95380
AMERICAN MEDICAL RESPONSE, INC.	801 10TH ST.	MODESTO	95354
AMERINE SYSTEMS, INC.	10868 CLEVELAND AVE	OAKDALE	95361
ANGEL R MARTINEZ, FARM LABOR	958 KANSAS AVE	MODESTO	95351
ART TO WEAR, INC	1420 GRANITE LN	MODESTO	95351
ASSOC. CLEANING SVC, INC	1312 KANSAS AVE	MODESTO	95351
ATTARCO MOTOR, INC.	1124 KANSAS AVE	MODESTO	
AUTO PARTS DISTRIBUTORS, INC	725 11TH ST	MODESTO	95354
AUTO PARTS DISTRIBUTORS, INC	725 11TH ST	MODESTO	95354

Figure 4.12: Sorting in ascending order can reveal missing data.

More difficult to detect are incorrect attribute values. They require familiarity with the original source maps and an understanding of spatial patterns. For example, if you were working with income data, you should select low income values and display them on a map. Does their spatial location make sense? Do the same with high income values. Nominal data sets can be displayed the same way. For example, select different land use classes, and see if they make geographic sense. Display all heavy industrial sites and look at their locations. If heavy industry appears in the middle of wealthy residential areas or they are not located along highways, railroads, or rivers (which they need for transportation purposes) than these values may be inaccurate. More information may be needed; try looking at web-based aerial photographs or field check odd values.

UNIT- 5

UNIT-5 Designing & Coding of QUBIS: Planning for the user requirement – preparation of spatial & Non spatial relational databases – QUBIS Designing – QUBIS Coding (Testing, Error handling, Monitoring, User interface Development.

How SDSS is different from QUBIS?

QUBIS:

The readymade queries built and kept in the form of GUI based buttons and menus can be operated by clicking mouse button so as to display the user required information automatically as map, very quickly. Spatial display of maps include simple / preferential retrieval of features and/or attribute data linked with spatial features through Join or Relate options in GIS and display them in a specified map layout.

SDSS:

On the contrary, in SDSS, apart from built-in queries linked with attribute tables and GUI options in windows, performing GIS based spatial analyses and spatial modelling lively and automatically at the background of GIS, with options for including latest temporal data or automated map generation with/without reclassification or manipulation, based on the recent data acquired from online sources and use them in analyses so as to generate up-to-date information in the form of basic thematic maps or derivative maps or action plan maps, immediately on strategic basis.

5.1 Planning for the user Requirement

Types of User Requirements for QUBIS-Query Based Information Retrieval System:

1. Query based map retrieval alone
2. Query based map and chart display
3. Query based map, chart and table display
4. Query based map, chart, table, report generation and display
5. Spatial Decision Support Systems for various Natural Resources Management and Conservation

6. SDSS for Natural Disaster Mitigation and Management
7. SDSS for identification of gap areas for Science and Technology Needs or Requirements
8. SDSS for Physical and Human Resources Sector Development and Management
9. SDSS with modeling capability
10. SDSS with networking capability
11. SDSS with Automated Capabilities such as Interpretation of Recent Satellite Images, Data Manipulation, Data Conversion, Data Classification, Layer Generation, Data Analysis and Action Plan Map Generation, etc.

Basic Ideas for QUBIS / SDSS Development

To be decided before SDSS generation:

1. Basic Display Unit
2. Levels of map display
3. Types of Menus and options
4. Number of Thematic, Action Plan and other Derivative maps to retrieve and display
5. Map handling tools for user interaction
6. Levels of help needed
7. Front page Design - based on the Type of SDSS users.

5.2 Preparation of spatial & Non spatial relational databases

1. List of **spatial databases** needed for SDSS to be generated first. For e.g.:
    a. **Base Layers**
        i. Administrative Boundaries
            1. Country
            2. State
            3. District
            4. Taluk / Block
            5. Village, etc.
        ii. Settlements
            1. Major – Cities, Towns, Headquarters, etc.
            2. Villages, Habitations, Huts, etc.

iii. Roads
    1. Major – Asian Highways, NH, SH, DH
    2. Village Roads, Unmetalled, Pack tracks, Cart Tracks, Foot paths, etc.
iv. Railway lines
    1. Broad Gauge
    2. Meter Gauge
    3. Narrow Gauge, etc.
v. Reservoirs, Lakes, Tanks, Rivers, Streams, Canals, Drainages, and other water bodies, etc.
vi. Hill and Reserved / Protected Forest Boundaries.

b. Basic Thematic Maps

i. Lithology
ii. Structural Trend lines and fold styles
iii. Lineaments / Fractures / Faults / Schistosity, Foliation, Shear systems
iv. Geomorphology
v. Slope
vi. Landuse / Land Cover
vii. Subsurface Lithology
    1. Thickness of Top soil
    2. Thickness of Weathered Zone
    3. Thickness of Fractured Zone
    4. Depth to Bed Rock
viii. Aquifer Characteristics
    1. Transmissivity
    2. Permeability
    3. Storage co-efficient
    4. Specific yield
ix. Groundwater level
x. Groundwater table
xi. Natural Recharge and discharge
xii. Groundwater Chemistry

1. Electrical Conductivity
2. TDS – Total Dissolved Solids
3. pH – Hydrogen ion concentration
4. Iron
5. Calcium
6. Sodium
7. Potassium
8. Fluoride
9. HCO3
10. Arsenic
11. Lead, etc.

xiii. Groundwater movement, etc.

c. **Specific Thematic Maps**
   - i. Areas prone for Soil Erosion
   - ii. Areas of Groundwater Pollution
   - iii. Areas prone for Landslides, etc.

d. **Derivative Maps**
   - i. Soil Erosion due to Geomorphology
   - ii. Groundwater Pollution due to industrial waste disposals
   - iii. Landslides due to sensitive toe removal along hill slopes
   - iv. Landslides due to dyke intrusion, etc

2. Similarly, a List of **non-spatial databases** to be included in SDSS has to be generated as mentioned below:

   a. Physical Resources
      - i. Agriculture Sectors
      - ii. Agricultural Engineering
      - iii. Aquaculture
      - iv. Animal Husbandry
      - v. Banks
      - vi. Civil Supplies
      - vii. Developmental Sector
      - viii. Electrical Sector
      - ix. Education
      - x. Fisheries

xi. Horticulture
xii. Highways
xiii. Industries
xiv. Irrigation
xv. Land Survey
xvi. Mining industries
xvii. Postal offices
xviii. Retired Officers Welfare
xix. Sericulture
xx. Small Industries
xxi. Transport
xxii. Etc.

b. Human Resources

i. Children – Male, Female
ii. Middle age People
iii. Elderly People
iv. Employed / Unemployed
v. Habit based classes, etc.

3. …etc.,…

After deciding the number and type of databases need to be generated, the structural formats for collecting, storing and retrieving the spatial and non-spatial data in SDSS by the users according to the requirement of the users/planners need to be established.

For example, if the user is interested to see the type of resources available in an area, then, the particular resource locations in terms of points / lines / polygons must have been digitized and the relevant resource details such as its areal extent, concentration, volume availability and exploitability, might have been worked out and entered in the attribute table.

Similarly for the non-spatial database, the planner may be interested in seeing the locations of different infrastructure / facilities such as hospitals, schools, postal facility availability in an area. In this case, their exact spatial locations as points or the villages having these facilities can be shown with polygon boundaries.

The distance to be travelled by the users / common people can also be shown as the planners are interested to know these details for identifying the grey areas where these facilities can be created and the funds can be allotted and sanctioned based on the priority basis.

Similarly, the database need to be classified or derivative information should be made readily available with attributes so as to retrieve them by the users whenever required. For example, the distance to be travelled to utilize the existing facility/infrastructure can be simplified based on the distance from various places such as, 0-2 km, 2.1 – 4 km, 4.1 – 6 km, >6 km, etc., and the same should be entered in the attribute tables so as to retrieve them by the planners / users.

5.3 QUBIS Designing

A welcome page can be shown to the user on clicking an icon of SDSS so as to introduce the containment and the use of the present system and the designing agency and contact information.

This can be followed by a Main Menu display. The main menu can contain several options with dropdown and slide down menus containing relevant features.

According to the requirement of the user/planner/administrator, the front page of QUBIS or SDSS should be designed incorporating all the necessary Main menus, Sub menus, tools for handling menus and maps in a user interactive and simple mode. Help menu or Tutor menu is most important as the user may be totally new to the handling operations of SDSS. The user should not get struggled by seeing so many tools and menus occupying the entire display area. By moving the mouse cursor over any area in the SDSS front page, he could get some helps or tips as guidance in terms of blinking text with highlighted colours on the next / immediate actions to be performed by the user.

Options could have been provided with simple designs so as to easily move around the SDSS and fetch the details required by the users.

The users may get confused while zooming into maps displayed and the area where they have zoomed. Hence, an index map / key map can be provided with highlighted rectangle showing the current area of the map displayed by the user.

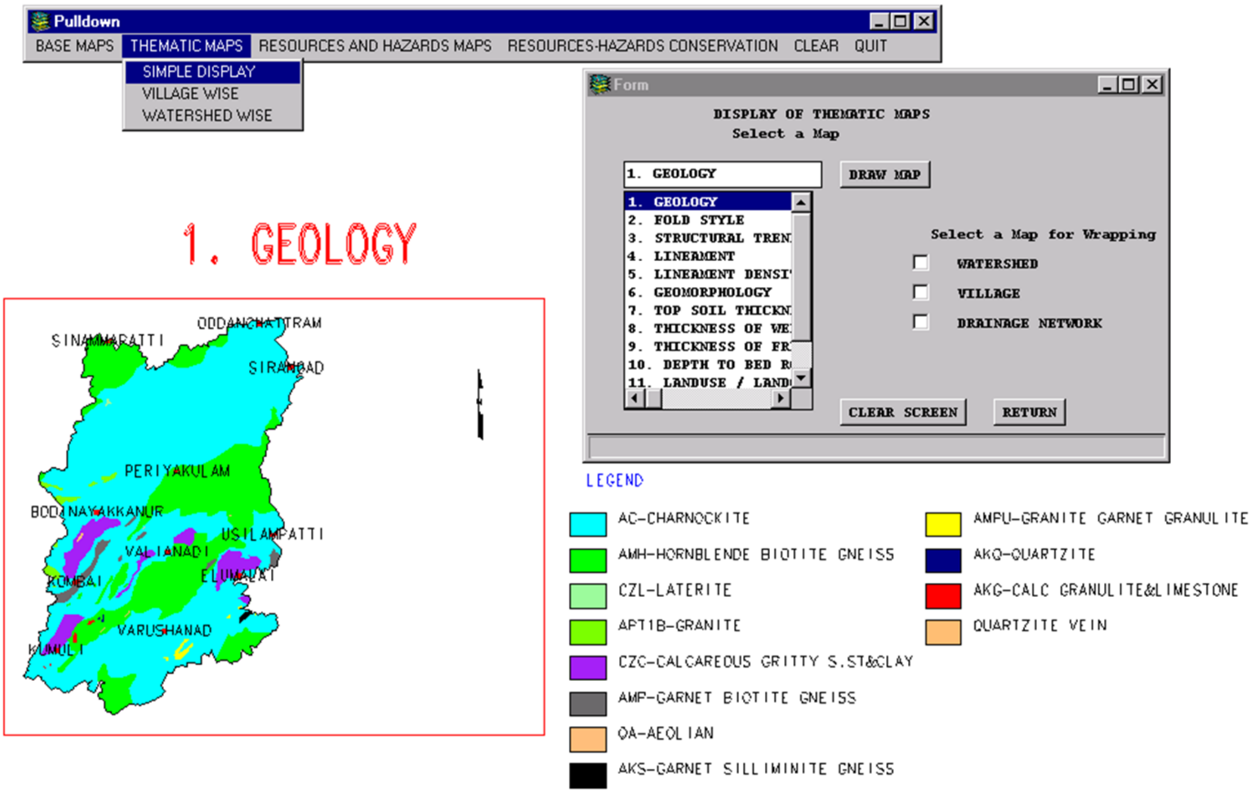
Similarly, the simple map handling tools such as measure tool to know the distance to the features, identify tool to know the clicked features in the display, Softcopy map tool for saving the required maps in digital format for further use by the user, print tool, etc.

