BUILDING A SOIL INFORMATION SYSTEM FOR MULTI-SOURCE DATA INTEGRATION

(A case study in Lake Naivasha area, Kenya)

BY

TILAYE BITEW BEZU

FEBRUARY, 2001

ITC, ENSCHEDE, THE NETHERLANDS

BUILDING A SOIL INFORMATION SYSTEM FOR MULTI-SOURCE DATA INTEGRATION

(A case study in Lake Naivasha area, Kenya)

BY

TILAYE BITEW BEZU

February, 2001

SUPERVISOR
Ir. BART KROL

CO-SUPERVISOR
Dr. DAVID ROSSITER

Thesis submitted to the International Institute for Aerospace Survey and Earth Sciences (ITC) in partial fulfilment of the requirements for the degree of Master of Science in Geoinformation Science and Earth Observation, with specialisation in Soil Information Systems for Natural Resources Management

Degree Assessment Board

Assoc. Prof. Dr D.G. Rossiter	(Head of Division)
Prof. Dr A.K. Bregt (WAU)	(External Examiner)
Mr C.M.J. Paresi (ITC)	(Internal Examiner)
Drs T. Loran (ITC)	(NRM Prog. Director)
Ir B.G.C.M. Krol (ITC)	(Main Supervisor)
Dr D.G. Rossiter (ITC)	(2nd Supervisor)
Dr A. Farshad (ITC)	(Student Advisor)

ITC

INTERNATIONAL INSTITUTE FOR AEROSPACE SURVEY AND EARTH SCIENCES
ENSCHEDE, THE NETHERLANDS

DISCLAIMER

This document describes the work undertaken as part of a programme of study at the International Institute for Aerospace Survey and Earth Sciences. All the views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the institute.

ABSTRACT

This study was aimed at the design of a Soil Geographic Database (SGDB), as part of a Soil Information System (SIS), for the improved management of soil and land resource data relevant to an area around Lake Naivasha in Kenya. Emphasis is given both to the rescue of existing data, available in different formats from a variety of sources, and to the improvement of user access to adequate soil information for multiple purposes. The methods adopted for SGDB data modelling and for SIS development are the relational data model and the iterative structured information system development method, respectively. Two components of the SIS were designed: a relational database structure having twelve relations, and a prototype information system architecture consisting of four subsystems.

The analysis of users' information requirements provides the input to the actual SGDB design process. The user need assessment showed that 94% of the respondents have indicated a clear need for soil information for their current and future resource management activities. A large majority (73%) of the users obtain their soil information by sending soil samples abroad for further analysis. Users have a clear preference for digital rather than analogue data format primarily because of the ease of handling and further analysis.

This study proposes a working approach for multi-source data integration and standardisation in a common database structure. Correlation tables have been developed to handle the multi-category issues when dealing with different soil survey approaches and/or soil classification systems. A nested database design approach is applied to integrate data resulting from surveys at different levels of detail and map scale.

The conversion of existing soil data from analogue to digital format involved a geometric correction process supported by additional ground control. For spatial data capture in the field two approaches have been followed. In those cases where not enough identifiable ground control points (GCP) could be selected on a geometrically-accurate base map, a directional compass traversing method was applied with distance measurement using a measuring tape and GPS readings. Whenever enough GCP could be identified, only GPS readings have been used. In both cases the spatial data have been successfully converted from analogue to digital format, using the same geometric definition. However, when the locational accuracy of the geometrically corrected soil maps is compared, the compass traversing approach provides better quality.

Both metadata documentation and user interface design has been applied to improve user access to the organised soil data. The Content Standard for Digital Geospatial Metadata (CSDGM) as developed by the Federal Geographic Data Committee (FGDC) of the USA has been applied. The user interface design has been done using Ms Access navigation screens.

Key words: Soil Information System Development, Soil Geographic Database Design, Multi-source Data Integration, Metadata, User Interface, Lake Naivasha, Kenya

ACKNOWLEDGEMENT

I have no words to express my feelings to the Almighty God who has the full power to create and pass the whole world only in his holly words. He gave me every thing what I was wishing but I have nothing to give him except thank you. Also I should say thank you for my holly mother Merry and Sent Michel whose prayer and providence brought me to this great event in my life. I hope and believe that they will continue to protect me against sin and devil's spirit and pave the way to heaven until my God, Jesus Christ, judges me. Amen.

I am highly indebted to my mamma W/ro Abunie Woldu Deres and my dad Ato Bitew Bezu Habitu for their love, care and support that they gave me since my childhood till now. It would have been impossible to be at this adventure without their prayer, care and support. My heart felt gratitude also goes to my sisters and brothers and Syrian Orthodox Church who have given me moral and spiritual strength during my study.

I am also highly indebted to Ato Beryihun Degu Head of my organisation, Planning & Economic Development Department, my organisation work mates, members of the Council of Awi Administrative Zone, Dr D.G. Rossiter Head of Soil Science Division, Drs Tom Loran Natural Resources Management Program Director, the Ethiopian Government, the Netherlands Fellowship Program, for your joint co-operation and support that made it possible to stay and study for my MSc degree at ITC. Thank you all for your invaluable help.

I would like to forward my special thanks to Ir B.G.C.M. Krol, my supervisor whose great effort and guidance turned my work from nothing to something visible. He did a lot for me. In addition to his close supervision he had arranged a new elective module together with Dr D.G. Rossiter and taught us only in favour of my research topic. So I say thank you very much.

My heart felt gratitude goes to my student advisor Dr Abbas Farshad for his kind consultation and advice, to the staff of soil science division Dr W. Siderius for his supply of working data for lake Naivasha area and valuable comments, Ir G.R. Hennemann for his support during the field work and presentation of different literatures concerning our study, Dr D.P. Shrestha and Prof. Dr J.A. Zinck whose consultation & support was highly mentionable. To all I say thank you.

Finally my thanks goes to my computer teachers Viet Anh Hoang (Vietnam) & Atiklt Girma (Ethiopia), who have helped me a lot in improving my computer skill. My class mates D. B. Ranatunga (Sri Lanka), Paul Simfukwe (Zambia), Fabiola Padilla (Bolivia) and Andre Nagelhout (Dutch), my country men Belayneh Demilew, Samson G/mariam, Getahun W/meskel, ITC community, people of Enschede and people of Naivasha thank you all for the knowledge we shared, for the moral & friendship we built, for the culture we exchanged, for the social life we enjoyed and many others which are all in my memory. I hope we will meet again at least once in a while.

Tilaye Bitew Bezu
February 01, 2001

DEDICATION

THIS PIECE OF WORK IS DEDICATED TO MY GRAND PARENTS
FOR WHOM I AM THE FIRST BORN GRAND SON

ATO WOLDU DERES BORN IN MAY, 1920 G.C/1912 E.C
&
W/RO ABEBA WORKIE BORN IN JULY, 1925 G.C/1917 E.C

WHERE THEIR SPIRITUAL PRAYER AND BLESSING IS SAVING ME
FROM DANGER AND UNFORESEEN CIRCUMSTANCES

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. Introduction

Soil is the largest and primary natural resource in this world in which no life can survive with out the direct and/or indirect use of it. The soil, one of the most precious natural resources is being exploited by human-beings at large extent without any logical planning. To ensure the soil resource to maintain its sustainability for the coming generations, man must seriously consider how to use it in the right way. For such reasons an organised soil information is highly necessary.

A Soil Information System (SIS) is a collection of people, activities and procedures involving systematic, geographically referenced soil data collection, storage, manipulation and retrieval using procedures of data processing. Information for specific purposes is derived from the data stored, using standard soil science conventions (Sadoviski and Bie 1978).

Soil Information Systems belong to the family of Geographical Information Systems (GIS). These types of systems differ from other information systems in the requirement that the data should be referenced geographically and temporally, thereby allowing for analysis, retrieval and display of the data on the basis of these criteria. "Information technology is the servant of information requirements, and information systems are the servants of organisations. Not the other way round"(Reeve and Petch, 1999).

Soil Information Systems are often called soil data banks or soil data processing units. "The term data bank reflects primarily the concept of a depository of data for future use, while soil data processing is the manipulation of small, special purpose data sets by individual researchers. Neither of these activities can be seen independent of the other. An information system involves both of these functions as well as that of systematic data collection" (Sadoviski and Bie 1978). This thesis puts emphasis on the development of a prototype soil information system for soil resource management in Lake Naivasha area.

1.1. Role of Soil Information Systems

Traditional, (manual) data handling methods has proven slow and ineffective in managing soil data. In most cases, the volume of data collected in a project was so immense that it required pre classification and grouping at very early stages of data collection. Invariably, this resulted in the loss of much primary data (Sandoviski and Bie, 1978).

Computer oriented data management overcomes certain of this deficiency where it provides a storehouse of basic data. Further, it permits the scientist to maintain contact with his basic data right up to the point of publication, there by precluding the necessity of early data classification. As new data is generated, the original store is expanded. This capacity facilitates the resorting and resummarisation of data in a variety of ways at different times, as maps, summaries or tables, in response to the multifarious (diverse) requests generated by environmental and land management agencies.

By developing links of communication and interfaces to other resource oriented data systems, Soil Information Systems contribute to the development of the comprehensive geographical databases necessary for information production for land evaluation, planning and management (Sandoviski and Bie, 1978).

A soil geographic database as the main component of SIS is a tool for the capture, storage, management, processing and display of soil data generated by soil surveys from original sources. Displayed information carries an aggregated value in relation to primary data because of the capabilities offered by the system for analysis, modelling and graphic combinations, but the output information is intrinsically/inherently only as good as the input data. Quality control of primary data is therefore a major concern to guarantee a reliable implementation of interpreted soil information for land use planning.

The use of soil database in a GIS, however, considerably enhances the possibilities of satisfying users' needs through the capabilities of data processing, automated data interpretation, map overlaying and building decision alternatives (scenarios) (Zinck and Valenzuela, 1990).

1.2. Basic Considerations in Designing Soil Information Systems.

The requirements of a SIS have already been reviewed by Dumanski et al (1975). These include orderliness in data collection and storage, accessibility, editing and updating, simplicity, flexibility and feedback. Of equal importance is the need to define the users of the system and their needs and expectations, as well as to develop a permanent relationship between users and specialists including those involved in data gathering and data processing.

A soil information system requires a high degree of standardisation on the part of those contributing data to the system. Definitions must be agreed upon within the geographic area within which the system operates, but such definitions must be in the context of international terminology. Failing this, data flow will be restricted to boundaries of immediate influence (Sadoviski & Bie, 1978).

There are two types of soil data from the standpoint of information systems. These are commonly called 'hard' or point data and map data. Each type requires system components that reflect the nature of the data (Sadoviski and Bie, 1978).

Data input for 'hard' includes the procedures for editing, conversion to a computer compatible form and submission to the system. The data collection document may be computer punch cards, notebooks, and specially prepared forms for keypunching or forms for optical character reading. In all systems concern for data quality, ease of use, and ease of access to the data are important considerations (Kloosterman, 1975).

From data storing point of view, in an information system, it is useful to divide the collected soil data in to point data and area data. Point/hard data are detailed soil profile descriptions in a pit, including chemical, physical and mechanical analysis and auger hole descriptions. In soil survey, what we consider to be point data is quite often a line description in the z dimension (profile). Soil maps and mapping unit descriptions (x, y, dimension) are regarded as area data (Bregt, 1992).

The function of data retrieval is to provide the capability of retrieving and summarising data in table, paragraph or map form. The ability to sort, select, summarise, combine, manipulate and compare data sets are important features of a retrieval system. These apply equally to 'hard' data and map data.

Some systems provide this capability interactively but some provide these features only in batch mode (Sadoviski and Bie, 1978). In this study too the map data and the hard data were treated separately in two different softwares, which is mostly called hybrid GIS approach.

Concerning analogue-digital data conversion costs, although, attention tends to focus up on initial hardware and software costs, experience of large GIS projects so far suggested that such costs will actually represent less than 20% of total cost over a project's life-cycle. Indeed, for many large GIS projects, the costs associated with data conversion alone are likely to dwarf/reduce hardware and software costs. Creating a worthwhile and accurate GIS database for an operational system, for example converting a local authority's manual terrier system to GIS, or establishing GIS database for a utility/public service, is a mammoth (enormous) task, which will soak up many months of labour costs. Conventionally it is estimated that database creation will account for around 80% of the total costs of a GIS project. The size and cost of the analogue-digital data conversion task appears often to be underestimated in GIS projects (Reeve and Petch, 1999).

1.3. Problem Statement

Traditionally, soil survey data collected for any purpose were stored in archives in paper (analogue) format where the accessibility of these archives is often quite limited, even by the survey organisation themselves. For users outside the survey organisation, the published profile descriptions and soil maps were often the only available soil data sources. Detailed descriptions of auger holes are not available or are very difficult to get hold of them. At this moment, a lot of collected soil data in the world are still stored in paper archives. Obviously this method of storage does not allow rapid analysis of the data and, furthermore, even hampers its use (Bregt, 1992).

This is very true mostly for developing countries especially that of Africa. Kenya is one of the developing countries in east Africa where the traditional (paper archive) system of data storage is prevailing at this moment, where data analysis and updating within a required time is inconceivable. On the other hand, in Kenya the request for soils information is unpredictably increasing from day today. The request of soil data is for various purposes of which some them are: for mixed agricultural investment, for commercial farming, for community development, for urban planning, for forest development, for animal husbandry, for catchment treatment (soil conservation), and for dam construction for irrigation. Due to these tangible reasons some soil data have been collected and stored since many years in some parts of the country.

The Lake Naivasha area is one of such areas where soils data have been collected mainly for commercial farming under Lake Water irrigation. These data include detailed soil surveys of farms at scale 1:5000 and 1:10 000 for the assessment of soil fertility, and semi-detailed soil survey of east, south east and south of the lake shore at scale 1:50 000 for scientific research purpose. The surveys were carried out by Kenya Soil Survey Institute who is the responsible entity for collection, analysis storage and distribution of information related to soils for the whole country at national level. In recent years, International Institute for Aerospace Survey & Earth Sciences (ITC) has carried out additional soil survey via its MSc program for academic learning in scientific research process.

However, most of the data collected for Lake Naivasha area are:

- Stored in analogue format (paper archive) which is difficult for updating, manipulation, retrieval and further analysis.
- Not organised and put in to a common database according to common standards
- Not easily accessible by different users according to their requirements, since the scale and the level of detail has been once fixed during its presentation, also it is bulky and difficult to find out from local archives.
- In different scales; different level of detail and with different survey approaches.
- In different classification systems of which the user can not easily be able to correlate them for interpretation purposes.

1.4. Research Objectives

This research focuses on the development of a GIS (geographic information system) environment to store, retrieve, manipulate and present soil information for multiple purposes. The storage includes the structuring of the relevant data in a geographic database both in spatial and non-spatial databases. The main specific objectives of this study are:

1. To rescue the existing soils spatial and attribute data available for Lake Naivasha area.
2. To organise and integrate available soil information for Lake Naivasha area from different sources and store them in a common database.
3. To build a preliminary Soil Information System for Lake Naivasha area.
4. To document working procedures for soil geographic database design and SIS development for Lake Naivasha area.
5. To suggest a solution to over come the problems which arise from the differences in classification systems and survey approaches during data integration.
6. To suggest some ideas and concepts on improving user accessibility.

1.5. Research Questions

As it has been stated in the aforementioned objective section, this research is more of an application of Geographic Information Systems Technology to rescue the relevant and available data for Lake Naivasha area with respect to soils. Moreover, the study looks aspects of suggesting ideas and solutions to some technical issues related to analogue digital conversion, data organisation and data integration. To perform these tasks, the following research questions have to be answered. These are:

1. Is Digital environment of Soil Information System preferred from that of the Analogue (Manual/Traditional)? Why?
2. How can the existing soil data (digital and/or analogue) be organised and documented in a database?
3. How can user access to the soil data set be improved?
4. What are the techniques or methods that are useful to change the analogue soil data to the digital environment?
5. How can the spatial data at different scales and the attribute data in different formats be combined so as to put in the same database with acceptable standards?

6. How can survey results from different classification systems and/or survey approaches be used for interpretation purposes, i.e. how these different classification system and survey approach results can be compared among each other?

1.6. Thesis Outline

The outline followed to write up this thesis consists of eight chapters, which are introduction, literature review, description of the study area, methods and materials, results and discussion, conclusions and recommendations, references and lastly appendices.

Chapter 1 deals with general introduction about soil information systems, their role and basic considerations in their design. In addition to these research problems, research objectives and research questions are explained under this chapter. This chapter lies the basis for the other seven chapters.

Chapter 2 is concerned with literature review where an extensive investigation of literature has been done about information system development methodologies, database design steps and data models, metadata standards, formats and tools, soil survey approaches, profile description and soil classification systems, types of soil maps and their scales.

Chapter 3 explains about the existing situation of the study area from social, economical and political perspective. This includes geographical location, climate, population and infrastructure, geology, geomorphology, soils, hydrology, agriculture, vegetation, wildlife, lake Naivasha and stakeholders.

Chapter 4 deals with materials and methods used in conducting this research process. The process was organised under seven sub chapters namely, materials and equipment, the research approach, the research methodology, information system development, data modelling, multi-source data integration and user access improvement. Each one of them is explained in detail.

Chapter 5 discusses about the results obtained from this research. The discussion includes about the responses to user need assessment, map scale and management area, the need for soils data, information system design, soil geographic database design, multi-source data integration, soil map accuracy assessment, user accessibility (metadata documentation and user interface design), and information system architecture. The discussion is accompanied by the presentation of the results.

Chapter 6 deals with conclusions and recommendations while **chapter 7** contains the list of references or citations that are used in writing this thesis. The last chapter that is **chapter 8** deals with appendices containing interview questionnaires, spatial data record formats, entities and their internal design aspects, data dictionary for processes and data stores, correlation tables and sample metadata document.

2. Literature Review

Introduction: This chapter deals with an investigation of different literatures available in order to solve the research problems mentioned in chapter 1. The research problems are implicitly embedded in research objectives and research questions. Therefore, in dealing with the research questions and objectives one has to look for methodologies to design and implement functions that are geared to improving or solving the present situation of soil information shortage at Lake Naivasha area. For this reason we reviewed literatures on issues related to information system development, database design, metadata documentation, and soils in general. Each of these topics is explained in more detail with its application in chapters 4 and 5 of this thesis document.

2.1. Information System Development

2.1.1. Definition of Terms

For a common understanding (for this study) on the basic terms in information system development process, the following most frequently used terms are defined as follows.

Information is an answer to a specific question. The question generally emerges in the context of problem solving in connection with managerial/decision making activities (Paresi, 1999).
System is a collection of people, means (technical, financial) and procedures organised to accomplish a specific set of functions (applications) (Paresi, 1999).
Information System is a system to transform data in information (including collection, processing, storing, retrieving, analysis, protection and communication) such that this information can be used as input for decision-making processes (Paresi, 1999).

Geographic Information is the answer to a question related to mutual relations existing between objects and/or phenomenon and their position on earth. Geographic information describes objects and/or phenomenon from the real world in terms of:

- Their position with respect to a known co-ordinate system
- Their attributes that are unrelated to position
- Their spatial interrelations with each other (topological relation)
- Their temporal relationships

Geographic Information System (GIS) is defined as a science which studies natural, social and economic geo-systems, their structure, relationship, dynamics, and function in time and space by means of computer modelling using databases and geographic knowledge. On the other hand GIS is a technology of collection, storage, transformation, display and dissemination of spatially related information with the aim of providing for decision making in the fields of geo-system inventories, optimisation and management (Berlyant, 1992).

Information System Development (ISD) is defined as the process of planning, designing, implementing, installing and maintaining an information system. The manner in which various techniques and tools are used to realise this process is referred to as **System Development Methodology (SDM).** A task for which an information system is developed is called an application (Paresi, 1999). From the term "development" it would seem rather straightforward that ISD should refer to improving up on a particular kind of system, referred to as an information system. In deed it does, but it is also often applied to mean establishing an information system to organise the existing data from various sources like that of this research (Mukumbura, 1995).

2.1.2. Components of an Information System and its Life Cycle

An information system in general but also GIS, for example an automated Soil Information System (SIS) is expected to consist of the following components.

1. Hardware – which is the computer it self.
2. Software – which includes programmes in a computer
3. Data – which includes input to the system to be processed and output from the system
4. Method – the steps in the process of information production
5. People – who performs the operations and uses the information.

When we look in to the life cycle of GIS, we can easily observe that it is an iterative in nature. This is because by now an iterative lifecycle is found to be better than a more linear approach. The merit of an iterative approach is that, while developing the system, one can have the option to go back to the former steps at any stage of the development phase. The same holds true in developing SIS, as it is obvious that SIS is grouped in the family of GIS. The development cycle is shown in figure 2.1.

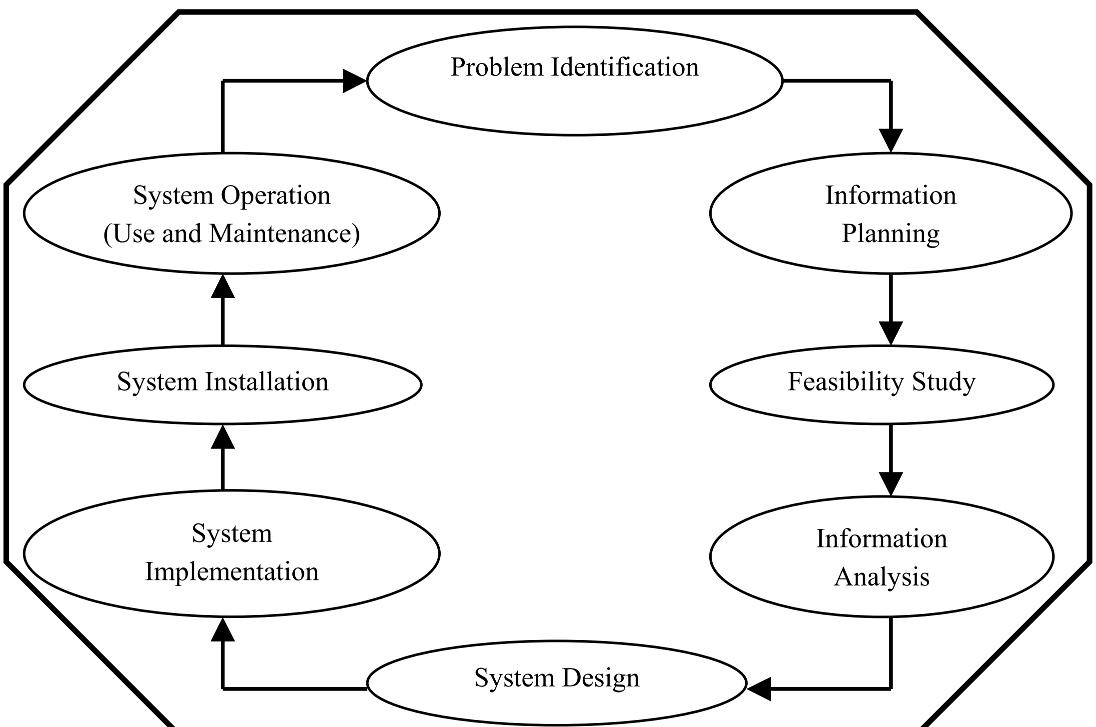


Figure 2.1 GIS/SIS Development Cycle
(Source: Adapted & modified from De By et al., 1999 and Paresi, 1999)

2.1.3. Information System Development Methodologies

An Information System Development Methodology (ISDM) is a recommended collection of philosophies, phases, procedures, rules, techniques, tools, documentation, management, and training for developers of information systems (Avison and Fitzgerald, 1988).

The most important factor for the need for methodologies is the limitation of the human mind to perceive and retain all information it requires to act on it promptly (Paresi, 1999). Benefits from methodologies include:

- High quality products which are easy to maintain and up grade
- Better control during design processes
- Reduced time to completion and lower costs of development

Most methodologies have techniques embedded within them. Techniques such as cost benefit analysis, prototyping and benchmarking are commonly found in the early feasibility stages of development methodologies. In subsequent data and process modelling stages, specialist analytical tools such as Normalisation, Entity Attribute Relationship Diagrams, Data Flow Diagrams, and Action Diagrams are often employed (Reeve and Petch, 1999).

Almost all the published accounts of GIS development strategies refer at some point to a user needs assessment (study). Essentially what this means is that the project team goes out in to the organisation and by means of interviews, questionnaires, observing existing work process and analysing data flows, attempts to determine the demand for GIS (Reeve and Petch, 1999).

As far as system development methodologies are concerned six approaches have been emerged from research and experience. These are:

1. ***Soft System Development Methodology (SSDM)*** – this methodology is process oriented, deals with fuzzy (not clear) types of problem situation, complex objectives, and covers only the strategy and analysis phases of system development. According to Lewis (1994), the 'track- record' of soft system methodology is impressive. It has been successfully used in several hundred interventions of different kinds and found to be a particularly flexible form of the system approach.

2. ***Socio – Technical System Development Methodology (STSDM)*** – takes the participatory design approach, gives emphasis on simultaneous design of social and technical subsystems. This method covers analysis and design phases in more details where as feasibility and implementation phases in less detail. According to Reeve and Petch (1999), the approach focuses on the changes involved in introducing new technology and has three main objectives. These are:
   - To establish a position which includes all users of a system in design
   - To enable groups involved in systems design to set job satisfaction objectives in addition to technical objectives and
   - To ensure that any new technical system is surrounded by a compatible functioning organisational system.
3. ***Structured System Development Methodology (StSDM)*** – is a proven methodology that provides techniques and tools, procedures and elaborated planning guidelines. It involves the participatory approach to analysis and is suited for automation. It covers all the phases of systems development,

but it separates data and process modelling and focuses on centralised systems. In the present situation the methodology is famous and widely used (Paresi, 1999). This methodology follows an iterative approach of system development and its schematic representation (it is some times called waterfall approach) is shown on figure 2.2 below.

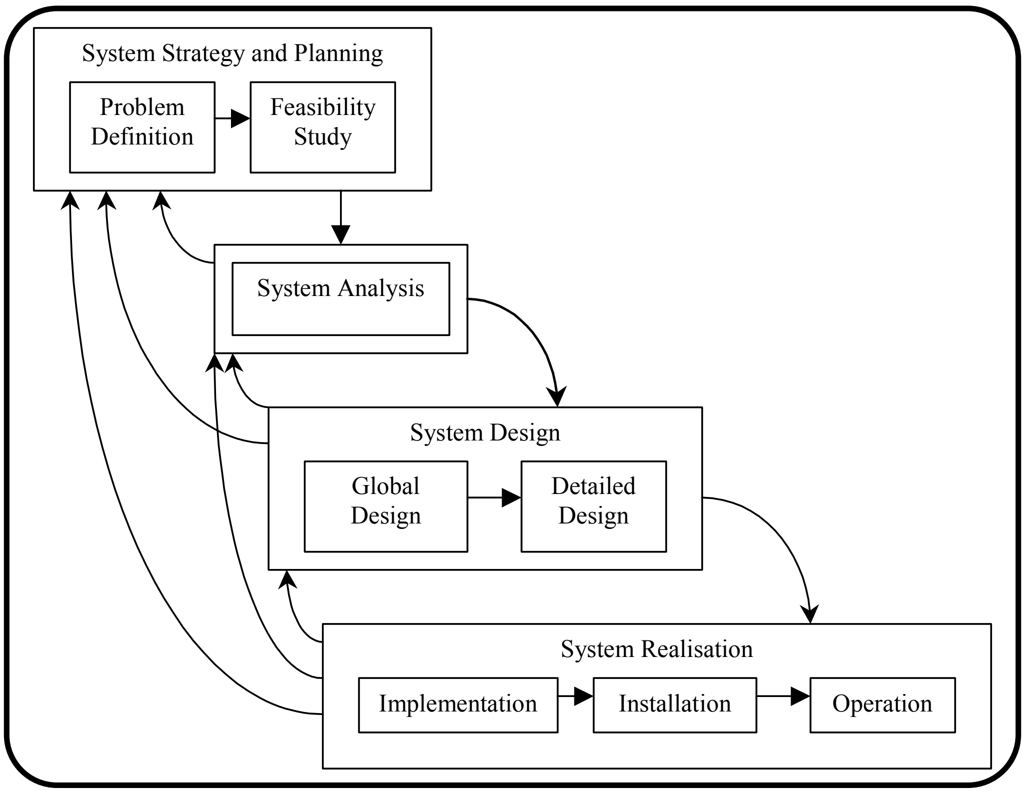


Figure 2.2 An Iterative Approach of Structured System Development Methodology
(Source: Adapted from Paresi, 1999)

4. ***Object – oriented System Development Methodology (OOSDM)*** – is based on object modelling, covers all system development phases and provides techniques and tools. It involves a participatory approach to analysis and is suited for automation. It is especially suitable for association or aggregation (e.g. parcel to house) and inheritance (e.g. building to house) relationships. The system is robust and easier to maintain. However, this method is at a stage of infancy (Paresi, 1999).

5. ***Formal System Development Methodology (FSDM)*** – is well adapted to software engineering and process control. It is based on the principles of formal specifications and verified design. However, this method does not cover all phases of system development (Paresi, 1999).

6. ***Evolutionary System Development Methodology (ESDM)*** – has the same potentials like that of OOSDM, but this method uses a cyclic/iterative and stepwise development approach. In addition to this there is user involvement in particular during process design/redesign and makes use of modern tools (CASE tools and software generators) (Paresi, 1999).

The choice of system development methodology depends on various factors, such as social factors, technical factors, relationship factors, and complexity of the information. In general the method applied should be problem solving. In practice system development can make use of a combination of these different approaches. According to Paresi (1999), an optimum mix of methodologies is recommended. The suggested mix is that of Soft and Structured methodologies for system strategy and planning phase and a choice between Structured and Object – Oriented methodologies for the other phases.

However, the selection of methodologies also depends on the availability of software that can handle the methodology and related tools, the clearness/fuzziness of the problem and its popularity. Therefore, the methodology that was chosen for the development of SIS for Lake Naivasha area is the structured system development methodology. This is because it fulfils all the requirements mentioned above. For more detail about this methodology see chapter 4 section 4.4.

2.2. Database Design

A number of commercial database management systems are available, but many of these were developed for business applications and are not well suited to the more complex structure of soil data. As a result, some groups have begun to develop systems adapted to the requirements of soil science. Similar efforts have been going on in related earth science disciplines such as Geology (Gordon, 1975). Since soil geographical database development is a growing theme, this study might contribute some additional ideas and concepts particularly, in organising (integrating) soil data from various sources.

2.2.1. Definition of Terms

Data are representations of facts, concepts or instructions in a formalised manner, suitable for communication, interpretation or processing by human or automated means. Only when placed in a context, data can produce information (Paresi, 1999).

Database is a collection of interrelated data stored together with as little redundancy as possible to serve one or more applications. Or a large computerised collection of structured data is what we call a database (De By et al., 1999). According to Howe (1989), a **Database** is a collection of non-redundant data shareable between different application systems.

Database design is the development of the structure of the database as well as the definition of its contents and the specification of the constraints to be placed on the validity of data (Hawryszkiewycz, 1998). A **Data structure** is the format in which the data will be organised within the database.

Database Management System (DBMS) is a software package that allows the user to set up, use and maintain a database (De By et al, 1999).

Database creation is critical within information system implementation, as it is known that the process is costly and time consuming. Setbacks/obstacles, false starts and time overruns during the database creation phase can quickly cause the support of the project management to decline. Hence, a wise project leader, therefore, will take particular care to establish appropriate and realistic data conversion and integration plans to over come the drawbacks. Within these plans, issues such as technology (scanning, digitising, line following, etc), quality (accuracy, documentation, validation, etc), staffing (internal sub

contractors, external agencies, etc) and scheduling (ensuring that some key products can be produced from a partially complete database) will need to be considered (Reeve and Petch, 1999).
The goals of database design include:

- To ensure that all data needed to satisfy users' requirements are stored in the database.
- To minimise redundancy.
- To provide a way to understand the organisation of the data to the users and
- To support the specific processing requirements and performance objectives.

The process of database design shall be an iterative type. This is for the reason that, some modifications to the solutions adopted in an early phase may be necessary or required while working in a later phase(s). Therefore, the database design approach followed by this study is an iterative one and can loop back as many times as needed. The steps followed in the design process traverse five phases where the first two (need assessment and conceptual) phases are DBMS independent while the other three (logical design, physical design and implementation) phases are dependent on the DBMS. The schematic diagram of the design process is shown on figure 2.3 below. The design steps are discussed in detail in chapter 4 section 4.5.

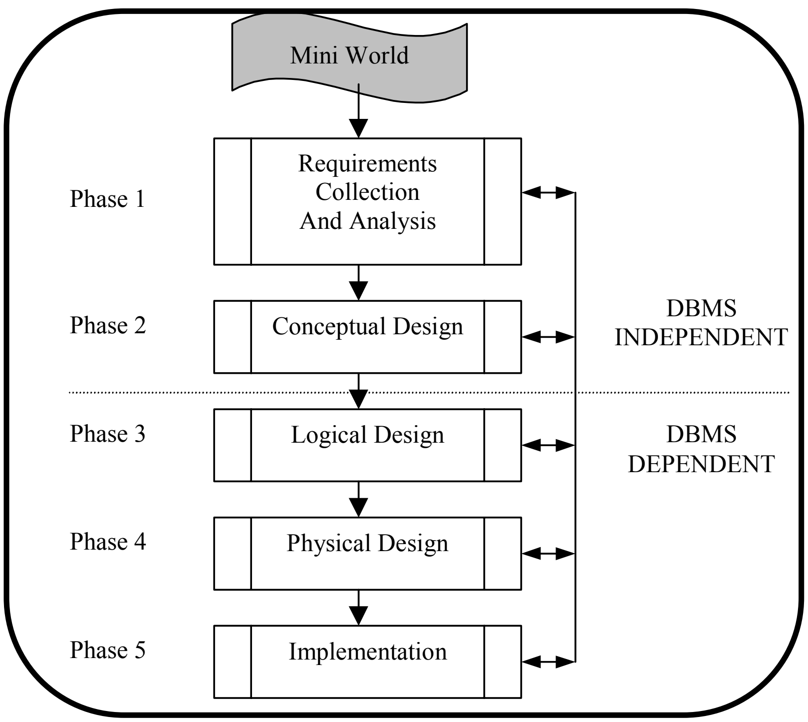


Figure 2.3 Typical steps for Database Design
(Source: Adapted From Elmasri and Navathe 1989)

2.2.2. Types of Data Models

A Data Model is a language with which one can define a database structure and manipulate the data stored in it (De By et al., 1999). According to De By et al (1999), five data model types have been identified. These are:

- Flat file attribute record data model
- Hierarchical data model

- Network data model
- Relational data model and
- Object oriented data model

De By et al (1999) further stated that among the five data models the relational data model is the most prominently used data model type. According to Worboys (1999), the three currently most important data modelling approaches are record-based, object-based and object-relational approaches where each of them are described briefly as follows.

1. **Record-based approach**

A record- based model structures the database as a collection of files of fixed format records. The records in a file are all of the same record type, containing a fixed set of fields (attributes). The early network and hierarchical database systems belong to the record-based data model. However, they proved to be too closely linked to physical implementation details, and the relational model has largely superseded them.

2. **Object-based approach**

The primary components of an object- based model are its objects or entities. The entity relationship attribute (ERA) model and the object-oriented (OO) models are the two main object-based modelling approaches. The ERA approach is attributed to Chen (1976) and has been a major modelling tool for relational database systems for about 20 years. Entities have explicit relationships with other entities. Entities are grouped in to entity types, where entities of the same type have the same attribute and relationship structure. The structure of data in a database may be represented visually using an ERA diagram (Worboys, 1999).

“For many application domains including GIS, ERA modelling has proved too limited and is being superseded by the OO approach. The OO approach is in use both as a method of semantic data modelling and as a model of data handled by object oriented programming and database management systems. The state of an object at any time is determined by the value of the data items within its wrapper/covering. These data items are referred to as instance variables, and the values held within them are themselves objects. This is an important distinction between objects (in the OO sense) and entities (in the ERA sense) which have a two-tier structure of entity and attribute” (Worboys, 1999). However, the OO approach is at the stage of infancy and not commonly used in designing geographic soil databases.

3. **Object-relational approach**

Object-relational models combine features of object-based and record-based models. They enhance the standard relational model with some object-oriented features as opposed to OODBMS that build database functionality around an OO programming language. Enhancements include complex, possibly user defined data types, inheritance, aggregation, and object identity (Worboys, 1999). The relational data model is one part of the object relational approach.

