Geographic Information Technology Reference Guide

With specific relevance to Statistical Agencies in Africa

Prepared by:

Geospace International (Pty.) Ltd.

Prepared for:

The World Bank, General Data Dissemination System, Socio-Demographic Statistics Project for Anglophone Africa

October 2009
Table of Contents

1. **SOME NOTES ON THE USE OF THE REFERENCE GUIDE** 6

2. **SPATIAL STATISTICS COLLECTION AND USE IN AFRICA** 6
   2.1 **Census Taking in Africa and around the world** 7
      2.1.1 A brief comparison of censuses from across the world 7
   2.2 **Census data sources - differences in what is asked where** 9
   2.3 **Basic African census methodology** 9
   2.4 **Challenges affecting the use of statistics and GIS in Africa** 10
      2.4.1 Accessibility of data 10
      2.4.2 Use of data 10
      2.4.3 Lack of funding and base data 10
   2.5 **The future of GIS in Africa** 10
   2.6 **A socio-economic perspective of Africa** 11
   2.7 **More reasons why statistical organizations need GIS** 12
      2.7.1 Obtaining the most from the GIS Unit 13

3. **WE ARE GIS** 14
   3.1 A day in the life of GIS 14

4. **GEOGRAPHIC INFORMATION SYSTEMS: BASIC PRINCIPLES** 15
   4.1 Factors aiding the rise of GIS 16
   4.2 **Philosophy of GIS** 16
   4.3 **Advantages of GIS** 16
      4.3.1 Planning of projects 16
      4.3.2 Making decisions 16
      4.3.3 Visual analysis 16
      4.3.4 Improving organizational integration 17
   4.4 **How does this relate to us?** 17
   4.5 **Components of GIS** 17
   4.6 **Maps and Spatial Information** 19
   4.7 **Characterizing Geographic Features** 20

5. **DATA ACCURACY AND QUALITY** 22
   5.1 **Accuracy** 22
   5.2 **Quality** 22
      5.2.1 Lineage (Metadata) 22
5.2.2 Positional Accuracy 23
5.2.3 Attribute Accuracy 23
5.2.4 Logical Consistency 23
5.2.5 Completeness 23

5.3 Error 23
5.3.1 Example of areas of uncertainty for overlaying data. 24

6. SOURCES OF DATA 24
6.1 The identification of Fundamental Geospatial Datasets within the African context 25

7. DATA INPUT TECHNIQUES 30
7.1 Digitizing 30
7.2 Automatic Scanning 30
7.3 Coordinate Geometry 31
7.4 Conversion of Existing Digital Data 31

8. DATA EDITING AND QUALITY ASSURANCE 31
8.1 Spatial Data Errors 32
8.2 Attribute Data Errors 33
8.3 Data Verification 33

9. DATA ORGANIZATION AND STORAGE 34
9.1 Organizing Data for Analysis 34
  9.1.1 Spatial Data Layers - Vertical Data Organization 34
9.2 Editing and Updating of Data 36
9.3 Data Retrieval and Querying 36

10. DATA ANALYSIS 37
10.1 Manipulation and Transformation of Spatial Data 37
  10.1.1 Coordinate Thinning 37
  10.1.2 Geometric Transformations 37
  10.1.3 Map Projection Transformations 37
  10.1.4 Conflation - Sliver Removal 37
  10.1.5 Edge Matching 38
  10.1.6 Interactive Graphic Editing 38
10.2 Integration and Modelling of Spatial Data 38
10.3 Integrated Analytical Functions in a GIS 39
  10.3.1 Retrieval, Reclassification and Generalization 39
  10.3.2 Topological Overlay 40
## 10.3.3 Neighbourhood Operations

## 10.3.4 Connectivity Analysis

## 11. IMPLEMENTATION ISSUES AND STRATEGIES

### 11.1 Current Options and Software Assessment

### 11.2 Justification and Expectations

### 11.3 Implementation Issues

#### 11.3.1 Common Pitfalls and Lessons Learnt

### 11.4 The Learning Curve

#### 11.4.1 GIS Productivity

## 12. IMPLEMENTING A GIS

### 12.1 Creating an awareness

### 12.2 Identifying System Requirements

### 12.3 System Evaluations

### 12.4 Justifying the System Acquisition

### 12.5 System Acquisition and Start Up

### 12.6 Operational Phase

#### 12.7 Implementation plan - a backbone structure

##### 12.7.1 Phase 1 - Awareness

##### 12.7.2 Phase 2 - System Requirements

##### 12.7.3 Phase 3 - System Evaluation

##### 12.7.4 Phase 4 - Develop Implementation Plan

##### 12.7.5 Phase 5 - System Start-up

##### 12.7.6 Phase 6 - Operational Phase

### 12.8 Country experiences regarding GIS implementation

#### 12.8.1 Hardware

#### 12.8.2 Software

#### 12.8.3 Human Resources and training

#### 12.8.4 Data and database

#### 12.8.5 Institutional Arrangement

#### 12.8.6 Administrative issues

## 13. DEVELOPMENTS AND TRENDS

### 13.1 New Data Sources

#### 13.1.1 Remote Sensing

#### 13.1.2 Global Positioning System (GPS) technology

## 14. THE USE OF POPULATION CENSUS DATA IN SMALL AREA STATISTICS

### 14.1 Small area statistics and poverty mapping

#### 14.1.1 What are poverty maps?
1. Some notes on the use of the reference guide

This reference guide serves to provide the user with insight as to how Geographic Information Technology (GIT), incorporating Geographic Information Systems (GIS), Global Positioning System (GPS) and Remote Sensing can assist statistical organizations in African countries to fulfil their mandates as official national statistical data collection, analysis and dissemination agencies. It also provides the user with information about the nature and principles of GIT and the collection, storage, analysis and manipulation of spatial information. Moreover, certain sections will focus on the planning and implementation of GIS within large organizations and the institutional challenges that accompany this endeavour. Background will also be provided on census taking in Africa, the challenges associated with it and how the use of GIT can assist statistical organizations to better plan and implement large scale surveys and census operations.