The entire SDSS design should incorporate three basic elements of display such as, display areas for (i) menus, (ii) content for the displayed maps / available attribute database and (iii) map display. Additionally, the area for displaying map handling tools, area for displaying live text in the form of hints for the users so as to proceed further and within the map display area, it necessary to allot area for displaying map title, legend, map scale and index map or key map should also be predefined and incorporated during designing stage of SDSS.

Further, it is necessary to allot area for displaying the options in the form of radio button of check boxes for different ‘Levels of Map display’. This is based on the user requirement at the instant. Hence, according to the user/planner during crisis or at any time, it is possible for them to display maps in any levels from regional to local levels such as for e.g. ‘Entire India’ as a whole or ‘Village’, a smallest local unit for map display.

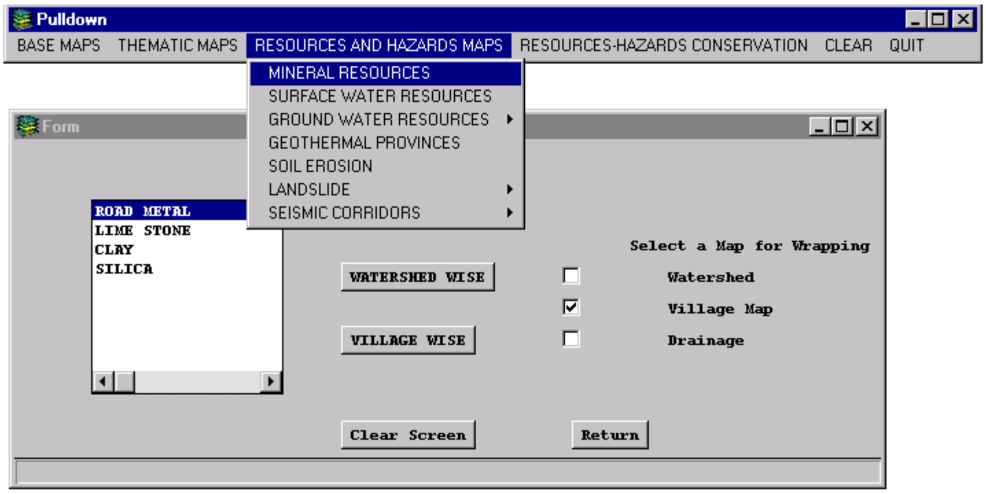
For example, WEGHARIS-QUBIS, a planner oriented information system developed during 2005 for the higher officials of Collectorate offices of Madurai, Theni and Dindigul districts of Tamil Nadu in order to retrieve all the basic thematic maps, resource maps and action plan maps showing management plans for resources and mitigation plans for natural disasters. Some of the screenshots of WEGHARIS-QUBIS are shown below.

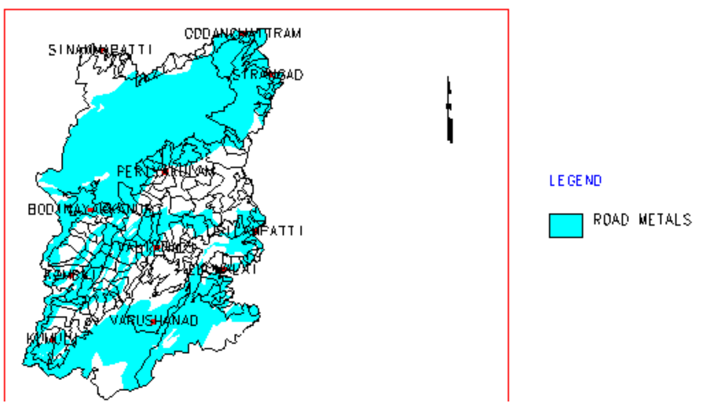
This WEGHARIS-QUBIS contains a simple Pull down menu with captions such as Base maps, Thematic maps, Maps on Resources and Hazards, Resources Conservation-Hazards mitigation, Clear and Quit.



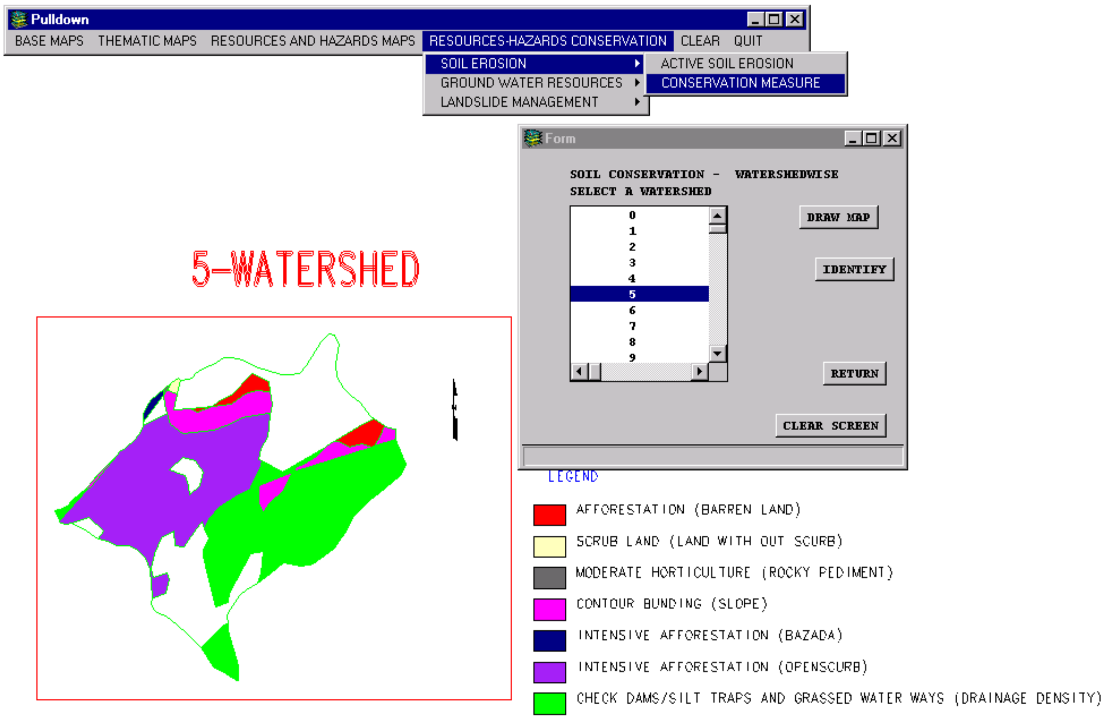
This QUBIS has been developed using Arc Macro Language, in short it is called as AML of NT ARCINFO GIS software. The user required layers on base maps, other thematic maps, maps on natural resources and natural hazards and action plan maps on natural resources conservation and natural disaster mitigation are prepared and relevant retrieval options are made ready using the command sequences linked with different menus predesigned in AML of NT ARCINFO GIS. Now, once the AML is played, then the Pulldown menu will get displayed as seen in the figure with the menu options. By moving the cursor over the menu options and clicking over it any of the buttons, a dropdown menu will appear listing the available display levels such as ‘Simple Display’ for the entire study area, ‘Village wise’ display for displaying map on village levels and ‘Watershed wise’ option to display the map based on the available sub-watersheds. Now, by clicking over any of these three options, the display will get set for that particular option clicked by the user followed by displaying a new window. Now, this new window named 'Form' is consisting of text boxes and check boxes and some buttons with captions such as 'Draw Map' ,'Clear Screen', 'Return', etc. The textbox is listing the layer names available for the study area, i.e. parts of Wester Ghats covering Theni, Dindigul and Madurai districts of Tamil Nadu. The user can select any of the listed layers for display by moving the cursor over the Textbox and by clicking over it. On clicking over the layer name, the particular layer will get displayed as a map by superposing the supporting base layers over the main theme. For example, while clicking the Geology Map in the Textbox, the main thematic layer Geology will get displayed first followed by the relevant other base layers such Major Settlements, Major Roads, etc., will get superposed over this Geology layer with the legend at the right hand as it was designed.

Similarly, the areas having Road Metal as a Natural Resource available in the study area, as displayed by the user is shown in the following screenshot.

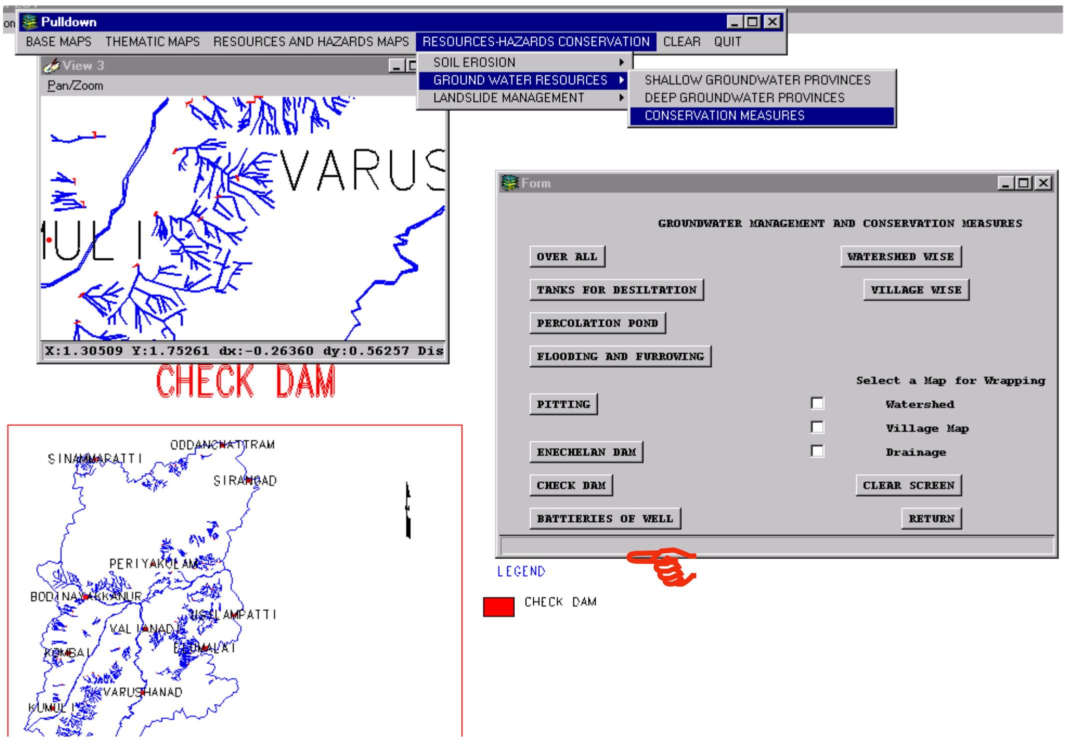




The following screenshot shows the display of watershed wise ‘Suitable areas for Soil conservation’ through suitable mechanisms in one of the sub-watersheds numbering 5.