The relational data model was introduced by Codd's seminal/original paper in 1970. The model is based on a simple and uniform data structure – the *relation* – and has a solid theoretical foundation, being predicate logic and set theory. The relational model has also become firmly established in the database application world, and there are many commercial relational database management system packages that implement it. Nowadays, it is considered as an appropriate model for internal schema definitions, where as the entity relationship model is used for conceptual purposes (De By, 1999).

A relational database is a collection of tabular relations each having a set of attributes where the data in a relation are structured as a set of rows. A row or a tuple consists of values one for each attribute. An attribute has a domain associated with it, from which its values are drawn. Most current systems require that values are atomic – for example they can not be decomposed as lists of further values. So a single cell in a relation can not contain a set, list or array of values (Worboys, 1999). Therefore, the most prevalent current database paradigm, the relational data model, was used for non-spatial data modelling for the case study of Lake Naivasha area.

2.3. Metadata

2.3.1. Definition of Terms

More generally the term 'Metadata' refers to background information about some thing. **Metadata** is data about data. In the context of digital spatial data, **metadata** is the background information, which describes the content, quality, condition, and other appropriate characteristics of the data set. Paper maps contain metadata, primarily as part of the map legend. In this form, metadata is readily apparent and easily transferred between map producers and map users. When map data are in digital form, metadata is equally as important, but its development and maintenance often require a more conscious effort on the part of data producers and the chain of subsequent users who may modify the data to suit their particular needs (Hart and Phillips, 1998).

Metadata is a simple mechanism to inform others of the existence of data sets, their purpose and scope. In essence/essentially, a metadata is expected to answer who, what, when, where, why, and how questions about all facts of the data set made available (De By et al., 1999).
A clearinghouse is a distributed network of spatial data producers, managers and users that are linked electronically together. It is a system of software and institutions that are to facilitate the discovery, evaluation, and downloading of digital spatial data and provides means to inventory, document and data sharing (De By et al., 1999).

2.3.2. Importance of Metadata

Metadata serves many important purposes, including data browsing, data transfer and data documentation. Metadata may be a simple listing of basic information about available data or detailed documentation about an individual data set (Hart and Phillips, 1998). Metadata can be used internally by the data provider to monitor the status of data sets, and externally to advertise to potential users through a national clearinghouse. Metadata are important in the production of a digital spatial data clearinghouse, where potential users can search for the data they need (De By et al., 1999).

At more detailed level, metadata may be considered as insurance. Metadata insures that potential data users can make an informed decision about whether data are appropriate for the intended use. Metadata also insures that the data holdings of an agency are well documented and that agencies are not vulnerable to losing all the knowledge about their data when key employees retire or accept other jobs. Metadata will also save the time spent on searching some ones else's data simply to discover its existence (Hart and Phillips 1998).

2.3.3. Forms of Metadata

The most common form of metadata is a file folder filled with notes on data sources and procedures used to build the data. While the less common form is complete, organised metadata such as the Wisconsin Department of Natural Resources' GIS Data Users Guide (Hart and Phillips, 1998).

Digital metadata may be created, stored, and used in a variety of formats. Among these:

- The most basic is an American Standard Code for Information Interchange (ASCII) text document. An ASCII document is easy to transfer to other users independent of the hardware/ software platform they use.
- Another common format is Hypertext Mark-up Language (HTML). HTML provides an attractive way to view metadata using a browser such as Netscape Navigator, Mosaic, or Microsoft Internet Explorer.
- Recently, there has been strong interest in creating metadata in Standard Generalised Mark-up Language (SGML). SGML provides an effective way to tag metadata elements.

Among the metadata forms mentioned above, the Hypertext Mark-up Language (HTML) was used to produce the metadata for this study. As an alternative both the Standard Generalised Mark-up Language (SGML) and the ASCII text formats were also produced. The user can choose either of them based on his/her preference for viewing from the same window.

2.3.4. Metadata Standards

The phrase 'data standard' refers to an agreed upon way of representing data in a system in terms of content, type and format. For metadata to be easily read and understood, standards create a common language for users and producers. Metadata standards provide appropriate and adequate information for the design of metadata (De By et al., 1999). Key developments in metadata standards are:

- The technical committee of the International Organisations for Standards (ISO standard 1504615 Metadata)
- The Federal Geographic Data Committee's (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM)
- Technical Committee of the European Standardisation (Comité Européen de Normalisation, CEN/TC 287)

The requirements for standards vary between different market sectors. It is thus important for all the 'actors' in the geographic information arena/stage to recognise that at a certain point in time and space, different standards may be required. Therefore the choice of which metadata standard to use depends on the organisation, ease of use and the intended purpose (De By et al., 1999; Salgé, 1999).

The compilation of the metadata for this study was done by using the Content Standard for Digital Geospatial Metadata (CSDGM) produced by Federal Geographic Data Committee's (FGDC) of the United States. The reason for selection is that, it is one of internationally accepted standard (i.e. what we call de jure standard), presence of editing and compiling software and experience of the researcher with this metadata standard.

2.3.5. Selection of Proper Metadata Tool

Deciding between holding the metadata in a database or to produce discrete metadata documents for each data set is somewhat dependent on the variety and volume of the data sets, as well as how often they (and the metadata) are updated. This decision will determine which metadata tools are appropriate to consider for use (Hart and Phillips, 1998).

The Federal Geographic Data Committee (FGDC) recommends that metadata be stored in a database if the data sets are subject to frequent change, or if some of the metadata is common to many of the data sets. If the data holdings have few metadata elements in common, then discrete metadata documents are simple ways to hold the metadata, and almost any tool can be used to produce it (Hart and Phillips, 1998).

The metadata creation tool selected to develop metadata for Lake Naivasha soil information provision was "Tkme" metadata editor, that is, structured documentation conforming to the Content Standard for Digital Geospatial Metadata (CSDGM) developed by the Federal Geographic Data Committee (FGDC). The editor shares a considerable amount of its code with its companion product, Xtme, which is specifically for Unix systems and is based on the Athena widget (graphic symbol) set of the standard X Window System distribution. Both Tkme and Xtme are closely allied with mp, a compiler for formal metadata, whose purpose is to verify that the syntactical structure of a file containing formal metadata conforms to the FGDC standard, and to re-express the metadata in various useful formats (Schweitzer, 2000).

The "mp" program is a compiler to parse (divide in to sections) formal metadata, checking the syntax against the FGDC Content Standard for Digital Geospatial Metadata (CSDGM) and generating output suitable for viewing with a web browser or text editor. It runs on UNIX systems and on PC's running Windows 95, 98, or NT. MP generates a textual report indicating errors in the metadata, primarily in the structure but also in the values of some of the scalar elements (i.e. those values which are restricted by the standard) (Schweitzer, 2000).

2.4. Soils

2.4.1. Definition of Terms

According to the Canadian Glossary of Terms in Soil Science (Schut, 1996), soil is defined as follows. **Soil** is naturally occurring unconsolidated material on the surface of the earth that has been influenced by parent material, climate (including the effects of moisture and temperature), macro- and micro-organisms, and relief, all acting over a period of time to produce soil that may differ from the material from which it was derived in many physical, chemical, mineralogical, biological, and morphological properties.

According to Soil Taxonomy, **"Soil** is the collective term used for the natural bodies, made up of mineral and organic materials, that cover much of the earth's surface, contain living matter and can support vegetation out of doors, and have in places been changed by human activity" (Rossiter, 2000b). In the writer's opinion, based on the concept of the above definitions, we can very briefly define **Soil** as the unconsolidated material on the immediate surface of the earth that serves as a natural medium for the growth of plants and whose formation is influenced by different factors.

2.4.2. Soil Survey Approaches

Various definitions of soil survey do exist amongst, the following definition was applied in this study. "**Soil survey**, or more properly, **soil resource inventory**, is the process of determining the pattern of the soil cover, characterising it, and presenting it in understandable and interpretable form to various consumers" (Rossiter, 2000a). Rossiter, (2000a) have defined five main classes of soil survey methods. Of these the Aerial Photo Interpretation (API) guided survey includes physiographic and geo-pedologic soil survey approaches.

2.4.2.1. Physiographic Analysis Approach

According to Goosen (1967), Physiographic analysis is defined as the analysis of processes, rather than of phenomena. Physiographic analysis of aerial photos is made based upon a thorough knowledge of the relation between physiography and soils, and upon the recognition of dynamic processes rather than of static elements. The geomorphological analysis of any terrain is the basic approach of the physiographic analysis for soil survey purposes (Goosen, 1967).

According to Farshad (1999), the term physiographic processes has been replaced by geomorphic processes where the understanding of these processes will be the guide to the delineation of the physiographic units which in turn will provide a good basis for the pattern of soil mapping units.
In a physiographic soil survey approach, three levels have been recognised:

- **Land type**: this is the higher level consisting large area. Example: Lacustrine plain (L).
- **Sub-land type**: this is the middle level containing more similar features under land types. Example: Lake flats (L1) and Swamps (L2), under lacustrine plain land type.
- **Sub-units/mapping units**: this is the lowest level containing very similar features under sub-land type. Example: Permanent swamp (L21) and seasonal swamp (L22), under swamp sub-land type (Farshad, 1999).

2.4.2.2. Geopedologic Analysis Approach:

The ITC geo-pedologic soil survey approach was developed by professor J Alfred Zinck and is essentially systematic application of geomorphic analysis to soil mapping. This approach is based on a strong integration of geomorphology and pedology, using geomorphology as a tool to improve and speed up the soil survey (Zinck, 1989).

The geo-pedologic approach can be used to cover large areas rapidly, especially if the relation between geomorphology and soils is close. The approach depends on two assumptions:

(1) Boundaries drawn by landscape analysis separate most of the variation in the soils.

(2) Sample areas are representative; their soil pattern can be reliably extrapolated to unvisited map units.

In addition, this approach has advantages in legend construction and structuring. It is a hierarchical system containing six levels (geostructure, morphogenetic environment, landscape, relief, lithology and landform). Once lines are drawn at one categorical level, they remain, even if the soils in adjacent units have the same classification. This is because of the many interpretations that are related to the geoforms (Rossiter, 2000a). The geopedologic hierarchical analysis structure with some description is shown on table 2.1. This structure, except levels 5&6 (which are not common for map legend making) has been used to form the soil map unit entity in the non-spatial database of this research.

Table 2.1 Synopsis of the geoform classification system based on the Geopedologic analysis approach

Level	Category	Generic concept	Short description
6	Order	Geostructure	Large continental portion characterised by a broad geologic structure (e.g. cordillera, geosyncline, shield)
5	Suborder	Morphogenetic environment	Broad type of biophysical medium originated and controlled by a style of internal and/or external geodynamics (e.g. structural, depositional)
4	Group	Landscape	Large portion of land characterised by a repetition of similar relief types or an association of dissimilar relief types (e.g. valley, plateau, mountain)
3	Subgroup	Relief/molding	Relief as determined by a given combination of topography and geologic structure (e.g. cuesta, horst). Molding as determined by specific morphoclimatic conditions or morphogenetic processes (e.g. glacis, terrace, delta)
2	Family	Lithology/facies	Petrographic nature of hard rocks (e.g. limestone, basalt) or origin/nature of soft cover formations (e.g. periglacial, lacustrine, alluvial)
1	Subfamily	Landform	Conspicuous basic geoform type, characterised by a unique combination of geometry, dynamics and history

(Source: Adapted from Zinck, 1989)

2.4.3. Soil Profile Description

Definition: **Soil profile** is defined as a vertical section of the soil through all its horizons and extending into the parent material (Schut, 1996). Soil maps are used and interpreted for a variety of purposes. Interpretations of soil maps are generally based on descriptions of representative profiles for each mapping unit. The quality of these interpretations depends entirely on the quality of the representative profile descriptions. A wrong representative profile description will automatically lead to wrong interpretations. Therefore, knowledge on the quality of these descriptions is very important (Bregt, 1988).

Soil profile descriptions form the basic data in all soil surveys. They provide a major part of the information required for correlation and classification of the soils of an area. They are essential for interpreting soils and for co-ordinating interpretations across state and regional boundaries. The soil descriptions and the soil map are the parts of a published survey having the longest life (Soil Survey Division Staff, 1993).

Field descriptions of soil profiles range from partial descriptions of material removed by a spade or by an auger to complete descriptions of pedons seen in three dimensions from intersecting pits as horizontal layers are removed sequentially from the surface downward (Soil Survey Division Staff, 1993). Concerning standard soil profile description forms, the two most widely used are:

- FAO Guidelines for Soil Description (FAO, 1990)
- USDA Field Book for Describing and Sampling Soils (Schoeneberger et al., 1998)

According to Soil Survey Division Staff (1993), field descriptions should include:

- Observed external attributes of the polypedon, such as landform and characteristics of slope
- Inferred attributes of the polypedon, such as origin of soil parent material and the annual sequence of soil-water states
- Observed internal properties of the pedon, such as horizon thickness, colour, texture, structure, and consistence
- Inferred genetic attributes of the pedon, such as horizon designations and parent material
- Inferred soil drainage class
- The classification of the pedon in the lowest feasible category
- The location of the site relative to geographic marks and in terms of landscape position
- The plant cove or the use of the site
- The date, time of day, and weather conditions and
- The name of the describer

2.4.4. Soil Classification Systems

The purposes of classification are manifold. "The purpose of any classification is so to organise our knowledge, the properties of the objects may be remembered and their relationship may be understood most easily for a specific objective" (Cline, 1949). When we classify the objects of a population, new relationships may become evident and predictions about the behaviour and best use of the objects are possible. Classification helps us to remember the significant properties of the objects and provides us with a 'language' by means of which people can exchange knowledge, ideas and information (Yifter, 1995).

The issue of soil classification systems is very important for any soil surveyor after conducting his/her survey activities. This is because different classification systems need different levels of detail for their classification. Among the different classification systems the common ones are the following. (Source: A Compendium of on-line Soil Survey Information on a World Wide Web site location http://www.itc.nl/~rossiter/research/rsrch_ss.html).

1. ***World Reference Base (WRB)*** is an internationally agreed upon standard soil classification system developed by an international working group. This classification system has 30 Reference Soil Groups and 121 qualifiers for two level subdivisions. This classification system is not mostly advised to be in use for detailed soil mapping. The reason is that many detailed soil properties that are important for land use and soil behaviour analysis are not specified in sufficient detail (Rossiter, 2000b).

2. ***Soil Taxonomy (ST)*** is a national soil classification system whose main aim has been to group soils of USA, however, by know it is pretending to be universal or international. It is being used widely by many countries. This classification system is hierarchical with six main levels, which are order, sub-order, group, sub-group, family and series. ST as compared to WRB looks detailed soil properties during its classification. Therefore, it is advisable to use soil taxonomy for detailed soil mapping (Rossiter, 2000b).

3. ***French system*** is another more or less widely used classification system among French language speaking countries. This classification system uses a set of 102 reference solums, which result from unique interpretations of morphology, genesis and function. A soil in the field can be assigned to one or more references, which fits with the idea of fuzzy classification (Rossiter, 2000b).

4. ***National classification system***, which is developed and used by a nation or nations. For example Great Britain, Australia, Canada, Brazil, and the Netherlands have their own national classification system (Rossiter, 2000b).

5. ***Interpretative classification*** is meant to classify soils according to their suitability for general or specific land uses. This type of classification includes for example: land suitability classification, land capability classification, hydrological classification, and wetland classification.

6. ***Indigenous classification*** is a local classification system used by the local people to classify their soil mainly for agricultural purposes. Such a classification system classifies soils for example in to good and bad for a certain use type. Oral explanation and/or experience sharing is often used to transfer indigenous classification knowledge (Rossiter, 2000b).

Among the aforementioned classification systems the first two were used to classify soils of Lake Naivasha area. In order to be able to interpret using both classification systems a correlation table was developed and added (including for diagnostic horizons) to the database. For more detail about correlation tables see chapter 5 section 5.4.1.2.

2.4.5. Types of Soil Maps

In soil survey maps are integral parts of the whole soil survey process and always presented as part of the report. Soil maps are made to show mainly the spatial distribution of soils. This geographic distribution of soils can be presented in different types of maps (Rossiter, 2000a). Some of these are:

1. ***Point soil maps:*** these are maps where the actual sample points are indicated along with their soil class and/or one or more properties.

2. ***Area-class polygon soil maps:*** these are maps, which are aggregates of polygons or delineations whose boundaries are precise lines. A delineation is labelled with a soil class name and each class name is described in map legend. These types of maps are useful for modelling of the real world as discrete model of spatial variation (DMSV).

3. ***Continuous-field maps:*** these are maps technically made by interpolation from point observations having fuzzy boundaries and showing inferred continuous distribution of a soil property. These

types of maps are useful for modelling the real world as continuous model of spatial variation (CMSV).

4. ***Continuous-field maps made by direct observation:*** these are maps showing the continuous distribution of soil properties like that of interpolated ones. But the difference is that in this case the maps are made from direct observations and measurements over the whole area of interest. These types of maps are very important for precision farming (Rossiter, 2000a).

Among the different map types mentioned above, the area-class polygon soil map is the most common map and almost all familiar soil survey maps are of this type, but the others can be better solutions in some cases (Rossiter, 2000a). It holds true for the Lake Naivasha area as well, because all the existing soil maps of this area are area-class polygon maps. The other reason is that with area-class polygon maps we can easily create topological relations, which are very important to link the spatial data with the non-spatial data in the database structure.

2.4.6. Map Scale

The publishing scale of any soil map greatly depends on the soil survey result that the users need. Therefore, it is important for the map producer to conduct user needs assessment before deciding on the map scale. This is unforgettable, because the level of detail for management purposes and for further information extraction is highly influenced by the published map scale.

According to Avery (1987) and Rossiter (2000a), soil map scales can be grouped in to three main groups. These are:

1. Large scale maps (1:2 000 – 1:25 000): These maps show all features of interest in detail and legibly.

2. Medium scale maps (1:50 000 – 1:150 000): These maps show most features of interest in detail, however, due to generalisation some finer features can not be shown.

3. Small scale maps (1:250 000 – 1:1 000 000): These maps give only an over view of an area and are less important for management and decision making purposes, because almost all features of interest are generalised.

Concerning the map scale issue for the Lake Naivasha area, while organising the existing spatial and non-spatial soil data in corresponding databases the user need assessment result was taken in to consideration. This is because the level of detail during retrieval will be greatly affected by the scale of generalisation during data input even if the digital environment is scale independent. Therefore, all the available data as it was obtained in the analogue format was converted in to the digital environment maintaining the level of detail as much as possible.

3. Description of the Study Area

3.1. Geographic Location

Lake Naivasha is a shallow (about 4.7 metres average depth and a maximum of 9 metres) smooth floored, fresh water lake (Clarke et al, 1990). It is about 100 km to the Northwest direction from Nairobi, in Rift Valley Province, Nakuru District. The geographic location of the study area lies between $36^0$15'E – $36^0$30'E longitude and $00^0$40'S – $00^0$53'S latitude with an altitude range of 1880 – 2300 masl. This geographic location includes the Naivasha Lake and its surroundings. The Lake Naivasha area is inscribed by Nyandaru Mountains in the east and the Mau escarpment in the west. The Rift valley width that Naivasha is located ranges from 45 to 70 km of horizontal distance (Stuttard et al, 1996). The actual location of the study area with respect to the spatial data captured is between $36^0$20'E – $36^0$27'E longitude and $00^0$43'S – $00^0$50'S latitude i.e. east and south east of the lake Naivasha shore (figure 3.2 top). But the user need assessment study included northern, eastern and southern parts of the lakeshore. The approximate location of Lake Naivasha area is shown in figure 3.1.

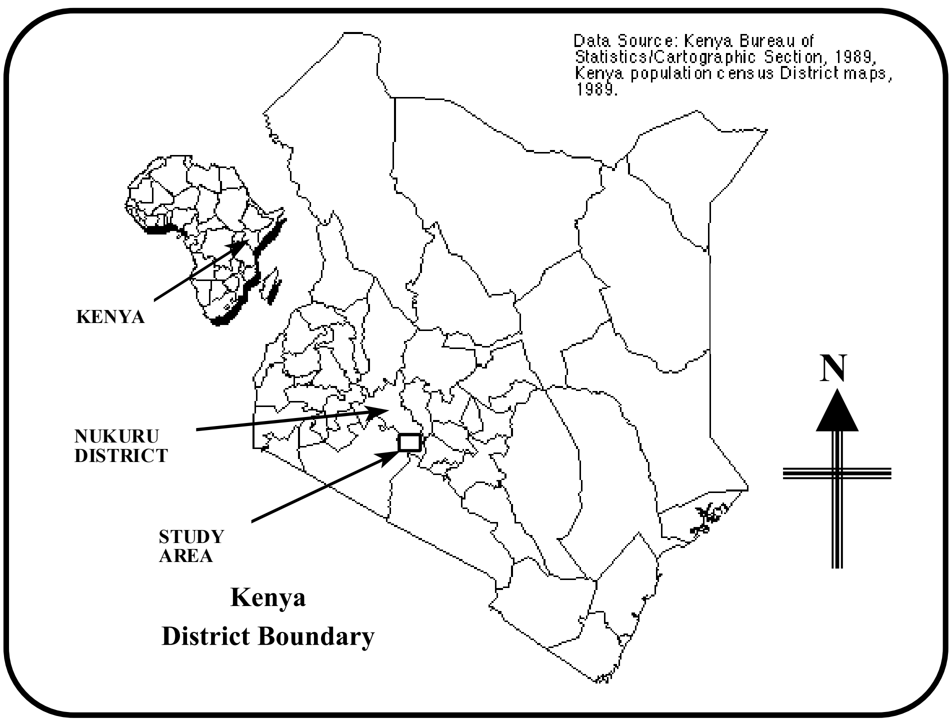
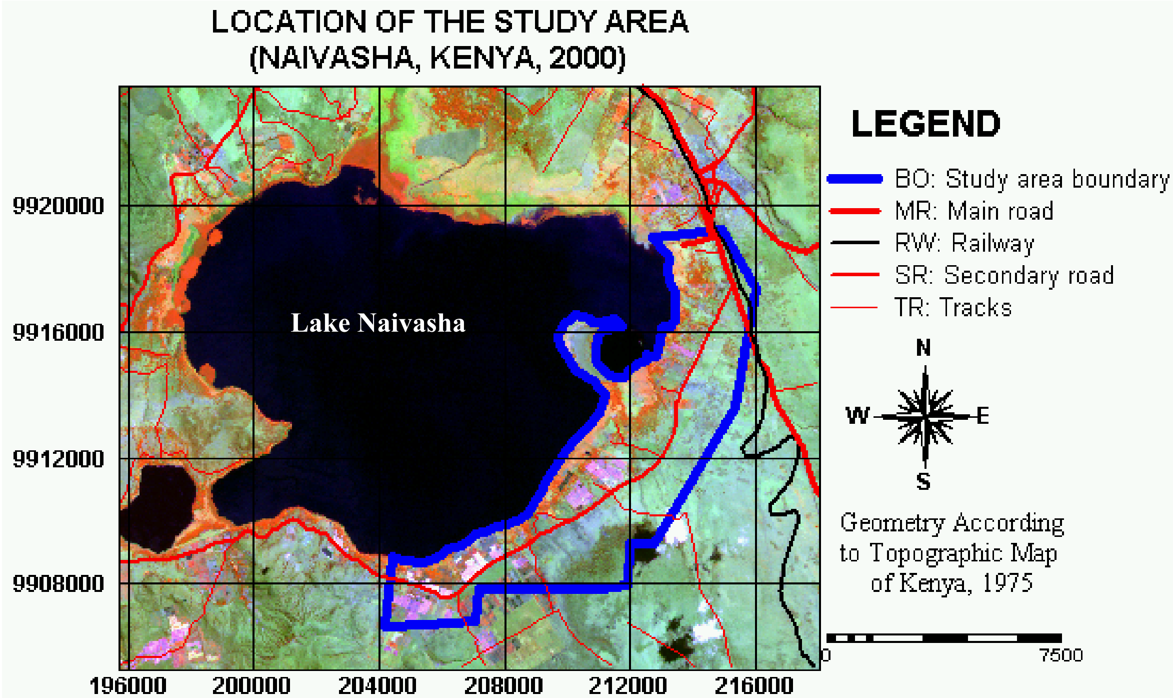


Figure 3.1 Approximate location of the Lake Naivasha area

Different views of the study area are shown in figure 3.2, which are the landscape view of Lake Naivasha area from the Kinangop plateau east, from Masai gorge north, from the Ol-Gane-Aro ridge south and the central part of the Hell's Gate. The Hell's gate is a narrow gorge, where before hundreds of years ago the lake was used to flow out through it from its current location. The exact location of the study area is also included with a background of a satellite image of March 2000.



Lake Naivasha area from Kinangop plateau, East

Central part of the Hell's Gate, South Naivasha

Lake Naivasha area from Masai Gorge, North

Lake Naivasha area from Ol-Gane-Aro Ridge, South

(Source: Photo by the writer and satellite image from WREM database, ITC)

3.2. Climate

The Lake Naivasha area has a semi-arid type of climate with an average annual precipitation ranging between 450 and 900 mm. The mean annual temperature ranges from 16^{0}C to 18^{0}C, while the maximum and the minimum temperature are 27.3^{0}C and 7.9^{0}C respectively. The mean annual, the mean max and the mean min temperature are 16.9^{0}C, 24.9^{0}C and 9^{0}C respectively (Kamoni, 1988).

The seasonal distribution of rainfall shows a longer rainy season from March to May and a shorter period of rainfall between October and November. The driest period lies between December and February. In general the rainfall pattern is not reliable for rain-fed agriculture, as it tends to fluctuate from year to year. Table 3.1 and figure 3.3 below show that it is only in April that the rainfall can satisfy crop water requirements. Hence, the rain fed agricultural practice is very negligible. Therefore, supplementary irrigation is required to cover the deficits during the rest of the year (Kamoni, 1988).

Table 3.1 Mean climatic data for Lake Naivasha area (1966 – 1980)

Month	Temp Max (^{0}C)	Temp Min (^{0}C)	Temp Mean (^{0}C)	Rainfall R (mm)	Evaporation Eo (mm)	Evapotrans. Et (mm)
Jan	27.3	7.9	17.6	24.0	118.0	79.0
Feb	27.3	8.1	17.7	39.0	178.0	119.0
Mar	27.2	9.4	18.3	59.0	190.0	127.0
Apr	25.0	11.0	18.0	113.0	149.0	99.0
May	23.6	10.6	17.1	84.0	132.0	88.0
Jun	22.8	9.2	16.0	41.0	120.0	80.0
Jul	22.4	8.6	15.5	34.0	125.0	83.0
Aug	23.0	8.6	15.8	44.0	142.0	95.0
Sep	24.5	7.9	16.2	44.0	158.0	105.0
Oct	25.5	8.9	17.2	47.0	183.0	122.0
Nov	24.5	9.1	16.8	59.0	134.0	89.0
Dec	25.7	8.3	17.0	39.0	158.0	105.0
Sum	298.8	107.6	203.2	627.0	1787.0	1191.0
Mean	24.9	9.0	16.9	52.3	148.9	99.3

(Source: Kamoni, 1988)

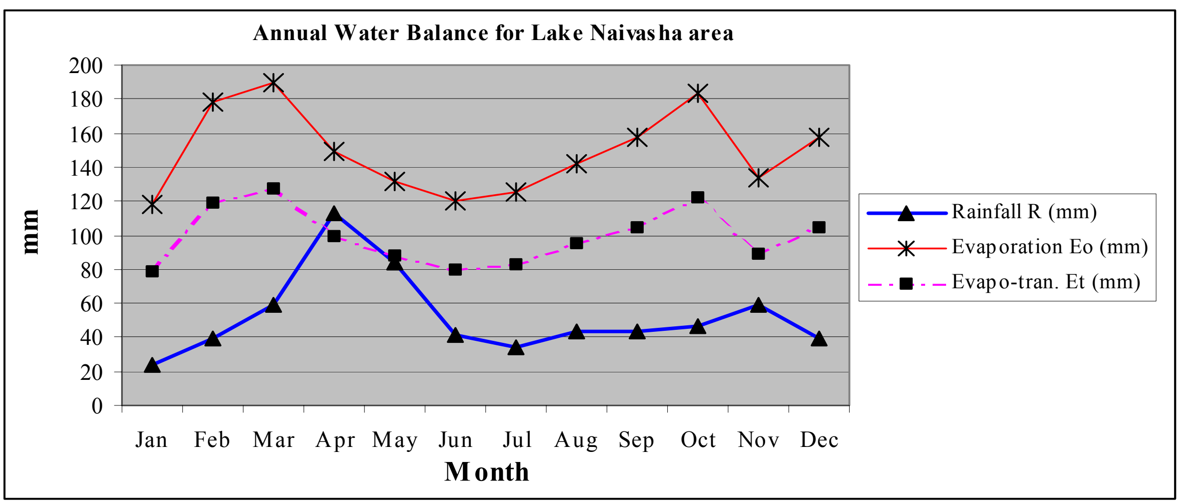


Figure 3.3 Mean annual water balance for Lake Naivasha area 1966 – 1980
(Source: Kamoni, 1988)

3.3. Population and Infrastructure

The human population of Naivasha town with its surrounding was estimated to be about 250 000 in 1995 (LNROA, 1995). The area has become industrially significant as the result of developments in commercial flower production for export, different horticultural practices and tourism industry with other various income-generating activities. Due to these reasons human population at Lake Naivasha area is continuously increasing.

As far as the road net work is concerned, the major tarmac road and railway line connecting Mombasa via Nairobi to Kisumu passes through Naivasha town. The major truck road from Mombassa to Kampala, Uganda also passes through this area. In addition to these there are other secondary all weather gravel roads connecting villages to villages and circumscribing the lake, and small air strips.

3.4. Geology

Two types of quaternary deposits, of mainly lacustrine and/or volcanic origin cover the study area. The older deposits vary in composition but largely comprise fine ashes with intercalation of pumaceous gravel deposited in lacustrine conditions during the various phases of the Gamblian Lake (Thopson and Dodson, 1963, Gatahi, 1986). The geological information for Lake Naivasha area is available on a map at scale of 1:00 000 which has been surveyed by (Clarke et al., 1988) and at scale 1:250 000 (Clarke et al., 1990).

The oldest rocks found insitu in the area have been dated to the tertiary era, and some rock fragments ejected by volcanoes may even be of an older age. The volcanic rocks in the area consist of tephrites, basalt, trchytes, phonolytes, ashes, tuffs, agglomerates, and the acid lava, rhyolites, comendite and obsidian (Thompson and Dodson, 1963).

3.5. Geomorphology

Thompson and Dodson (1963), distinguished three types of landscapes in the Naivasha area. These are Kinangop Plateau, Mau Escarpment and the Rift Floor. The study area is found at the rift floor and in the geopedological soil survey terminology, one major landscape type i.e. plain and two subdivisions of this landscape (lacustrine plain and volcanic plain) have been identified.

The lacustrine plain occurs around the shore of Lake Naivasha and extends from about 1880 – 1920 masl. According to relief in the lacustrine plain a number of terraces, such as low terrace, middle terrace and high terrace are recognised. These terraces have been formed due to fluctuation of the lake water level at times. The topography of this area is flat to gently undulating, with slopes up to 4%, non-dissected with no distinct drainage channels (Thompson & Dodson, 1963; Kwacha, 1998).

The volcanic plain occurs in close association with the lacustrine plain. The feature of the surface in general is non-dissected to slightly dissected with slopes about 5%. The volcanic plain is the result of lava flow from Longonot volcano and wind deposition of pyroclastic materials (Thompson and Dodson, 1963). The topography of this landscape consists slightly undulating lava flow, long ridges & extensive non to slightly dissected land surface (Kwacha, 1998).

3.6. Soils

From the complexity of the marine and volcanic activities, we can deduce and predict that the Lake Naivasha area will have different soil types, as relief, parent material, climate, time and organisms are known to be the factors affecting soil formation.

Different soil surveyors have surveyed the soils of Lake Naivasha area at different levels of detail for different purposes. These are Siderius (1977), Siderius (1980), Somboek et al. (1980), Gatahi (1986), Kamoni (1988), and Kwacha (1998). According to Somboek et al. (1980), the distribution of soils in the area is complex, having been influenced by the extensive variation in relief, climate, volcanic activity and underlying rocks. The soils are derived mainly from weathered volcanic and basement rock system. Generally soils of the study area can be grouped in to two: soils developed on the lacustrine plain and soils developed on the volcanic plain.

Soils developed on the lacustrine plain are moderately well drained to well drained, very deep, very dark greyish brown to pale brown, silty clay to clay loam. According to USDA (1994), soils were classified as Andic Xerorthents, Typic Xerochrepts (clay over sandy), Typic Xerochrepts (fine loamy over clay), Andic Haploxeralfs, Typic Eutrochrepts, Typic Haploxeralfs and Calcaric Haploxeralfs (Kwacha, 1998).

Soils developed on the volcanic plain are well drained, moderately deep to very deep, dark brown to pale brown, with non-calcareous to moderately calcareous topsoil, and moderately to strongly calcareous subsoil. According to USDA (1994) the soils were classified as Andic Xerochrept, Calcaric Xerorthents and Lithic Xerorthents (Kwacha, 1998).

On the other hand, on the exploratory soil map of Kenya (Sombroek et al, 1980), soils developed on the lacustrine plain were classified as undifferentiated Solenetz, saline phase and soils developed on the volcanic plain were classified as complex ando-haplic Phaeozems, gleyic Cambisols (fragipan), ando-calcaric Regosols and Lithosols.

In general when we look in to the soil moisture regime and soil temperature regime, the study area was classified as having ustic moisture regime and isothermic temperature regime (Siderius, 1977).

3.7. Hydrology

Lake Naivasha catchment has an internal drainage system with under ground water inflows and outflows. Lake Naivasha with a total area of 13,255 hectares, receives drainage water from Malewa river, draining the Nyandarua Mountains with a drainage area of about 1730 km2 and Gilgil river draining the Rift Valley floor from the North with a drainage area of about 420 km2 (Harper et al, 1990).

Other sources of water inputs in to the lake include rainfall that occurs directly over the lake and through underground movements from the catchment. The outputs from the lake are direct evaporation from the lake surface, transpiration from the swamp and other aquatic vegetation, under ground seepage and water extraction by human activities.

3.8. Land Use and Vegetation

3.8.1. Agriculture

The Massai tribe pastoralists who were using the land for grazing and the lake water for watering their cattle formerly used to occupy the Lake Naivasha area, but now the situation is drastically changing (Harper et al., 1990). With the arrival of white settlers in the area considerable changes in land use had occurred. Among these: they introduced beef and dairy farming, high tech irrigated agriculture with horticultural crops and flowers. Flower (roses, carnation) and horticultural (vegetable) productions for export are the main agricultural activities that are dominating the shores of Lake Naivasha area and the sector absorbs the highest number of employees.

3.8.2. Vegetation

As far as the semi-natural vegetation of the study area is concerned, three main types of vegetation are observed. These are papyrus mixed with grassland, acacia woodland, and wooded grassland. A large part of the natural vegetation has been removed and replaced by high tech agricultural farming and pasture. Except for some parts, the increasing population of human being in the area has disturbed most part of the natural ecosystem.

3.8.3. Wildlife

In Lake Naivasha area a number of game parks and/or reserves with a variety of wildlife are located. For example, Hell's gate wildlife reserve, the Crater Lake wildlife reserve, Elsamere Conservation Centre and Kenya wildlife service centre. The types of wildlife that are predominantly observed in the study area are giraffe, zebra, hippopotamus, impala, water buck, monkey, buffalo and warthog. These wildlife animals are mainly concentrated around the lake (riparian zone) and watering points.

3.8.4. Lake Naivasha

Lake Naivasha is the only fresh Water Lake in the rift valley province approximately covering an area of 13255 ha, i.e. 132.6 square kilometres. Fishery is another major activity next to agriculture where a considerable number of lives depend on. Main species of fishes of lake Naivasha include Tilapia zilii, Barbus amphigramma bigar, Louisiana red (Cry fish) and Procambarus clarkii (Harper et al., 1990). The products have good markets at Naivasha, Nakuru, and Nairobi City.

3.9. Stake Holders

Since Lake Naivasha area is economically, socially and politically becoming very important, the number of user groups in other words stakeholders is increasing in time. Some of the major stakeholders/groups involved in the management of the natural resources of the lake and its surroundings are the following (KWSTI, 1996).

- Ministry of Land Reclamation and Regional Water Department
- Water Resource Assessment Project
- Kenya Power Company
- Kenya Association of Tourism Organisation

- Horticultural Crops Development Authority
- Livestock Industry (Delamere Farm)
- Ministry of Environment and Natural Resources
- Naivasha Town Development Council
- Elsamere Conservation Centre
- Lake Naivasha Fishermen Co-operative Society
- Mirera Water project
- Commercial Farmers
- Lake Naivasha Pastoralists
- Kenya Wildlife Service
- Lake Naivasha Riparian Owners Association
- Universities and Colleges
- Research institutes
- Labourers of the Naivasha area
- Urban population of Naivasha town

4. Materials and Methods

The materials and methods chapter deals with the specific procedures and methods followed to meet the objectives of the research by solving the problems and answering the research questions that are stated under chapter 1. The chapter has been divided and presented in seven sections, which are materials and equipment, the research approach, the research methodology, information system development, data modelling, multi-source data integration and improving user access to soil information.

4.1. Materials and Equipment

The following materials and equipment were used while carrying out this research. These are:

1. Soil survey reports and research results
   - ♦ Detailed soil survey of a part of Quarantine Farm-National Animal Husbandry Research station, Naivasha. Report No. D45, 1988. Scale 1:5000, by P.T. Kamoni.
   - ♦ Soil conditions at Kulia Farm, Naivasha, 1980. Scale 1:10 000, by W. Siderius.
   - ♦ Detailed soil survey of the Nini Farm, Naivasha. Report No. D27, 1986. Scale 1:5000, by M.M. Gatahi.
   - ♦ Vulnerability of soils to change in agricultural use around Lake Naivasha. MSc. thesis, 1998. Scale 1:50 000, by C.P.H.J. Kwacha.
2. Topographic map of Kenya, 1975 Scale 1:50 000 (Index No. 133/2 Naivasha, 133/4 Longonot), by Survey of Kenya
3. Satellite images of Landsat, March 2000.
4. Aerial photos: year 1984 Scale 1:12 500 and Year 1991 Scale 1:10 000
5. Geological map of Longonot Volcano, the greater Olkaria and Eburru Volcanic Complexes and adjacent areas, 1988. Scale 1:100 000, by Clarke et al, 1988.
6. Garmin 12XL GPS receiver
7. Suunto directional compass and protractor
8. Measuring tape 50 m length, ruler
9. Suunto slope meter (Clinometer)
10. FAO 1990, Guidelines for soil description
11. Scanner, digitizer

4.2. The Research Approach

The methodological approach comprises the following three phases.

Phase 1. Pre-field work: this includes

- ♦ Exploring and collection of existing spatial and non spatial data from previous works about Lake Naivasha area
- ♦ Extensive literature review on titles related to the research mainly, on soil information system development methods and data models.
- ♦ Preparation of questionnaire for the needs assessment of selected user groups related to SIS applicability, about major management problems, about the required level of detail for organising the information (see Appendix A)

- Preparation of other different formats that are useful to register data or information for capturing spatial data during the fieldwork (see Appendix B).
- Preparation of preliminary database structure
- Set-up of soil information system design approach

Phase 2. Field work: this includes

- Identifying user groups around Lake Naivasha area
- Interviewing selected user groups and filling of the questionnaires
- Visiting KSS (which is responsible for Lake Naivasha area with respect to soil data collection analysis and management) to collect information on organisational structure and work tasks, inflow and out flow of data inside and outside the divisions.
- Defining the Universe of discourse or what we call system boundaries.
- Spatial data capture (compass traverse & GPS reading) for analogue digital conversion

Phase 3. Post-field work: this includes

- Analysis of the information gathered by the questionnaires on user need assessment
- Analysis of the information from the interview on KSS organisational structure
- Information system modelling
- Spatial and non-spatial data modelling
- Spatial and non-spatial data integration
- Designing friendly user interface to improve user accessibility
- Metadata documentation
- Thesis writing

4.3. The Research Methodology

The methodological approach of integrating/organising existing soil information for the Lake Naivasha area involves the following major activities.

a) Problem identification and planning to solve the problem
b) Selection of system design methodology
c) Information system modelling
d) Data modelling
e) Data integration

Further explanation on problem identification and planning, selection of design methodology, information system modelling, spatial and non-spatial data modelling and on data integration topics has been given in the following sections. The explanation also includes the reasons why the methods were selected from the others. The schematic flow diagram of the research methodology is shown on next three consecutive pages i.e. in figure 4.1, figure 4.2 and figure 4.3. The first figure shows the problem identification and method selection process while the second figure deals with information system modelling process. The third figure shows spatial and non-spatial data modelling and analogue-digital conversion processes.

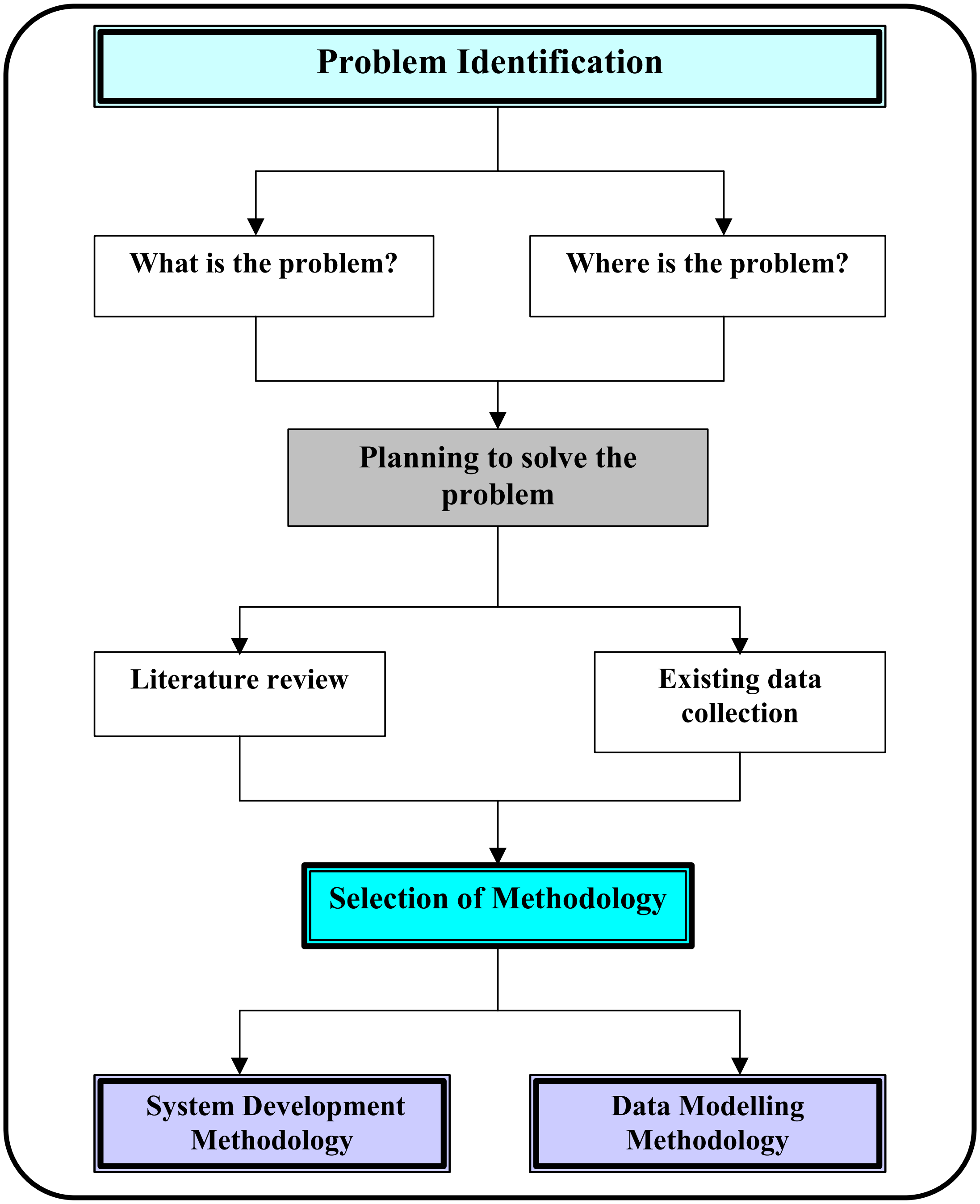


Figure 4.1 Schematic flow diagram of the Research Methodology (Part one)

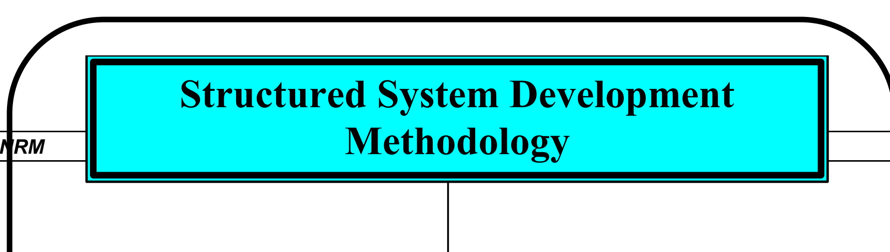


Figure 4.2 Schematic flow diagram of the Research Methodology (Part two)

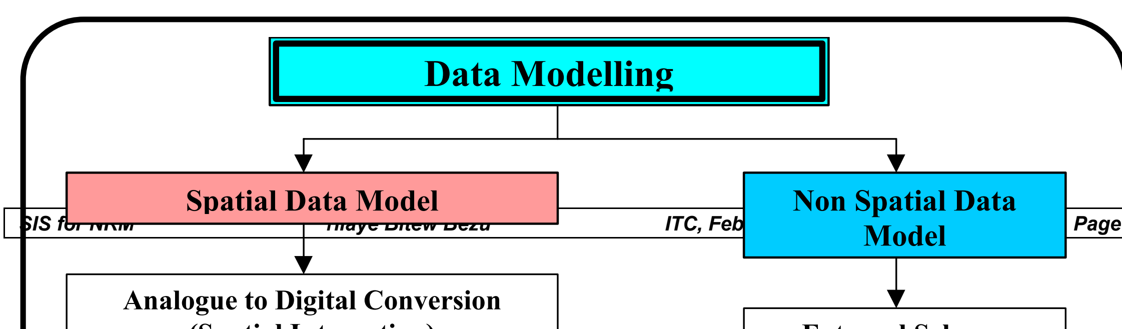


Figure 4.3 Schematic flow diagram of the Research Methodology (Part three)

4.4. Structured Information System Development Methodology

Structured system development methodology is one of information system development methodologies. To realise the design and introduction of a new information system for an organisation a certain methodology must be chosen (Krijger, 1995). According to Paresi and Essien (1992), a methodology defines a step-by-step approach in the process of development. Each step is related to a phase of the process and specifying what activities are involved, which techniques can be used and how one phase is related to other phases of the process in an information system development.

In this research a Structured System Development Methodology is used to develop a soil geographic information system. The justification for selecting this methodology includes the following reasoning.

1. It seeks to address the failure of conventional approaches when systems become more complex.
2. It has particular consideration to the reduction of the cost of rectifying changes in user requirements i.e. by its iterative nature any change in user requirement can be included at any stage of its design phases. Such changes would remain minimal if these requirements were considered extensively during the analysis phase.
3. In the present situation this methodology is well known, tested and widely used.
4. User participation is highly recognised in information requirement determination and verification at every stage or phase of development wherever necessary.
5. It is based on the life-cycle approach for project management, with clear and well-defined phases and activities.
6. It is iterative/evolutionary in nature where one can return back at any stage of development when the work is found incomplete on the previous phase (Paresi, 1999).

4.4.1. Information System Strategy and Planning

This phase involved studying goals, organisational framework, and operational environment of the intended user of the information system i.e. KSS with a view to determine the exact nature of the information system problem. Kenyan Soil Survey (KSS) is the organisational entity for soil data collection, analysis and information dissemination around Lake Naivasha area. For this study KSS is taken as the main data processor. At this phase questionnaires were prepared and used for the assessment of main system problems, system goals and system feasibility for realisation.

Problems of the system:

- The main problem of KSS in providing soil information is being unable to satisfy the needs of the users according to their needs (e.g. detailed soil survey results for private farms within a short period of time) due to traditional systems used in the organisation.
- Users are not able to get soils information according to their requirements.
- Weak link between the users and KSS (there is no well structured way for information dissemination)

Goals of the system:

- An inventory of information about soils, land and environment and their interrelationships with agricultural production for the whole country.
- Providing a service of survey or advice as required for specific purposes of development or research.
- Conducting research on the most effective methods for collecting and presenting soils information and on the characteristics of the main soils and environmental conditions.

Feasibility of the system:

- Even though a detailed investigation for feasibility study have not been conducted, it was clear and understood from the interview that Kenyan Soil Survey is planning to commercialise and bring itself to competition. Therefore, an efficient and improved, relatively faster system is in need to be established.

4.4.2. Information System Analysis

The main objective of this phase is analysing the existing system under review as a basis for an improved design. For this phase, the socio-technical SDM was more appropriate to analyse the system. Because the socio-technical approach is the one that recognises the interaction of technology and people and produces work systems that are both technically efficient and have social characteristics that lead to high job satisfaction (Paresi, 1999).

However, the case study area has no information system already existing. Therefore, the task was to analyse the information requirements for a new system. In order to plan for the new system a technique has to be decided upon how to determine data needs of the users. Davis and Olson, (1985) distinguished four different strategies for determining information requirements:

1. Asking or what we call interviewing
2. Deriving from an existing information system (data analysis approach)
3. Synthesising from characteristics of present system
4. Discovering from experimentation with an evolving information system

However, in practice combinations of these strategies are used in most of the situations. In this research also we have used the combinations of some of the above strategies. The short descriptions of the two methods used in this study in determining the information requirements are the following.

A. **Asking/interviewing:**
   Information requirements were obtained by asking the potential users of lake Naivasha area what their requirements are. This strategy was used during the fieldwork to collect most of the data by means of formal and informal interviews with the selected users (commercial farms, Urban planners and soil researchers are expected users of the soil survey results around Lake Naivasha area) by using a standard questionnaire. The information obtained from the interviews was used as a basis for the specification of data needs in designing the database to be coupled in the proposed information system. Different data requirements were identified for different group of users. These were structured and introduced in to the database.

B. **Deriving from an existing information system:**
   This method is mainly based on data analysis approach. The existing traditional information system of KSS with its organisational structure was thoroughly investigated. The result of this was used as an anchor in designing the new information system. Data inflows and outflows from the system and within the system i.e. between different sub-systems were studied. Different files were also studied to see how the current manual system is operating and where it could better be automated, what new processes are to be introduced and which processes are to

be merged so as to bring better efficiency in the system. The existing organisational structure of KSS is shown on figure 4.4.

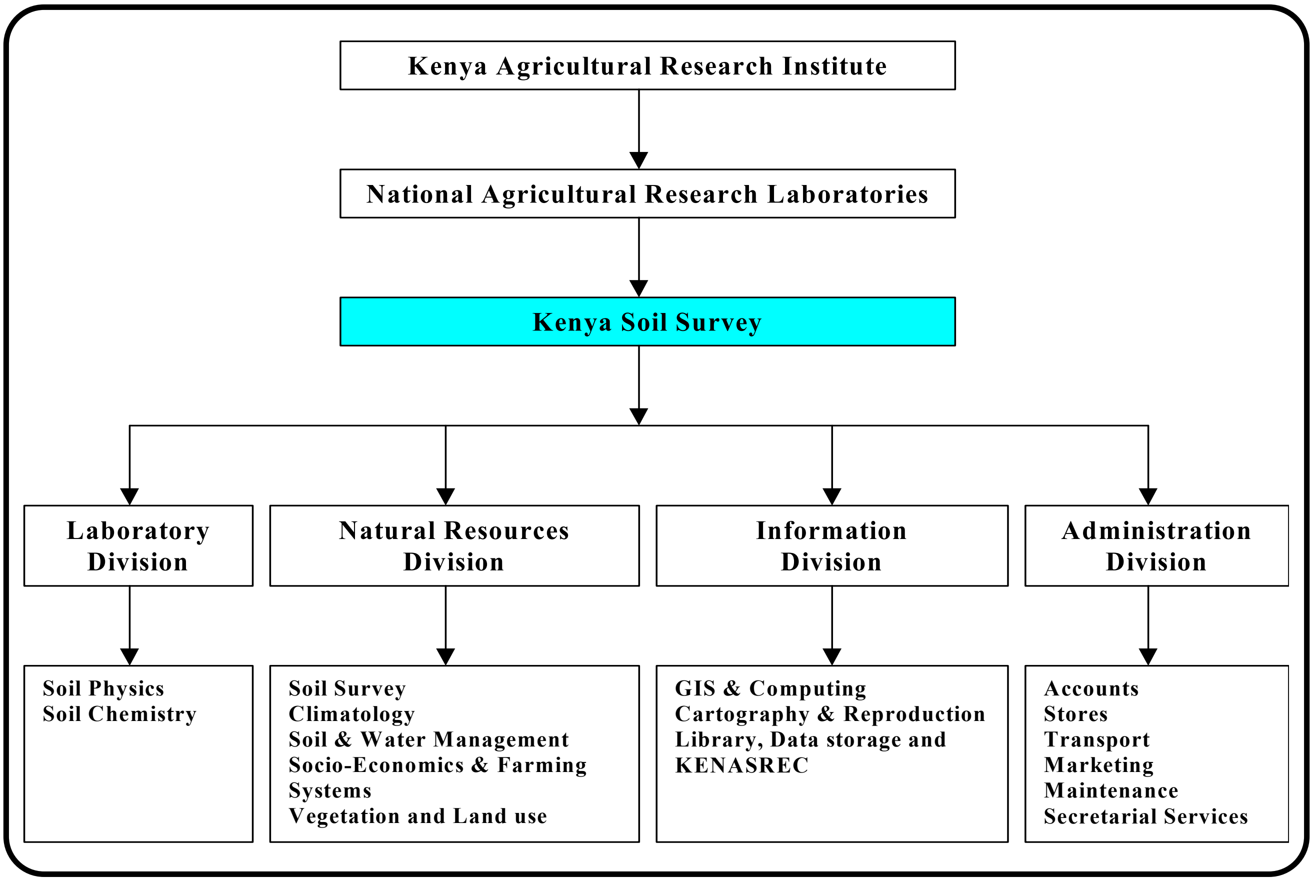


Figure 4.4 The existing Organisational Structure of Kenya Soil Survey (KSS)
(Source: KSS, 1998)

(**Note:** For the users' requirement assessment and analysis results refer to chapter 5 sections 5.1 & 5.2 specifically for the list of required some soil parameters see table 5.8 udder the same chapter)

4.4.3. Information System Design

System design aims at selecting the engineering design needed to implement the system, specifying changes to existing programs and databases, and specifying detailed user procedures that describe how they would use the system. Design is first made in the broad perspective (global design) i.e. context level and then in a more detailed form i.e. top level (Paresi, 1999). At this phase translation of information requirements (which are obtained from phase 2) in to information subsystems of the newly proposed system have been made based on the procedures required to produce the information. For the results of global and detailed system design refer to chapter 5 sections 5.2.2 & 5.2.3.

4.4.4. Information System Realisation

This step is the last phase in structured information system development methodology where by the following major activities are to be done.

- System Implementation
- System Installation
- System Operation and
- System Evaluation

Among the above mentioned major activities the last three are not to be discussed here. The reason is that these activities are beyond the scope of this research. Concerning system implementation, this phase involves creating of databases, acquiring hard/software, integrating the separately developed system components, documenting the system and train the staff of the organisation (Paresi, 1999). As far as software is concerned Micro-soft Access (Ms Access) was used to implement the relational attribute data model, Integrated Land and Water Information System (ILWIS) was used to handle the spatial data model and System Development Workbench (SDW) was used to model the information system. The relational data modelling has been discussed thoroughly further on the coming sub-chapters/sections. The proposed SIS architecture has been presented in chapter 5 section 5.7.

4.5. Data Modelling

For spatial and non-spatial data modelling, we usually find two approaches. The hybrid GIS and the integrated GIS approaches. The integrated approach is on the process of development where as the hybrid approach can be implemented in most of GIS softwares. The hybrid approach separates the spatial from the non-spatial data models where data integration is only possible at the level of data analysis. Even if, logically the integrated GIS approach is most likely to be preferable, the hybrid approach was used for data modelling in this study. This is because the integrated approach was not familiar to the researcher at the moment of study.

4.5.1. Non-Spatial Data Modelling

The "Entity-Relationship (E-R) Model, that is a popular high-level conceptual data modelling method" and the relational data model (De By, 1999) have been selected as most suitable data modelling methods to organise the existing multi-source soil data of lake Naivasha area and to improve user accessibility to the data/information. The entity relationship diagram modelling was done using Flow 4 charting software that is more suitable for such type of graphical representations, while the internal entity design was handled in Ms Access database management software. Some reasons for the selection of the relational data model over the others are:

- ♦ The model is based on a simple and uniform data structure – the relation
- ♦ There are many commercial DBMS packages that can implement it
- ♦ It is the most prevalent database paradigm

4.5.1.1. External Schema/Design

Schema is defined as a data view that is specified in a language, which database management system software is designed to understand (Howe, 1989). On the database design process the first step is user requirements collection and analysis i.e. external schema. An external schema describes the local view of the database required by an application. The expected out come of this phase is a listing of user needs or requirements. During the fieldwork three major soil information user groups were identified around Lake Naivasha area. These are:

- Commercial farms
- Urban planners and
- Soil researchers

The method used to assess the above mentioned users' requirements was asking, in a formal language interviewing. A standard questionnaire was prepared and used during interview. The interviewees were mostly people who are directly and in some cases indirectly involved in using soil data/information. One of the results of the interview containing the listing of user requirements of the aforementioned users is shown in chapter 5 table 5.8. The listing in the table does not fully contain the data that is captured in the database. For the full listing refer to appendix C section C2 that gives soil parameters list captured in the database per each entity type.

4.5.1.2. Conceptual Schema/Design

The conceptual schema is a concise description of the data requirements of the users and includes detailed descriptions of the data types, relationships and constraints. The high-level conceptual schema can also be used as a reference to ensure that all users' data requirements are met and that the requirements do not include any conflicts. This approach enables the database designers to concentrate on specifying the properties of the data without being concerned with storage details (De By, 1999).

At this phase the results of the external schema (user requirement list) were used as a basis to design the conceptual database by grouping them in to entities and attributes correspondingly. Consequently, six hierarchical entities (soil map, soil mapping unit, soil component, soil site observation, soil horizon and soil delineation) with extremely different size of attributes were identified.

Entity-relationship modelling:
Entity: An entity is a thing (object, concept) which the enterprise recognises as being capable of an independent existence, and which can be uniquely identified. For example, Soil Horizon is an entity.
Attribute: An attribute is a property of an entity. For example, pH is an attribute of soil horizon.
Relationship: A relationship is an association between two (or more) entities (Howe, 1989). For example, an entity Soil mapping unit is associated with an entity Delineation via a relation Contains.
There are two approaches for entity-relationship modelling so far identified. These are bottom-up data modelling and top-down data modelling, where the later one was applied for this research. The reason to select it was its simplicity to apply. The sequence followed for top-down data modelling was:

a) Selecting the entities and the relationships between them, which are of interest to the enterprise concerned.
b) Assigning attributes to these entities and relationships in such a way that a set of fully normalised tables is obtained.

E – R diagram is defined as a diagram or a representation which shows individual entity occurrences and their relationships (Howe, 1989). Similarly, De By (1999), defines **E – R diagram** as a method to show how the schema (conceptual structure) for a specific database application can be displayed by means of graphical notation. The conceptual model for this study has been proposed by integrating the models developed by Zinck & Valenzuela (1990), and Rossiter (1999). The E-R diagram model is more convenient and clear to represent entity relationships as compared to table occurrence and other models. The conceptual Entity Relationship diagram was constructed using Flow 4 software.

The conceptual database design is carried out based on an enterprise rule that governs the relationship between the entities. Of course, the conceptual model is independent of DBMS. For the enterprise rules and for the graphical representation of the conceptual design refer to the results and discussion chapter sections 5.3.1.1 & 5.3.1.2 respectively.

4.5.1.3. Logical Schema/Design

Both design steps (external design and conceptual design) are independent of the database management system (DBMS), where as logical and physical database design steps are DBMS dependent (Elmasri & Navathe, 1989). Most currently available commercial DBMS use an implementation data model, so the conceptual schema is transformed from the high-level data model in to the implementation data model. This step is called logical database design or data model mapping, and its result is a database schema in the implementation data model of the DBMS (De By, 1999).

Normalisation:
Normalisation is the process of minimising redundancy in the database. **Redundancy** is unnecessary duplication. Duplicate data is present when an attribute has two or more identical values. A data value is redundant if it can be deleted without information being lost (Howe, 1989). In the interests of logical simplicity a number of restrictions were applied to the entities (tables). These are:

1. The ordering of rows is not significant; i.e., the rows can be interchanged without affecting the information content of the entity (table).
2. The ordering of columns is not significant. We ensured that this is so by insisting that each column within a table has a distinct attribute type name.
3. Each row/column intersection contains a single attribute value. Multiple values are not allowed.
4. Each row in a table must be distinct; no two rows can have the same attribute values throughout. The significance of this rule is that quoting an appropriate combination of attribute values can always uniquely identify a row.

An entity (table) which satisfies the above restrictions (rules) is said to be a normalised entity (table). If not it is said to be unnormalised entity (table) (Howe, 1989). These normalisation rules were applied in the design of the logical schema in this study too.

The Five Normal Forms:
The normalisation process is often described in terms of stages known as first, second, third, fourth, and fifth normal forms which are symbolised as 1NF, 2NF, 3NF, 4NF & 5NF respectively. The general idea is that at each successive stages of normalisation certain undesirable features are eliminated from the initial unnormalised entity (table). Most commonly by most people, normalisation is used to be done up to the third normal form where the same was followed as most people do for this research. The five normal forms with short description are presented here.

1. 1NF = First normal form is obtained by eliminating repeating groups.
2. 2NF = Second normal form is obtained by eliminating non-identifier attributes which are not functionally dependent on the whole of the identifier.

3. 3NF = Third normal form is obtained by eliminating functional dependency between non-identifier attributes. The third normal form similar to Boyce/Codd normal form which states that, 'Every determinant must be a candidate identifier', but does not deal satisfactorily with overlapping candidate identifiers.
4. 4NF = Forth normal form deals with multi-valued determinancies.
5. 5NF = Fifth normal form deals with a rather unusual situation known as join dependency which is of little practical significance (Howe, 1989 & De By, 1999).

For this study the normalisation process was carried out up to the third normal form because the fourth and fifth normal forms have little importance. For the results of the logical design refer to chapter 5 section 5.3.1.3.

4.5.1.4. Internal Schema/ Physical Design

Physical database design is the last phase in database design, during which the internal storage structures and file organisation for the database is, specified (De By, 1999). On this study, ten entities and two associate/combination entities with a number of lookup (help) tables have been implemented. The entity code and entity name, with respective primary and foreign keys are listed in table 4.1. For more information on internal design specification of the entities see appendix C.

In physical design phase emphasis was given in obeying the basic data integrity rules which are very important for data consistency, retrieval and manipulation. The **entity integrity** constraint is specified on individual relations, where as the **referential integrity** constraint is specified between two relations and is used to maintain the consistency among tuples of the two relations.

Rules of data integrity:

a) **Key Uniqueness:** the primary key value of any tuple (record, row) in any relation instance must be different from that of any other tuple in the same relation instance. **A Primary key** is an attribute or a combination of attributes that uniquely identifies a tuple/row in a relation (relation here means an entity).

b) **Key integrity:** the value of any primary key attribute of any tuple (tuple here means record/row) in any relation instance should never be null.

c) **Referential integrity:** the value of a foreign key is either null (for all its attributes), or it is the primary key value of an existing tuple in the relation that the foreign key refers to. In a more plain language referential integrity constraint states that: a tuple in one relation that refers to another relation must refer to an existing tuple in the relation. **A Foreign key** is an attribute in a relation, which is used to link that relation with other relation(s) (De By et al, 1999 & De By, 1999).

Table 4.1 Entities/relations of the designed database with their corresponding keys

No.	Entity Code	Entity Name	Primary Key	Foreign Key
1	SMAP	Soil Map	Map_Id	Map_Id
2	SMU	Soil Map Unit	SMU_Id	SMU_Id
3	SPOLY	Soil Polygon	Poly_Id	SMU_Id
4	SMUCOMP	Associate/Combination entity	SMU_Id, COMP_Id	SMU_Id, COMP_Id
5	SCOMP	Soil Component	COMP_Id	COMP_Id

6	SOBS_SITE	Soil Observation at Site	OBS_Id	COMP_Id, OBS_Id
7	SOBS_HOR	Soil Observation at Horizon	OBS_Id, HOR_No	OBS_Id
8	SULMAC	Sulmac Farm lab-data (plot level)	Plot_code	SMU_Id
9	TSMUCOMP	Associate/combination entity	SMU_Id, TCOMP_Id	SMU_Id, TCOMP_Id
10	TCOMP	Top-Soil Component	TCOMP_Id	TCOMP_Id
11	TOBS_SITE	Top-Soil Observation at Site	OBS_Id	TCOMP_Id
12	TOBS_LAB	Top-Soil Observation at Lab	OBS_Id	TCOMP_Id

With the design of the entities there is mostly a need to design what we call help tables. Help tables are tables designed to contain domains that are later used by one or more entities. For this study such tables are named as lookup tables.

The Entity Design Window:

The entity/table design window consists of two areas. These are:

- The field entry area and
- The field properties area

The field entry area is for entering each field's name, data type and an optional description about the field name that will be displayed on the status bar while working in the datasheet view. The property area is for entering more options for each field, called properties. These include field size, format, input mask, alternate caption for forms, default value, validation rules, validation text, required, zero length for null checking, and index specifications. The design of the lookup tables is exactly the same as we do for the entities. The only difference is in the way we use the table (Prague & Irwin, 1997).

An example of the design view of one of the entities is shown in figure 4.5 next page. The field properties area in addition to the general specifications, it gives the option to link the lookup/help tables during the design process. This can be done after entering the necessary parameters for general properties, by clicking on the lookup button just to the right of the button for general. The final implemented database structure is presented in chapter 5 section 5.3.1.4. In addition to this data dictionary for entities, internal design parameters and specifications of the entities are presented in appendix C.

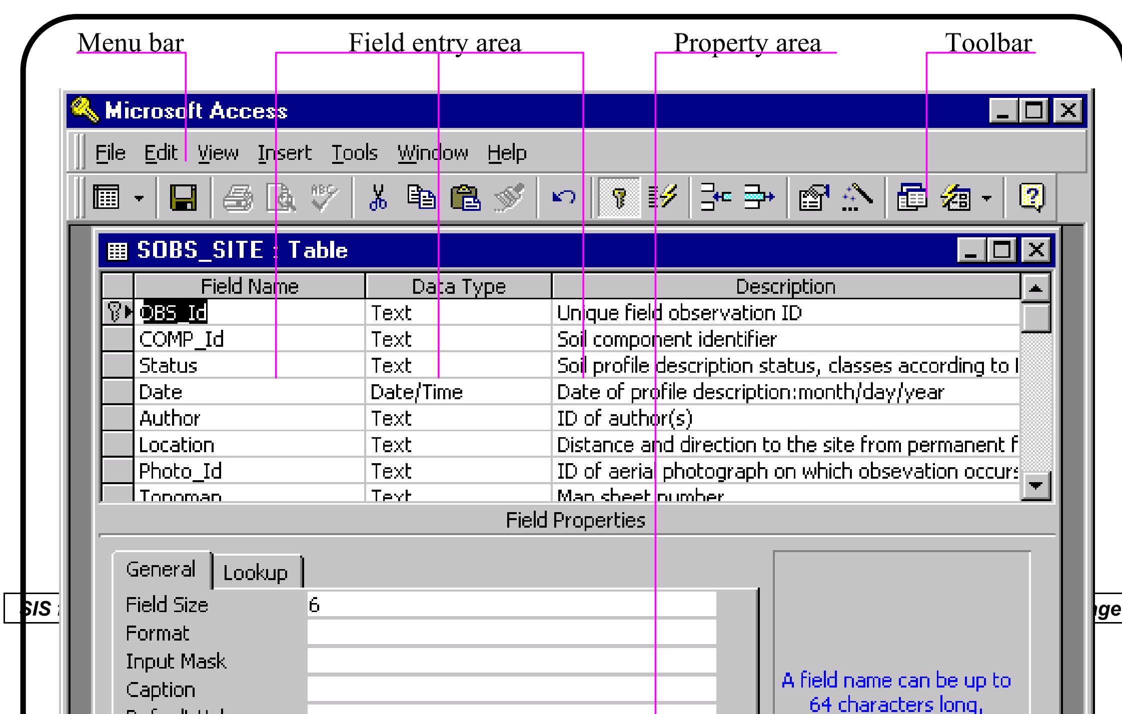


Figure 4.5 Soil Observation at Site (SOBS_SITE) entity/table design window

4.5.2. Spatial Data Modelling

A GIS database organises specific kinds of geographical data in a useful way. A database can be viewed as a representation of a relevant part of reality developed for some application. There are quite a few software dependent spatial database management systems employed for GIS because as there is no generally accepted solution yet (De By et al, 1999). For spatial data modelling first we differentiate between two concepts, spatial data models and spatial data types.

4.5.2.1. Spatial Data Models

As far as spatial data model is concerned, in spatial databases, we distinguish between vector and raster data models, and between zero, one, two, and three-dimensional primitives (geometric shapes). The data stored in one model can be converted in to representations of the other model using GIS. However, converting back to the first model may not always lead to a result that is identical to the original data. For this study both models were used where necessary. The vector model was used to handle desecrate (polygon) soil maps, for example map showing soil types; where as the raster model was used to represent continuous (interpolated) soil maps, for example map showing average pH.

- **The raster model:** this model subdivides space in to regular pieces, mostly square tessellations of dimension two (pixels) or three (voxels). These pieces are called raster cells, usually have no direct semantic bearing to the phenomena one wants to represent. A raster based system displays, locates, and represents geographic data by a matrix of cells. A unique reference co-ordinate pair is associated with each pixel, predefined either as one of its corners or as its centroid. In turn, each cell may have a thematic attribute value associated with it. The raster data resolution may vary from sub-meter to kilometres.

- **The vector model:** the vector approach does not apply a priori subdivision of space but uses the location, direction, orientation, boundary and/or shape of the objects of interest as the means to split up space. A vector based system displays geographic data as points, lines, curves, or areas, possibly with extra attributes. Cartesian co-ordinates, i.e., (x, y) pairs define points in a vector system. Lines or arcs are represented as a series of ordered points. Regions or polygons are also stored as ordered list of points, but by making the start and end points the same; the shape is closed and turned in to a two-dimensional feature. Vector data requires less computer storage space as compared to raster data.

4.5.2.2. Spatial Data Types

A data type is concerned with the way, how the data is stored in GIS. There are four spatial data types so far known up to now. These are tessellation data, point data, line data and area data. On this study soil profile description points were represented as point data, soil boundaries were represented as line data, soil types and terrain units were represented as area data in vector model and all of them correspondingly were represented as tessellation data in raster model in ILWIS software.

4.5.2.3. Spatial Data Dapture (Analogue-Digital Conversion)

The procedure and technique used for converting the existing analogue (sketch) soil maps of Kulia farm, Nini farm, Quarantine farm and Kwacha's soil map, in to the digital environment, for various advantages and reasons were the following. The nature and type of these sketch maps have been explained in detail in chapter 5 section 5.4.2.1.

During field work:
Condition one: For soil sketch maps (Kulia, Nini, and Quarantine) where enough ground truth points both on the sketch map and on the ground for taking GPS reading were not possible to identify, the following procedure was followed.

1. The location of the farm was traced by asking the residents of the area (Lake Naivasha).
2. Looking around the farm to compare its present position with respect to the sketch map in the report document.
3. Using the scale of the sketch map, converting the map distance in to ground distance and measuring the bearing by protractor.
4. Selecting a reliable start point (benchmark) that can be identified both on the sketch map and the actual ground, and then taking GPS reading.
5. Marking the point on the sketch map where GPS reading was taken.
6. Measuring the ground distance (as converted from the sketch map) and the bearing from the first start point to the next point. Taking GPS reading for the second point.
7. Marking the second point on the sketch map
8. Continue measuring the ground distance and bearing from the second point to the third point; from third point to fourth point and so on; … until the traverse is closed by coming back to the start point.
9. Continuing marking all the points on the sketch map where GPS reading was taken.
10. Registering all the necessary information on data collection format.