In the similar way, the areas favourable for constructing Check dams in the study area so as to conserve groundwater resources of the study area is shown below.

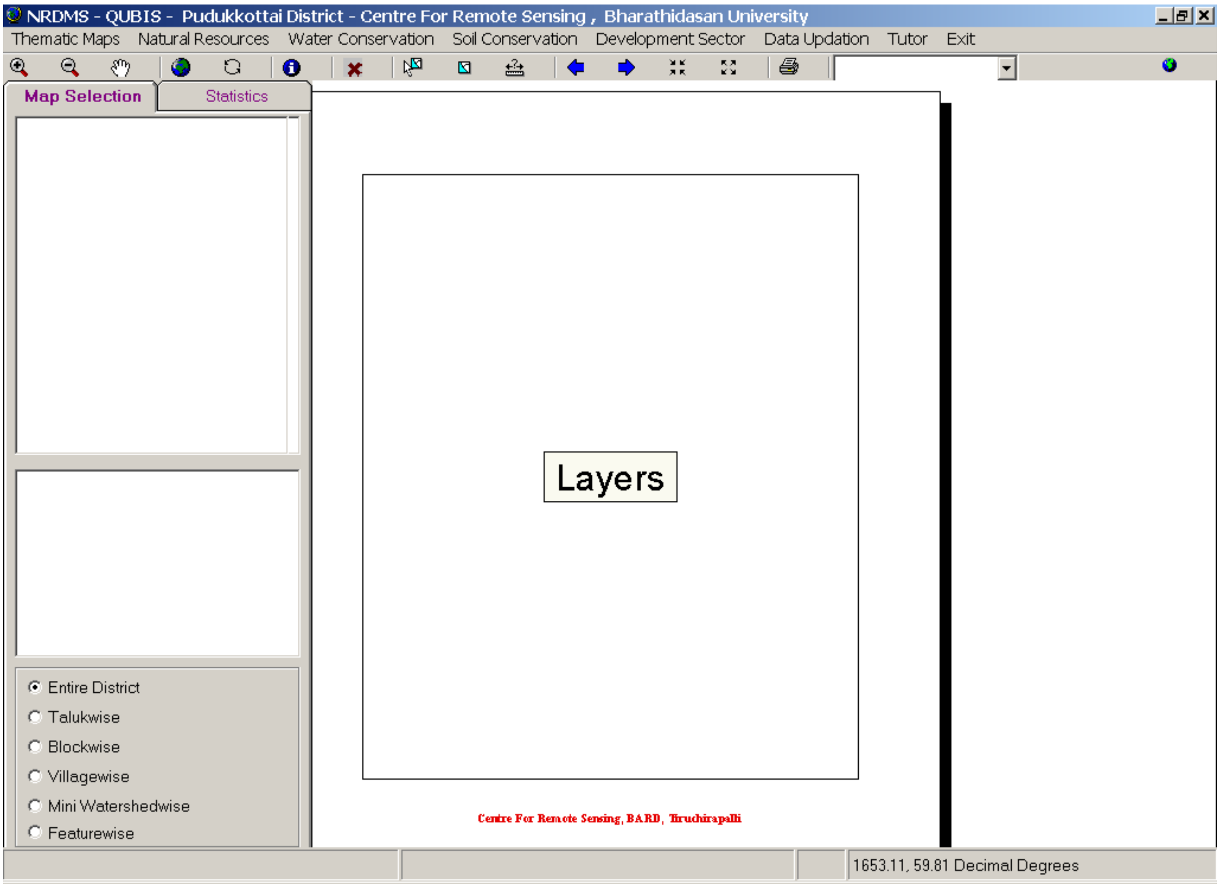


Similarly, some of the screenshots of NRDMS-QUBIS (Project 'Natural Resources Data Management Systems' sponsored by DST, i.e., Department of Science and Technology, Government of India, New Delhi) generated for Pudukkottai district of Tamil Nadu are shown in the following figures.

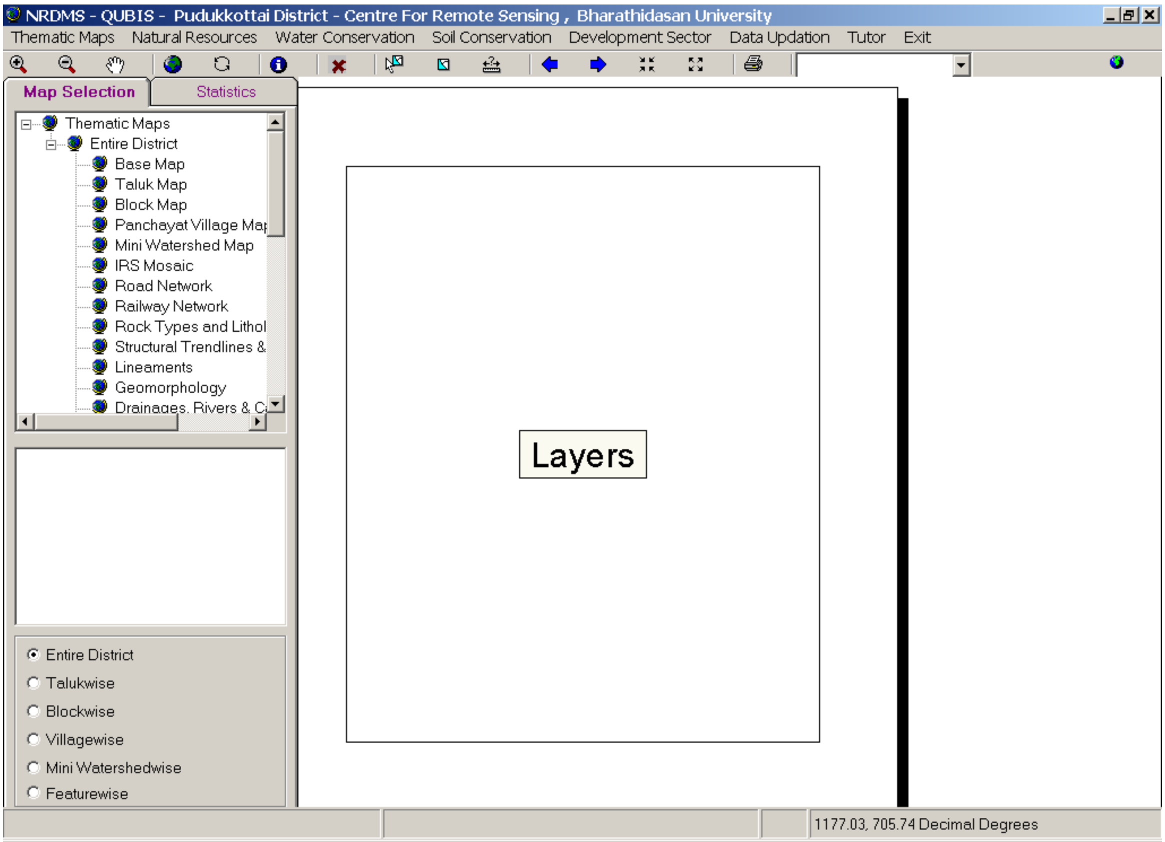
NRDMS QUBIS has been designed with a Main Menu consisting of following options:

- Thematic Maps
- Natural Resources
- Water Conservation
- Soil Conservation
- Development Sector
- Data Updation
- Tutor and
- Exit.

The basic display unit has been kept as Village polygon for displaying all the data in this SDSS. The different levels of displaying spatial data are on the basis of Taluks, Blocks, Panchayat Villages, Mini Watersheds and Features available in Pudukkottai district.



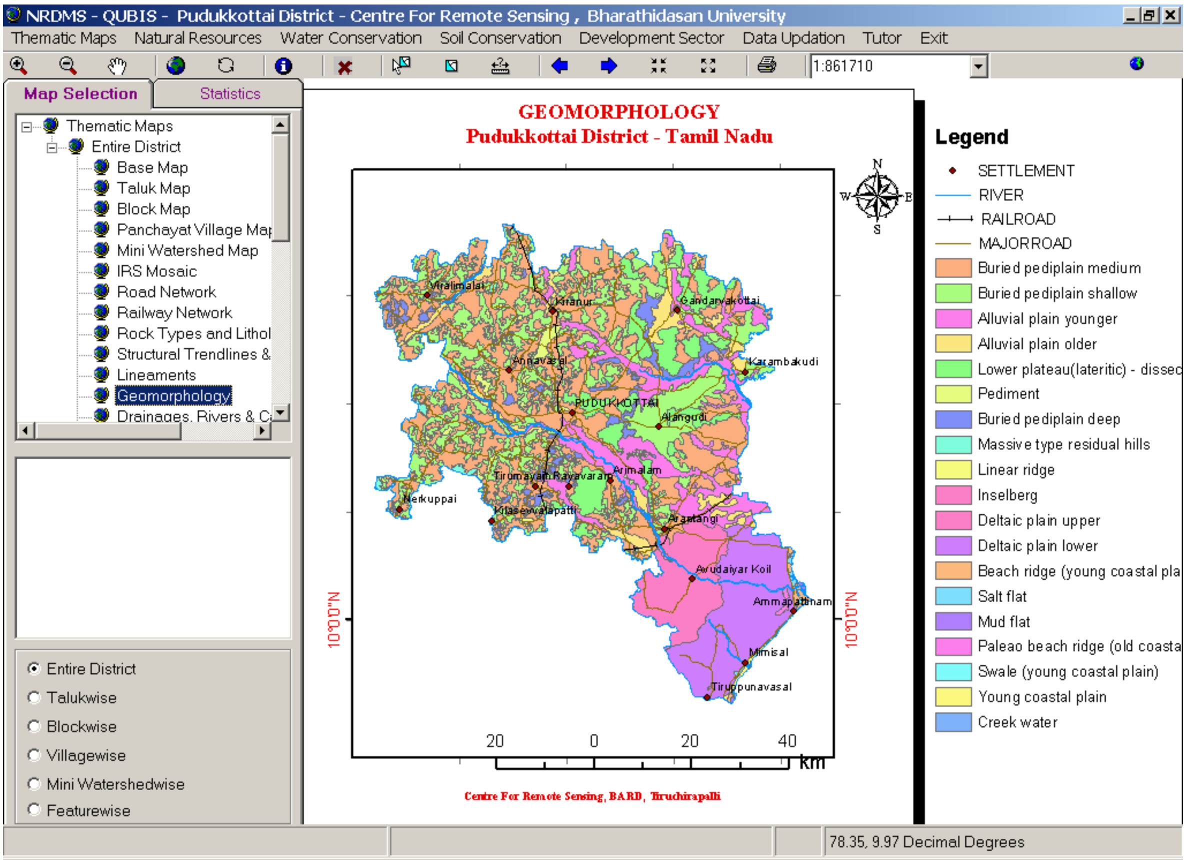
By clicking ‘Thematic Maps’ option in the Main menu bar, and the user can view a list of all available thematic maps in the TreeView control which is placed on the left margin of the window.



`Then, on clicking any desired theme, the same will get displayed immediately on the Map display area of the computer screen in a map layout format with the necessary map elements such as Title, Subtitle, Legend, Scale bar, North Arrow, etc., and also with the supporting details such as Major Roads, Major Settlements, Rivers, etc. For example, one of the Basic thematic maps (Geomorphology) for the ‘Entire District’ are shown in the following figure.

If the user is interested in displaying the same map with different colour ramp for the features, then he can click again one more time on the same Map listed in the TreeView Control, so that the particular map will be displayed with the next default color ramp available in the GIS software with new Legend.

The user can use the Map Handling Tools kept below the Main menu at the top of the screen so as to zoom the displayed map, pan, measure the distance between objects/features, etc., interactively as per the requirement. An option is also kept to save the displayed map as a digital map in different formats such as .bmp, .jpeg, .png, .img, etc., which can be used by the planners latter.

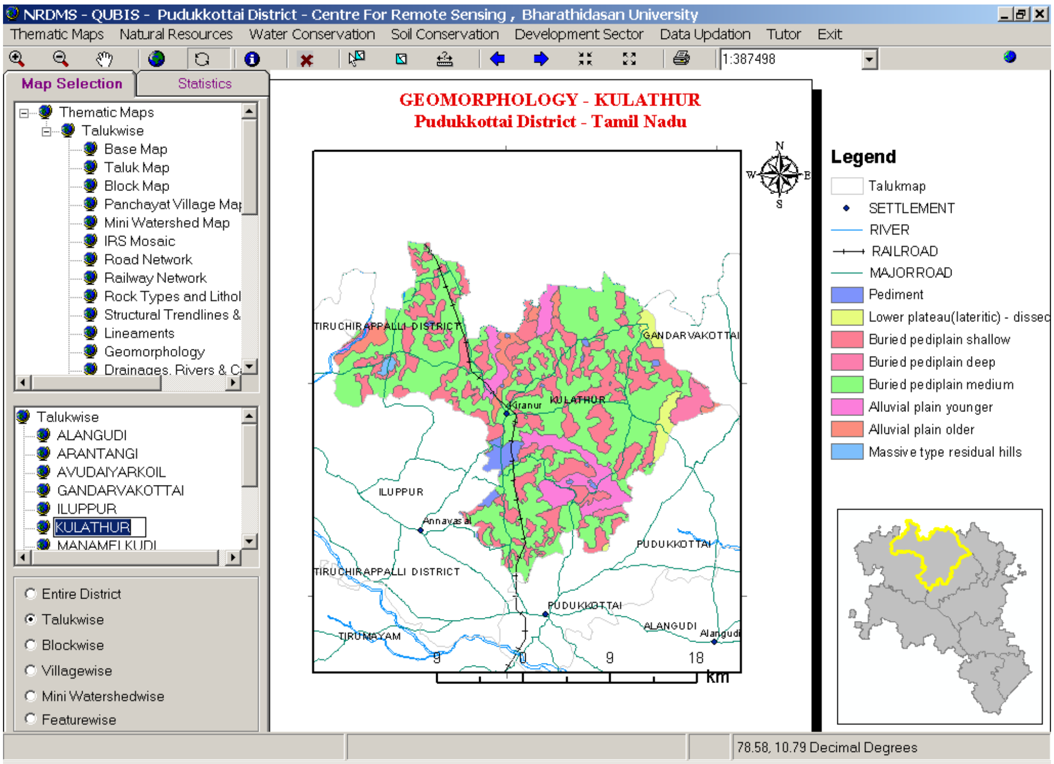


Advanced options available in NRDMS-QUBIS:

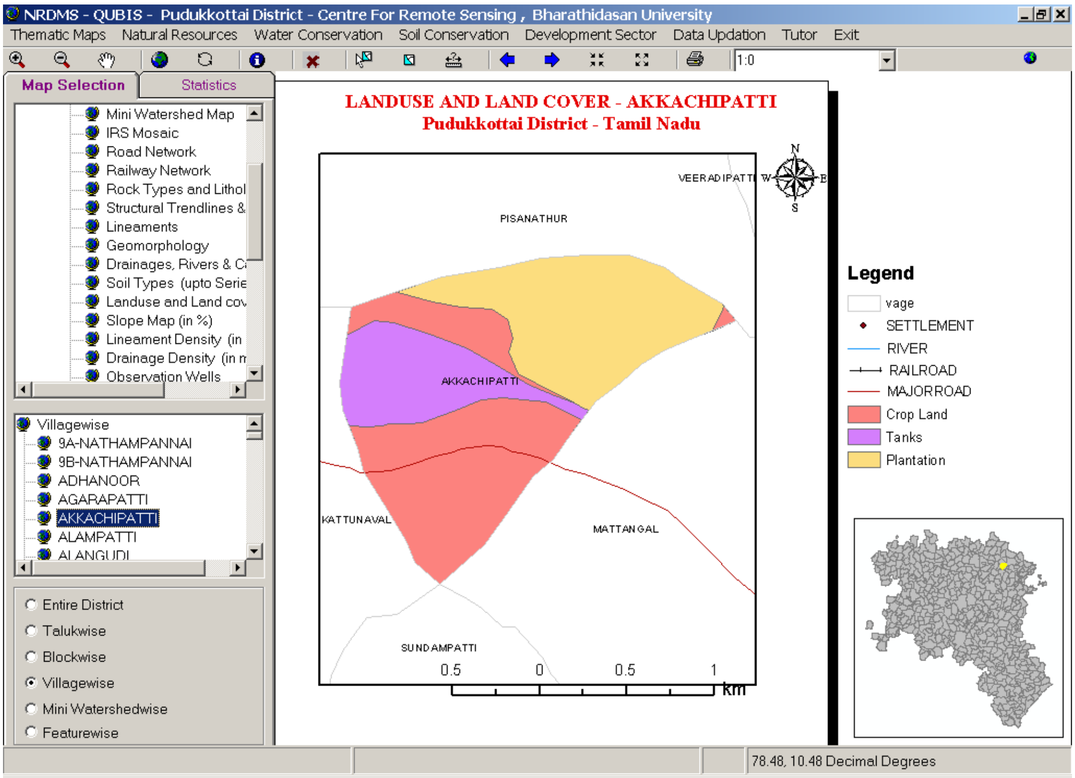
Many more user friendly options are kept available in the form of menus and tools. On the event of cursor movement in the screen in different areas, different types of activities are tagged and programmed in VBA using GIS. The feature labels are displayed on moving the cursor over the map that is displayed currently. The spatial location detail, i.e., longitude and latitude are displayed at the bottom right side area for the cursor location over the displayed map. Similarly while moving the cursor over the Tree View Controls, then the half-hidden names of the different maps will get highlighted with their full texts respectively and the action need to be done by the user will also get displayed at the text box located at bottom left side area of SDSS window. The full text form of the map handling tools and the menus will get displayed with the cursor on moving it over them respectively.

A processing bar at the bottom right side and a processing globe at the top right side of Map handling tools area are also kept in this design so as to indicate the user that the SDSS system is working and he/she needs to wait till the action is completed by the system, whenever the menus or tools are clicked.

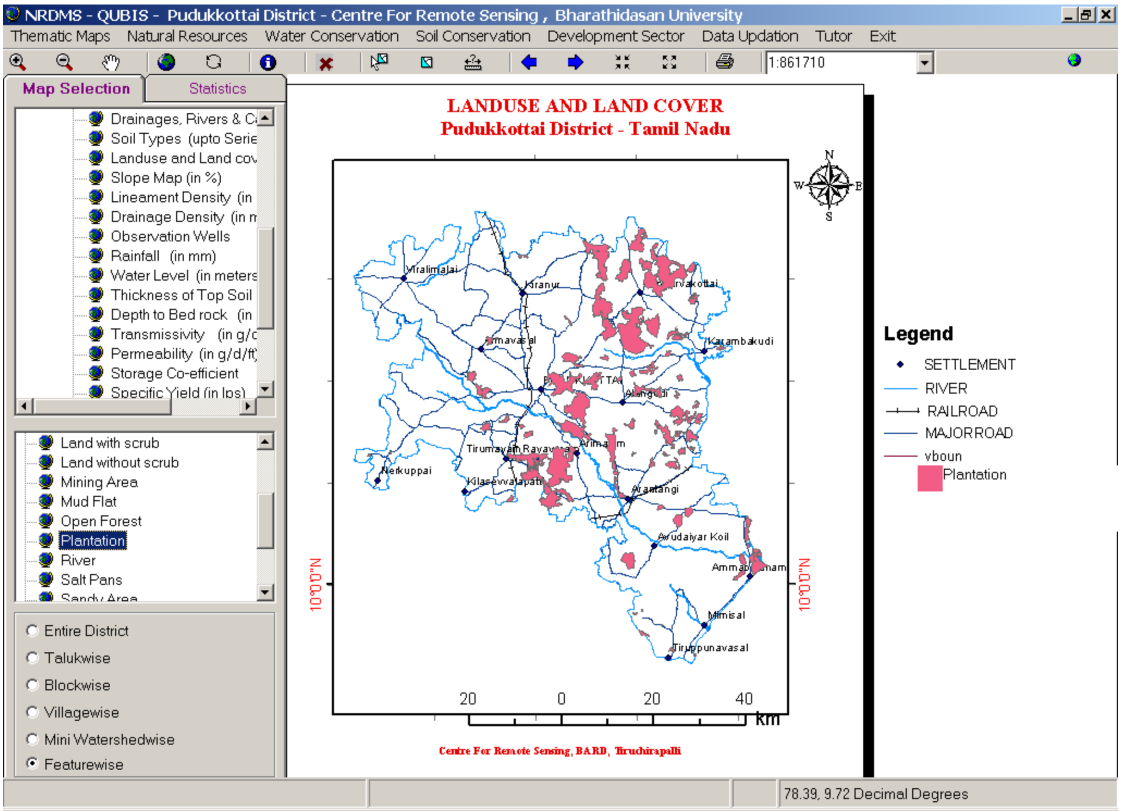
The other thematic /resource maps retrieved on different display levels are shown in the following figures.