Condition two: For soil map (Kwacha's study area i.e. south east of the lakeshore) where enough ground truth pints (such as road junctions, road crossings) both on the map and on the ground for taking GPS reading were possible to identify, this procedure was followed.

1. The topographic map of Kenya at scale 1:50 000 was selected as a reference to find ground control points.

2. Identification of ground control points on the soil map and on the ground.
3. Taking GPS readings for all identified ground control points.
4. Marking all the GPS points on the analogue soil map and
5. Registering the corresponding co-ordinates for all identified points on the data collection format.

An example of tabular data that was used for spatial data capture during the fieldwork at a situation of condition one is shown in table 4.2. For the data capture formats see appendix B.

Table 4.2 Datasheet for recording spatial data captured by compass traverse and GPS measurements for Quarantine farm

Station	Forward Bearing In degree	Slope In %	Map Distance In (mm)	Ground Distance In (M)	X Co-ordinate IN (M)	Y Co-ordinate In (M)	Remark
1 – 2	268	4	36	180	215026	9917722	SE fence corner
2 – 3	268	2	39	195	214843	9917716	
3 – 4	268	1	41	205	214646	9917708	
4 – 5	268	0.5	20	100	214443	9917701	
5 – 6	268	0.5	10	50	214338	9917697	
6 – 7	349	0	53	265	214287	9917697	SW fence corner
7 – 8	349	0	61	305	214247	9917959	
8 – 9	76	0.5	16	80	214189	9918239	KWS gate
9 – 10	76	4	107	535	214266	9918276	
10 – 11	160	3	68	340	214772	9918416	NE farm corner
11 – 1	160	1	75	375	214892	9918083	

(Source: Field survey by the researcher)

Post fieldwork:

After enough information has been collected from the field for all available spatial data for Lake Naivasha area, the following procedure was followed to capture the digital soil data under the digital environment.

1. Scanning the sketch maps with GPS points marked on, with 300dpi, black and white for sketch maps and RGB for the colour maps, as Tif format, on A-3 size scanner.
2. Importing the scanned image in to ILWIS as .mpr format
3. Creating one common co-ordinate system based on the Geometric Definition of Topographic map of Kenya, scale 1:50 000, published on 1975.

Co-ordinate System Name = “Naiv”
Co-ordinate System Type = “Projective”
Co-ordinate System Boundary
Min X = 166 000; Min Y = 9889400
Max X = 221750; Max Y = 9972350
Projection = UTM
Datum = Arc 1960
Datum Area = Mean
Ellipsoid = Clarke 1880
Hemisphere = Southern
UTM Zone = 37 M

4. Creating geo-reference for each individual imported image
5. Geo-referencing the images/sketch maps using Geo-reference tie points option
6. After geo- referencing by screen digitising, creation of segment map that is indicating the soil boundaries.
7. Similarly by screen digitising creation of point locations indicating observation points.
8. Creating a domain for each segment map having polygon identifier.
9. Polygonsing the segment maps using the domain created above and editing the representations.
10. By creating different domains different attribute maps were calculated. For example, map showing soil types.
11. Creating a polygon histogram table, and taking the area and the perimeter to the non-spatial database and inserting in the soil polygon table. On the same way the SMU_Id was brought from Ms Access and inserted to the table associated to the map so as to have indirect link between the spatial data and the non-spatial data.
12. By synthesising the data stored in the non-spatial database and importing in to the spatial database different attribute maps were produced based on the users' requirements. For example, maps showing pH, nutrient level, organic carbon content and soil texture.

4.6. Multi-Source Data Integration

According to Oxford Advanced Learners' Dictionary of Current English, "integrate means to combine two things in such a way that one becomes fully a part of the other"(Hornby, 1995). Multi-scale data integration is one of the most important topics of research among which the issue is not yet resolved. Different scientists of whom the synopses of some of these ideas are mentioned below for comparison have followed different paradigms.

Even when data from adjacent sources can be successfully merged in to a common co-ordinate system, the problem of scale variability remains. Specifically, the number and detail of map features, such as rivers and roads, is a function of the scale at which they were surveyed and the scale at which they are to be displayed. The number of wiggles in a coastline, and hence the number of discrete points required to represent them, is far greater at large scales than at small scales. In practice a common solution to this problem is to store data for the same area at **different scales or levels of detail**. This results in a considerable data volume and data management overhead, due to the practical duplication of features from one scale to another. An alternative approach is to store only the most detailed, large scale version of the data, and to generalise with appropriate algorithms at the time of retrieval (Jones, 1985).

Input data for spatial databases span a large range of different scales. Topographic maps usually have a scale range from 1:5 000 to 1:200 000 and more. One of the many sources of error in spatial data processing is the use of data that come from different scales. It is not very wise to overlay a map of political boundaries that were digitised from 1:100 000 scale maps with parcels that come from 1:1000 scale maps. The results are boundaries that do not match and many sliver polygons. Spatial DBMS should warn the user when critical operations are attempted, or **even deny access** to data that should not be combined (Kainz, 1992).

"It is unlikely that any single database at a given scale could be developed to serve all conceivable uses. However, **nested databases** can be developed at different scales to address different requirements. By nesting the data, the information can be used at various scales without violating its integrity"(Dumanski et al, 1993). As stated above the ideas to the problem of multi-scale data integration are quite different. Therefore, a decision must be taken to choose an approach to solve the existing problem of data integration. The **nested database** concept was adopted from Dumanski et al (1993) to handle multi-source data integration for this study.

4.6.1. Non-Spatial Data Integration

Nonetheless, society perceives pressing problems predominantly at these smaller scales, and solutions for such problems must be developed. This requires that databases be developed at several scales, but with scientifically sound procedures to ensure the integrity and validity in the use of the data. The concept of **nested databases** was developed to address this dual problem of multiple scale databases while ensuring data integrity among the levels of the database. Geographic Information Systems with relational database management systems are normally used to manage these data (Kirkwood et al., 1983)

According to Dumanski et al. (1993), data congruence is the main task to consider in designing the relational database structure for data integrity. Even if most of the secondary soil data available for Lake Naivasha area are congruent, still there are some issues to be resolved before integrating the data. These are survey approaches and classification systems.

The method used to resolve the differences in classification systems and survey approaches was developing correlation tables for both of them. The correlation tables have been developed for FAO & USDA classification systems and for physiographic & geopedologic soil survey approaches. Based on these correlation tables the non-spatial data was integrated following the **nested database** approach. These correlation tables are presented in chapter 5 section 5.4.1.

4.6.2. Spatial Data Integration

According to Hornby, 1995 the term "nest" is defined as a group or set of similar things of different sizes made to fit inside each other. Detailed aspects of the quality of stored geometry in terms of the sources, accuracy and processing history are stored in the metadata-base (Jones et al, 1996), where as for this study metadata document was produced.

4.6.2.1. Geometric Discrepancies

Geometric discrepancies between spatial data sets may result from three major sources. These are:

1. Geometric error (for example, projection error, poor geometric control and interpretation error).
2. Different definitions of spatial objects (how to define a geographic object is one of the oldest problems in geography).
3. Differences in generalisation during mapping (interpreters will often have different experience and opinions when drawing a line or a polygon).

It is critical to identify the sources of the difference before the discrepancy is resolved (Wang and Howarth, 1994). In general, the geometric discrepancy can be categorised in to one of the two types of error (systematic error or random error).

Systematic Error: this is usually introduced by external sources (for example, projection error, poor geometric control, and other human errors).
Random Error: this is frequently called "Uncertainty" in the GIS literature, usually represents a spatial variation of the spatial object, particularly when the definition of the spatial object is fuzzy.
So understanding the sources of discrepancy should be the first step in spatial data integration process and then we look for possible solutions to alleviate the discrepancy. In this case study the spatial data were subjected to both types of errors. For the results see chapter 5 section 5.4.2.

4.6.2.2. Possible Solutions

After the source of geometric discrepancy or boundary mismatch is known the second step is looking for methods to resolve the mismatch. This can be done in a variety of ways, for example making use of statistical methods (detecting land cover changes); GIS techniques such as sliver removal, and remote sensing techniques, such as aerial photo and image reinterpretation.
1. Use of statistical methods (detecting land cover changes).
1.1. Kruskal-Wallace test: this test is used to compare more than two populations. The assumptions of the Kruskal-Wallace test are that the samples are drawn randomly from a continuous population, and the populations have approximately the same shape (Skidmore, 2000).

The null hypothesis is: Ho: $\eta 1 = \eta 2 = \ldots = \eta k$; where η is the median of the sample population.
The alternate hypothesis is: Ha: not all equal
This test have no direct result on soil delineations, but from the result of land cover change information we may be able to infer for soil boundaries.

1.2. Bayesian probabilistic method for change detection with area-class maps: the methodology proposed is based on the assumption that having identified either equivalent boundary segments, or the presence of sliver polygons, at some specified confidence level. The two maps will be conflated to produce a new map, which merges the equivalent items of geometry from the source map. For this method the whole area is expected to be surveyed/interpreted in both cases (Jones et al, 2000). In this case also there is no direct relation between the statistical results and the soil delineations, but we can infer from cover change information about the mismatch on the soil boundaries.

However, the land covers of the areas concerned have not been recorded at the times of soil survey (three is no any information on land cover). In addition to this the soil delineations are not closed and interpretation have not been done for the whole area of interest. So, it is not possible to use the statistical methods to resolve the soil boundary dispute among the maps mentioned above.

2. Sliver Removing:
For sliver removing method, the basic steps include: Gluing polygons, Area numbering/labelling, rasterising, Calculate the area and decide a threshold area below that will be removed, then run the removal with appropriate command. This holds true if the soil boundaries are closed and the interpretation covers the same area if possible all. But when we see the spatial data only the soil map made by

Kwacha has complete/closed polygons while all the others do not. So, this approach is also not useful for this particular case to solve the discrepancy.

3. Reinterpretation:
A research study on an investigation of uncertainty in field habitat mapping and the implications for detecting land cover change has found out that over all spatial correspondence between maps was found to be only 44.4%. A maximum of 14.4% of the total area surveyed was found to have undergone land cover change. The remaining discrepancies, equivalent to 41.2% of the total area, were attributed primarily to differences of land cover interpretation between surveyors (Cherrill and McClean, 1995).

According to the study of Cherrill and McClean (1995) about 55.6% of the survey area was subjected to boundary mismatch of which 26% of this discrepancy is due to real land cover change and 74% of the discrepancy is due to differences on interpretation/perception of the surveyors. The same is true for the discrepancy occurred on the spatial data of Lake Naivasha area. Therefore, for this research, **reinterpretation** with a reliable background (base map) has been taken as a solution to resolve the spatial boundary dispute and integrate the data. The procedure followed was the following.

1. An ortho-photo mosaic covering the study area was produced using aerial photos of year 1984 at scale 1:12 500 and year 1991 at scale 1:10 000 in ILWIS software. (Source: Mr Atiklt Girma who is doing his MSc research at ITC on soil phenoforms and genoforms 1999/2000)
2. The segment maps of Kulia, Nini, and soil map of Kwacha on the one hand and Quarantine and soil map of Kwacha on the other hand were overlaid together with different representations on the same photo mosaic as a base map for all of them.
3. The boundaries of each segment map were thoroughly investigated by comparing them to the common background map.
4. Soil boundaries nearest to the natural boundaries, terrain boundaries, vegetation boundaries etc. were taken as true boundaries and for this step the aerial photo interpretation rules by Goosen (1967) were applied.
5. For soil boundaries, which do not match to any one of the analysis elements on the background photo, were reinterpreted.
6. After finishing reinterpretation, by screen digitising an integrated new segment map was produced.
7. By polygonising the segment map, final integrated soil polygon map was created. The polygon identifier of the map was used to link the spatial data to the integrated/nested attribute database in Ms Access.

The results of the assessment of the nature of the maps to be integrated, over all discrepancies and reinterpretation map are presented in results and discussion chapter section 5.4.2. Also accuracy assessment of the soil maps has been computed using root mean square error (RMSE) method and the results are presented in the same chapter section 5.5.

4.7. User Access to Soil information

The last topic in relation to the methodological chapter of this study is to find out a method how to improve user access to the digital data set organised for Lake Naivasha area. The user access to the organised spatial and/or non-spatial data set can be improved by a variety of ways. Among these variety of ways two approaches have been selected for this study to improve user accessibility. These are:

- Providing enough information about the data stored and
- Making the path of browsing to the data set easier

These two approaches were realised by providing a well-documented metadata and by designing a user-friendly interface that can allow easy access to the data set required by the user. Both of the approaches are explained as follows.

4.7.1. Metadata

Metadata is data about data, which explains what has been contained in the data set? How the data set was made and organised? Who was the creator of the original and the present data set? What is the accuracy of the data set? How the data set can be obtained? Whom to contact about the data set? And when the data set was made? All these questions must be answered in a standard metadata document. Indeed, the above questions have been answered in this study too.

The metadata for this research was edited using "Tkme" version 2.3.4 metadata editor and compiled by "mp" version 2.4.10 metadata compiler. This software was developed by Peter N. Schweitzer, who is the staff member of United States Geological Survey (USGS) and is freely available in the internet. The metadata was compiled in two formats, which are notepad text and hypertext mark-up language (HTML) formats. The metadata with HTML format was developed as a question and answer and outline format and can be viewed using the Internet explorer, which is very suitable and easy for browsing. For a sample metadata document refer to appendix G.

4.7.2. User Interface Design

The second approach to improve user accessibility to the database is by designing a user-friendly interface. The purpose of the user interface is to make it easier for the user to access the system. The use of the system requires a certain amount of knowledge, which has to be distributed between the use and the user interface. Consequently, the greater the knowledge embedded in the user interface, the smaller is the knowledge that the user needs (Yifter, 1995).

Usually system design is made up of three activities; database design, user procedure design that includes the user interface, and program development. The goal of interface design is to provide the best way for people to interact with computers, or what is commonly known as human computer interaction (HCI). Provision of good interface is becoming more important because of its impact on most organisations. Many people believe that improving interaction between people and computers is one of the most important activities in system design. People are no longer interested in the technology behind the computer; they simply want a tool that is easy to use and can help them with their problems. They don't want to spend a lot of time learning about computer software; they just want computers to make their own work easier. A good interface certainly helps to satisfy this goal (Hawryszkiewycz, 1998).

In the process of user interface design we have two options. The first one is to choose the transactions in the process to be supported by interfaces. This will define the broad interface requirements in terms of what information is input and output through the interface during the transaction. The second option is the design of the actual screen presentation, including its layout, and in fact the sequence of screens that may be needed to process the transaction (Hawryszkiewycz, 1998).

The user interface design for soil information users of Lake Naivasha area follows the second option, i.e. the use of menus/actual screen presentation. The first advantage of menu/screen presentation is that it provides a familiar format and a clear set of choices, which are well understood by the user. The second advantage is that it reduces learning time and the probability of making mistakes and it is easy to learn and more user friendly. Of course, the disadvantage is its inefficiency for the application programmers and the danger of having too many menus/screens if their design is not carefully thought over, which can lead to confusion and the sense of being lost within the system.

The interface design was implemented on Ms Access 97 software where the non-spatial data set has been organised. The user need assessment from the selected users (commercial farms, urban planners and soil researchers) has been used as a basis to the start of the interface design. The schematic representation of the flow of the user interface is shown in chapter 5 section 5.6.2 figure 5.11. The general step followed for the interface design process is the following.

1. Selection of attributes according to the specific user needs
2. Querying the attribute values relevant to the specific user from the database, i.e. by making use of structured query language (SQL) and query by example (QBE) and the development of a query library
3. Development of reports from queries which are going to be viewed by the user at the final destination of the screen navigation
4. Development of forms or what we call screens for navigation with pop-up buttons containing clear and legible statements about the information that the user will obtain after clicking the pop-up button
5. Design of macros to guide, order and control the navigation process
6. Linking macros to reports and/or queries and to navigation screens via pop-up buttons

4.7.3. Query for Information Retrieval

The word query comes from the Latin word "*quoerere*", which simply means to ask or inquire. Over the years the word query has become synonymous with quiz, challenge, inquire or question. Therefore, we can think of a query as a question or inquiry posed to the database about information found in its tables (Prague & Irwin, 1997).

Types of queries:
Ms Access supports many different types of queries. They can be grouped in to six basic categories. These are select, total, action, crosstab, SQL and top (n). The select query selects information from one or more tables (based on specific criteria) and displays the information in a dynaset that we can use to view and analyse. Total queries give us the capability to sum or produce totals (such as count) in a select query. Action queries let us to create new tables or change data (delete, update, and append/add) in existing tables. In action queries, changes can be made to many records during a single operation (Prague & Irwin, 1997).

Crosstab queries can display summary data in cross-tabular form like a spreadsheet with row and column headings based on fields in the table. Concerning the structured query language (SQL) query, it is used for advanced SQL databases manipulation. We can create these queries only by writing specific SQL commands. Top (n) query limiter lets us specify a number or percentage of the top records we want to see in any type of query (Prague & Irwin, 1997).

The query types are mostly interrelated to each other. Most of them are used in combination while performing a single query. Results of sample queries using SQL have been presented in chapter 5 section 5.6.3.

5. Results and Discussion

This chapter deals with results and discussion part of the thesis. Results are generally statements of outcomes of a scientific test when we speak of scientific research, while discussion is an explanation with respect to the results obtained. The results of this research are geared to address the problems and objectives of the study, which are stated at the beginning of this document in chapter 1 sections 1.3 and 1.4. From this standpoint the under mentioned outcomes have been obtained.

Of course, most of the cases were thoroughly discussed in the methodology chapter. For instance the methodological approach and working procedures for the design of SIS mainly information system planning and information system analysis have been well discussed in chapter 4 sections 4.4.1 & 4.4.2. Under these two sections the major system problems, system goals, system feasibility, techniques to determine data needs and the existing organisational structure of KSS has been presented.

The results and discussion chapter then deals with outcomes related to summary of user need assessment, which is the foundation for information system specification (section 5.1), information system design (section 5.2), and soil geographic database design (section 5.3). Furthermore, multi-source soil data integration process, map accuracy assessment for spatial data quality, a proposal to improving user accessibility to soils data, and finally the proposed soil information system architecture have been presented under this chapter in sections 5.4, 5.5, 5.6, and 5.7 respectively.

5.1. Analysis of User Need Assessment

The user need assessment section deals with the results of the inquiry made to the potential soil information users of lake Naivasha area and to Kenya Soil Survey organisation who is the proposed soil information supplier. The main topics of this section are summary of responses to user need assessment, preference of data format, map scale and management area, the reason behind for the need of soils data/information, and most commonly required soil parameters.

5.1.1. Summary of Responses to User Need Assessment

The population of the interviewees was that of the potential users of the soil information around Lake Naivasha area. Similarly the proposed producers of the required soil information were also targets of the assessment. The total number of the interviewees was 21 from which 17 of them are key persons representing the information users' environment while 4 are from the producers' environment.

The results of the user need assessment that was conducted during the fieldwork are summarised in table 5.1. The table also shows the proportion of the responses in percentages. In general 94% of the respondents need soil information in the future, have an experience in using soils data and they are familiar with the digital environment (computer system).

The study showed that most of the users prefer field observation and lab data at plot/field level as compared to at point level i.e. 82% versus 41% of the cases. On the other hand about 88% of the respondents have an experience in interpreting raw soil data and in the future they need both raw and processed data so that they can compare and improve their skills of soil data interpretation.

Concerning the preference of soil map data 53% of the cases prefer both types of maps while 24% of the cases prefer conventional map data and 18% of the cases prefer interpolated (continuous) map data. The need of soil information as map data is low as compared to tabular data in the study area. This is because the users have a very little experience in using soil map data i.e. only 29% of the respondents have experience in using soil map data.

Table 5.1 Summary of the responses of the interviews to selected soil information users

No	Parameters that were considered for assessment	Response		Percentage	
		Yes	No	Yes	No
1	The user have experience in using soils data	16	1	94	6
2	The user is receiving soil data in map form only	0	17	0	100
3	The user is receiving soil data in point/plot data form only	12	5	71	29
4	The user is receiving soil data in both map and point data form	5	12	29	71
5	The user have an experience in interpreting soil data	15	2	88	12
6	The user have an experience in using computer	16	1	94	6
7	The user shares the soil data with others	6	11	35	65
8	The user needs soil data in the future	16	1	94	6
9	In the future the user need raw soil data only	0	17	0	100
10	In the future the user need processed soil data only	2	15	12	88
11	In the future the user need both raw and processed soil data	15	2	88	12
12	The user prefers conventional map data only	4	13	24	76
13	The user prefers continuous map data only	3	14	18	82
14	The user prefers both conventional and interpolated map data	9	8	53	47
15	The user prefers field observation and lab data at point level	7	10	41	59
16	The user prefers field observation and lab data at plot/field level	14	3	82	18
17	In the future the user needs the soil data in digital format only	4	13	24	76
18	In the future the user needs the soil data in analogue format only	2	15	12	88
19	In the future the user needs the soil data in both data formats	10	7	59	41
20	The primary choice of the user (+KSS) for information exchange is digital over that of the analogue/traditional system	15	6	71	29
21	The primary choice of the user (+KSS) for information exchange is analogue/traditional over that of the digital system	5	16	24	76

The user need assessment also revealed that most of the existing soil information around Lake Naivasha is obtained by sending soil samples abroad mainly to Holland and receiving the analytical results through email, fax and postage. From the organisations interviewed 46% of them completely depend on raw and interpreted soil data from abroad, 27% both from abroad and their own lab analysis, 9% from abroad and KSS, 9% from KSS and Ministry of Public Works and 9% from their own field observation and lab analysis. From this we can conclude that the proportion of soil information users around Lake Naivasha area that depend on soil data from abroad and their own lab constitutes 73% from the total users population (mainly commercial farms). This summary is obtained from table 5.2 below indicating the existing sources of soil information and communication systems used with respect to each organisation.

Table 5.2 Existing sources of soil information & means of communication around Lake Naivasha

Name of Organisation/Farm	Existing Sources of Soil Information	Type of Communication
Longonot Horticulture	Holland	Email, Postage, Fax
Longonot Farm	Holland	Email, Fax
Naivasha Town Municipality	KSS, MoPW	Postage
Home-grown Flamingo Farm	United Kingdom, Own lab	Email
Kijabe Limited	Holland	Email
Sher Agencies	Holland	Email, Postage, Fax
Oserian Development Company LTD	Holland, Own lab	Email, Postage, Fax
Three Point Farm	Holland, KSS	Email, Postage
Soil Research Offices	Field Survey	Field Observation
Panda Flowers Company	Holland	Email
Sulmac Company Limited	Holland, South Africa, Own lab	Email, Fax
Note	KSS = Kenya Soil Survey, Nairobi, Kenya	
	MoPW = Ministry of Public Works, Nairobi, Kenya	

On the other hand, the analysis of the existing conditions and major problems on soil information producer's side (KSS), showed that in most of the cases there is a data gap between what the user needs and the actual data supplied by KSS. These data gaps include for example, map scale, soil physical characteristics such as bulk density, soil moisture, and soil temperature. The reasons for the data gap are:

1. The Kenyan soil survey does not carry out formal user need assessments regularly before planning for soil data collection (except for two occasions i.e. in 1995 & 1996)
2. The Kenyan soil survey is biased towards fertility status/appraisal tasks as compared to other soil physical parameters.
3. High cost and time constraint to conduct detailed and complete survey and
4. Skilled man power and financial shortage

Hence, the survey mapping scale is decided up on KSS independent of the minimum decision or management area (MDA) that is identified by the user. Therefore, from this situation we can conclude that the link between the information user and Kenyan soil survey organisation is very weak.

The main purpose of soil data collection by KSS is for soil suitability assessment for rainfed and/or irrigated crop cultivation, for rangeland management, for construction, for farm planning and for multipurpose and specific land use planning. The over all comment suggested by the respondents to improve the KSS organisation's performance includes aggressive marketing of products and services, full commercialisation of the organisation, structural improvement, more dedication and hard work. The synopsis of the parameters considered for the assessment and the responses obtained are summarised in table 5.3.

Table 5.3 Synopsis of the responses to the interview to KSS

No	Parameters considered for the assessment	Responses
1	Most frequently required soil data	Fertility appraisal, soil physical & chemical, present land use, and soil type data
2	Major sources of primary soil data	Field observation and samples from clients

3	Methods used for soil data acquisition	Field survey and remote sensing
4	Most common users of the soil data	Small farmers, commercial farms, extension workers, NGOs, private companies, GOV organisations, international institutions, researchers & consultants
5	Experience in conducting user need assessment, when, how, to which user	Only twice i.e. in 1995 & 1996, through workshops and questionnaires to farmers and agricultural extension staff
6	Experience in using MDA for map scale definition	No experience for MDA assessment for soil map scale determination
7	Presence of data gap between the supply and user requirement	Yes there is data gap. For example, mapping scale, soil physical characteristics such as bulk density, soil moisture, soil temperature
8	Reasons for the data gap	There was a bias towards fertility appraisal , high cost and time constraint for detailed and complete survey
9	Data handling/storage format	Both digital and analogue, but in most cases analogue format
10	Ways of access to the soil data	Through library service, by marketing soil report documents and thematic maps
11	Problems encountered in data acquisition	Skilled man power & financial shortage
12	Problems encountered in data handling	Shortage of trained man power digital data
13	Problems encountered in information dissemination	Shortage of funding for mobility and skilled man power
14	Over all comment/suggestion to improve the organisation's performance	Aggressive marketing of the products and services, full commercialisation of the organisation, organisational structure improvement, more dedication and hard work

5.1.2. Preference of Data Format

One of the research questions dealt within this study is the preference of the data format (digital or analogue) for information exchange/dissemination between data users and producers. The data used for the validation of the preference was collected through interviews using a standard questionnaire. The interview data was analysed statistically using two-way tables and chi-square test to prove whether there is a significant relation between the data format and choice of the format by users or not. The data set and the analytical results are shown on tables 5.4, 5.5, and 5.6 below.

The population considered for sampling was the potential soil information users of Lake Naivasha area only. The sampling method was a systematic random sampling approach with 20% of the sampling units from soil information producers' environment and 80% of the sampling units from the direct soil information users' environment. The statistical test was done for a 95% confidence level.

Question: Is digital environment of soil information system preferred from that of the Analogue?
Hypothesis:

Ho: There is no relation between data format and user preference

Ha: There is relation and user preference is for a digital data format

Table 5.4 Absolute frequency of the data set

Data format type	Response to preference		
	Yes	No	Total
Digital format	15	6	21
Analogue format	5	16	21
Total	20	22	42

Table 5.5 Relative frequency of the data set

Data format type	Response to preference		
	Yes	No	Total
Digital format	71.4	28.6	100
Analogue format	23.8	76.2	100

Table 5.6 Summary of statistical values

Parameter	Value
Significance level (α)	0.05
Chi-square critical value at α = 0.05	3.84
Chi-square statistic (X^2)	9.545
Degree of freedom (df)	1
Probability (p) value	0.002

Conclusion: (1) The statistical test revealed that there is a significant association/relation between data format type and user preference (X^2-test; X^2 = 9.545; DF = 1; α = 0.05; p =0.002). (2) From the relative frequency it is obvious that the digital environment is preferred as compared to the traditional (71.4% versus 23.8% of the cases). Therefore, we are 95% confident that the digital format is most likely to be preferred as compared to the analogue format for soil information interchange. The reasons for the preference of the digital format as replied by the user are less storage requirement, favourable for analysis, retrieval and updating, easier for transportation and exchange.

5.1.3. Map Scale and Management Area

The mapping scale for spatial data presentation greatly depends on the minimum decision/ management area (MDA) that is handled by the user. The results of the interviews showed that the map scale preferred by the users lie between 1:5000 – 1:27386. The map scale was calculated based on the Cornell's definition of minimum legible delineation (MLD), where it is defined as = 0.4 cm^2 or 40 mm^2 on map regardless of the scale used (Rossiter, 2000a). Therefore, a map should be prepared at a scale where the minimum legible area (minimum decision area) is at least 0.4 cm^2. The management area and the required map scale for soil data around Lake Naivasha is summarised in table 5.7.

The values of minimum decision area (MDA), maximum decision area (MaxDA) and the total farm size that are included in the table are those values provided by the interviewee. The required map scale was calculated using the mathematical formula:

$$\text{Scale Number} = [\sqrt{(\text{MDA (in ha)} \times 250)}] \times 1000$$

Table 5.7 Required map scale and management area of Lake Naivasha area

Farm/Organisation Name	MDA (ha)	MaxDA (ha)	Total farm size (ha)	Map scale required
Longonot Horticulture	0.25	1	100	1:7905
Longonot Farm	0.1	0.5	33	1:5000
Naivasha Town Municipality	0.3	28	1080	1:8660
Home-grown Flamingo Farm	0.25	1	114	1:7905
Kijabe Limited	0.25	13	40	1:7905
Sher Agencies	0.5	35	400	1:11180
Oserian Development Company Limited	1.2	20	110	1:17320
Three Point Farm	3	6	297	1:27386
Panda Flowers Company	0.5	0.5	10	1:11180
Sulmac Company Limited	0.4	3.5	1000	1:10000

When the required map scales in table 5.7 are grouped in to scales of multiples of five, it can be concluded that among the soil data users of Lake Naivasha area over 90% of them need a detailed soil map (i.e. large-scale map). The percentages that are shown below were calculated by dividing the counts satisfying the stated condition to the total count of organisations. Hence:

- 60% of the users need soil map at scale >= 1:10 000
- 80% of the users need soil map at scale >= 1:15 000
- 90% of the users need soil map at scale >= 1:20 000
- 100% of the users need soil map at scale >= 1:30 000

5.1.4. The Need for Soils Data/Information

The reasons why the users need soils data differ according to the purpose of the data to be used. The three-selected user groups (commercial farms, urban planners and soil researchers) of Lake Naivasha area have given different reasoning during the interview, which was conducted for the user needs assessment in the field. The major ones are the following.

Commercial Farms:

- For fertiliser and irrigation water application programming
- For soil suitability assessment for crops and control of soil born diseases
- To know macro and micro-nutrient levels in soil
- To know effects of fertiliser on water quality
- To know the soil fertility depreciation

Urban Planners:

- To know soil limitation factors for building and construction
- To know the load bearing capacity of the soil
- To know the effluent absorption capacity and ease of excavation
- To plan for different engineering purposes such as house and road construction

Soil Researchers:

- To classify and map soils, for wind erosion assessment
- To conduct research on different soil parameters/characteristics
- To give advice to planners and farmers about soil status and possible potential uses
- To carry out research at the level of soil series and soil phenoforms

5.1.5. Soil Parameter/Characteristic Requirements

One of the results of the user need assessment is a list of soil characteristics or parameters that are most important and commonly required by the users. Of course, the user need is not a static element rather it is dynamic in nature. Therefore, the soil parameters list which are shown in table 5.8 are those parameters collected by the interview which are parts of the broad dynamic user needs. These data are used in information system design mainly for soil geographic database design.

Table 5.8 Most commonly required soil parameters list

Comercial Farms	Soil Researchers	Urban Planners
Electrical conductivity (EC)	Electrical conductivity (EC)	Engineering soil classification
pH (soil acidity, alkalinity)	pH (soil acidity, alkalinity)	- AASHTO group index
Soil texture	Soil texture	- Unified classification
Cation exchangeble capacity	Cation exchangeble capacity	Particle size distribution
Organic carbon content	Organic carbon content	Atterberge limits
Bulck density	Bulck density	- Liquid limit
Field capacity	Consistency	- Plastic limit
Soil structure	Soil structure	- Shrinkage limit
Soil type	Soil type	- Plasticity index
Nitrate, Nitrite levels	Soil colour	Relative compaction
Infiltration rate	Infiltration rate	- Field bulck density
Water holding capacity	Gypsum	- Laboratory bulck density
Percolation rate	Clay minerals	California bearing ratio test
Ion constitution of the soil	All sand separates	Permeability
Moisture content	Clay percentage	Shrink-swell potential
Na (sodium) level	Na (sodium) level	Soil collaps
Ca (calcium) level	Ca (calcium) level	Soil resistence
N (nitrogen) level	N (nitrogen) level	pH (soil acidity, alkalinity)
P (phosphorus) level	P (phosphorus) level	General site conditions
K (potassium) level	K (potassium) level	- Slope
Mg (magnesium) level	Mg (magnesium) level	- Flooding
Mn (manganse) level	S (sulfur) level	- Depth to bed rock
B (boron) level	Al (aluminium) level	- Depth to ground water
Fe (iron) level	Fe (iron) level	
Phosphate levels	Penetration resistance	
Porosity	Organic matter content	
Soil compaction	Bicarbonates content	
Permeability	Ammonium content	
Organic matter content	Histirical data of the soil	
Bicarbonates content	Soil microbial analysis	

(Source: Interview to the users by the researcher)

5.2. Information System Design

The process of information system design most commonly traverses four main phases. These are information system strategy and planning, information system analysis, information system design and information system realisation. The first two phases have been thoroughly discussed in chapter 4 section 4.4. Therefore, in this chapter our focus will be on the third phase that is system design phase.

Under information system design phase we have the global system design and detailed system design stages. These two stages of system design and system requirements are presented as follows.

5.2.1. System Requirements

From the analysis of the interviews and from the existing traditional information system of KSS, it was found out that the major system requirements to process and produce the basic soil information as per the needs of the users include:

- Aerial photo and satellite image interpretation for soil mapping
- Field data collection (primary and/or secondary data) and archiving
- Laboratory analysis of (chemical and physical) collected samples
- Making soil interpretations based on the data collected from the field and from laboratory results
- Putting all the soil data and information available in to GIS environment.
- Disseminating the soil information to the users as per their need.

5.2.2. Global System Design

The main objective of global/broad perspective design is to design the system to a level at which sub-systems can be designed separately, to design interfaces (data stores) between sub-systems to define a modified organisational structure, and to define the system boundaries or what we call terminators or universe of discourse. Organisational boundary indicates the limit of the system such that beyond it the system has no mandate. From the data analysis it was found out that the universe of discourse includes:

1. Commercial farms
2. NGOs & Int. organisations
3. Consultants
4. Private companies
5. Decision-makers
6. Govt ministries
7. Research officers
8. Irrigation schemes
9. Universities and colleges
10. Farmers
11. Survey of Kenya
12. Regional Remote Sensing (RS) Centre
13. Department of Resource survey and RS
14. Public libraries
15. Mines & Geology department
16. Meteorological department
17. Urban planners
18. International Livestock Research Institute

In designing the context level system process model with its system boundaries KSS Soil Information Provision is defined as the main process model in the system. KSS is taken as the core of the proposed soil information system where at the end of the day system realisation is expected to take place. Among the terminators mentioned above the database design was done based on the requirements of three users. These are commercial farms, urban planners and research officers. The reason for the selection of these three users is that these are the ones that are in need of soil data/information around Lake Naivasha area. The context level data flow diagram of the system, which is shown on figure 5.1, however, includes all the terminators that have a link with KSS in soil information provision process.

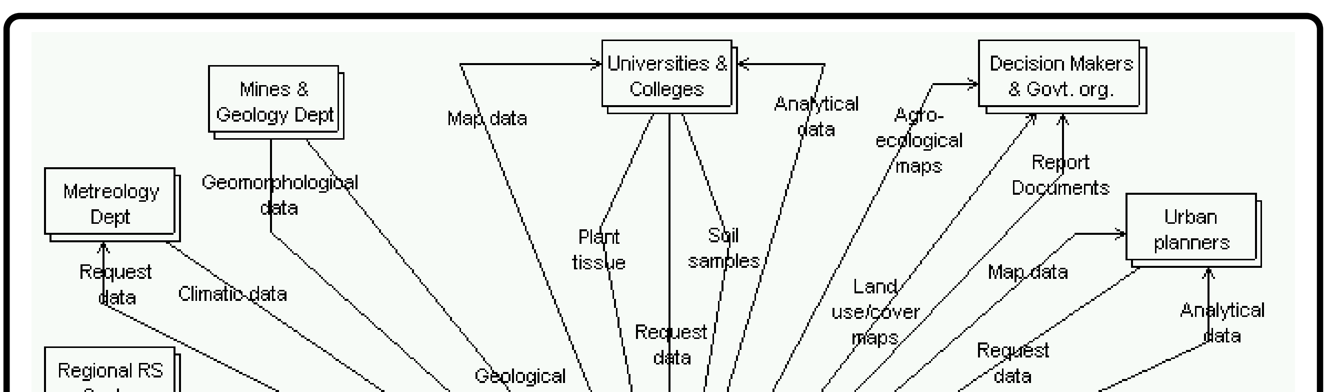


Figure 5.1 Context-level data flow diagram

5.2.3. Detailed System Design

The main objective of this stage is to refine designs and specifications to a level that permits to start implementing specific system components and to propose a detailed new/modified organisational structure. Therefore, at this phase detailed consideration of the system at a level of sub-systems was taken in account to design top-level data flow diagram. Prior to the start of working on the top-level diagramming, a thorough investigation was made on the organisational structure of KSS. The existing structure with slight modifications was used to propose a new structure, where this forms the corner stone to define the sub-systems.

The slight modification on the structure was done solely to suit the handling of the soil data in the system with clear and simple structure. The modified and simplified organisational structure is shown in figure 5.2. (Note: In a real case of system design the proposed/modified structure should be discussed with the organisation i.e. KSS before implementation. In the context of this study this was not done).

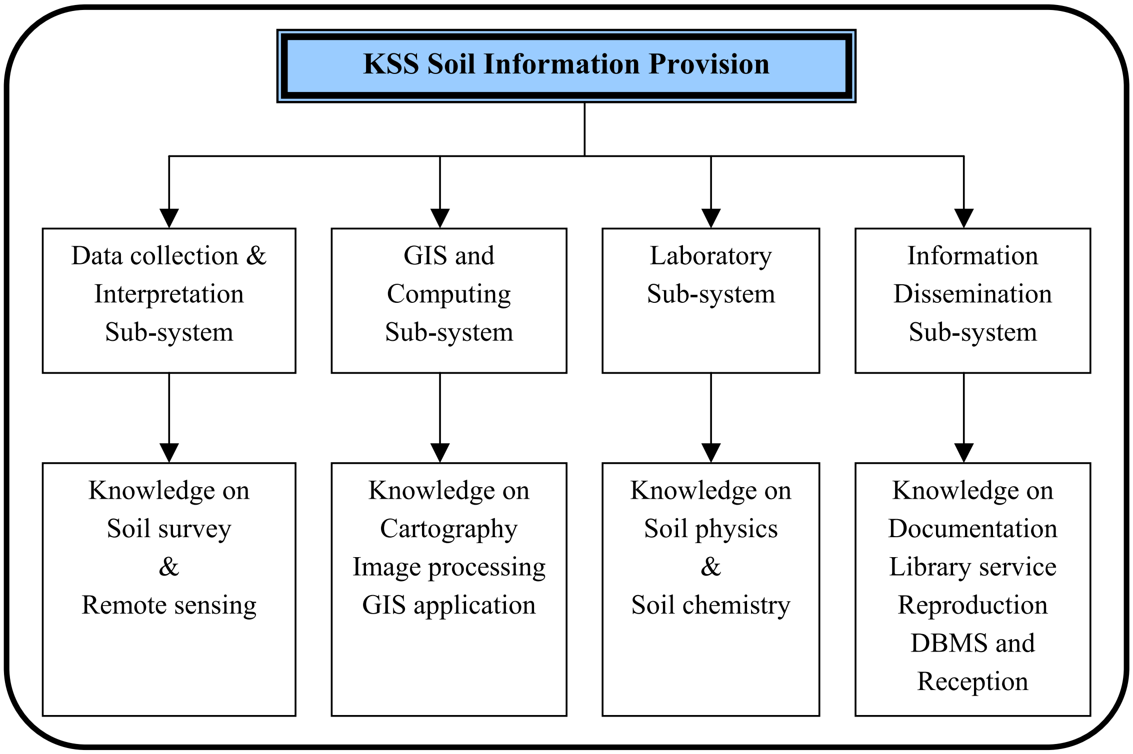


Figure 5.2 Modified and simplified organisational structure for soil information provision

On the structure shown above, a new sub-system namely Information Dissemination has been introduced to the system. This is because information dissemination sub-system would play a role in administering and channelling the information going out from the system and the request data coming in to the system. It would also be easy for the users whom to contact and where to get the information ready. This was one of the main problems in the existing traditional information system of KSS.

Based on the modified organisational structure the detailed top-level data flow diagram was designed. The top-level diagram contains four process models (data collection and interpretation, laboratory, GIS and computing and information dissemination sub-systems) and four data stores (survey data registry, lab data registry, final data registry and request data registry). The data stores are used as interfaces between the sub-systems. The schematic design of data flow in to the sub-systems, out of the sub-systems, in to the data stores, out of the data stores and the mandate to update the data store is shown in the figure 5.3. For data dictionary of process models & data stores and for data inflow & out flow with respect to the selected user groups refer to appendices D.

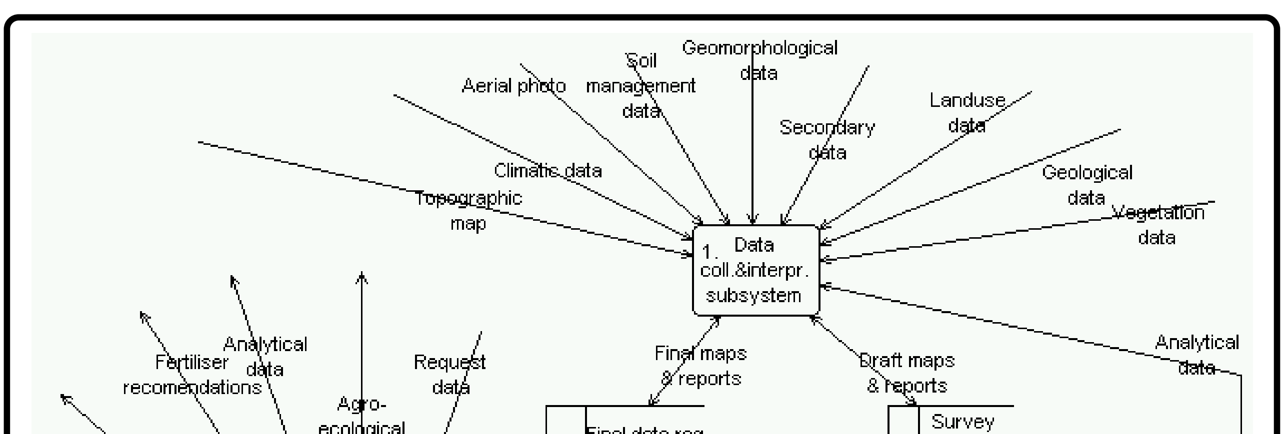


Figure 5.3 Top-level data flow diagram

5.2.4. Function and Mandate of Sub-systems

1. **Data Collection and Interpretation Sub-system:** Within this sub-system appropriate knowledge and skill level on remote sensing, soil survey and quite good experience of soil data interpretation is required. The expertise are expected to collect primary data such as soil samples, site and profile descriptions from the field, and secondary data from literature and interviews. Soil experts are also expected to interpret the data by combining the lab analysis result with site observation data for different purposes. The sub-system produces report documents and draft soil interpretation maps with necessary information.

2. **Laboratory Sub-system:** For this sub-system appropriate knowledge on soil physics and soil chemistry is a requirement. The sub-system would be expected to analyse the soil samples collected from the field by soil survey experts or brought by the clients themselves for characterisation of physical and chemical properties of soils. The product of this sub-system is analytical data, which will be used by data collection and interpretation sub-system for finalising the reporting and

by GIS & computing sub-system to produce different soil thematic maps according to user preference.

3. **GIS and Computing Sub-system:** This is the sub-system where analogue - digital conversion is carried out. For this sub-system basic knowledge on cartography, image processing, remote sensing, database management and GIS application is required. The main task of this sub-system is to produce discrete and/or continuous spatial models (what we call maps) by integrating the analytical data or information from the laboratory sub-system with draft soil interpretation maps from data collection & interpretation sub-system.

4. **Information Dissemination Sub-system:** This is a newly introduced sub-system in addition to the existing ones. The main task of this sub-system is to process the request for data and information from the users and channel the data/information to and fro. For this sub-system basic knowledge on database management, documentation, library service, reception and reproduction techniques are required. The sub-system is expected to provide all the necessary information concerning soils to the clients as per their need, unless otherwise the user seeks additional explanation from other sub-system expertise. For more information refer to appendix D.

5.3. Soil Geographic Database Design

Soil geographic database design is not an independent task when one deals with information system development. Rather it is the core and an integral part of system design. As of other geographic databases, soil geographic database contains spatial and non-spatial components.

5.3.1. Non-Spatial Database Design

Like that of system design here also we have four main phases in database design process. These are requirement collection and analysis (external design), conceptual design, logical design and physical design. These phases are interrelated to each other and they are briefly explained in chapter 4 section 4.5.1. However, in this chapter the logical and the physical design phases are presented in detail.

5.3.1.1. The Enterprise Rule

According to Oxford Advanced Learner's Dictionary of Current English, an **Enterprise** is defined as a project or an activity, especially one that is difficult or requires effort (Hornby, 1995). Therefore, the project for this study is Soil Information Provision for Lake Naivasha area.

An Enterprise rule is defined as any rule which is applicable to the conceptual model of an enterprise's data (Howe, 1989). These rules are the ones that govern the relationships among the entities while developing the conceptual model. The enterprise rules defined for the project Soil Information Provision for Lake Naivasha area are:

'A soil map must contain at least one soil map unit.'
'A soil map unit can occur, at most, in one soil map.'

'A soil map unit may contain many soil components.'

‘A soil component may occur in many soil map units.’

‘A soil map unit may contain many delineations.’
‘A delineation can occur, at most, in one soil map unit.’

‘A soil component may be described by many observations.’
‘An observation can define, at most, one soil component.’

‘An observation can be represented by, at most, one point location.’
‘A point location can represent, at most, one observation.’

‘An observation may contain many horizons.’
‘A horizon can occur, at most, in one observation.’

5.3.1.2. Entity Relationship Model

The entity relationship model was defined based on the enterprise rule that is stated in the earlier section. The model consists of six entities, which are independent objects in reality. The entities are soil map, soil map unit, soil component, soil observation, soil polygon and soil horizon.

These entities with their respective relationships are shown in figure 5.4. The entities are in a form of hierarchy where one entity is more general than the other that means the extent of their spatial coverage differs accordingly. For example a soil map covers a very large area while a single point represents one observation only.

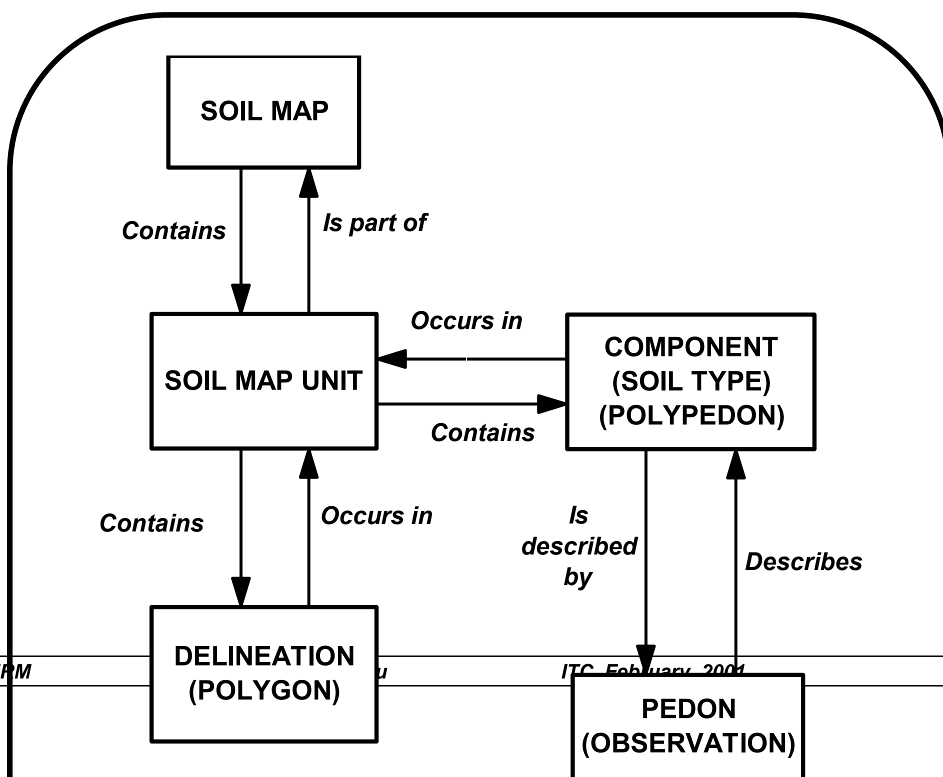


Figure 5.4 Conceptual soil geographic database design (SGDBD) for Lake Naivasha area
(Source: Adapted and modified from Zinck & Valenzuela, 1990 & Rossiter, 1999)

5.3.1.3. Logical Design

The logical database design follows the conceptual design where the logical design is dependent and the later is independent of database management system (DBMS). Therefore, in logical design normalisation is the main activity to prepare the data with respect to the DBMS rules. The normalisation process has been explained in detail in chapter 4 section 4.5.1.3. Normalisation is necessary to bring the data in to the format that is suitable for the database management system. Many to many relationship between SMU and SCOMP entities was normalised in to two one to many relationships by introducing a combination entity called SMUCOMP. Attributes at horizon level were grouped to SOBS_HOR entity while those attributes recorded at site level were grouped to SOBS_SITE entity. Other entities & attributes were grouped on the same way.
After normalisation the data model mapping was done on Ms Access 97 database management system software, where as the graphical notation of the logical design was made using Flow 4 Charting software. Figure 5.5 shows the graphic notation of entity relationship diagram where the codes 1 – 1 & 1 – M represent the relationships one to one and one to many respectively. The logical model consists of two components the spatial component (soil delineations and point locations) and non-spatial component (soil map, soil map unit, polypedon, pedon and soil horizon). For more information on entity description refer to appendix C.

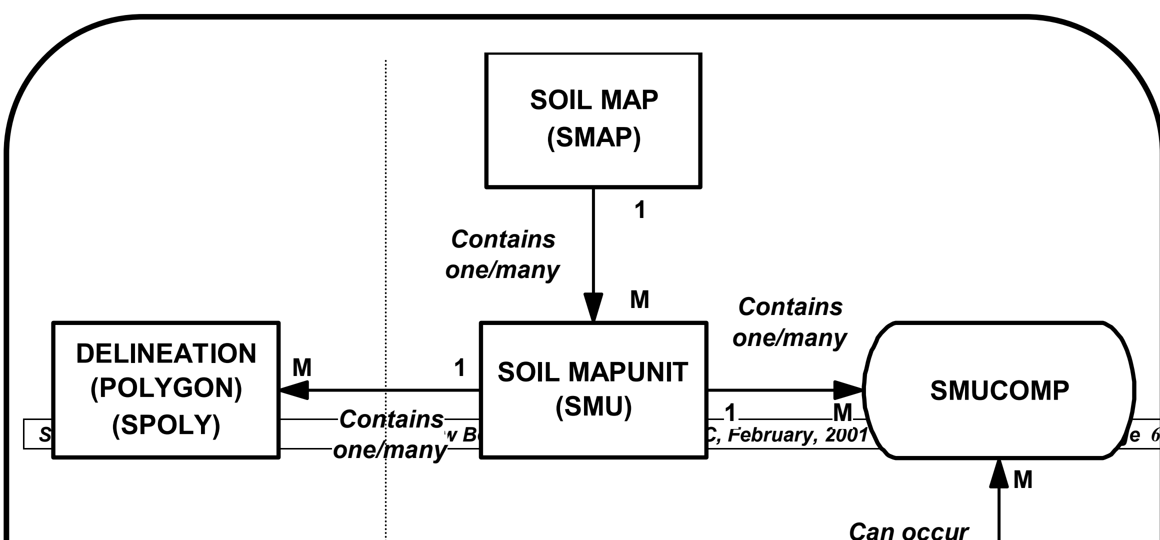


Figure 5.5 Logical soil geographic database design (SGDBD) for Lake Naivasha area
(Source: Adapted and modified from Zinck & Valenzuela, 1990 & Rossiter, 1999)

5.3.1.4. Physical Design

The non-spatial database has been organised in Ms Access 97 software following the nested database approach. This structure was selected to give the maximum benefit to the data user by supporting the access to the original data. This was realised by designing four individual databases and one nested/integrated database with a clear and standard metadata document for each database. The meta-data explains both the spatial and non-spatial data sets. The nested database is a result of integration of those independent databases. This integration involved the generalisation of data to the level of the least detailed database i.e., corresponding to 1:50 000 map scale. The non-spatial database designed for Lake Naivasha area includes the following.

1. Kulia soil geographic database (at scale 1:10 000)
2. Kwacha soil geographic database (at scale 1:50 000)
3. Nested soil geographic database (at scale 1:50 000)
4. Nini soil geographic database and (at scale 1:5 000)
5. Quarantine soil geographic database (at scale 1:5 000)

All the databases have the same structure containing six individual entities and one combination entity, which are: soil map, soil map unit, soil polygon, soil component, soil observation at site, soil observation at horizon and soil map unit component entity. However, the nested/integrated database structure in addition to the aforementioned ones contains other five entities. These entities are topsoil component, topsoil observation at site, topsoil observation at lab, topsoil map unit component (combination entity) and topsoil data of sulmac farm at plot level.

The attributes of the additional entities are the same (of course with some additional attributes that are specific for the topsoil study) as the corresponding entities, which are common to all. For example topsoil component entity has the same attributes like that of soil component entity. The reason for the design of these additional entities is to handle the topsoil data that was collected from Lake Naivasha area for the first 0 - 30 cm of the soil depth irrespective of the soil horizon difference. The data was collected by Mr Atkilt Girma, (reference) who investigated about soil phenoforms and genoforms for his Msc research at ITC.

The detailed physical database design process, data dictionary and the design parameters of the entities are explained in chapter 4 section 4.5.1.4 and appendix C. In addition to these for each of the nested databases a metadata document has been compiled in html and text formats describing about data sources, process steps, data quality, data organisation and accessibility issues. For more detail see an example in appendix G.

The designed database structure containing the entities with the referential integrity enforced is shown on figure 5.6 next page.

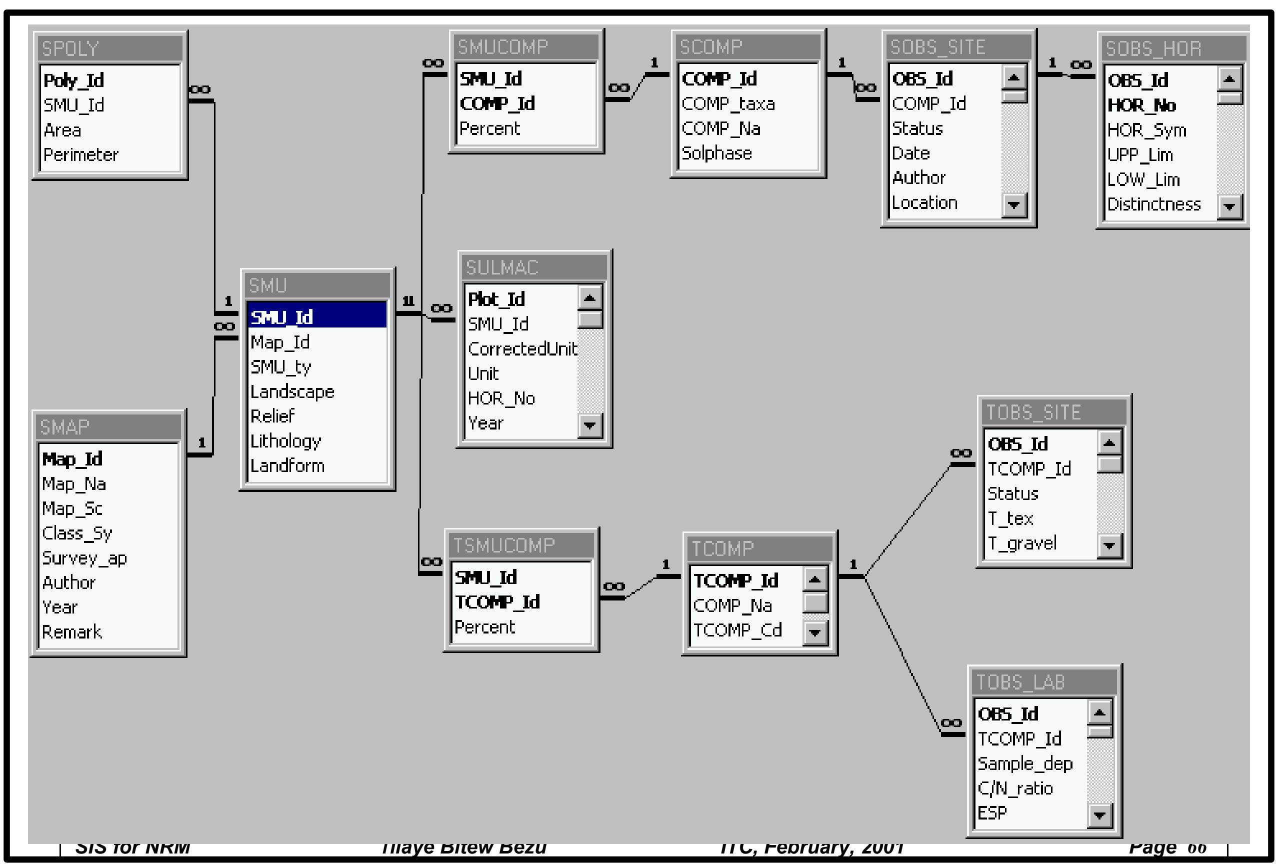


Figure 5.6 Non-spatial Database Structure as implemented in Ms Access

5.3.2. The Spatial Database

The spatial database has been organised in ILWIS GIS software and it contains digitised soil maps (as converted from the existing analogue format) and other relevant attribute maps generated from these basic spatial data sources. The approach followed in map conversion from analogue to digital format and data capture has been discussed in detail in chapter 4, section 4.5.2.

The spatial databases are organised in the nested approach with the same manner like that of the non-spatial databases. The type and content of the data set contained in these different spatial databases is basically similar, for example soil type maps, observation point maps and single valued attribute maps.

The spatial data organisation, the information sources, the process steps, the data quality and other necessary information are described in a metadata document for each data set independently. For a sample metadata document refer to appendix G. Two examples of spatial data showing the major soil types of the Nini farm and Quarantine farm which are obtained from the Nini and quarantine spatial databases, are shown in figures 5.7 and 5.8 respectively. The soil map of Kulia farm and a thematic map showing the average Calcium content in the farm as extracted fro the non-spatial Kulia soil geographic database and interpolated using ILWIS has been presented in appendix F.

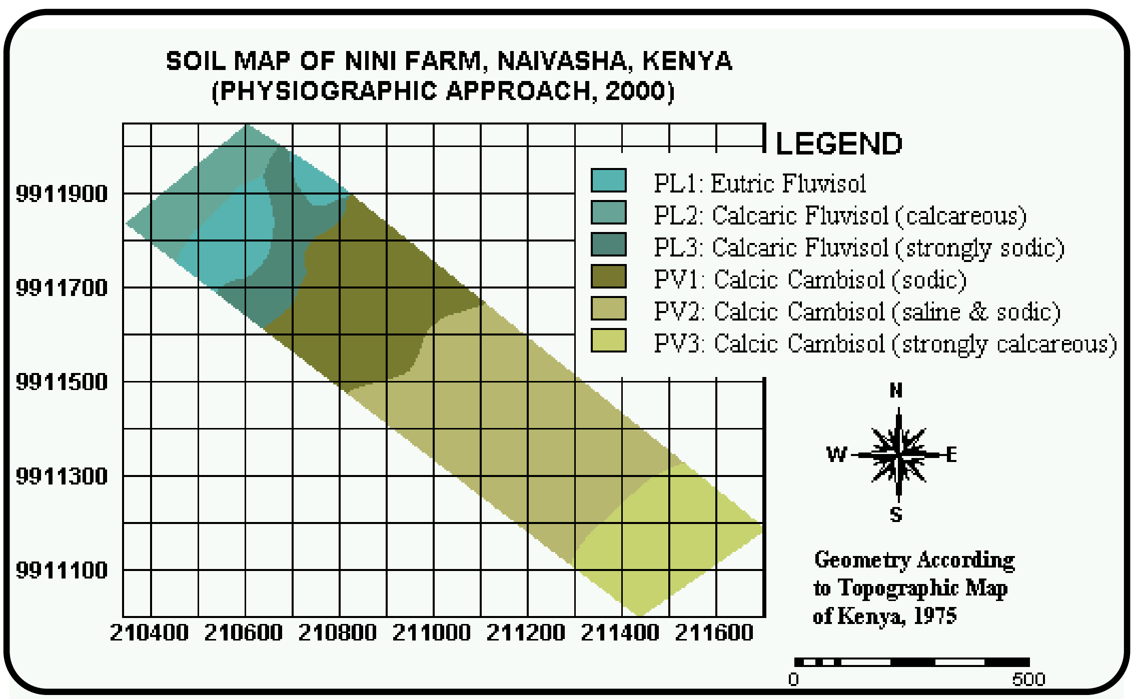


Figure 5.7 Soil map of Nini farm, converted from analogue to digital format
(Source: Analogue soil sketch map by Gatahi, 1986)

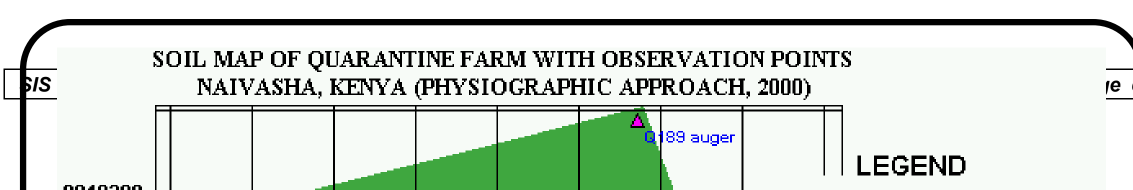


Figure 5.8 Soil map of Quarantine farm, converted from analogue to digital format
(Source: Analogue soil sketch map by Kamoni, 1988)

5.4. Multi-Source Data Integration

In the process of multi-source soil data integration, problems from differences in classification systems and survey approaches may arise. Therefore, before one puts the categorical data together in the same database, first these differences should be resolved. On the other hand, The geometric boundary of soil maps (common lines) is another issue that has to be resolved particularly in integrating the spatial data from different sources. "Common lines" in soil maps refer to those geometric soil boundaries, which are expected to be the same for two or more soil maps, irrespective of the scale used to produce them.

5.4.1. Non-spatial Data Integration

In this study two classification systems (FAO & USDA), and two survey approaches (physiographic & geo-pedologic) were encountered. The problem was resolved by creating a correlation table that matches the corresponding equivalents. These correlation tables were used as a basis to develop an integrated database for Lake Naivasha area from different sources.

5.4.1.1. Correlation in Survey Approaches

Among the four-soil survey documents available for Lake Naivasha area three of them (Quarantine, Kulia and Nini) were surveyed following the physiographic analysis approach, while one of them (soil map made by Kwacha) was surveyed with geo-pedologic approach. (See the difference between the two survey approaches on chapter 2, section 2.4.2)

Despite the fact that no literature was found explaining the direct correlation between physiographic and geo-pedologic survey approaches, it was possible to deduce from literatures written independently and by sharing past experience from soil scientists.

Based on the explanations given by Farshad (1999) on both physiographic and geo-pedologic approaches, Goosen (1967) on physiographic approach and Zinck (1989) on geo-pedologic approach, we have proposed a correlation between the two approaches. The proposed correlation table was used to integrate the non-spatial data of this study on nested soil geographic database. The proposed correlation table for the two survey approaches is shown in table 5.9.

Table 5.9 Correlation table between Physiographic and Geo-pedologic Soil Survey approaches

Survey Approach	Terrain units			
	Higher unit	Middle unit	Parent material	Smallest unit
Physiographic	Land type	Sub-land type	Geology	Mapping unit
Example	Lacustrine plain	Flat	Lacustrine sediments	Almost level
Geopedologic	Landscape	Relief	Lithology	Landform
Example	Lacustrine plain	Terrace	Alluvium	Tread

Note: For Geo-pedologic approach there are two more high-level terrain units (Geo-structure and Morphogenetic environment), but mostly they are not used in map legend making. For detail descriptions, see chapter 2 table 2.1.

5.4.1.2. Correlation in Classification Systems

The difference in classification systems is another second issue to be resolved before one starts to integrate the data in a common database. Different types of classification systems have been discussed briefly on this paper in chapter 2 section 2.4.4. Among the classification systems, FAO and USDA classification systems are more widely used around the world, including the study area concerned.

Fortunately, all the soil survey report documents available for Lake Naivasha area have followed FAO classification system and in addition to this one of them has added equivalent USDA classification. However, for the benefit of the user groups, the writer of this paper has added a correlation table for diagnostic horizons and soil types between USDA and FAO classification systems. The correlation is not a direct fit, but an equivalent is given. Part of the correlation table for diagnostic horizons are shown in table 5.10 and similarly for soil types it is shown in table 5.11. For the complete correlation tables see appendix E.

The correlation table was produced by classifying the same soil type in both classification systems where the classification was done based on FAO/Unesco 1988, soil map of the world, revised legend and USDA 1994, soil taxonomy.

Table 5.10 Correlation table for Diagnostic Horizons

Serial No:	Approximate Equivalents of Diagnostic Horizons	
	FAO	USDA
1	Histic H horizon	Histic epipedon

2	Mollic A horizon	Mollic epipedon
3	Umbric A horizon	Umbric epipedon
4	Argic B horizon	Argillic horizon
5	Cambic B horizon	Cambic horizon

(Source: Adapted from Buol et al., 1997)

Table 5.11 Correlation table for Soil Types

Serial No:	Approximate Equivalents of Soil Types	
	FAO	USDA
1	Eutric Cambisols	Typic Xerochrepts, Typic Eutrochrepts
2	Haplic Luvisols	Andic Haploxeralfs, Typic Haploxeralfs
3	Haplic Andosols	Andic Xerochrepts
4	Calcaric Fluvisols	Typic Xerothents
5	Eutric Fluvisols	Andic Xerorthents

(Source Kwacha, 1998)

5.4.2. Spatial Data Integration

As far as the writer of this thesis has investigated, there was no any literature found that explains how to identify soil boundaries (common lines) on different soil maps of the same area with out referring back to some kind of base map.

However, Goosen (1967) explained the relation between soil boundaries and aerial photo interpretation elements by giving comparative categorical measures high, medium and low. Based on this general concept, the soil maps of Lake Naivasha area were reinterpreted to resolve the discrepancy at common soil boundaries by using a common ortho-photo mosaic as a base map. The general rules for the elements of analysis and their comparative coincidence with soil boundaries are given in table 5.12.

Table 5.12 General rules to identify common lines of soil boundaries

Elements of Analysis	Visibility in Stereo image	Relation to Soil condition	Coincidence with Soil boundaries
Land type	High	High	High
Relief	High	High	High
Slope form	High	High	High
Drainage condition	Medium	High	Medium
Constructional drainage system	High	High	High
Destructional drainage system	High	High	Medium
Natural vegetation	High	High	Medium
Parent material	Low	High	High
Colour/tone	High	Low	Low
Land use	High	Medium	Low

(Source: Adapted from Goosen, 1967)

5.4.2.1. Assessment of the Nature of Maps to be Integrated

The assessment of the nature of soil sketch maps available around Lake Naivasha area was done by referring back to the original sources which have resulted on the sketch map we are concerned. For example, type of base map used, purpose of the survey and the status of profile description. The results of the assessment are presented for each map as follows.

a) Kulia soil sketch map:
  1) The Kulia soil sketch map was the result of a soil survey conducted to assess soil fertility status of Kulia estates (for commercial cultivation of strawberries under irrigation) in 1980.
  2) The map was produced at scale 1:10 000 where all soil polygons/delineations are open, for the sake of completeness the boundaries were closed by the farm boundary.
  3) The soil description was made using auger holes only for fixed depths i.e. 0 – 20 cm, 50 – 60 cm, and 80 – 100 cm and the information rather is incomplete for soil classification.
  4) No one complete profile was described.
  5) There is no information about the type of the base map used as background to delineate the soil boundaries. From the review of the document it was concluded that the surveyor have not used any reliable base map.
  6) The survey was not concerned on soil mapping (only one soil type was identified for the whole survey area).
  7) The total area covered by this survey was about 200 ha which is covering only 4.3% as compared to the soil map surveyed by Kwacha.

b) Nini soil sketch map:
  1) The Nini soil sketch map was the result of a soil survey conducted to assess the soil conditions of the Nini farm and their suitability for cultivation of some horticultural crops using sprinkler irrigation in 1982 and reported in 1986.
  2) The map was produced at scale 1:5000 where all soil polygons/delineations are open, for the sake of completeness the boundaries were closed by the farm boundary.
  3) The soil description was made for 6 full pit profiles and 53 auger holes, where the information for these auger holes was not included in the report document.
  4) An uncontrolled map of the farm at a scale of 1:1000, made by the Provincial Irrigation Unit, Nakuru, was used as a base map (which is more liable to geometric in accuracy).
  5) The total area covered by this survey was about 46 ha which is covering only 1% as compared to the soil map made by Kwacha.

c) Quarantine soil sketch map:
  1) The Quarantine soil sketch map was the result of a soil survey conducted to evaluate the suitability of the soils of Quarantine farm for growing irrigated vegetables in 1986 and reported in 1988.
  2) The map was produced at scale 1:5000 where all soil polygons/delineations are open, for the sake of completeness the boundaries were closed by the farm boundary.
  3) The soil description was made for 3 full pit profiles and 30 auger holes, where the information for these auger holes was not included in the report document.
  4) A topographic survey of the survey area at scale 1:50 000 that was carried out by B.G. Mwangi and S. Wataka both from Kenya Soil Survey was used as a base map.

5) The total area covered by this survey was about 43 ha which is covering only about 1% as compared to the soil map made by Kwacha.

d) The Kwacha's soil map:
   1) The kwacha's soil map was the result of a soil survey conducted to assess vulnerability of soils to change in agricultural use around Lake Naivasha for the purpose of Msc thesis research in 1998.
   2) The map was produced at scale 1:50 000, which is medium scale.
   3) Soil description was made for a total of 28 soil profiles among which 16 were mini-pits and 12 were auger holes.
   4) Aerial photos of 1984 at scale 1:12 500 and 1970 at scale 1:50 000 were used to make the interpretation base map. In addition, topographic map of Kenya, 1975 at scale 1:50 000 and the geological map of Kenya, 1988 at scale 1:100 000 were used for reference.
   5) The total area covered by this survey was about 4613 ha which is quite a big area as compared to the Kulia, Nini, and Quarantine.

5.4.2.2. Summary of Discrepancies

Based on the individual assessments on the nature of soil sketch maps, a summary of the differences that are expected to be causes for boundary discrepancy are presented here. Therefore, the major discrepancies that have resulted for the mismatch of the soil boundaries among the aforementioned soil maps include:

1. Difference on the purpose of the survey
2. Difference on the use of reliable base map
3. Difference on surveyors' perception (interpretation)
4. Difference on production scale i.e. level of generalisation
5. Poor geometric control points for geo-referencing
6. Uncertainty, due to the fuzzy nature of soil boundaries
7. Cumulative error while converting from analogue to digital

5.4.2.3. Reinterpretation of Common lines

The general rules to identify common lines of soil boundaries by Goosen (1967) were applied to resolve soil boundary discrepancies for this study. After resolving the soil boundary discrepancy among the different soil maps that are to be integrated by reinterpretation, an integrated and generalised soil map was produced. The discrepancy resolution was made between the soil maps Kulia (green), Nini (blue) and Kwacha's soil map (red), then by reinterpretation an integrated soil map (magenta colour) has been created as shown in figure 5.9.

The geometric shape of the integrated soil map is almost similar to that of the soil map made by Kwacha with some differences at areas of intersection with other soil maps. The reason for the geometric similarity is that the Kwacha's soil map covers the same spatial extent to this study area while the other three maps cover only about 5.3% of it. The complete integrated soil map of east and south east of Lake Naivasha shore is shown in figure 5.10 next page.