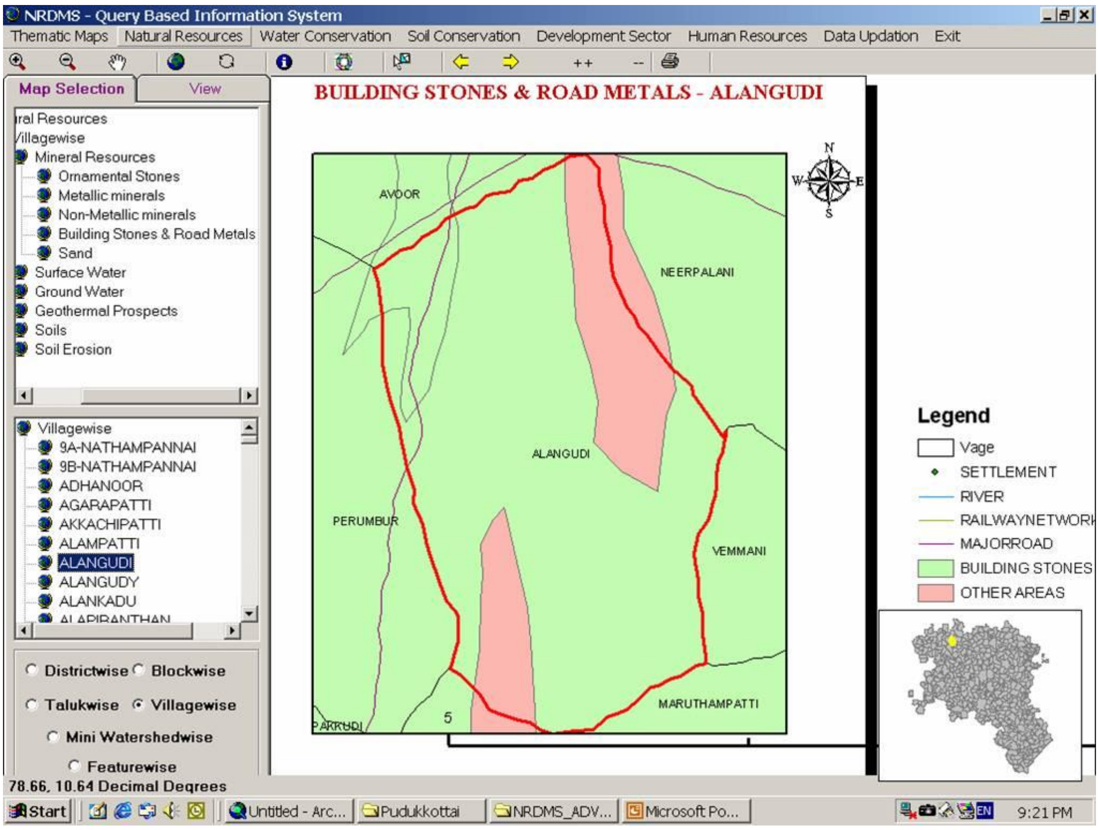
Geomorphology map displayed for Kulathur Taluk with Key map



The Landuse / land cover map displayed for Akkachipatti village with Key map showing the location of the village in Pudukkottai district.



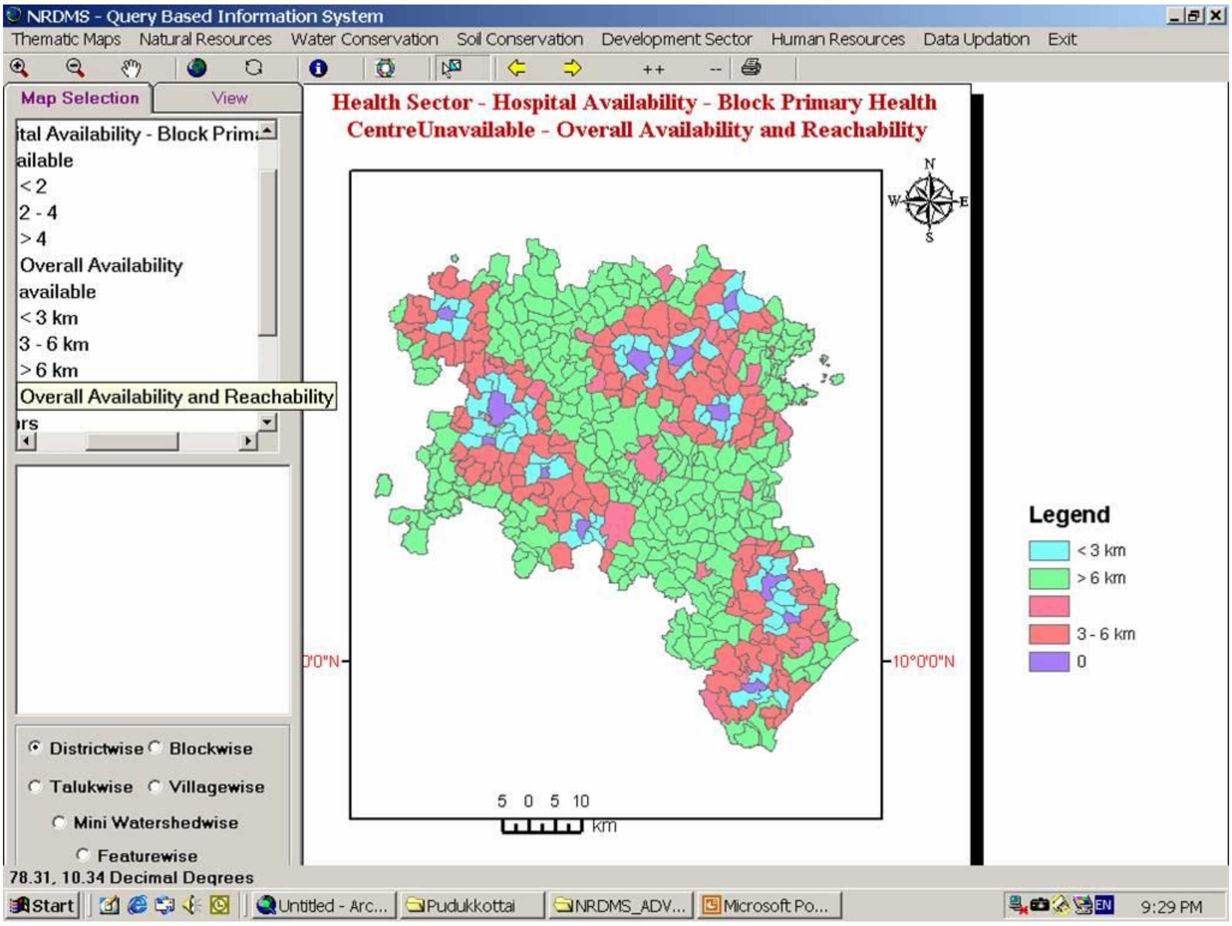
Feature wise display of Plantation retrieved from Landuse/land cover map



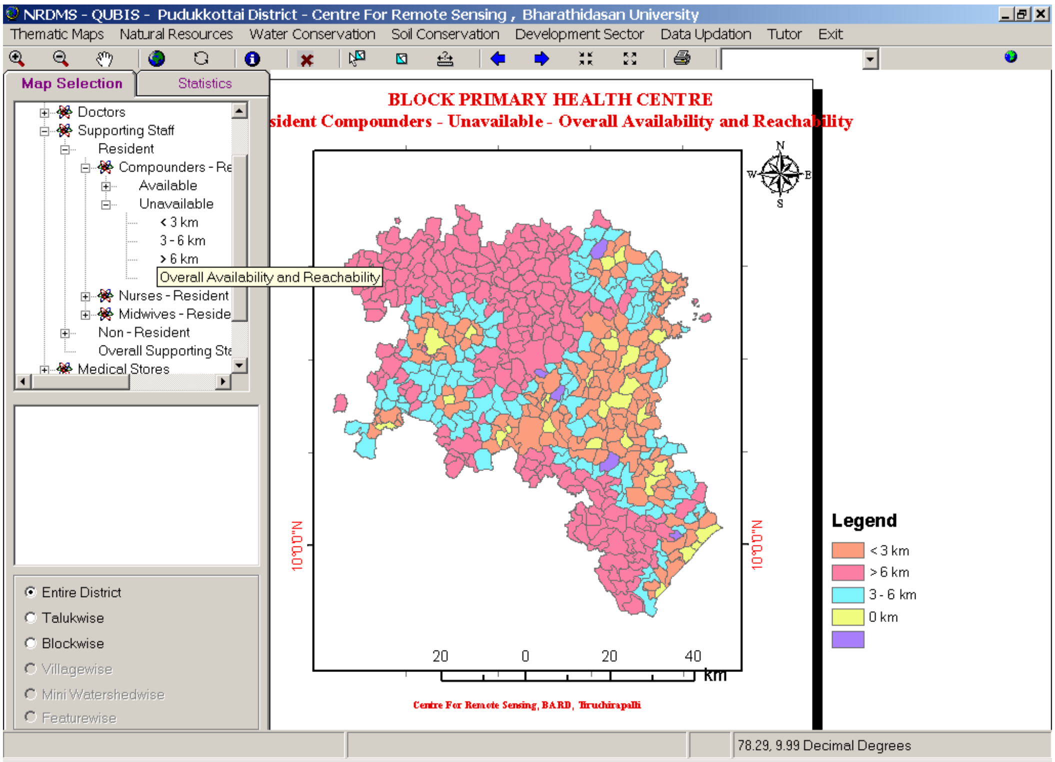
Building stone resource available in Alangudi village with its key map

Retrieval of attribute data and Spatial Display of Non-spatial data:

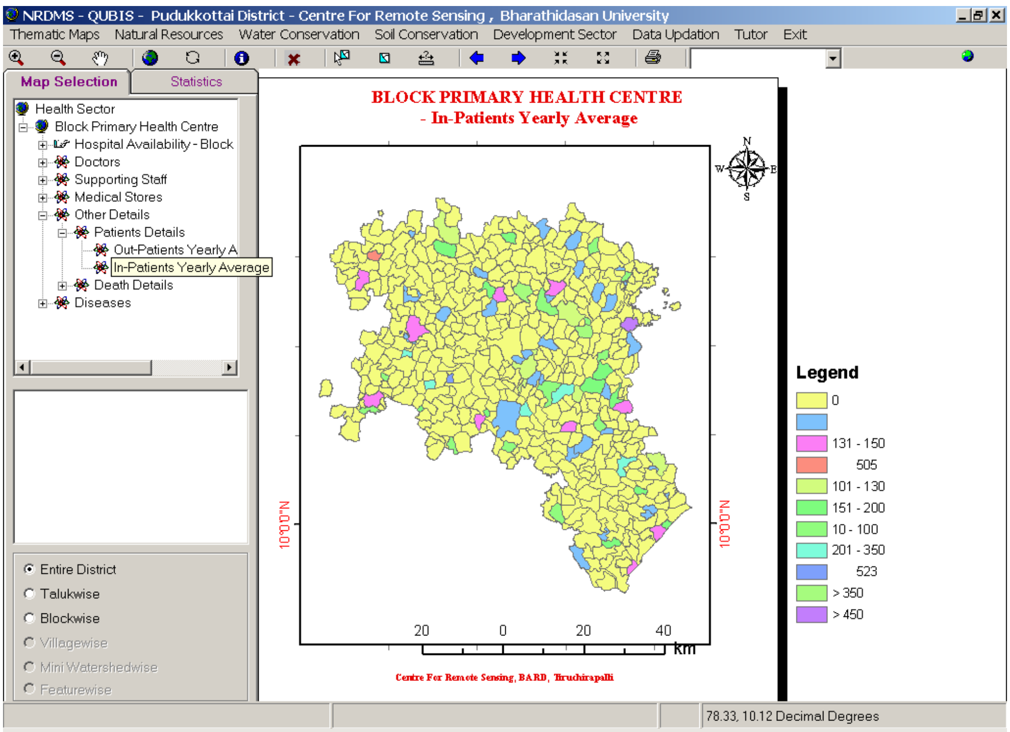
The cooked / reclassified non-spatial data are kept ready within the system in the specified folder location in the computer system. Whenever the user is interested to display them, then the particular option has to be clicked by him. In this case, it is Development Sector. Under this option kept in the main menu, there are several sectors will get listed in the drop down menu. As the user has clicked on the Health sector and the 'Overall availability and Reachability of Block Primary Health Centres' in Pudukkottai district are shown with their reachable distances in different colours as shown below.