Before reinterpretation After reinterpretation

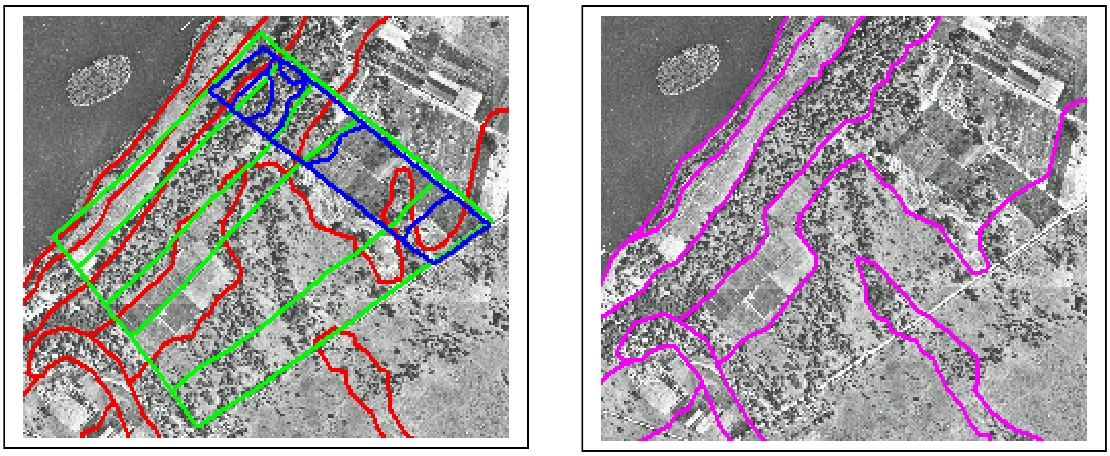


Figure 5.9 Resolving soil boundary discrepancies by reinterpretation

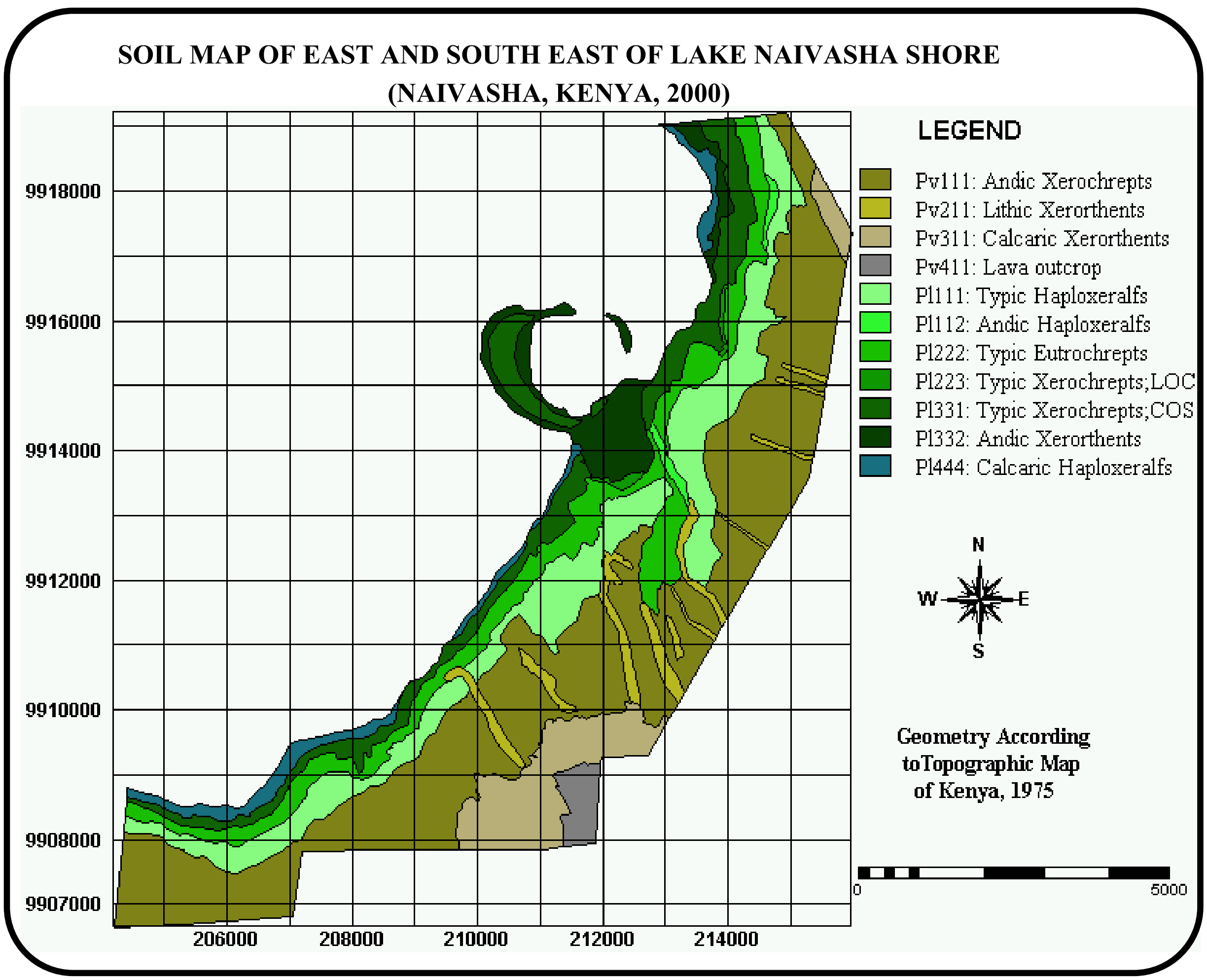


Figure 5.10 Integrated soil map of Lake Naivasha area converted from analogue to digital
(Source: Gatahi, 1986, Kamoni, 1988, Kwacha, 1998, and Siderius, 1980)

5.5. Map Accuracy Assessment

An accuracy assessment of soil maps was carried out in support of (1) analogue digital conversion (2) map registering to a common geometry and (3) for the assessment of the spatial data quality of the multi-source information for data integration. The locational accuracy assessment of the soil maps for this study was computed using the root mean square error (RMSE) and the results are presented in the following sub-sections.

5.5.1. Accuracy of the Garmin 12 GPS Personal Navigator

The procedure followed how to capture and convert from analogue to digital has been explained in chapter 4 section 4.5.2. The equipment used to collect the geographic co-ordinates for this purpose was the Garmin 12XL GPS receiver. Its receiving accuracy also contributes to the quality of the spatial data captured using this equipment. Therefore, the assessment of the equipment's receiving accuracy as compared to a reference map is essential. The topographic map of Kenya, 1975 at scale 1:50 000 was taken as a reference map for this accuracy assessment. The RMSE value in table 5.13 tells us that on average the location of a point can be displaced by +/-25 meters from the actual ground position (according to the map) when one uses the GPS receiver for capturing co-ordinates of a point.

Table 5.13 Accuracy of Garmin 12XL GPS receiver

No	Co-ordinates from Topographic map		Co-ordinates from GPS reading		Component error RMSE in (M)		Vector error RMSE in (M)
	X	Y	X	Y	X	Y	(X + Y)
1	215125	9917500	215137	9917500	12	0	12
2	214475	9916175	214493	9916198	18	23	29
3	213975	9915150	213979	9915118	4	32	32
4	213463	9914050	213481	9914067	18	17	25
5	213350	9913750	213369	9913743	19	7	20
6	212925	9912150	212950	9912154	25	4	25
7	212575	9911750	212607	9911755	32	5	32
8	210250	9910188	210275	9910189	25	1	25
9	209950	9909988	209975	9909998	25	10	27
10	206325	9907575	206301	9907576	24	1	24
11	213350	9913550	213369	9913569	19	19	27
	Sum				221	119	279
	Mean				20	11	25

5.5.2. Accuracy of Geo-referencing

The accuracy of geo-referencing can be evaluated from the pixel size and the value of sigma that can be obtained from the transformation used during geo-referencing. The sigma value refers to the distance between the actual position of the point and its assigned position while geo-referencing that is measured in terms of number of pixels. The accuracy was measured as the product of the pixel size and the sigma value for each map. As it has been indicated in table 5.14 the average accuracy of geo-referencing is nearly +/-10 meters.

Table 5.14 Accuracy of transformation/geo-referencing

No:	Map Name	Pixel size In (M)	Sigma value In Pixels	Variation In (M)	Transformation
1	Nini farm soil map	0.260	16.334	4.2	Affine
2	Kulia fram soil map	0.864	12.899	11.1	Affine
3	Quarantine farm soil map	0.425	16.337	6.9	Affine
4	Kwacha's study area soil map	7.197	2.648	19.1	Affine
	Sum	8.746	48.218	41.4	
	Mean	2.187	12.055	10.3	

5.5.3. Accuracy of Soil Maps

The soil map accuracy was computed using the RMSE as used earlier for the GPS receiver. The co-ordinates that are used for the calculation of the root mean square error are those co-ordinates captured by the GPS receiver in the field and which are not used for geo-referencing in the analogue digital conversion process. The corresponding co-ordinates from each map were read after the maps have been digitally geo-coded.

The location accuracy i.e. the mean vector error of the soil maps organised for this study range from the minimum +/-10 meters for Nini soil map to the maximum +/-26 meters for soil map of Kwacha's study area (Kwsoil). The over all accuracy of the integrated soil maps is +/-18 meters. The RMSE computation result for each soil map including the component errors is shown in table 5.15 below.

Table 5.15 Location accuracy of soil maps of Lake Naivasha area

Map Name	Point No.	Co-ordinates from GPS reading		Co-ordinates from Soil Map		Component error RMSE in (M)		Vector error RMSE in (M)
		X	Y	X	Y	X	Y	(X + Y)
Nini	1	210463	9911750	210456	9911749	7	1	7
Nini	2	211525	9911317	211534	9911323	9	6	11
Nini	3	210786	9911935	210788	9911927	2	8	8
Nini	4	210689	9912001	210695	9911988	6	13	14
Nini	5	210681	9912009	210676	9912000	5	9	10
Nini	Sum					29	37	51
Nini	Mean					6	7	10
Kulia	1	211004	9910676	211004	9910691	0	15	15
Kulia	2	211160	9911031	211143	9911028	17	3	17
Kulia	3	210835	9911370	210811	9911357	24	13	27
Kulia	Sum					41	31	60
Kulia	Mean					14	10	20
Quarantine	1	214338	9917697	214346	9917692	8	5	9

Quarantine	2	214719	9918294	214717	9918276	2	18	18
Quarantine	3	214307	9918010	214305	9918002	2	8	8
Quarantine	4	214629	9917950	214637	9917951	8	1	8
Quarantine	Sum					20	32	44
Quarantine	Mean					5	8	11
Kwsoil	1	213979	9915118	213997	9915097	18	21	28
Kwsoil	2	213369	9913743	213386	9913740	17	3	17
Kwsoil	3	212607	9911755	212582	9911775	25	20	32
Kwsoil	4	210998	9910668	210970	9910695	28	27	39
Kwsoil	5	210275	9910189	210252	9910198	23	9	25
Kwsoil	6	207690	9908751	207706	9908741	16	10	19
Kwsoil	7	204972	9907981	204983	9908000	11	19	22
Kwsoil	Sum					138	109	181
Kwsoil	Mean					20	16	26

The locational accuracy of Nini and Quarantine soil maps is twofold better than that of the Kwacha's study area (Kwsoil) soil map i.e. +/-10 & +/-11 versus +/-26 meters respectively. For Nini and Quarantine soil maps the data in the field was captured using compass traversing method with distance measurement and GPS reading while for Kwsoil only GPS reading was taken.

Therefore, from this difference in accuracy result we can conclude that compass traversing method is better for spatial data capture in the field as compared to taking only GPS readings for analogue to digital conversion. The compass traverse method is most useful for the conditions where enough ground control/truth points could not be identified for geo-referencing. The drawback of this method is it is time consuming and needs more labour as compared to taking only GPS readings.

5.6. User Accessibility to Soil Information

On the aforementioned sections and sub-sections of this chapter we tried to explain how the spatial and the non-spatial data sets were obtained and organised, furthermore, their accuracy/quality was also briefly discussed. Hence, the remaining part of this chapter is to answer the research question of how to improve user accessibility to the organised data set. To answer this question two approaches metadata documentation and user interface design were applied. The approaches are explained as follows.

5.6.1. Metadata Documentation

The metadata standard implemented for this study was that of Content Standard for Digital Geo-spatial Metadata (CSDGM) developed by the Federal Geographic Data Committee (FGDC), which is one of international metadata standards. More detail about metadata forms, standards, and tools have been explained in chapter 2 section 2.3.

Metadata documentation was done for each of the five data sets (Kulia, Kwacha's study area, Nini, Quarantine and Nested/Integrated soil geographic databases) independently. An example of these metadata documents is presented in appendix G and the basic standard information included in each metadata document comprises:

- Identification information

- Data quality information
- Spatial data organisation information
- Spatial reference information
- Entity and attribute information and
- Metadata reference information

5.6.2. User Interface Design

The user interface design is meant to help the information user during browsing for the data in the database. The designed interface is a continuous screen presentation asking the user for some standard questions where in reply (s)he will click a button which leads to an other screen until the required information or data is reached. The schematic representation of the flow of the user interface as implemented in Ms Access is presented in figure 5.11.

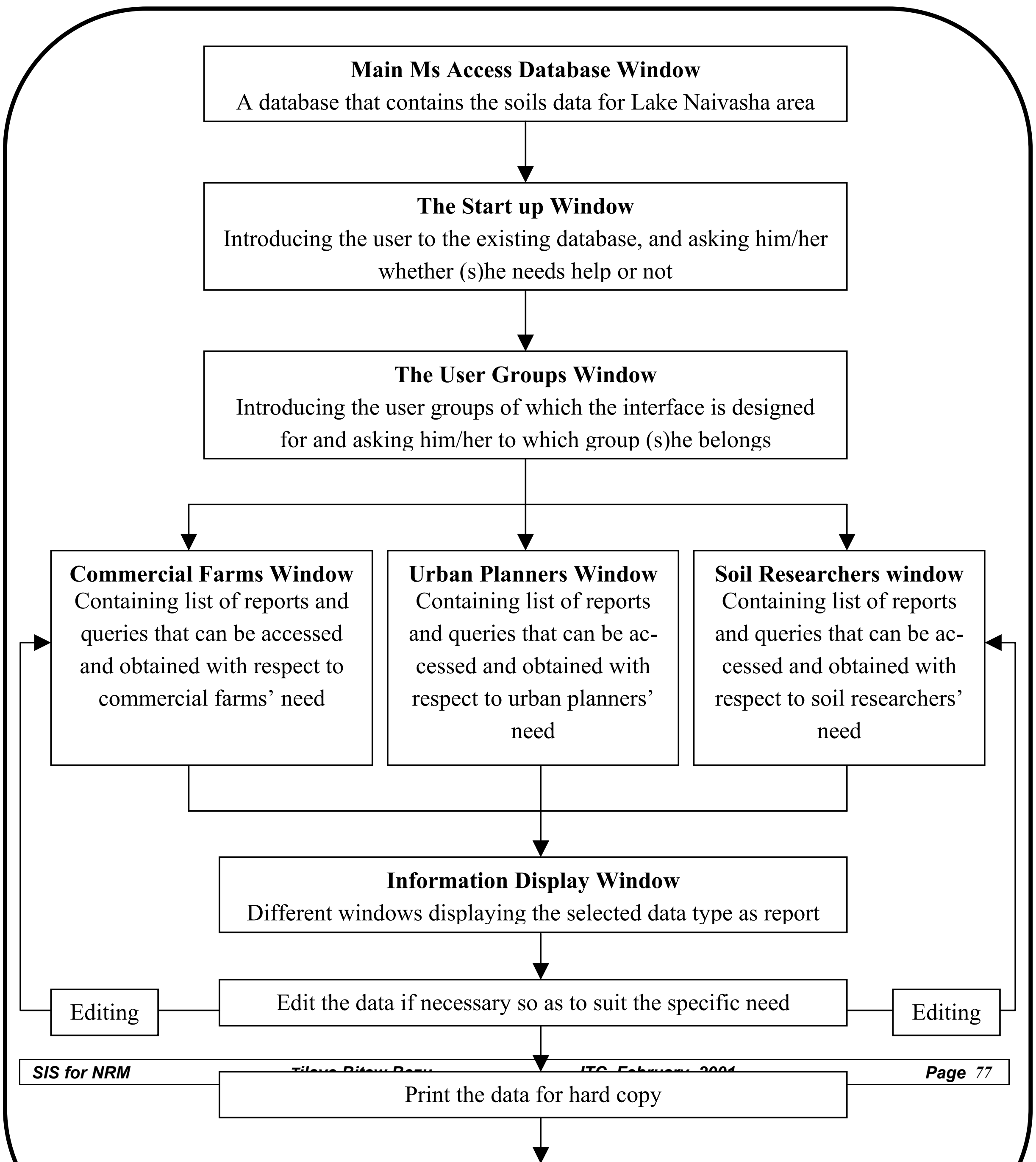


Figure 5.11 Schematic representation of the user interface as implemented in Ms Access

The presentation screens are of similar in their basic structure. This includes, name of the screen, back, forward and close/exit buttons, a question and/or a statement, and buttons with a list of choices. The first navigation screen that the user gets immediately while entering the database is presented in figure 5.12, which is named as the start up window. The interface is designed so that it automatically displays the start up window when one opens the database and then follows the instructions.

Figure 5.12 The start up window navigation screen as implemented in Ms Access

5.6.3. Use of Query for Information Retrieval

The query window:

The query window has two modes, the design mode and the datasheet mode. The design mode is where we create the query and the datasheet mode is where we display the query's dynaset. The query design window consists of two panes.

- The table/query entry pane and
- The query by example (QBE) design pane

The table/query entry pane is where tables and or queries and their structures are displayed. The QBE pane is used for the fields and criteria that the query will display in the dynaset. Each column in the QBE design pane contains information about a single field from a table or query in the upper pane.

Sample queries by SQL:

SQL offers a high level of data independence. The SQL architecture has three basic elements, which are select, from and where. These elements are combined in the general structure of the language as SELECT a particular attribute/field FROM a specific table/entity WHERE a predefined value is found. Sample queries from the nested database are shown below.

Query 1:

One of the needs of the commercial farms is to know their soils nutrient level. To supply this information five entities were internally joined to run the query where as the required information is retrieved only from three entities. The SQL command language query used to extract the nutrient level of the top soil (first horizon) which includes Ca, Mg, K, Na, Mn, N, C, OM, P, at the level of mapping units with respect to each observation point is the following.

SELECT

SMU.SMU_Id, SOBS_SITE.OBS_Id, SOBS_SITE.Easting, SOBS_SITE.Northing, SOBS_HOR.HOR_No, SOBS_HOR.Ca, SOBS_HOR.Mg, SOBS_HOR.K, SOBS_HOR.Na, SOBS_HOR.Mn, SOBS_HOR.N, SOBS_HOR.C, SOBS_HOR.OM, SOBS_HOR.P

FROM

((SCOMP INNER JOIN (SMU INNER JOIN SMUCOMP ON SMU.SMU_Id = SMUCOMP.SMU_Id) ON SCOMP.COMP_Id = SMUCOMP.COMP_Id) INNER JOIN SOBS_SITE ON SCOMP.COMP_Id = SOBS_SITE.COMP_Id) INNER JOIN SOBS_HOR ON SOBS_SITE.OBS_Id = SOBS_HOR.OBS_Id

WHERE

(((SOBS_HOR.HOR_No)=1));

Result:

The result of the query is a tabular data containing fourteen attributes/fields and fifty-two dynamic sets of records (dynaset) where only nineteen of them are shown here. The tabular data result of the query as it has been taken directly from the query datasheet view in Ms Access format is shown in table 5.16.

Table 5.16 Soil nutrient level of the first horizon as extracted from the nested database

SMU_Id	OBS_Id	Easting	Northing	HOR_No	Ca	Mg	K	Na	Mn	N	C	OM	P
PL111	10AB	214755	991756	1	18.7	1.38	3.2	1.93	0.6	0.25	2.6	4.472	68
PV111	10KJ	211959	991119	1	3.75	2.46	2.48	74	0	0.04	1.29	2.219	38
PL331	1AB	214208	991751	1	12.1	3.63	3.95	1.7	0.13	0.18	1.34	2.305	101
PL331	1KJ	210571	991201	1	16.2	3.4	2.35	0.85	0.27	0.17	1.71	2.941	393
PV111	1LN	209139	990838	1	5.75	4	1.45	0.65	0.23	0.08	0.78	1.342	95
PL331	1SH	205266	990827	1	15.2	4.68	3.25	1.6	0.14	0.12	0.39	0.671	213
PL223	2AB	214351	991750	1	13.5	3.52	3.55	1.6	0.34	0.16	1.64	2.821	103

PL222	2KJ	210688	991190	1	12.2	4.16	3.5	0.9	0.41	0.07	1.83	3.148	61
PV111	2LN	208817	990829	1	4.25	4.23	2.5	0.75	0.12	0.09	0.66	1.135	89
PL222	2SH	205291	990813	1	15	5.01	2.6	1.3	0.2	0.07	2.06	3.543	205
PL222	3AB	214471	991746	1	17.5	3.32	3.65	2.15	0.13	0.19	1.88	3.234	177
PL111	3KJ	210823	991178	1	13.5	1.52	2.8	3.1	0.03	0.06	1.35	2.322	336
PL111	3SH	205304	990795	1	2.75	2.37	1.3	0.58	0.13	0.02	0.87	1.496	151
PL112	4AB	214600	991745	1	10.5	3.42	3.65	2.15	0.13	0.14	1.67	2.872	126
PL111	4KJ	211008	991166	1	7	4.7	2.3	1.43	0.09	0.05	1.98	3.406	99
PL331	4SH	205676	990824	1	11.5	3.93	2.8	1.85	0.03	0.01	0.69	1.187	83
PL111	5AB	214756	991744	1	11.2	3.78	3.7	1.73	0.3	0.42	1.64	2.821	114
PL111	5KJ	211236	991147	1	7.75	4.93	3.65	1.4	0.11	0.05	2.44	4.197	99
PL222	5SH	205701	990804	1	6.25	3.91	3	1.7	0.07	0.02	0.63	1.084	87

Query 2:

This query deals with urban planners. Urban planners need soil particle size distribution for several engineering purposes. For example they need soil particle size distribution for determining soil compatibility, load-bearing capacity, shrink-swell potential, soil collapse and soil permeability. Therefore, for local roads and dwellings with out basement construction, soil particle size information for the first 70 cm of the soil depth is required. Hence, to supply this information five entities were internally joined to run the query where as the required information is retrieved only from three entities. The SQL command language used to extract the soil particle size distribution of sand, silt and clay for the first 70 cm of the soil depth at the level of mapping units with respect to each observation point is the following.

SELECT

SMU.SMU_Id, SOBS_SITE.OBS_Id, Avg(SOBS_HOR.Sand) AS AvgOfSand, Avg(SOBS_HOR.Silt) AS AvgOfSilt, Avg(SOBS_HOR.Clay) AS AvgOfClay

FROM

(SMU INNER JOIN ((SCOMP INNER JOIN SMUCOMP ON SCOMP.COMP_Id = SMUCOMP.COMP_Id) INNER JOIN SOBS_SITE ON SCOMP.COMP_Id = SOBS_SITE.COMP_Id) ON SMU.SMU_Id = SMUCOMP.SMU_Id) INNER JOIN SOBS_HOR ON SOBS_SITE.OBS_Id = SOBS_HOR.OBS_Id

WHERE

(((SOBS_HOR.LOW_Lim)<=70))

GROUP BY SMU.SMU_Id, SOBS_SITE.OBS_Id

ORDER BY SOBS_SITE.OBS_Id DESC;

Result:

The result of the query is a tabular data containing five attributes/fields and fifty-two dynamic sets of records (dynaset) where only nine of them are shown here. The tabular data result of the query was used to process the report in Ms Access. The report was again published with Ms Word, which is internally linked with Ms Access and finally copied to this document as shown in table 5.17.

Table 5.17 Particle size distribution of the top 70 cm soil depth

SMU_Id	OBS_Id	AvgOfSand	AvgOfSilt	AvgOfClay
PL222	Q221	30	22	48
PL331	Q220	32.6	35.3	32
PL111	Q219	30.6	37.3	32
PV111	N187	32	36	32

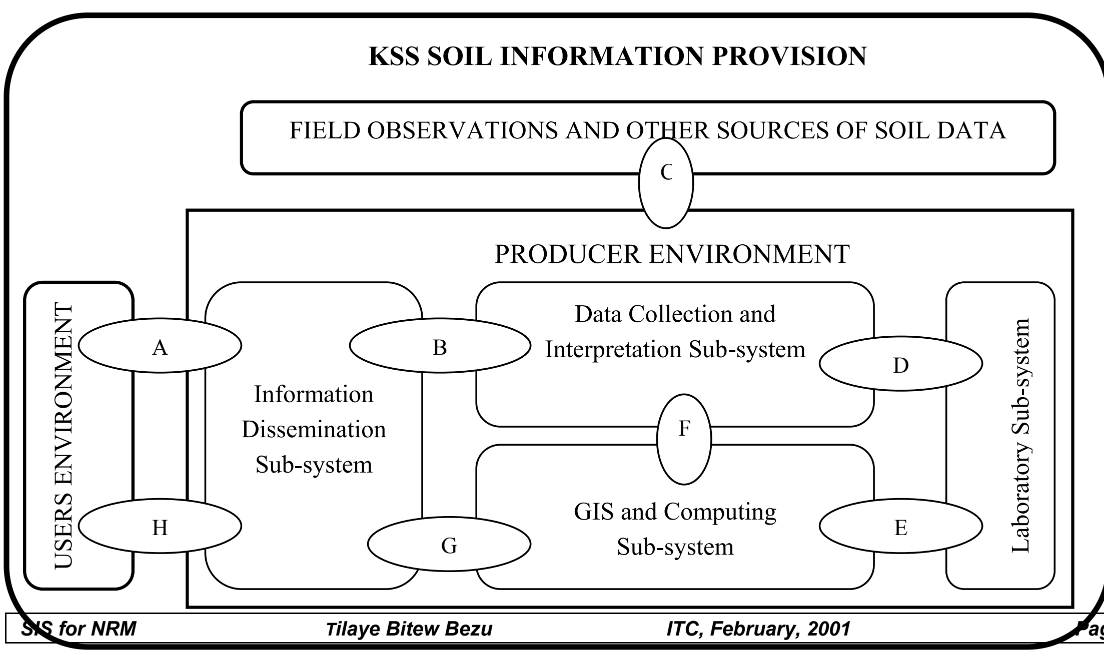
PL111	N186	48	29	23
PL111	N185	56	36	8
PL222	N184	57	32	11
PL222	N183	48	32	20
PL222	N182	60	30	10

5.7. Information System Architecture

Among the objectives of this research, designing a soil geographic database with a preliminary soil information system was the one that has been dealt with more emphasis. Soil information system (SIS) is under the family of geographic information systems which deals both with spatial (the soil) phenomenon and their corresponding non-spatial aspects. Information system is a method to convey data/information with phases of data collection, data processing, data storing, data retrieving and data analysis that is useful for decision making.

System architecture is a combination of system structure and its functionality, where the system structure refers to the combined components of the system while the functionality refers to the process that is carried out in the component. The information system design process has been dealt in detail in chapter 4 section 4.4 and in this chapter section 5.2 where as the theoretical background was discussed in chapter 2 section 2.1. The proposed new/modified system architecture of Kenya Soil Survey Soil Information Provision has four subsystems, which are data collection and interpretation, GIS and computing, laboratory and information dissemination sub-systems.

The interaction of these sub-systems with the customers and the outside environment as well as between themselves is shown in figure 5.13 with a simplified way of presentation. The data dictionary for the processes (sub-systems) and data stores is explained in appendix D.



A: User request for soil information
B: Planning for new soil survey
C: Collection of primary survey data
D: Samples for laboratory analysis
E: Results of laboratory analysis
F: Draft maps and report document
G: Required product
H: Information delivery

Figure 5.13 Architecture of SIS for providing soil information for Lake Naivasha area

6. Conclusions and Recommendations

This research tried to demonstrate the steps required in developing a soil information system for Lake Naivasha area. This was done based on the main objectives of the study i.e. to rescue the existing soils spatial and non-spatial data available for Lake Naivasha area and to document working procedures for soil geographic database design and SIS development. At the end of this study we came up with the following conclusions and recommendations.

6.1. Conclusions

The conclusions here are presented in accordance to the research questions posed at the start of the research. The conclusions include the following.

1. The preference of data format for soil information interchange between users and producers was statistically tested and revealed that the digital format is highly preferred ($p = 0.002$) from that of the analogue format. Where for 71.4% of the cases the digital is preferred while only for 23.8% of the cases the analogue format is preferred. The main reasons for the preference are less storage requirement, convenient for manipulation, analysis, retrieval and updating, easier for information exchange and transportation.

2. In organising soils data from different sources, the Nested database design approach is the best approach due its capability to maintain the level of categorical and cartographic details as same as the original source. In nested database approach the user has an option to access to both the original data source and to the integrated data set.

3. User access to the soil data set can be improved by providing a well-documented metadata that contains all the details about the data set stored in the database including spatial accuracy assessment results. Moreover, user interface design using forms in Ms Access is able to improve the troubles in the process of browsing to reach the required information. However, the use of navigation screens/forms gives the user a limited support and access i.e., only to those predefined queries and reports.

4. For analogue-digital conversion of spatial data the techniques applied are scanning, georeferencing, digitising (screen or tablet) and geo-coding where all performed adequately. However, for spatial data capture in the field at a condition where enough ground truth points both on the analogue map and on the real ground are not identifiable, the use of directional compass traverse method with distance measurement using a measuring tape and GPS reading is the best solution. For the conditions where enough ground truth points are identifiable only GPS reading suffice.

5. In the case of multi-source data integration, the problem of differences in classification systems and survey approaches can be resolved by developing correlation tables. A four level correlation table between physiographic and geopedologic survey approaches has been suggested by this study. On the other hand the spatial data dispute at common soil boundaries (common lines) can

be resolved by reinterpretation using an ortho-photo mosaic as a background (base map), and applying the rules on the relation between soil boundaries and aerial photo interpretation elements.

6. This study has proposed a prototype information system architecture for soil information provision based on the soil information required around Lake Naivasha area. Hence, this architecture can be used as a corner stone for KSS to restructure its present organisational structure in order to improve its performance in supplying soil information.

6.2. Recommendations

In the process of soil information provision and multi-source data integration the following issues must be taken in to consideration.

1. Mostly people in developing countries (like in Kenya) are insisted in collecting new data for every specific task required by the information user, which might has had been done in one of the former surveys. Therefore, to prevent this unnecessary time and money loss that is also prevailing in the study area, we advise the data collection & processing expertise at KSS to put all the available soil data digitally and then update when additional information is obtained. From this updated database it is possible to retrieve the information required for different specific applications. Hence, staff training for efficient use of the digital environment (system) is crucial.

2. In planning to collect new soil data for an area, it is essential to identify the ultimate users of the information and conduct user need assessment. This is because the mapping scale and categorical detail should not be decided upon the will of the information producer but it should originate from the minimum decision/management area of the information user.

3. There is a high need for soil information around Lake Naivasha area where at the time of this study, most of the commercial farms (73% of the respondents) are getting their soils data from abroad with high foreign currency expenditure. Therefore, the Kenyan Soil Survey Organisation should make further study on this issue and organise it self to supply the requirements of the users and alleviate the shortage of detailed soil information around the lake.

4. Further research on resolving soil boundary discrepancies that arise from differences in surveyors interpretation, mapping scale, survey purpose, and base map use is essential. In addition to this the proposed correlation table between the two survey approaches and the designed information system architecture needs a detailed investigation for further development.

References

1. Avery, B.W., 1987. Soil survey methods: a review. Technical Monograph No. 18. 1987, Silsoe: Soil Survey & Land Resources Centre.86.

2. Avison, D.E. and G. Fitzgerald, 1988. Information Systems Development: Methodologies, Techniques and Tools. Oxford: Blackwell.

3. Berlyant, A.M., 1992. Geoinformatics: A science, A Technology, A Subject. EGIS 92. (pp 1288-1293)

4. Buol, S.W., F.D. Hole, R.J. McCracken, and R.J. Southard, 1997. Soil Genesis and Classification. Fourth Edition. Iowa State University Press, Ames, Iowa 50014, USA.

5. Bregt, A.K., 1988. Quality of representative profile descriptions for predicting the land quality moisture deficit at different scales. In: Land qualities in space and time. Proceedings of a symposium organised by International Society of Soil Science (ISSS). Edited by J. Bouma and A.K. Bregt. Wageningen, The Netherlands.

6. Bregt, A.K., 1992. Processing of soil survey data. PhD thesis. Wageningen Agricultural University, Wageningen, The Netherlands.

7. Chen, P.P-S, 1976. The entity – relationship model – towards a unified view of data. ACM Transactions on Database Systems 1, (pp 9 – 36).

8. Cherrill, A. and C. McClean, 1995. An investigation of uncertainty in field habitat mapping and the implications for detecting land cover change. In Landscape Ecology, Vol. 10, No. 1, pp 5-21. SPB Academic Publishing bv, Amsterdam, Netherlands.

9. Clarke, M.C.G., D.G. Woodhall, D. Allen and G. Darling, 1990. Geological, volcanological and hydrogeological controls on the occurrence of geothermal activity in the area surrounding Lake Naivasha, Kenya. Ministry of Energy, Nairobi, Kenya.

10. Cline, M.G., 1949. Basic principles of soil classification. In: Soil Science, Vol. 67, pp 81-91. Rutgers University, New Brunswick, New Jersey 08903. USA.

11. Davis, G.B. and M.H. Olson, 1985. Management Information Systems; Conceptual Foundations, Structure and Development. 2nd Edition. McGraw-Hill, Singapore.

12. De by, R.A., 1999. Spatial Information Theory & Applied Computer Science, Course Notes Databases. Database Design Reader. ITC, Enschde, Netherlands.

13. De By, R.A., G.T.M.T. Dam and C.J.V. Westen (eds.), 1999. Principles of Geographic Information Systems. ITC core module Textbook. ITC, Enschede. Netherlands.

14. Dumanski, J., B. Kloosterman and S.E. Brandon, 1975. Concepts, objectives and structure of the Canada Soil Information System. Can. J. Soil Science Vol. 55: pp 181-187.

15. Dumanski, J., W.W. Pettapiece, D.F. Acton and P.P. Claude, 1993. Application of agro-ecological concepts and hierarchy theory in the design of databases for spatial and temporal characterisation of land and soil. In Geoderma, Vol. 60: pp 343 – 358.Elsevier Science Publishers B.V., Amsterdam.

16. Elmasri, R. And S.B. Navathe, 1989. Fundamentals of Database Systems. The Benjamin/Cummings publishing C, Inc. Redwood California.

17. Farshad, A., 1999. Aerial Photo Interpretation for Soil Survey. Earth Resources and Environmental Geosciences, Module 6, Lecture Note. ITC, Enschede, The Netherlands.

18. Food and Agriculture Organisation of the United Nations (FAO), 1990. Guidelines for Soil Description. 3rd edition (revised). FAO, Rome, Italy.

19. Gatahi, M.M., 1986. Detailed soil survey of the Nini farm, Naivasha, Nakuru District, Kenya. KSS, Detailed soil survey report No. D27.

20. George, W., 1999. Developing an optimised information system to support the operations of survey and mapping department in Uganda. IFA, ITC, Enschede. The Netherlands.

21. Goosen, D., 1967. Aerial Photo Interpretation in Soil Survey. FAO, Soils Bulletin No.6, Rome, Italy.

22. Gordon, T. 1975. Data management: general systems and particular needs. In: Soil Information Systems. Proceedings of the Meeting of the ISSS Working Group on Soil Information Systems, Wagningen, the Netherlands. Sept. 1-4 (ed. S.W. Bie), pp 43-47. Pudoc, Wagningen.

23. Harper, D.M., K.M. Mavuti and S.M. Muchiri, 1990. Ecology and Management of Lake Naivasha, Kenya, in Relation to climatic change, Alien species Introduction and Agricultural Development. Environmental conservation, Vol. 17, No. 4.

24. Hart, D. And H. Phillips, 1998. Metadata primer, a 'How to' Guide on Metadata Implementation. Web site. Http://rat.lic.wise.edu/metadata/metaprim.htm

25. Hornby, A.S., 1995. Oxford Advanced Learner's Dictionary of Current English. Fifth Edition. Edited by J. Crowther, K. Kavanagh and M. Ashby. Oxford University Press, England.

26. Howe, D.R. 1989. Data Analysis for Database Design. Second Edition. Jw Arrowsmith Ltd, Bristol, Great Britain.

27. Hawryszkiewycz, I., 1998 (pp 332-350). Introduction to Systems Analysis and Design. Fourth Edition. Prentice Hall Australia Pty Ltd, Sydney.

28. Jones, C.B., 1985. Database design for variable-scale, worldwide cartography. Presented at Computer Graphics '85: Online Publications, pp 117-129. Pinner, UK.

29. Jones, C.B., J.M. Ware, and D.R. Miller, 2000. Bayesian Probabilstic Methods for Change Detection with Area-Class Maps. Fourth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. July 2000, Amsterdam, Netherlands.

30. Kainz, W., 1992. Multi-scale Databases. In: Computer, Environment and Urban Systems, Vol. 16, pp 271-279. Pergamon Press Ltd. USA.

31. Kamoni, P.T. 1988. Detailed soil survey of a part of Quarantine farm-National Husbandry Research Station. Naivasha, Nakuru District. Kenya. KSS, Detailed soil survey report No. D45.

32. Kenya Wildlife Service Training Institute (KWSTI) 1996. Lake Naivasha Management Implementation Committee Stakeholders Workshop, 8 October 1996. Executive Summary. Naivasha, Kenya.

33. Kenya Soil Survey Staff, 1998. Strategic Plan of Kenya Soil Survey 1998 – 2004. Action plan document. KSS, Nairobi, Kenya.

34. Kirkwood, V., J. Dumanski, K.B. MacDonald, R.B. Stewart, and B. Gronas, 1983. The Land Potential Database. Users' Handbook. Research Branch, Agriculture Canada, Ottawa, pp 65.

35. Kloosterman, B., 1975. Input of data to soil information systems. In: Soil Information Systems, Proceedings of the Meeting of the ISSS Working Group on Soil Information Systems. Wageningen, the Netherlands, Sept. 1-4 (ed. S.W. Bie), 22- 30. Pudoc, Wageningen.

36. Koranteng, A.P., 1998. A Meta-Database System Shell for Analytical GIS. (An application case on AgNPS-GIS linkage for surface Runoff/Erosion Estimation). MSc Thesis. WAU, Wageningen, the Netherlands.

37. Krijger, B., 1995. The development of a Geographical Information System to support environmental management: a case study on the "Biologische Station, Zwillbrock e.v." at Zwillbrock, Germany. MSc thesis, ITC, Enschede, The Netherlands.

38. Kwacha, C.P.H.J., 1998. Vulnerability of Soils to change in Agricultural use around Lake Naivasha, Kenya. MSc Thesis, ITC, Enschede. The Netherlands.

39. Lake Naivasha Riparian Owners Association (LNROA) 1995. Report Document. Naivasha, Kenya.

40. Lewis, P.J. 1994. Information Systems Development. Systems thinking in the field of information systems. Pitman Publishing. Great Britain.

41. Mukumbura, L., 1995. Information System Development for Land Use Change Analysis. Case study on citrus cultivation in Alora, Spain. Msc Thesis, ITC, Enschede. Netherlands.

42. Natalie, M.R., 2000. Improving User Access to Soil Information Systems: the construction of user interfaces for multiple user groups. IFA, ITC, Enschede. The Netherlands.

43. Paresi, C.M.J., 1999. Information System Development. Lecture Note (Reader). (pp 4 - 47) GFM 28, ITC, Enschede. Netherlands.

44. Paresi, C.M.J., and O.U. Essein, 1992. Structured information systems development methodologies, a review. ITC, Enschede, The Netherlands.

45. Prague, C.N. and M.R. Irwin, 1997. Access 97 Bible. IDG Books Worldwide, Inc. An International Data Group Company. Foster City, CA 94404. USA.

46. Reeve, D.E. and J.R. Petch, 1999. GIS, Organisations and People. A Socio-Technical Approach. (pp 13-115) Taylor & Francis Ltd, London, UK.

47. Rossiter, D.G., 1999. Soil Geographic Databases. (SOL. SGDB). Lecture Note. ITC, Enschede, Netherlands.

48. Rossiter, D.G. 2000a. Methodology for Soil Resource Inventories. Lecture Notes and Reference. 2nd Revised Version. Soil Science Division. ITC, Enschede, Netherlands.

49. Rossiter, D.G., 2000b. Soil Classification for Soil Survey. A Compendium of on-line Soil Survey Information. http://www.itc.nl/~rossiter/research/rsrch_ss.html

50. Sadoviski, A.N. and S.T. Bie (eds.), 1978. Developments in Soil Information Systems. Proceedings of the ISSS Working Group on Soil Information Systems, Varna/Sofia, Bulgaria. Centre for Agricultural Publishing and Documentation, Wageningen, the Netherlands.

51. Salgé, F., 1999. National and international data standards. In: Geographical Information Systems. Volume2 (pp 693-706). Management Issues and Applications. Second Edition. Edited by: P.A. Longley, M.F. Goodchild, D.J. Maguire, and D.W. Rhind, 1999. John Wiley & Sons, Inc. New York, USA.

52. Schoeneberger, P.J., D.A. Wysocki, E.C. Benham and W.D. Broderson, 1998. Field book for describing and sampling soils. Natural Resources Conservation Service, USDA, National Soil Survey Centre, Lincoln, NE.

53. Schut, P.H., 1996. Glossary of Terms in Soil Science. Canada Department of Agriculture, Ottawa. Publication 1459, pp 44. Web Site. http://res.agr.ca/CANSIS/GLOSSARY/soil.html

54. Schweitzer, P.N., 2000. Tools for creation of formal metadata. Mail Stop 918, National Centre, U.S. Geological Survey, Reston, VA 20192. On web site: http://geology.usgs.gov/tools/metadata/tools/doc/tkme.html

55. Siderius, W., 1998. Background Information on the Soils, Geology and Landscapes of the Lake Naivasha Area. Handout, ITC, Enschede, the Netherlands.

56. Siderius, W., 1980. Soil Conditions at Kulia Farm (Naivasha). Soil survey report, Naivasha, Kenya.

57. Siderius,W., 1977. Soils and Environmental Conditions of Agricultural Research Stations in Kenya. KSS, soil survey report No. M5.

58. Skidmore, A.K., 2000. Introduction to Non-parametric Statistics and Their Potential. Lecture note for module 12 on research methods. ITC, Enschede, Netherlands.

59. Soil Survey Division Staff, 1993. Soil Survey Manual. United States Department of Agriculture (USDA), Handbook No. 18, pp 266. USDA, USA.

60. Sombroek, W.G., H.M.H. Braun and B.J.A. Vander Pouw, 1980. Exploratory soil map and agro-climatic zone map of Kenya, at scale 1:1000 000. Kenya soil survey, Nairobi, Kenya.

61. Stuttard, M.J. et al 1996. Use of GIS to assist Hydrological Modelling of Lake Basins in the Kenya Rift Valley-conclusions.

62. Thompson, A.O., and R.G. Dodson, 1963. The geology of the Naivasha area. Geology and mines report No. 55, Nairobi, Kenya.

63. Wang, M. and P.J. Howarth, 1994. Multi-source spatial data integration: Problems and some solutions. In: Canadian Journal of Remote Sensing. Vol. 20, No. 4, December, pp 360-367.

64. Worboys, M.F., 1999. Relational databases and beyond. In Geographical Information Systems. Volume1 (pp 373-384). Principles and Technical Issues. Second Edition. Edited by: P.A. Longley, M.F. Goodchild, D.J. Maguire, and D.W. Rhind, 1999. John Wiley & Sons, Inc. New York, USA.

65. Yifter, K.F., 1995. Soil and Landscape Information System from Users Requirement Perspective. (A case study in the Alora Region South of Spain). MSc Thesis, Wageningen Agricultural University, Wageningen. Netherlands.

66. Zinck, J.A. and C.R. Valenzuela, 1990. Soil geographic database: structure and application example. ITC Journal 1990-3: pp 270-294. Enschede, the Netherlands.

67. Zinck, J.A., 1989. Physiography and Soils. Soil Survey Courses. Subject matter: K6. SOL 41 Lecture Note. ITC, Enschede, The Netherlands.

Appendices

Appendix A: Standard Interview Questionnaires

A1: Questionnaire for user needs assessment for Soil Geographic Database design

Objectives of the questionnaire:

- To identify users of soil data/information around Lake Naivasha
- To assess user needs for soil data/information around Lake Naivasha area
- To identify the type of soil data/information mostly required by the user(s)
- To know the level of detail that can be used in organising the soil data of Lake Naivasha area
- To get an impression how efficiently farms utilise soil data/information

1. Date of interview ______________________________
2. Name of interviewer ______________________________
3. Interviewee: 3.1 Name ______________________________
   3.2 Occupation ______________________________
4. Name of Organization/Association/Farm
   ____
5. The smallest field/plot size (Minimum Decision/Management Area) in ha ______________
6. The largest field/plot size (Maximum Decision/Management Area) in ha ______________
7. Total farm size in ha ______________________________
8. Does your farm/organisation have an experience in using soils data/information?
   Yes/No ______________________________
9. For what purpose do you use the data/information?
   ____
   ____
   ____
10. Where do you get soil data/information? ______________________________
    ____
11. How do you get the data/information? ______________________________
    ____
12. In what form do you receive the data/information?
    -In the form of map? Yes/No ______________________________
    -In the form of point/plot data? Yes/No ______________________________
    -In the form of both map and point/plot data? Yes/No ______________________
    -Other (specify) ______________________________
13. Do you have an experience in interpreting soils data? Yes/No ______________________
14. Does your farm have an experience in using computer? Yes/No ______________________
15. Does your farm share soil data/information with others? Yes/No ______________________
    If yes, with whom? ______________________________
    ____
16. Does your farm/organisation need soils data/information in the future? Yes/No______________

____

17. If your answer is yes for question 15, what type of soil data does your farm need?
    - -Do you need raw/unprocessed data/information? Yes/No ______________________________
    - -Do you need processed data/information? Yes/No ______________________________
    - -Do you need both raw and processed? Yes/No ______________________________
18. For what purpose does your farm need soils data/information?

____
____
____

19. Which soil parameters/characteristics are most relevant to your farm? (Make tick mark).
    - -Field observation data at point level ______________________________
    - -Field observation data at plot/field level ______________________________
    - -Laboratory data at point level ______________________________
    - -Laboratory data at plot/field level ______________________________
    - -Map data presented in conventional mapping method (Discrete approach) ________________
    - -Map data presented by interpolation method (Continuous approach) ________________
    - -Other (specify) ______________________________

List some of the main parameters/information you need:

____
____
____
____
____

20. Which data/information format is convenient for your farm?
    1. Do you prefer digital format i.e. in computer system? Yes/No ________________________
    2. Do you prefer analogue format i.e. in paper format? Yes/No ________________________
    3. Do you prefer both formats? Yes/No ___________ if your answer is yes, rank them according to your preference ______________________________
21. Would you explain why you chose _____________ format(s)?

____
____
____

A2: Questionnaire for KSS unit (data processor) for Lake Naivasha area

The objective of this questionnaire is:

- To get background information about the existing traditional working system of KSS for designing Soil Geographic Database for Lake Naivasha area
- To get basic information for the functional analysis of the soil survey organization to develop a preliminary soil information system to provide soil data for lake Naivasha area

1. Date of interview ______________________________
2. Name of interviewer ______________________________
3. Information source: 3.1 Name ______________________________
   3.2 Status/Title ______________________________
4. Name of organization ______________________________

5. What are the main roles & responsibilities (objectives) of your organization?
____
____
____

6. What is your organizational structure? (Top - bottom structure)
____
____
____

7. What are the functional sub units of your organization? (Internal structure)
   1. ____
   2. ____
   3. ____

8. What are most frequently required soil data/information?
____
____
____

9. Where do your organization get soil data/information? (Referring to source)
____
____
____

10. If your organization is collecting primary and/or secondary data, what method(s) do you use for acquisition?
____
____
____

11. For what purpose do you collect soil data/information?
____
____
____

12. Who are the end users of your soil data/information?
____
____
____

12.1 Does your organization conduct a user needs assessment? Yes/No ____________________
   If your answer is yes, when ____
   How ____
   To which group ____
   If your answer is no, why not? _______________________________________
12.2 What is their Minimum Decision Area? _____________________________________
   What is their Maximum Decision Area? ____________________________________

12.3 In what format do the users want the data to be?
   1. Analogue/digital or both ____
   2. Point data/map data or both_______________________________________
   3. Raw/interpreted data or both______________________________________

13. Is there any data gap between the requirements of the end users and what your organization is supplying? Yes/No ____

13.1 If yes, what are this data gaps?

____

____

____

13.2 Why your organization doesn't fill these data gaps?

____

____

____

14. How do your organization store or handle soil data/information?

1. Do you store mostly in digital format? Yes/No ______________________
2. Do you store mostly in analogue format? Yes/No______________________
3. Do you store in both formats? Yes/No ______________ If your answer is yes, please rank your preference ______________________________
4. Other (specify) ______________________________

15. Do you believe that your present soil data/information is efficiently utilized? Yes/No _______

If yes, how ____

If No, why ____

16. Is your soil data/information easily accessible by the users? Yes/No _______________

If Yes, how? ____

If No, why? ____

17. Do you think that for soil data/information handling, digital environment is preferred from that of manual/traditional one? Yes/No ______________________________

Why? ____

____

18. What are the problems that your organization is facing with respect to?

18.1 Data acquisition/collection ______________________________

____

18.2 Data handling/storage______________________________

____

18.3 Data/information dissemination ______________________________

____

19. What possible solutions do you suggest with respect to:

19.1 Data acquisition/collection ______________________________

____

19.2 Data handling/storage______________________________

____

19.3 Data/information dissemination ______________________________

____

20. What is your over all comment/suggestion to improve your organisations' performance better?

____

____

____

A3: Data flow within and/or to and/or out of the sub-units

1. Name of functional/sub unit ____
2. Main responsibilities/objectives of the sub unit

____

____

____

3. What data or information does the sub unit use/demand for its functional process and who is the data/information source?

Input data/information name	Source	Data/Information type/format

4. What information does the sub unit produce/supply and who is the user of the data?

Out put data/information name	User	Data/Information type/format

5. How do you communicate with the data source and/or the data user?

____

____

6. What problems do you face during the communication?

____

____

7. What possible solutions do you suggest?

____

____

Appendix B: Spatial Data Recording Formats

B1: Datasheet for recording compass traverse measurement

1. Date of surveying ____
2. Name of surveyor ____

3. Status of survey: 1. Using compass and measuring tape ___________________________________
2. Using compass and pacing (pace length _________ m) ___________________
3. Other (specify) ____
4. Site/ Area/Farm name ____

Station No:	Forward Bearing In degree	Slope In %	Map Distance In (mm)	Ground Distance In (M)	X Co-ordinate In (M)	Y Co-ordinate In (M)	Remark

Foot Note: ____
____
____

B2: Datasheet for recording co-ordinates of observation points and site locations

1. Date of observation ____
2. Name of observer ____
3. Status of observation: 1. Map measurement-using ruler with precision of _________________ mm
2. Using GPS receiver, single fix Garmin 12XL precision of ___________ m
3. Other (specify) ____
4. Site/Area/Farm name ____

Code (station)	Observation number	X Co-ordinate In (M)	Y Co-ordinate In (M)	Status	Remark Description of location)

Foot Note: ____
____
____
____

Appendix C: Entity Description

C1: Data Dictionary for Entities

Entity Name And Code	Characteristics (Attributes)	Description	Formation
Soil	Horizon symbol, horizon no, pH,	A vertical cross-section of a	It is the result of

Horizon (SOBS_HOR)	colour, consistency, texture, structure, porosity, contents of N, K, Mg, Na, P, mottles, horizon thickness, etc.	pedon. A layer in the soil profile from which soil characteristics are derived in the field as well as in the lab.	soil formation. (Pedogenesis)
Pedon (Observation) (SOBS_SITE)	Observation identifier, surface stoniness, parent material, topography, vegetation, land use, soil depth, internal drainage, external drainage, moisture regime, temperature regime, slope class, easting, northing, flooding, etc.	The smallest three - dimensional representation of a soil individual where soil profile description takes place. The conventional size of a pedon is one meter square to ten meter square.	It is the result of soil formation, which contains one or more soil horizons.
Polypedon (Soil type), (Soil Component) (SCOMP)	Soil component name, soil component identifier, soil component taxonomic code, and soil phase.	A three dimensional representation of the soil body containing one or more similar pedons.	It is formed from the aggregation of several pedons of the same type.
Delineation (Polygon) (SPOLY)	Polygon identifier, area, perimeter	A terrain unit with two-dimensional spatial extent containing same soil type all over that terrain area.	It is formed from the geomorphic processes.
Soil Map Unit (SMU)	Soil map unit identifier, soil map unit type, landscape, relief, lithology, and landform.	A two dimensional representation of the soil body containing one or more delineation.	It is formed by grouping similar polygons that contain same terrain units.
Soil Map (SMAP)	Map identifier, map name, map scale, classification system, survey approach, author, year of publication, and remark.	A tow dimensional representation of the soil continuum that contains one or more soil map units.	It is a higher level aggregation or grouping of delineations.
Soil Map Unit Component (SMUCOMP)	Soil map unit identifier, component identifier, and component percentage i.e. proportion.	A combination or association entity which is not an object in reality to break many to many relation.	It is a combination of SMU and SCOMP entities.

C2: Internal Design Parameters and Specifications of Entities

Entity Name: Soil Map (SMAP) Entity Type: Object in reality

No	Field Name	Data Type	Field Size	Required	Allow Zero Length	Indexed	Description
1	Map_Id	Text	20	Yes	No	Yes	Unique map identifier

2	Map_Na	Text	50	No	Yes	No	Map name
3	Map_Sc	Text	50	No	Yes	No	Map scale
4	Class_Sy	Text	50	No	Yes	No	Classification system
5	Survev_ap	Text	50	No	Yes	No	Survey approach
6	Author	Text	50	No	Yes	No	Author(s) of the map
7	Year	Number	Integer	No	Yes	No	Year of publication
8	Remark	Text	150	No	Yes	No	Additional remark

Entity Name: Soil Polygon (SPOLY) Entity Type: Object in reality

No	Field Name	Data Type	Field Size	Re-quired	Allow Zero Length	In-dexed	Description
1	Poly_Id	Text	10	Yes	No	Yes	Unique polygon identifier
2	SMU_Id	Text	10	No	Yes	Yes	Foreign key
3	Area	Number	Long Integer	No	Yes	No	Area of the polygon
4	Perimeter	Number	Long Integer	No	Yes	No	Perimeter of the polygon

Entity Name: Soil Map Unit (SMU) Entity Type: Object in reality

No	Field Name	Data Type	Field Size	Re-quired	Allow Zero Length	Indexed	Description
1	SMU_Id	Text	10	Yes	No	Yes	Unique soil map unit identifier
2	Map_Id	Text	20	No	Yes	Yes	Foreign key
3	SMU_ty	Text	20	No	Yes	No	Soil map unit type
4	Landscape	Text	50	No	Yes	No	Landscape type
5	Relief	Text	50	No	Yes	No	Relief type
6	Lithology	Text	50	No	Yes	No	Lithology/facies type
7	Landform	Text	50	No	Yes	No	Landform type

Entity Name: Soil Map Unit Component (SMUCOMP) Entity Type: Non-Object in reality

No	Field Name	Data Type	Field Size	Re-quired	Allow Zero Length	Indexed	Description
1	SMU_Id	Text	10	Yes	No	Yes	Primary & Foreign key
2	COMP_Id	Text	20	Yes	No	Yes	Primary &Foreign key
3	Percent	Number	Single	No	Yes	No	Percentage of composition

Entity Name: Soil Component (SCOMP) Entity Type: Object in reality

No	Field Name	Data Type	Field Size	Re-quired	Allow Zero Length	Indexed	Description
1	COMP_Id	Text	20	Yes	No	Yes	Unique component identifier
2	COMP_taxa	Text	20	No	Yes	No	Coded taxonomic name
3	COMP_Na	Text	50	No	Yes	No	Component name
4	Solphase	Text	50	No	Yes	No	Soil phase

Entity Name: Soil Observation at Site (SOBS_SITE) Entity Type: Object in reality

No	Field Name	Data Type	Field Size	Re-quired	Allow Zero Length	In-dexed	Description
1	OBS_Id	Text	6	Yes	No	Yes	Unique observation identifier
2	COMP_Id	Text	20	No	Yes	Yes	Foreign key
3	Status	Text	10	No	Yes	No	Status of soil profile description
4	Date	Date/ Time	Short date	No	Yes	No	Date of profile description
5	Author	Text	50	No	Yes	No	Name of the surveyor
6	Location	Text	150	No	Yes	No	Narrative location description
7	Photo-Id	Text	6	No	Yes	No	Photograph identifier number
8	Topomap	Text	20	No	Yes	No	Topographic map name/number
9	UTM	Text	6	No	Yes	No	UTM zone
10	Easting	Number	Long integer	No	Yes	No	Longitudinal co-ordinate of the observation point
11	Northing	Number	Long integer	No	Yes	No	Latitudinal co-ordinate of the observation point
12	Elevation	Number	Long integer	No	Yes	No	Height of the observation point above sea level
13	FAO88	Text	50	No	Yes	No	FAO 1988 soil classification
14	WRB_DH	Text	50	No	Yes	No	WRB diagnostic horizons
15	WRB_2ND	Text	50	No	Yes	No	WRB 2nd level qualifiers
16	WRB_RSG	Text	50	No	Yes	No	WRB reference soil group
17	KST_DH	Text	50	No	Yes	No	KST diagnostic horizons
18	KST_SG	Text	50	No	Yes	No	KST sub group
19	KST_GG	Text	50	No	Yes	No	KST great group
20	KST_F	Text	50	No	Yes	No	KST family
21	Moist_reg	Text	50	No	Yes	No	Soil moisture regime
22	Temp_reg	Text	50	No	Yes	No	Soil temperature regime
23	Parent_m	Text	50	No	Yes	No	Soil parent material
24	Over	Text	50	No	Yes	No	Parent material underlying rock
25	Rock-ty	Text	50	No	Yes	No	Rock type in the site
26	Topography	Text	50	No	Yes	No	Topography of the site
27	Micro-top	Text	50	No	Yes	No	Micro-topography of the site

28	Slope_cl	Text	50	No	Yes	No	Slope class of the surface
29	Slope_for	Text	50	No	Yes	No	Slope form of the surface
30	Slope_gr	Text	50	No	Yes	No	Slope gradient of the surface
31	Slope_len	Text	50	No	Yes	No	Slope length of the surface
32	Landuse	Text	50	No	Yes	No	Land use of the site
33	Crops	Text	50	No	Yes	No	Crops covering the surface
34	Human_inf	Text	50	No	Yes	No	Human influence on the site
35	Vegetation	Text	50	No	Yes	No	Vegetation type of the site
36	Grass_cov	Text	50	No	Yes	No	Grass cover of the site (percent)
37	Eff_soil_dep	Text	50	No	Yes	No	Effective soil depth
38	Rock_out	Text	50	No	Yes	No	Rock outcrop cover
39	Rock_dis	Text	50	No	Yes	No	Avg. distance b/n rock outcrops
40	Stoniness	Text	50	No	Yes	No	Surface stoniness
41	Stonin_si	Text	50	No	Yes	No	Surface stoniness size
42	Erosion_ty	Text	50	No	Yes	No	Surface erosion type
43	Erosion_ar	Text	50	No	Yes	No	Total area affected by erosion
44	Erosion_deg	Text	50	No	Yes	No	The degree of erosion
45	Sealing	Text	50	No	Yes	No	Soil surface sealing
46	Cracks_wd	Text	50	No	Yes	No	Surface cracks width
47	Cracks_dis	Text	50	No	Yes	No	Distance between cracks
48	Salt_cov	Text	50	No	Yes	No	Salt cover at the surface
49	Drainage	Text	50	No	Yes	No	Soil drainage class
50	Int_dr_sat	Text	50	No	Yes	No	Internal drainage saturation
51	Hyd_con	Text	50	No	Yes	No	Permeability/hydraulic conduct.
52	Ext_drain	Text	50	No	Yes	No	External drainage of the site
53	Flood_fr	Text	50	No	Yes	No	Flooding frequency
54	Flood_du	Text	50	No	Yes	No	Flooding duration
55	Flood_dep	Text	50	No	Yes	No	Flooding depth
56	Gr_wat_dep	Text	50	No	Yes	No	Ground water depth
57	Mois_con	Text	50	No	Yes	No	Moisture conditions
58	Remark	Text	150	No	Yes	No	Additional remarks of the site
Note:	USDA = United States Department of Agriculture						
	UTM = Universal Transverse Mercater Projection						
	FAO = Food and Agriculture Organisation of the United Nations						
	WRB = World Reference Base for Soil Classification						
	KST = keys to Soil Taxonomy according to USDA Soil Classification						

Entity Name: Soil Observation at Horizon (SOBS_HOR)Entity Type: Object in reality

No	Field Name	Data Type	Field Size	Re-quire d	Allow Zero Length	In-dexed	Description
1	OBS_Id	Text	6	Yes	No	Yes	Unique observation identifier
2	HOR_No	Number	Byte	Yes	No	Yes	Horizon number

3	HOR_Sym	Text	6	No	Yes	No	Horizon symbol
4	UPP_Lim	Number	Byte	No	Yes	No	Horizon upper limit
5	LOW_Lim	Number	Byte	No	Yes	No	Horizon lower limit
6	Distinctness	Text	6	No	Yes	No	Horizon boundary distinctness
7	Topobound	Text	6	No	Yes	No	Horizon boundary topography
8	Color_dry	Text	10	No	Yes	No	Soil colour dry
9	Color_moi	Text	10	No	Yes	No	Soil colour moist
10	Mottle_ab	Text	50	No	Yes	No	Abundance of mottles
11	Mottle_col	Text	50	No	Yes	No	Colour of the mottles
12	Mottle_cont	Text	50	No	Yes	No	Colour contrast b/n mottles
13	Mottle_si	Text	50	No	Yes	No	Size class of mottles
14	Texture_cl	Text	50	No	Yes	No	Soil texture classes
15	Rock_fr_ab	Text	50	No	Yes	No	Abundance of rock fragments
16	Rock_fr_si	Text	50	No	Yes	No	Size of rock fragments
17	Rock_fr_sh	Text	50	No	Yes	No	Shape of rock fragments
18	Structure_ty	Text	50	No	Yes	No	Type of soil structure
19	Structure-gr	Text	50	No	Yes	No	Soil structure grade
20	Structure_si	Text	50	No	Yes	No	Soil structure size
21	Consist_dr	Text	50	No	Yes	No	Consistence when dry
22	Consist_mo	Text	50	No	Yes	No	Consistence when moist
23	Plasticity	Text	50	No	Yes	No	Soil plasticity when wet
24	Stickiness	Text	50	No	Yes	No	Soil stickiness when wet
25	Pore_ab	Text	50	No	Yes	No	Abundance of pores/voids
26	Pore_si	Text	50	No	Yes	No	Size of pores/voids
27	Porosity	Text	50	No	Yes	No	Total volume of voids/pores
28	Cutans_ab	Text	50	No	Yes	No	Abundance of cutanic features
29	Cement_na	Text	50	No	Yes	No	Nature of cementation
30	Nodule_ab	Text	50	No	Yes	No	Abundance of mineral nodules
31	Root_ab	Text	50	No	Yes	No	Abundance of plant roots
32	Root_si	Text	50	No	Yes	No	Size of plant roots
33	Biolog_act	Text	50	No	Yes	No	Abund. of biological activities
34	$CaCO_3$_field	Text	50	No	Yes	No	Calcium carbonate content
35	pH_field	Number	Single	No	Yes	No	Acidity measured in the field
36	pH_KCl	Number	Single	No	Yes	No	Acidity measured in KCl
37	pH_H_2O	Number	Single	No	Yes	No	Acidity measured in water
38	H_2O_clay	Number	Single	No	Yes	No	Water dispersible clay
39	Sand	Number	Integer	No	Yes	No	Proportion of sand content
40	Silt	Number	Integer	No	Yes	No	Proportion of silt content
41	Clay	Number	Integer	No	Yes	No	Proportion of clay content
42	$CaSO_4$	Number	Single	No	Yes	No	Calcium sulphate content
43	$CaCO_3$	Number	Single	No	Yes	No	Calcium carbonate content
44	Ca	Number	Single	No	Yes	No	Calcium content in the soil
45	Mg	Number	Single	No	Yes	No	Magnesium content in the soil
46	K	Number	Single	No	Yes	No	Potassium content in the soil

47	Na	Number	Single	No	Yes	No	Sodium content in the soil
48	Mn	Number	Single	No	Yes	No	Manganese content in the soil
49	Sum_cation	Number	Single	No	Yes	No	Sum of cations in the soil
50	N	Number	Single	No	Yes	No	Nitrogen content in the soil
51	C	Number	Single	No	Yes	No	Organic carbon content in soil
52	OM	Number	Single	No	Yes	No	Organic matter content in soil
53	P	Number	Single	No	Yes	No	Phosphorus content in the soil
54	P-olsen	Number	Single	No	Yes	No	Phosphorus content by olsen
55	EC	Number	Single	No	Yes	No	Electrical conductivity
56	ESP	Number	Single	No	Yes	No	Exch. sodium percentage
57	CEC	Number	Single	No	Yes	No	Cation exchangeable capacity
58	CEC_clay	Number	Single	No	Yes	No	Cation exch. capacity in clay
59	BS	Number	Single	No	Yes	No	Base saturation
60	Remark	Text	150	No	Yes	No	Additional remark

Appendix D: Data Dictionary for Processes and Data Stores

D1: Data Dictionary for Data Processes

Process No:	Process Name	Process Description	Input data	Output data	Process Frequency
1	Data Collection & Interpretation Subsystem	Primary soil data collection from the field, interpretation of the soil data, preparation of soil, thematic maps and interpretation maps.	Topographic data Climatic data Aerial photo data Management data Geomorphology data Secondary data Land use data Geological data	Draft maps and reports, Final report documents	High

			Vegetation data Analytical data		
2	GIS and Computing Subsystem	Analogue to digital conversion of draft maps, thematic map calculation, and image processing	Topographic data Digital elevation data Aerial photo data Satellite image data Analytical data Draft thematic maps and reports	Final thematic maps and report documents	Very High
3	Laboratory Subsystem	Analysis of the samples in the laboratory for chemical and physical characteristics of soils.	Water samples Soil samples Plant tissue	Physical and chemical data	High
4	Information Dissemination Subsystem	Processing of the request data from clients, disseminating data or information to users via digital and/or analogue format	Request data, Map data Report document Analytical data	Advice on soil management, report documents, water quality data, request data, thematic map data, interpretation data, fertility appraisal reports, fertiliser recommendations, analytical data,	High

D2: Data Dictionary for Data Stores

Data Store Name	Store composition / description	Updated by process No:	Consulted by process No:	Store organisation and security	Data volume
Survey Data Registry	Draft thematic maps and reports. This includes all the primary and secondary data collected from the field, aerial photo interpretation maps, draft thematic soil maps, soil interpretation maps, and narrative report documents data.	1	2, 3,	Some in archives in analogue format and some in digital format, accessed by staff members	High
Laboratory Data Registry	Analytical data. This contains physical and chemical analysis of soil characteristic data.	3	1, 2, 4,	Digital data, accessed by staff members	Medium
Final Data	Finalised maps and report documents i.e. different soil related	1, 2	4	Digital data accessed by staff	High

Registry	thematic maps and descriptive report documents.			members.	
Request Data Registry	Request data i.e., different requests /appeals for soils data or information by clients/users.	4	1, 2, 3,	Digital and analogue data that can be accessed by external users.	High

D3: Data flow into the system and out of the system with respect to the selected users

No	User/Terminator	Out flow data from the system	Inflow data to the system	Data format
1	Commercial farms	Water quality data, nutrient level in plants, soils and water data, fertiliser recommendation data, soil and other thematic map data, narrative report document data,	Request data Soil sample Water sample Plant tissue	Digital and or analogue and real objects.
2	Urban planners	Physical and chemical soil analytical data and detailed soil map data with narrative report document.	Request data Soil sample Water sample	Digital and or analogue and real objects.
3	Soil researchers And consultants	Soil and other thematic maps and physical and chemical analytical data.	Request data Soil sample Water sample Plant tissue	Digital and or analogue and real objects.