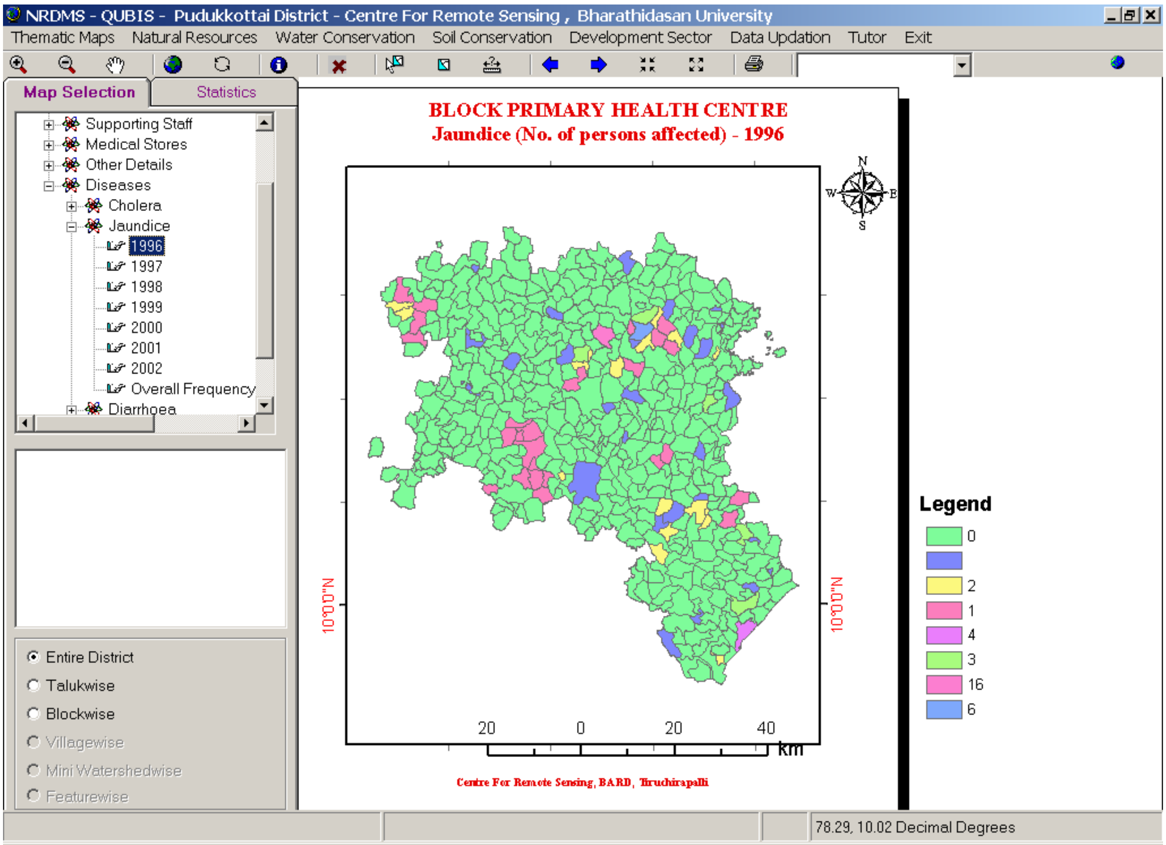


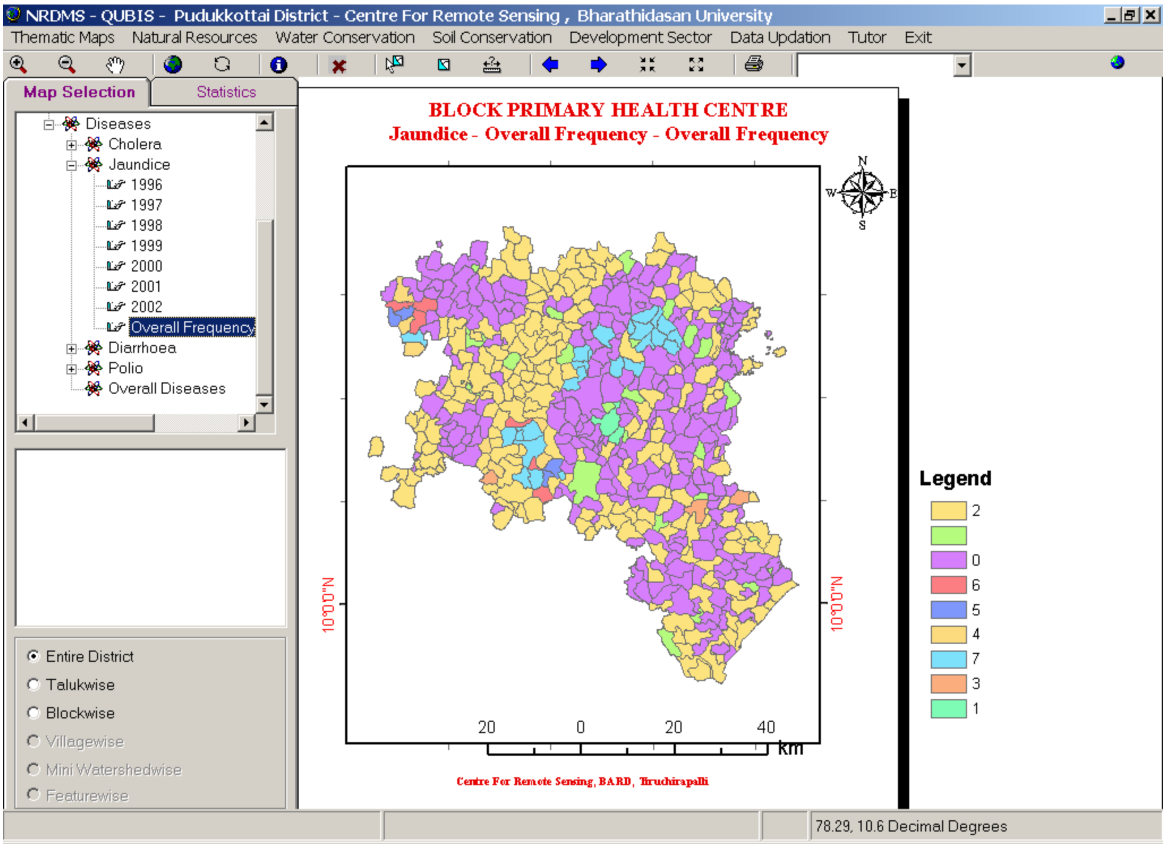
In the same way, the user can also click over the other facilities and staff availability and see the required details. One such display of 'Overall availability and reachability' detail of Resident Compounders in Block Primary Health Centres of Pudukkottai district. This will not only help the users to know about the staff availability in the district hospitals, but also to know the distances to reach those hospitals. Further, this can help the planners to identify 'Grey areas' where fund can be allotted and mobilized to create more such facilities.



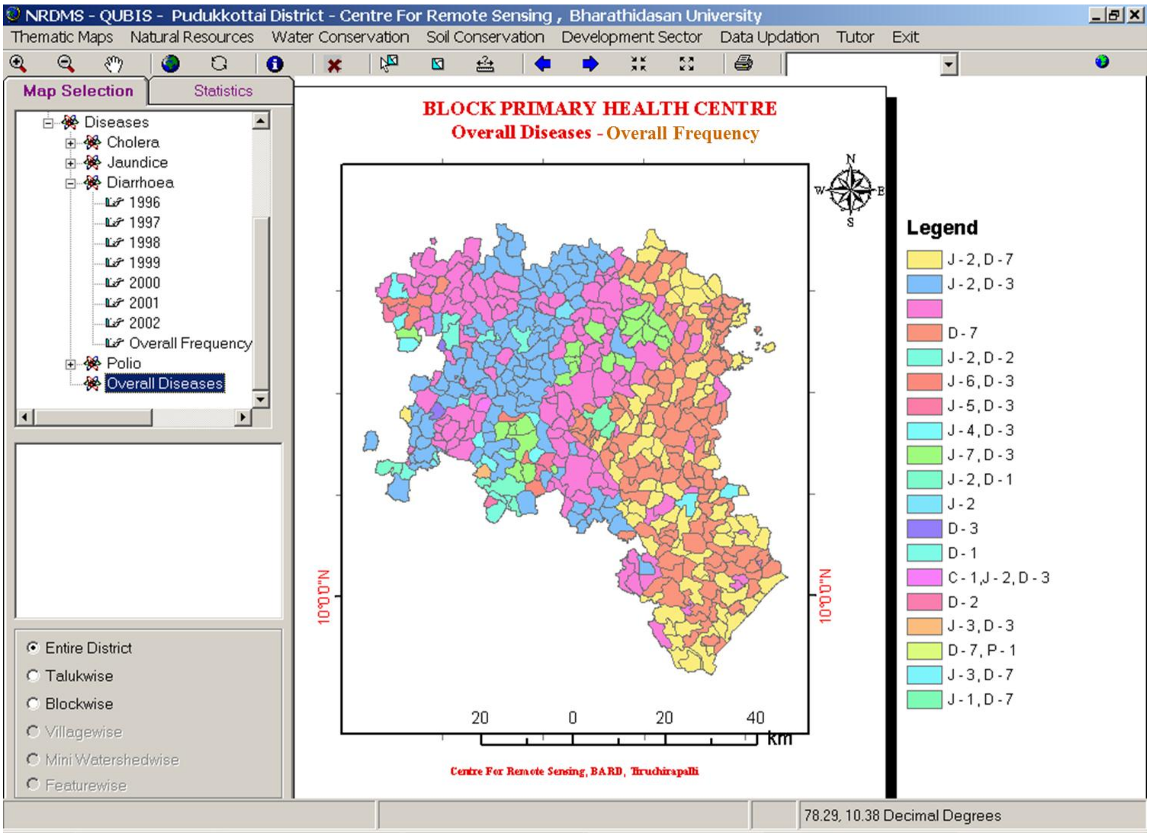
Subsequently, the other details such as patients admitted and patients affected and admitted due to different diseases such as Jaundice, Diarrhea, etc., and the frequency of diseased areas, etc., can also be get displayed as per the requirement of the planners / users for their further planning as shown in the following figures.

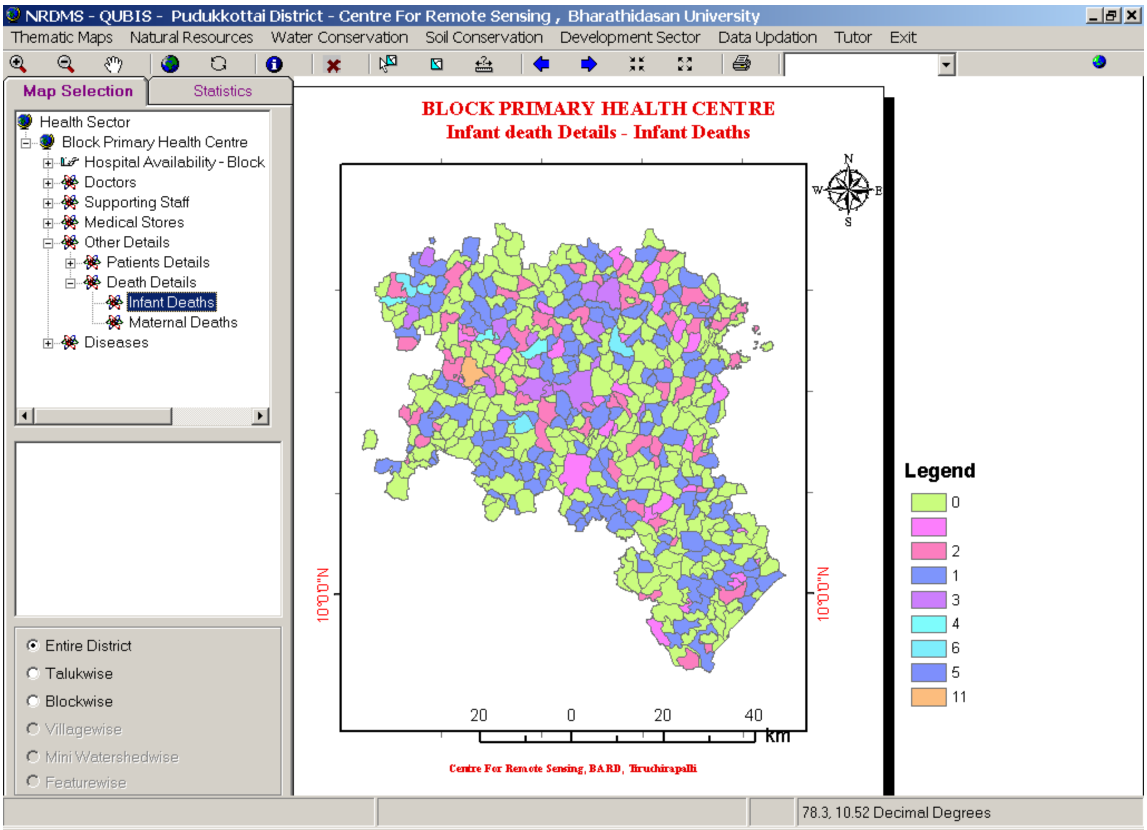






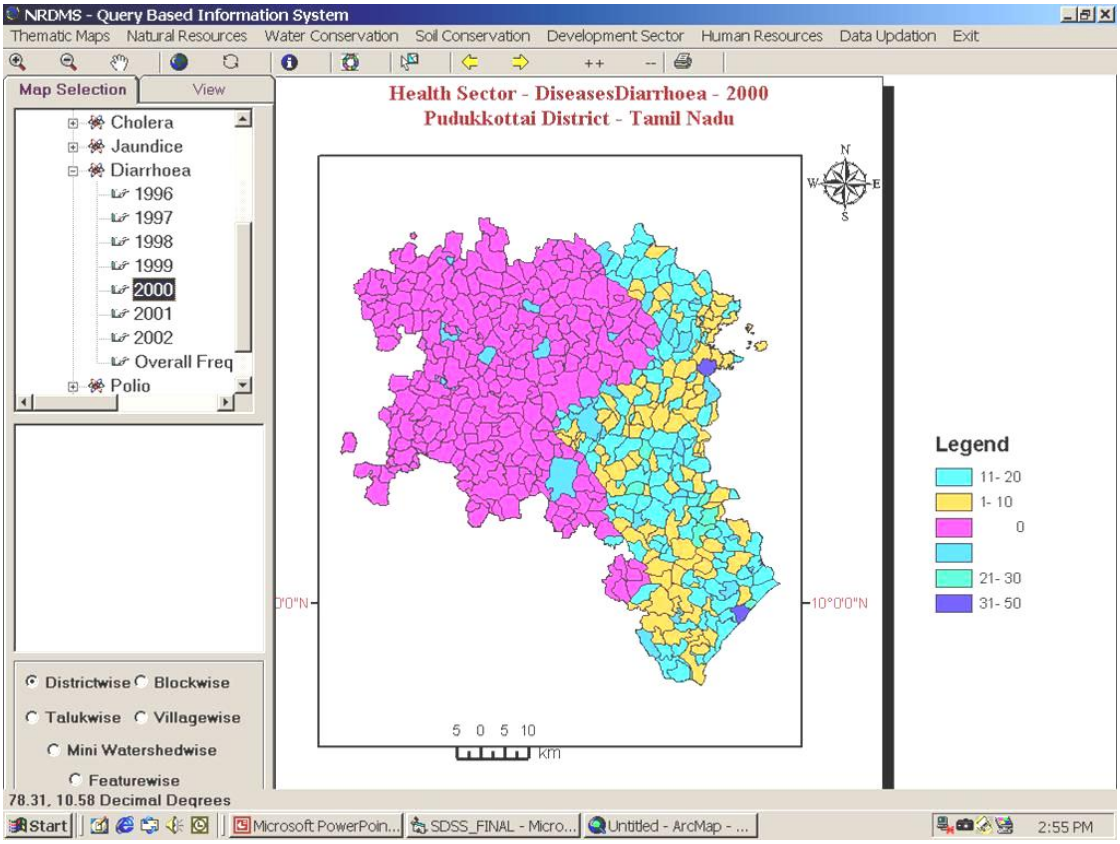
Overall frequency of Jaundice affected areas during 1996 to 2002





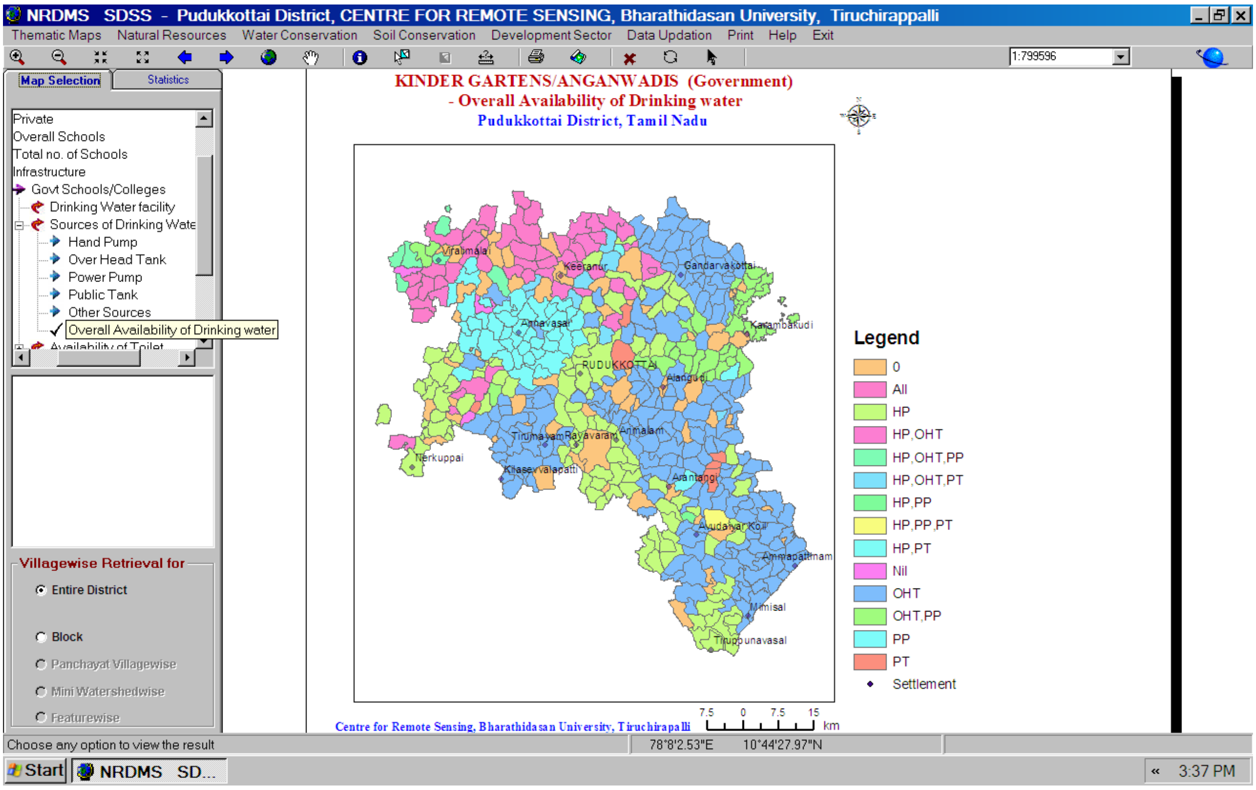
Spatial display of non-spatial data on infant death during a particular year in Panchayat villages of Pudukkottai district is shown in the figure above.

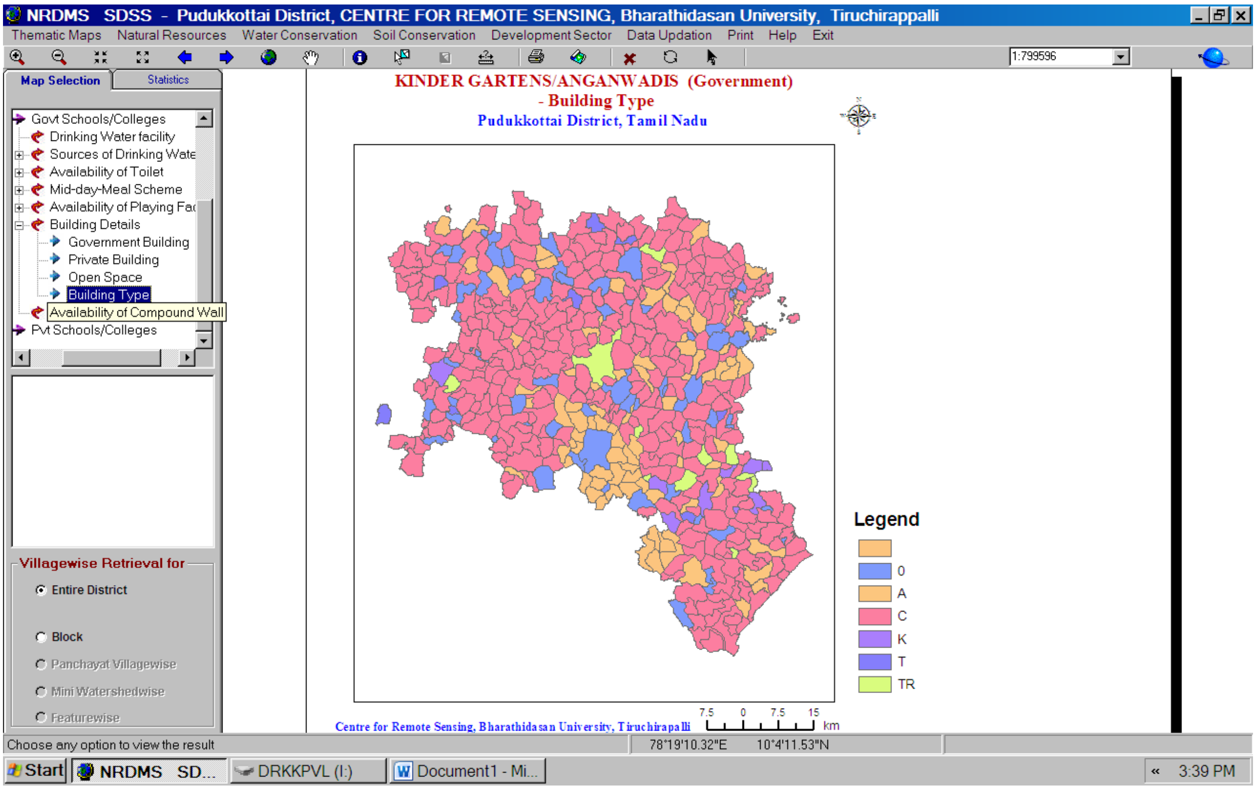
The pattern of diarrhea affected areas in Pudukkottai district shown below seems to have clear relations with the sedimentary terrains having more number of fluvial systems and water bodies. Hence, the water borne diseases can be easily diagnosed for their sources once their areal locations are plotted and analyzed using some specialized QUBIS systems like this kind.



In this way, the other attribute data collected from several other sectors have been entered, cooked, classified and kept ready for linking and display using GIS. The spatial display of Education sector details on the types of buildings available for different types of schools, drinking water sources, infrastructure, lab facility availabilities, compound walls availability, toilet facility for students and teachers separately, and several other details can be done easily in QUBIS.