Appendix E: Correlation Tables

E1: Correlation Table for Diagnostic Horizons and Properties

No	Approximate Equivalents of Diadnostic Horisons & Properties	
	FAO	USDA
1	Histic H horizon	Histic epipedon
2	Mollic A horizon	Mollic epipedon
3	Fimic A horizon	Plaggen & Anthropic epipedon
4	Umbric A horizon	Umbric epipedon
5	Ochric A horizon	Ochric epipedon
6	Argic B horizon	Argllic epipedon
7	Ferralic B horizon	Oxic and Kandic horizons
8	Natric B horizon	Natric horizon
9	Cambic B horizon	Cambic horizon
10	Spodic B horizon	Spodic horizon
11	Calcic horizon	Calcic horizon

12	Petrocalcic horizon	petrocalcic horizon
13	Gypsic horizon	Gypsic horizon
14	Petrogypsic horizon	Petrogypsic horizon
15	Sulfuric horizon	Sulfuric horizon
16	Albic E horizon	Albic materials
17	Abrupt textural change	Abrubt textural change
18	Andic properties	Andic soil properties
19	Calcareous	Strong effervescence with 10% HCl
20	Calcaric	Calcareous throughout 20 to 50 cm layer
21	Continuous hard rock	Lithic contact
22	Ferralic properties	Apparent CEC <24 cmol/kg clay
23	Ferric properties	Course red mottles
24	Fluvic properties	Irregular C content decrease w/depth
25	Geric properties	ECEC 1.5 cmol(+) /kg or less
26	Gleyic & stagnic properties	Aquic conditions
27	Gypsiferous	5% or more gypsum
28	Interfingering	Interfingering of Albic materials
29	Nitic properties	30% or more clay; angular blocky structure with shiny ped faces
30	Organic soil material	Organic soil material
31	Permafrost	Permafrost
32	Plinthite	Plinthite
33	Salic properties	Salic horizon
34	Slickensides	Slickensides
35	Smeary consistence	Thixotropic soil materials
36	Sodic properties	CEC has 15% or more Na or 50% or more Na + Mg
37	Soft powdery lime	Identifiable secondary carbonates
38	Strongly humic	More than 1.4g organic carbon per 100 g soil in 100cm of depth; assumes 1.5 Mg/m^3 C
39	Sulfidic materials	Sulfidic materials
40	Tonguing	Interfingering of Albic material (thicker)
41	Vertic properties	Some slickensides or parallelepiped structure but not a Vertisol
42	Weatherable minerals	Weatherable minerals

(Source: Buol et al., 1997 pp 212 (obtained from FAO 1988 and USDA 1994))

E2: Correlation Table for Soil Types

No:	Approximate Equivalents of Soil Types	
	FAO	USDA
1	Eutric Cambisols	Typic Xerochrepts
2	Eutric Cambisols	Typic Eutrochrepts
3	Eutric Cambisols	Typic Xerorthents
4	Haplic Luvisols	Andic Haploxeralfs

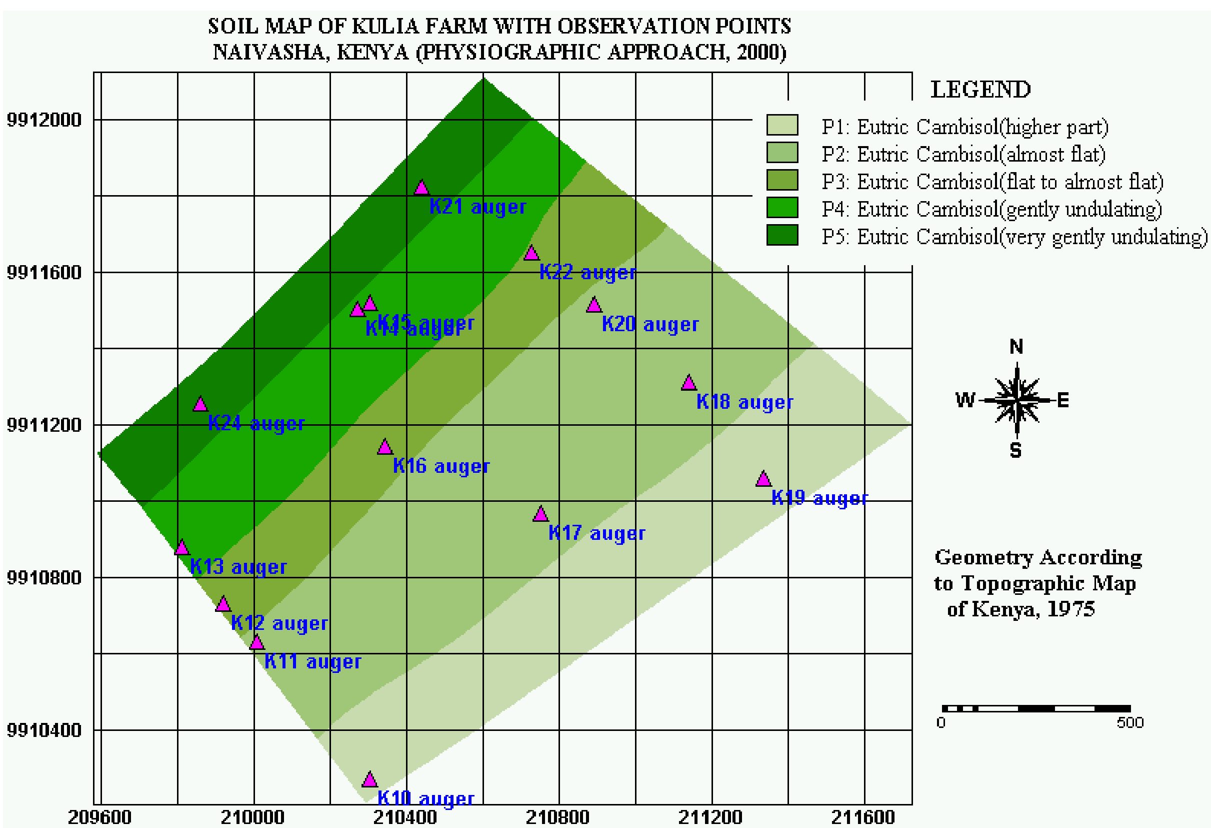
5	Haplic Luvisols	Typic Haploxeralfs
6	Haplic Fluvisols	Andic Xerorthents
7	Haplic Andosols	Andic Xerochrepts
8	Lithic Regosols	Lithic Xerorthents
9	Ando-calcaric Regosols	Calcic Xerorthents
10	Calcaric Fluvisols	Calcic Haploxeralfs
11	Luvic Calcisol	Typic Calcixerepts
12	Calcic Luvisol	Calcic Haploxeralf
13	Epi-calcic Luvisol	Lithic Hapolxeralf
14	Petric Calcisol	Petrocalcic Haploxeralf
15	Calci-rhodic Luvisol	Calcic Rhodoxeralf
16	Hypocalcic Luvisol	Typic Haploxeralfs
17	Haplic Calcisol	Typic Calcixerepts
18	Lepti-calcaric Cambisol	Lithic Hapolxeralf
19	Stagni-calcaric Cambisols	Fluventic Haploxerept
20	Calcaric Cambisol	Typic Haploxeralfs
21	Vertic Calcisol	Vertic Calcixerepts
22	Stagni-calcaric Cambisols	Fluvaquentic Endoaquepts
23	Calcaric Cambisol	Calcic Haploxerept
24	Leptic Cambisol	Lithic Hapolxeralf
25	Rhodi-skeletic Luvisol	Typic Rhodoxeralf
26	Rhodi-cutanic Luvisol	Vertic Rhodoxeralf
27	Hypercalcic Calcisol	Typic Calcixerepts
28	Verti-calcaric Luvisol	Vertic Haploxerepts
29	Calci-skeletic Cambisol	Typic Haploxeralfs
30	Fluvi-mollic Cambisol	Fluventic Haploxerept
31	Haplic Calcisol	calcic Haploxeralf
32	Verti-fluvic Cambisol	Vertic Endoaquepts
33	Hypocalci-rhodic Luvisol	Typic Haploxeralfs
34	Rhodi-cutanic Luvisol	Typic Rhodoxeralf
35	Skeleti-haplic Calcisol	Typic Calcixerepts
36	Cutanic Luvisol	Typic Haploxeralfs
37	Hypocalcic Calcisol	Typic Calcixerepts
38	Gleyi-sodi-calcic Solonchaks	Typic Halaquepts
39	Hypocalcic Luvisol	Calcic Haploxeralf
40	Petric Calcisol	Petrocalcic Calcixerepts
41	Orthicalcic Calcisol	Typic Calcixerepts
42	Epipetric Calcisol	Petrocalcic Calcixerepts
43	Petric Calcisol	Lithic Calcixerepts
44	Rhodic Cambisols	Typic Haploxerepts
45	Epi-gleyic Cambisol	Aquertic Entrudepts
46	Luvi-hypocalcic Calcisol	Typic Calcixerepts
47	Calcic Cambisol	Calcic Haploxerept
48	Gleyic Cambisol	Aquertic Entrudepts

49	Chromic Cambisol	Typic Hapludox
50	Eutric Fluvisol	Andic Xerorthents

Source: Kwacha, 1998 and Antequera Soil Database (FAO 1988 and USDA 1994).

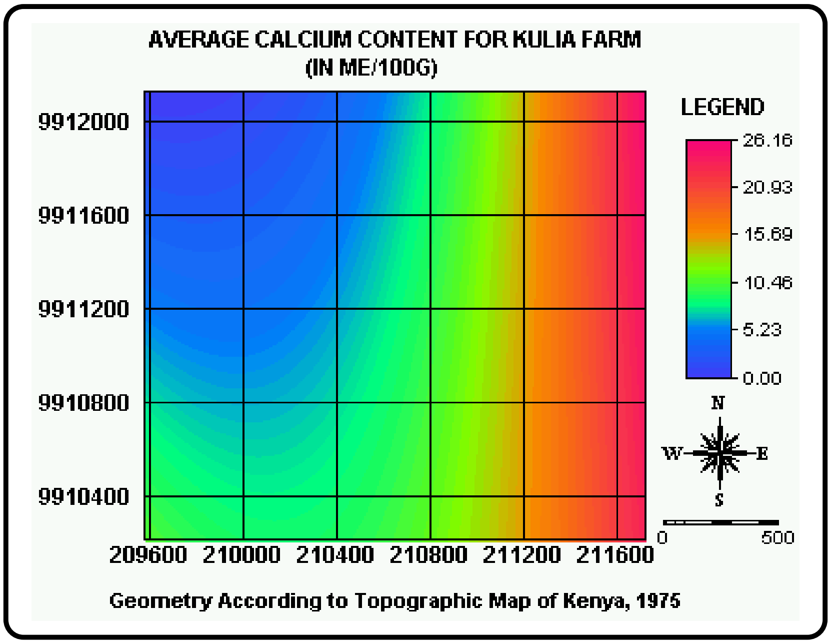
Appendix F: Additional Thematic Maps

F1: Soil map of Kulia Farm with Observation Points



(Source: Analogue soil sketch map by Siderius, 1980)

F2: An Example of Interpolated/Continuous Map Indicating Ca Content Distribution



Appendix G: Metadata

Soil Geographic Database for Lake Naivasha Area

Available as [Questions & Answers] - [Outline] - [Parseable text] - [SGML] - [XML] - [DIF]

Frequently-anticipated questions:


Soil Geographic Database for Lake Naivasha Area.htm

- What does this data set describe?
    1. How should this data set be cited?
    2. What geographic area does the data set cover?
    3. What does it look like?
    4. Does the data set describe conditions during a particular time period?
    5. What is the general form of this data set?
    6. How does the data set represent geographic features?
    7. How does the data set describe geographic features?
- Who produced the data set?
    1. Who are the originators of the data set?
    2. Who also contributed to the data set?
    3. To whom should users address questions about the data?
- Why was the data set created?
- How was the data set created?
    1. Where did the data come from?
    2. What changes have been made?
- How reliable are the data; what problems remain in the data set?
    1. How well have the observations been checked?
    2. How accurate are the geographic locations?

  3. How accurate are the heights or depths?
  4. Where are the gaps in the data? What is missing?
  5. How consistent are the relationships among the data, including topology?
- How can someone get a copy of the data set?
  1. Are there legal restrictions on access or use of the data?
  2. Who distributes the data?
  3. What's the catalogue number I need to order this data set?
  4. What legal disclaimers am I supposed to read?
  5. How can I download or order the data?
- Who wrote the metadata?

What does this data set describe?

Title: Soil Geographic Database for Lake Naivasha Area
Abstract:

The soil geographic database of Lake Naivasha area covers east, southeast and some part from the south of the lakeshore. The soil surveys were conducted at 4 different times and situations.
1. In October, 1997, for the purpose of academic learning in research. The total area covered by this survey was 4613 ha (i.e. Kwacha's study area). The surveyor used topographic map of Kenya, 1975 at scale 1:50000 and aerial photos of 1984 at scale 1:12500 to prepare an interpretation map which later he used as the base map to produce the soil map of the area at scale 1:50000. The result was finalised on April 1998.

2. In December 1986, covering an area of 43 hectares (Quarantine farm). At this survey the topographical survey map at scale 1:50000, produced by B.G. Mwangi and S. Wataka (both from KSS) was used as the base map to produce the soil sketch map of the farm at scale 1:5000. The survey report was completed in 1988.

3. In February 1982, covering an area of 46 ha (Nini farm). For this survey an uncontrolled map of the farm at scale 1:1000, produced by the Provincial Irrigation Unit, Nakuru, was used as a base map to produce the soil sketch map of the farm. The farm was revisited on 25th May 1983 to investigate an alleged non- significant response to fertiliser applications by crops on the farm. The final report was released in 1986.

4. In May 1980, covering an area of 200 ha (Kulia farm). The type of base map used for this survey was not known. Furthermore, no full or mini pit profile was described during the survey. Therefore, the subdivisions of the mapping units were made based on augerhole descriptions only.

From the above survey results an integrated map was produced by reinterpretation using an ortho-photo mosaic map as a common base map for reinterpretation. Most spatial boundaries are similar to that of Kwacha's study area soil map.

Most soils of the study area are moderately well drained to well drained, moderately deep to very deep, but also shallow soils are found mainly over the raised ridges in the volcanic plain.

The top soil varies from a very dark brown to dark brown, silty clay loam to sandy clay, and the subsoil varies from a dark greyish brown to yellowish brown, which is silty clay to sandy loam, and the soil pH varies from 4.7 to 10.6. According to USDA soil classification system soils identified belong to three orders. These are Alfisols, Inceptisols and Entisols.

1. How should this data set be cited?

Tilaye Bitew Bezu (ITC), 2001, Soil Geographic Database for Lake Naivasha Area: International Institute for Aerospace Survey & Earth Sciences (ITC), Enschede, the Netherlands.

Online Links:

The data can be accessed through Internet by the path
NetworkNeighbourhood/EntireNetwork/ha2/Pc4014-03/Naivasha/Nested

Other_Citation_Details:

This digital data was captured from four different Citations.

1. From the analogue soil survey report document entitled as "Detailed Soil survey of a part of Quarantine farm-National Husbandry Research Station, Naivasha, Nakuru district". Detailed soil survey report No. D45, 1988, by P.T. Kamoni, staff at KSS.

2. From the analogue soil survey report document entitled as "Soil conditions at Kulia farm" 1980, Naivasha, by Dr. Walter Siderius, staff at ITC.

3. From the analogue soil survey report document entitled as "Detailed Soil survey of the Nini Farm, Naivasha (Nakuru District)", Detailed soil survey report No. D27, 1986, by M.M. Gatahi, staff member of KSS.

4. From the analogue MSc thesis document entitled as "Vulnerability of Soils to Change in Agricultural use Around Lake Naivasha, Kenya". MSc thesis, 1998, by C.P.H.J. Kwacha. ITC, Enschede.

2. What geographic area does the data set cover?

West_Bounding_Coordinate: 36.342 degrees
East_Bounding_Coordinate: 36.448 degrees
North_Bounding_Coordinate: -0.730 degrees
South_Bounding_Coordinate: -0.845 degrees

3. **What does it look like?**

4. Does the data set describe conditions during a particular time period?

Calendar_Date: 27-Dec-2000

Currentness_Reference: Publication date

5. What is the general form of this data set?

Geospatial_Data_Presentation_Form:
Vector digital data, Raster digital Data & Tabular digital data

6. How does the data set represent geographic features?

a. How are geographic features stored in the data set?
Indirect_Spatial_Reference:
The geometry of the spatial data set is according to the Geometric definition of the Topographic map of Kenya, 1975
This is a Vector data set.
b. What co-ordinate system is used to represent geographic features?

Grid_Coordinate_System_Name: Universal Transverse Mercator
Universal_Transverse_Mercator:
UTM_Zone_Number: -37
Transverse_Mercator:
Scale_Factor_at_Central_Meridian: 0.9996
Longitude_of_Central_Meridian: 39.0
Latitude_of_Projection_Origin: 0.0
False_Easting: 500000
False_Northing: 10000000

Planar coordinates are encoded using Coordinate pair
Abscissae (x-coordinates) are specified to the nearest 0.01 meter
Ordinates (y-coordinates) are specified to the nearest 0.01 meter
Planar coordinates are specified in meters

7. How does the data set describe geographic features?

Entity_and_Attribute_Overview:
The spatial database of Lake Naivasha area (east, southeast & some part from south of the lakeshore) contains point location map of auger and pit observations, a soil polygon map, a raster map showing pH, and different nutrient levels. All the maps contain the word "**integ"** on their names if they are concerned with the integrated soil map of Lake Naivasha area.

The non-spatial soil geographic database of Lake Naivasha area contains a total of seven entities where one of them is a combination entity. In addition to the entities a number of help (lookup) tables are also included. These help tables are already linked to corresponding entities, mainly to SOBS_SITE and SOBS_HOR entities. Summary of the information contained in each entity data set is the following.

1. Entity Name = Soil Map, Code = SMAP The soil map entity contains 8 attributes. These are Map_Id, map name, map scale, classification system, survey approach, author, and year of publication and general remark.

2. Entity Name = Soil Map Unit, Code = SMU The soil map unit entity contains 7 attributes. These are SMU_Id, Map_Id, SMU type, landscape, relief, lithology and landform.

3. Entity Name = Soil Polygon, Code = SPOLY The soil polygon table contains 4 attributes. These are Poly_Id, SMU_Id, area and perimeter. The area and the perimeter are indicated for each polygon or what we call delineation.

4. Entity Name = Soil Component, Code = SCOMP The soil component entity contains 4 attributes. These are Comp_Id, Comp_tax, and component name and soil phase. Component taxa contain coded taxonomic names in standard classification system where as component name contains full taxonomic soil component name.

5. Entity Name = Soil Observation at Site, Code = SOBS_SITE The soil observation entity at site level contains more than 60 attributes. For example: OBS_Id, COMP_Id, status, date, location easting, northing, diagnostic horizon, soil classification name, moisture regime, temperature regime, parent material, slope class landuse, human influence, soil depth, surface sealing, vegetation, drainage, ground water depth, salt cover, etc.

6. Entity Name = Soil Observation at Horizon, Code = SOBS_HOR The soil observation entity at horizon level contains more than 50 attributes whose values are recorded for each horizon in a soil profile. The entire record is uniquely identified by the combination of observation number and horizon number. Some of the attributes included in this entity are: OBS_Id, HOR_No, HOR_Sy, horizon upper limit & lower limit, soil colour, soil texture, soil structure, mottles, pH, macro and micro nutrient level, carbon content, phosphorus content, electrical conductivity, etc.

7. Entity Name = Combination/associate, Code = SMUCOMP The soil map unit - Component combination entity contains 3 attributes, which are SMU_Id, COMP_Id and percentage composition of the soil component in a mapping unit. This entity has no more information, but it has been introduced to break the many to many relationship between SMU and SCOMP entities in to one to many relationship.

Note: (Useful Information): On Lake Naivasha area soil geographic database the following remarks should be considered before using the data.

A. DATA OBTAINED FROM SIDERIUS, 1980 (ABOUT KULIA FARM):

1. The profile description was not continuous from top to down. The soil samples were taken at fixed depths, e.g. 0-20cm, 50-60 cm, and 90-100cm. In some cases even only two horizons. All the corresponding record values in the database are for these depths. Therefore, for the convenience of the data store in the database the fixed depths were labelled as A, B, and C horizons irrespective of the horizon symbol designation. The same was done for the horizon number.

2. The analytical data stored in the Kulia database covers only for these fixed soil depths. No more information out of these depths is available.

B. DATA OBTAINED FROM GATAHI, 1986 (ABOUT NINI FARM):
1. Even if a total of 53 augerholes were described, no one of them was included in the database. The reason for this is that the descriptions and respective data for the augerholes were not included in the reference report document D27, 1986.

2. The analytical results for composite samples for fertility analysis which were taken at a depth of 0-30cm from the vicinity of each profile pit were recorded as values of first horizon together with the data from the corresponding pit. These attributes are C%, N%, Mn (me/100g), P (ppm) and P-Olsen (ppm).

C. DATA OBTAINED FROM KAMONI, 1988 (ABOUT QUARANTINE FARM):
1. Even if a total of 30 augerholes were described, no one of them was included in the database. The reason for this is that the descriptions and the respective data for the augerholes was not included in the reference report document D45, 1988.

2. The analytical results for composite samples for fertility analysis which were taken at a depth of 0-30cm from the vicinity of each profile pit were recorded as values of first horizon together with the data from the corresponding pit. These attributes are C%, N%, Mn (me/100g), P (ppm) and P-Olsen (ppm).

D. DATA OBTAINED FROM KWACHA, 1998 (ABOUT SOUTH EAST LAKE SHORE):
1. Most of the chemical analytical data obtained from the lab analysis was for fixed soil depths. These depths are at topsoil, at 50cm depth and at 120cm depth. However, in the database the values are recorded for the horizon whose depth interval includes the corresponding fixed depth.

Entity_and_Attribute_Detail_Citation:

The citation was made from four sources. These are:

1. Siderius, W., 1980. Soil conditions at Kulia Farm, Naivasha. Soil survey report document with soil sketch map, scale 1:10000. KSS, Nairobi, kenya.

2. Gatahi, M.M., 1986. Detailed soil survey of the Nini farm, Naivasha, Nakuru district. Detailed soil survey report No. D27, 1986 with soil sketch map, scale 1:5000. KSS, Nairobi, kenya.

3. Kamoni, P.T., 1988. Detailed Soil Survey of a part of Quarantine Farm-National Husbandry Research Station, Naivasha, Nakuru District. Detailed soil survey report No. D45, 1988 with soil sketch map, at scale 1:5000. KSS, Nairobi, Kenya.

4. C.P.H.J. Kwacha, 1998. Vulnerability of soils to change in agricultural use around lake Naivasha, Kenya. MSc thesis. ITC, Enschede, The Netherlands.

Who produced the data set?

1. **Who are the originators of the data set?** (may include formal authors, digital compilers, and editors)

   Tilaye Bitew Bezu (ITC)

2. **Who also contributed to the data set?**

   The data set credit goes to the following individuals, staff members and institutions.

   1. To Mr. C.P.H.J. Kwacha (who carried out the survey for an area covering about 4613 ha mainly east of the lake Naivasha shore), to ITC library and soil science division staff who made possible the availability of the copy of the original paper data.

   2. To Mr. P.T. Kamoni who carried out the survey for Quarantine farm and to Dr. W. Siderius who made possible the availability of the copy of the original paper data.

   3. To Mr. M.M. Gatahi who carried out the survey for Nini farm and to Dr. W. Siderius who made possible the availability of the copy of the original paper data.

   4. To Dr. W. Siderius who carried out the survey for Kulia farm and made possible the availability of the copy of the original paper data.

3. **To whom should users address questions about the data?**

   Soil Science Division, ITC
   C/o Dr. D.G.Rossiter
   Head of Soil Science Division
   ITC, Hengelose straat 99, 7500 AA
   Enschede, Overijsel province, 06
   The Netherlands
   +31 (0) 53 487 4444 (voice)
   +31 (0) 53 487 4400 (FAX)
   rossiter@itc.nl

   Hours_of_Service: Monday to Friday, 8:30 AM to 5:00 PM
   Contact_Instructions:
   Any body who is interested on this data set can contact the division head at any time by any of possible ways within the service hours mentioned.

Why was the data set created?

The main purpose of this data set is to capture the existing multi-source soils data of the area in digital format and make available for different users around Lake Naivasha.

The other purpose, why the primary soil survey carried out for, was to evaluate the suitability of the farms/land for different agricultural crops under irrigation.

How was the data set created?

1. Where did the data come from?

Kwacha 1998 (source 1 of 9)

Mr. C.P.H.J. Kwacha, "(comp)", 1998, Vulnerability of Soils to Change in Agricultural use Around Lake Naivasha, Kenya: International Institute for Aerospace Survey & Earth Sciences, ITC, Enschede, the Netherlands.

Online Links:

- No online linkage

Type_of_Source_Media: Paper
Source_Scale_Denominator: 50000
Source_Contribution:

1) The paper soil map was used as a primary analogue soil map to be captured in the digital environment. 2) The text document with its analytical data was used to develop the digital attribute database.

Naivasha Sheet 133/2; Longonot Sheet 133/4 (source 2 of 9)

Survey of Kenya, 1975, East Africa, 1:50000 scale Topographic maps: Survey of Kenya, Nairobi, Kenya.

Online Links:

- No online linkage

Type_of_Source_Media: Paper
Source_Scale_Denominator: 50000
Source_Contribution:

This map was used as a reference to correct and geo-reference the Aerial Photo interpretation base map.

AP1984; AP1991 (source 3 of 9)

Survey of Kenya, 1984; 1991, Aerial Photo: Survey of Kenya, Nairobi, Kenya.

Online Links:

- No online linkage

Type_of_Source_Media: Paper
Source_Scale_Denominator: 12500, 10000

Source_Contribution:

1) Base material to make mosaic of the entire study area
2) Material for stereoscopic photo-interpretation

AP1970 (source 4 of 9)

Survey of Kenya, 1970, Aerial Photo: Survey of Kenya, Nairobi, Kenya.

Online Links:

- No online linkage

Type_of_Source_Media: Paper

Source_Scale_Denominator: 50000

Source_Contribution:

1) Basic material to make a mosaic of the entire study area
2) Material for stereoscopic photo-interpretation as used by Kwacha in 1998.

Kulia 1980 (source 5 of 9)

Dr. Walter Sidrius, "(comp)", 1980, Soil Conditions at Kulia Farm, Naivasha: Kenya Soil Survey, Nairobi, kenya.

Online Links:

- No on line linkage

Type_of_Source_Media: Paper

Source_Scale_Denominator: 10000

Source_Contribution:

1) The sketch map was used as a primary analogue soil map to be captured in the digital environment. 2) The text document with its analytical data was used to develop the digital attribute database.

D27, 1986 (source 6 of 9)

Mr. M.M. Gatahi, "(comp)", 1986, Detailed Soil Survey of the Nini Farm, Naivasha, Nakuru District: Kenya Soil Survey, Nairobi, Kenya.

Online Links:

- No on line linkage

Type_of_Source_Media: Paper

Source_Scale_Denominator: 5000

Source_Contribution:

1) The sketch map was used as a primary analogue soil map to be captured in the digital environment. 2) The text document with its analytical data was used to develop the digital attribute database.

D45, 1988 (source 7 of 9)

Mr. P.T. Kamoni, "(comp)", 1988, Detailed Soil Survey of a Part of Quarantine Farm, Naivasha: Kenya Soil Survey, Nairobi, Kenya.

Online Links:

- No online linkage

Type_of_Source_Media: Paper
Source_Scale_Denominator: 5000
Source_Contribution:

1) The sketch map was used as a primary analogue soil map to be captured in the digital environment. 2) The text document with its analytical data was used to develop the digital attribute database.

Nini Irr. map (source 8 of 9)

Provincial Irrigation Unit of Nakuru, year of publication unknown, Uncontrolled map of Nini farm, Naivasha: Provincial Irrigation Unit, Nakuru, Kenya.

Online Links:

- No online linkage

Type_of_Source_Media: paper
Source_Scale_Denominator: 1000
Source_Contribution:

Used as base map to produce the soil sketch map of Nini farm in 1986 by M.M. Gatahi.

Topo Quarantine (source 9 of 9)

B.G. Mwangi and S. Wataka both from KSS, publication year unknown, Topographic Survey of Quarantine farm: Kenya Soil Survey, Nairobi, Kenya.

Online Links:

- No online linkage

Type_of_Source_Media: Paper
Source_Scale_Denominator: 50000
Source_Contribution:

Used as base map to produce the soil sketch map of quarantine farm in 1988 by P.T. Kamoni.

2. What changes have been made?

Date: Apr-1980 (change 1 of 9)

The procedure followed for primary soil data collection in the field was divided in to four parts. The procedures used for each part are indicated as steps 1-4 under Phase 1.

Phase 1. Step 1. Field survey of the soils of Kulia farm:

A field soil survey was conducted on 17th May 1980 and about 34 soil samples were collected and analysed at soil laboratory of Kenya soil survey. Based on the field survey and the lab analysis a soil sketch map in paper format was produced at scale of 1:10000. The type and quality of map used as a base map for the production of the soil sketch map is not mentioned in the source document.

Date: Apr-1983 (change 2 of 9)

Phase 1. Step 2. Field survey of the soils of Nini farm:

1. The field survey was carried out in February 1982 and a revisit was made on 25th May 1983. During the survey, an uncontrolled contour map of the provincial irrigation unit, Nakuru, at scale 1:1000 was used as base map on to which the observations were plotted. Augerhole observations to a depth of 150-200cm were made following a 100m grid system and where necessary additional augerholes were made to check uncertain soil boundaries. Further augerhole observations were made in areas where some crops were performing poorly during a revisit.

2. The soil and site information was recorded on standard augerhole description sheets. Representative sites for each soil-mapping unit were selected and profile pits made. The profiles were examined, described and sampled per horizon for both physical and chemical analysis at NAL (National Agricultural laboratory). Composite samples for fertility analysis were taken from the vicinity of each profile pit.

Date: 19-Dec-1986 (change 3 of 9)

Phase 1. Step 3. Field survey of the soils of Quarantine farm:

1. The field survey was carried out in December 1986. During the survey, a topographical survey map of the area as surveyed by B.G. Mwangi and S. Wataka (both from KSS), at scale 1:50000 was used as the base map in the production of the Quarantine soil sketch map.

2. Augerhole observations were made along transects spaced at 200m apart. The distance between augerholes per transect was 100. Observations were made to a depth of 120cm where possible. Soil physical properties like texture, colour, consistency, etc. were recorded for each observation.

3. A representative profile pit was dug in each of the three soil mapping units from which soil samples for chemical and physical analysis were collected. Composite soil samples for the fixed depth (0-30cm) were collected from the vicinity of the corresponding pits and used for fertility analysis.

Date: Mar-1998 (change 4 of 9)

Phase 1. Step 4. Field survey of the soils of Southeast Lake shore (Kwacha's study area):

1. The first step for this phase was doing an interpretation of aerial photos and the elaboration of an intensive legend prior to the fieldwork. In the interpretation legend information from the geological and existing soil maps was incorporated.

2. Next, the free survey method was applied for conducting the field survey using the aerial photo interpretation map as the base map and to select sample areas. Three sample areas were identified covering two landscapes (lacustrine & volcanic Plain)
3. A field check was carried out by opening mini-pits and by additional soil augering. A total of 28 observation points were visited of which 12 are mini-pits and 16 are augerholes. From these observations soil samples were collected at fixed depths (at topsoil, at 50cm and at 120cm) for chemical and physical analysis. The soil samples were analysed at Kenya soil survey laboratory. Based on these information and the base map a paper soil map covering an area of 4613 ha was produced as final product with a geopedologic legend.

Date: 03-Oct-2000 (change 5 of 9)

Phase 2. Step 1. Data capture for analogue-digital conversion:
The procedures followed in the field to capture the spatial data of Quarantine, Nini and Kulia farms, so as to be able to convert from analogue to digital were:

1) The location of the farm was traced by asking the residents of the area (Lake Naivasha). (2) Looking around the farm to compare its present position with respect to the sketch map in the report document. (3) Using the scale of the sketch map, converting the map distance in to ground distance and measuring the bearing by protractor. (4) Selecting a reliable start point (benchmark) that can be identified both on the sketch map and the actual ground, and then taking GPS reading. (5) Marking the point on the sketch map where GPS reading was taken. (6) Measuring the ground distance (as converted from the sketch map) and the bearing from the first start point to the next point. Taking GPS reading for the second point. (7) Marking the second point on the sketch map. (8) Continuing measuring ground distance and bearing from the second point to the third point, from third point to fourth point and so on … until the traverse was closed by coming back to the start point. (9) Continuing marking all the points on the sketch map where GPS reading was taken. (10) Registering all the necessary information on data collection format.

Date: 03-Oct-2000 (change 6 of 9)

Phase 2. Step 2. Data capture for analogue-digital conversion:
The procedures followed in the field to capture the spatial data of south east lake shore (Kwacha's study area) so as to be able to convert from analogue to digital were:

1) The topographic map of Kenya at scale 1:50 000 was selected as a reference to find ground control points. (2) Identification of ground control points on the soil map and on the ground. (3) Taking GPS readings for all identified ground control points. (4) Marking all the GPS points on the analogue soil map and registering the corresponding co-ordinates for all identified points on the data collection format.

Date: 23-Dec-2000 (change 7 of 9)

Phase 3. Analogue-Digital Conversion:
The procedures followed in the office to convert the sketch maps of Quarantine, Kulia, and Nini farms and Kwacha's study area soil map from analogue to digital after enough information has been captured during the field work were:

1) Scanning the paper map with 300dpi, with RGB representation, as Tif format, on A-3 size scanner. (2) Importing the scanned image in to ILWIS as .mpr format (3) Creating one common co-ordinate system based on the Geometric Definition of Topographic map of Kenya, scale 1:50 000, published on 1975. (4) Creating geo-reference for each individual imported image (5) Geo-referencing the images/sketch maps using Geo-reference tie points option (6) After geo- referencing by screen digitising, creation of segment map indicating the soil boundaries. (7) Similarly by screen digitising creation of point locations, indicating observation points. (8) Creating a domain for each segment map having polygon identifier. (9) Polygonsing the segment maps using the domain created above and editing the representations. (10) By creating different domains different attribute maps were calculated. For example, map showing soil types. (11) Creating a polygon histogram table, and taking the area and the perimeter to the non-spatial database and inserting in the soil polygon table. On the same way the SMU_Id was brought from Ms Access and inserted to the table associated to the map in the spatial database so as to have a link between the spatial data and the non-spatial data. (12) By synthesising the data stored in the non-spatial database and importing in to the spatial database different attribute maps were produced based on the users' requirements. For example, maps showing pH, nutrient level, organic carbon content and soil texture.

Date: 23-Dec-2000 (change 8 of 9)

Phase 4. Step 1. Spatial Data Integration:

The method followed to integrate the spatial data was reinterpretation. The procedures used for the reinterpretation process were:

1) An ortho-photo mosaic covering the study area was produced using aerial photos of year 1984 at scale 1:12500 and year 1991 at scale 1:10000 in ILWIS software. (2) The segment maps of Kulia, Nini, and soil map of Kwacha on the one hand and Quarantine and soil map of Kwacha on the other hand were overlaid together with different representations on the same photo mosaic as a base map for all of them. (3) The boundaries of each segment map were thoroughly investigated by comparing them to the common background map. (4) Soil boundaries nearest to the natural boundaries, terrain boundaries, vegetation boundaries etc. were taken as true boundaries. (5) For soil boundaries, which do not match to any one of the analysis elements on the background photo, were reinterpreted. (6) After finishing reinterpretation, by screen digitising an integrated new segment map was produced. (7) By polygonising the segment map, final integrated soil polygon map was created. The polygon Identifier of the map was used to link the spatial data to the integrated attribute database in Ms Access.

Date: 23-Dec-2000 (change 9 of 9)

Phase 4. Step 2. Non-Spatial Data Integration:

The method followed to integrate the non-spatial data was the nested database approach. The procedures used for the integration process were:

1. An investigation of the soil survey approaches adopted and preparation of a correlation table for the approaches used.
2. An investigation of the soil classification systems adopted and preparation of correlation tables for the systems used.

3. Entering the available point data in to the database based on the correlation tables previously created.
4. Inserting the correlation tables in to the database containing the integrated/nested data set.

How reliable are the data? What problems remain in the data set?

1. How well have the observations been checked?

In Nested attribute database 9 full, 16 mini-pit profiles and 27 augerhole descriptions in total 52 observations with corresponding site characteristics and other necessary data have been stored. All the information included in the database was collected and stored according to the general soil survey procedures. However, some violations have been made. These are:

1. For some soil descriptions made only at fixed depths, the horizon number and symbol has been given according to the sequence of these fixed depths.

2. Composite soil sample analysis results for the topsoil depth 0-30cm was recorded as the first horizon in the data set irrespective of the actual horizon depth.

3. Where no analytical data was recorded for corresponding horizons in an observation, in the source document, the values are shown empty in the records of the database.

2. How accurate are the geographic locations?

The accuracy of these digital data greatly depends upon the accuracy of the original individual soil maps and on the photo mosaic which was used for reinterpretation as base map (i.e. aerial photo of 1984 at scale 1:12500 and 1991 at scale 1:10000). A common co-ordinate system named "Naiv" which was created based on the Geometric Definition of Topographic map of Kenya, scale 1:50 000, published on 1975, and was used to maintain the geometric accuracy of all related maps before integration. The geometric definition includes Projection UTM, Datum Arc 1960, Datum area Mean, Ellipsoid Clarke 1880, and UTM zone 37M.

The horizontal positional accuracy was tested using the root mean square error (RMSE) test. The co-ordinates were measured using a millimetre ruler on the topomap to the nearest 0.25 mm and using Garmin 12XL GPS receiver in the field to the nearest 1 meter. The over all accuracy of the spatial data was calculated to be +/-18 meters.

3. **How accurate are the heights or depths?**

4. Where are the gaps in the data? What is missing?

The data has been converted from its analogue format to digital format and integrated at a scale of 1:50000. While integrating, the categorical detail was maintained at the same level to the analogue, but the cartographic detail was generalised in to scale 1:50000 by reinterpretation process. However, the detailed spatial data is available for each component on separate four databases.
In this digital data the map unit is taken as a collection of areas defined and named having same names in terms of their soil and/or non-soil areas. Each map unit differs in some characteristics from all others in a survey area and is uniquely identified. Each individual area in a map unit is a

delineation. Each map unit also consists of one or more delineations, and one or more soil components.

5. **How consistent are the relationships among the observations, including topology?**
The data set in the non-spatial database can be manipulated or retrieved using SQL and/or QBE. The entities are well related according to the rules of data integrity. The database has been tested by both data extraction languages mentioned above.

In the spatial database all soil delineations or polygons are closed and satisfy topological requirements that are necessary for the vector format. On this data set a point can be located to the nearest 0.25 mm (12.5 m) on hard copy and 0.1 mm (5 m) on the digital copy.

How can someone get a copy of the data set?

Are there legal restrictions on access or use of the data?

Access_Constraints: None
Use_Constraints:
ITC (International Institute for Aerospace Survey and Earth Sciences) the organisation and the compiler B.B. Tilaye should be acknowledged/mentioned as the data source in any product derived from this data set.

Who wrote the metadata?

Dates:
Last modified: 28-Dec-2000
Last Reviewed: 29-Dec-2000
To be reviewed: 28-Dec-2001

Metadata author:

ITC, Soil Science Division, Editor B.B. Tilaye
C/o Dr. D.G. Rossiter
Head of Soil Science Division
ITC, Hengelose straat 99, 7500 AA
Enschede, Overijsel province, 06
The Netherlands
+31 (0) 53 487 4444 (voice)
+31 (0) 53 487 4400 (FAX)
rossiter@itc.nl

Hours_of_Service: Monday to Friday, 8:30 AM to 5:00 PM
Contact_Instructions:
Any body who is interested on this data set can contact the division head at any time by any of possible ways within the service hours mentioned.

Metadata standard:

Federal Geographic Data Committee (FGDC) Content Standards for Digital Geospatial Metadata (CSDGM) (FGDC-STD-001-1998)

Generated by mp version 2.4.10 on Tue Jan 09 22:22:49 2001