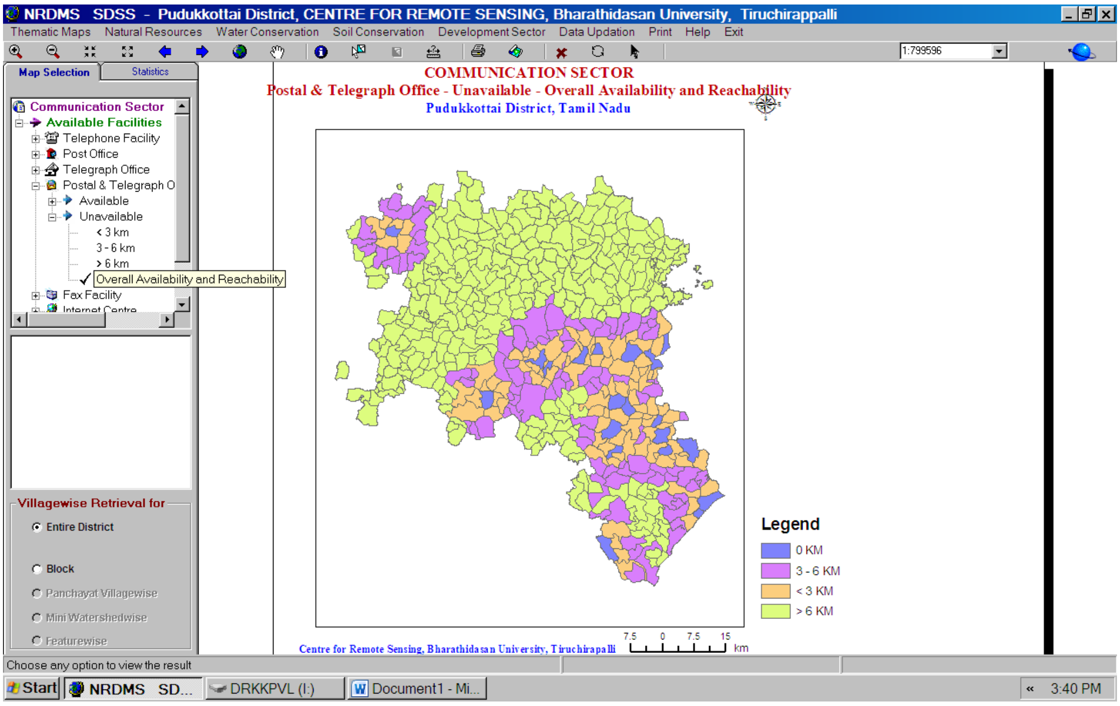
The following two figures showing such spatial display of drinking water sources and type of buildings available for Kinder Gartens of Pudukkottai district.



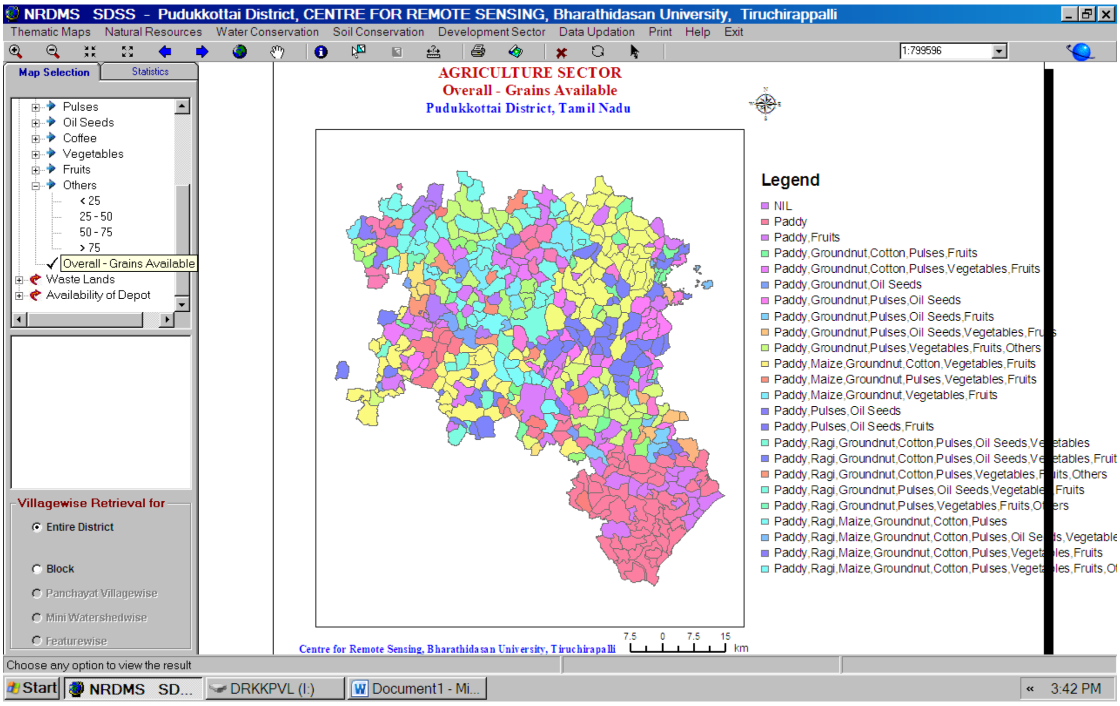


NRDMS – QUBIS screenshot image showing the Building details of Kinder Gartens located in Pudukkottai district, Tamil Nadu.

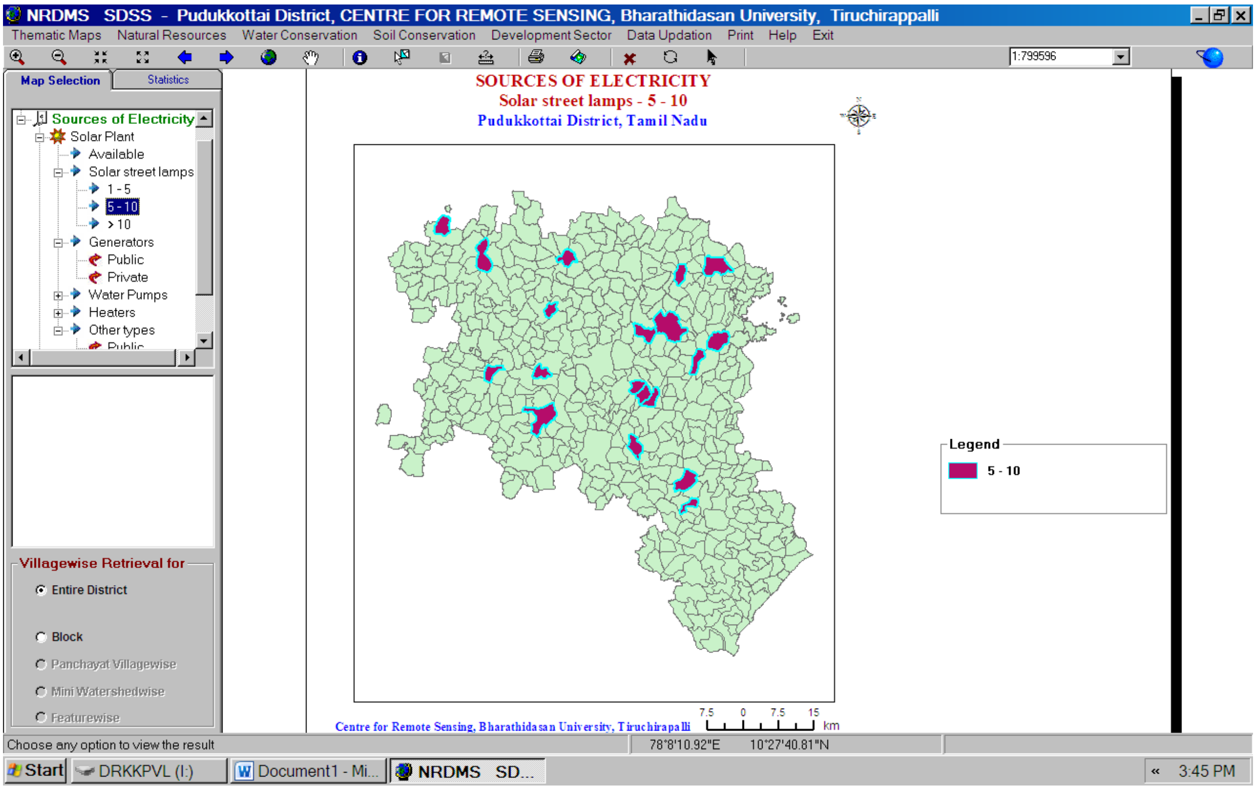
Spatial Display of one of the communication sector details that is, overall avaiability and reachability of Postal and Telegraph facility in Pudukkottai district is shown below.



In agriculture sector, the overall food grain producing lands available with reference to the aerial coverage of each Panchayat village is shown in the following figure.



As far as the electricity sector is concerned, the data such as types of usage and amount of electricity usage, electrified villages, sources of electricity, etc., are collected from TNEB and entered in DBMS softwares and kept readily with cooked and derived information in tabular form as attribute table with unique identifier so as to provide linkage between this attribute table and the spatial feature database whenever it is required by the user through QUBIS. The NRDMS-QUBIS developed with Electricity sector data has been used to display the villages having 5 to 10 numbers of solar street lamps in Pudukkottai district and the same is shown below.



Data Updation in QUBIS:

A simplified menu is formulated and incorporated with NRDMS-QUBIS design so as to update the existing both kind of databases such as spatial and aspatial data by the authorized users. A separate option called 'Data Updation' is kept in the main menu. On clicking this option, a dropdown menu showing the two types of databases available in this QUBIS, i.e.,Spatial Data and Non-spatial Data will get displayed. On selecting any of these two, a new login window will lead the users to enter 'User Name' and 'Password'. On checking the authenticity of these user inputs, the users will be permitted to access the database so as to add a new, delete, change or update the existing databases in QUBIS.

User Defined Query Builder in QUBIS:

A query builder is simply designed and kept available in QUBIS for the users to make their own simple and complex queries easily according to their requirement. The answerable features satisfying the query raised by the users will get displayed with specific legend and title with map layout.

Multiple Layer Wrapping options for Comparison of Several Themes :

According to the situation raised, the user may be interested to see the displayed features with some other information such as road / rail accessibility, nearby settlement locations, drainages that are flowing along those already retrieved/displayed features, etc. In that case, it can be made available in QUBIS by writing program in VBA / .net using GIS software, for the process called 'Simple superposition / overlay of layers in GIS'.

Programming in VBA / .net :

It is necessary for the QUBIS/SDSS developers and GIS programmers to know the available controls in GIS and their applications to execute different actions on the click event of mouse. SML, i.e., Simple Macro Language in Arc/Info DOS versions, AML in NT Arc/Info versions, AVENUE in ArcView, ArcObjects in ArcGIS, ArcSDE in the latest versions of ArcGIS are useful to generate such QUBIS and SDSS platforms for the benefit of the users community.

The GUI based options available for designing the main display page which will provide simplified user interaction by simple click events through mouse. Hence, on compilation of such VBA or .net programs under the backdrop of GIS software, it is possible for the non-GIS users to utilize the fullest capabilities of GIS with simplified menu options for interactions through QUBIS / SDSS.