It is important to understand that the purpose of this guide is not to be a comprehensive manual, but rather to provide the user with basic knowledge regarding the implementation and use of geographic information technology in their specific domain, with references to literature which they themselves can access via the web or through existing books and manuals.

2. Spatial statistics collection and use in Africa

It is worthwhile to have an idea of some of the difficulties facing statistical agencies regarding data collection, analysis and dissemination in Africa, since the data generated by statistical agencies forms the base dataset for the development and socio-economic upliftment of the continent. Firstly, we must define why relevant spatial and statistical data is so important to the development of the continent. The usefulness of spatial and statistical data when not being linked to each other is severely limited. For argument’s sake, we can say that census attribute data linked to relevant small area spatial units such as place names or neighbourhoods forms a geo-statistical database which can act as the platform for socio-economic decision making and development within a country. An accurate census database without a doubt forms the backbone of any developmental initiative.

One of the mandates of a statistical organization within a developing country can therefore be seen as creating a geo-statistical database in order to facilitate proper development planning, implementation and monitoring of a range of various development and quality of life issues facing developing countries, such as:

- Land information management
- Poverty reduction and economic empowerment
- The development of basic transport and tele-communications infrastructure
- The provision of electricity, water and sanitation
- The effective and accurate provision of primary health care to especially rural communities
- Fighting the HIV/AIDS pandemic, malaria and tuberculoses (TB)
- Implementing sustainable agricultural methods in order to curb food insecurity
- Granting disadvantaged individuals access to land, markets, welfare and social services
- Maintenance of democracy through better planned and more transparent elections

Moreover, development in Africa is guided by the 8 Millennium Goals.

These goals are:

- To eradicate poverty and hunger
- To achieve universal primary education
- To promote gender equality and the empowerment of women
- To reduce child mortality
- To improve maternal health
- To combat HIV/AIDS and other diseases
- To ensure environmental stability
- To develop global partnerships for development
None of these goals can be achieved without accurate and relevant statistics to qualify, quantify and spatially reference the problem. Poverty and hunger, for example, has unique demographic and geographic attributes that need to be analyzed in a spatial context to fully assess the problem and then plan and monitor intervention processes. Geodemographic classification techniques can assist with poverty mapping and service delivery programmes.

Accurate spatial information about certain phenomena is also essential. If you can't measure it, you can't manage it. This is true with regard to the planning, implementation and monitoring of development and economic processes, intervention strategies and projects. The geographic or spatial component concerning the analysis and dissemination of data is of utmost importance in order to understand causal effects and underlying trends revealed by statistics.

Moreover, since most of the socio-demographic and some economic statistics are collected using area-based methodologies, it makes sense that this data needs to be analyzed with the spatial component forming an integral part of the analysis.

The accurate creation of administrative and statistical area boundaries that forms the geographic frame within which data is collected, analyzed and disseminated will create an accurate base for the delivery of effective and timely statistics.

A disturbing fact is that although many Asian countries have made inroads into these goals, some African countries have actually gone backwards. This is mostly due to social and political unrest and poorly conceived and implemented socio-economic policies and corrupt governance.

There are many drivers for development. However, there is one driver that is essential to all other drivers. That is information. Information has two components, the statistical and the spatial component. Both statistical and spatial information is needed to facilitate development.

Most African countries have an acute lack of both statistical and spatial data. In countries where data does exist, it is not standardized, it is disparate, of varying accuracy and currency and not available across various governmental and private domains.

Information about relevant issues, such as land information and mapping, is lacking in most countries, especially where the rural areas are concerned. Without land information, it is impossible to implement land restitution and management or land taxation.

Development programs and initiatives are difficult to plan, implement and monitor due to the lack of relevant and accurate geo-statistical information.

### 2.1 Census Taking in Africa and around the world

Since a part of this guide focuses on the use of GIT for census operations, it is prudent to provide some background on current census methodologies being implemented around the world and in Africa and specifically to focus on the challenges experienced in Africa which also impacts on the implementation and sustainable use of GIT and the creation of relevant statistical data.

#### 2.1.1 A brief comparison of censuses from across the world

Most countries undertake population and housing censuses at least once every 10 years. The UN statistical commission estimates that 165 countries have conducted a census in the course of the last decade. These censuses have covered approximately 95% of the world's population. The methods for conducting a census cover a wide spectrum, ranging from the traditional questionnaire based census through those carried out using administrative registers. In between are combinations of the two, supplemented in some cases by various types of sample survey, either in conjunction with the census or as separate activities.

The traditional census involves the distribution of census questionnaires which are then collected by enumerators or by a combination of enumerators and 'post back'.
Advantages of the traditional census

- The national statistical organization has control over the operation which means that it can be managed to maximise the statistical potential of the data collected.
- The traditional census tends to have wide topic coverage
- The public awareness generated for official statistics in general engenders a sense of national participation and highlights the importance of objective and timely statistics for society at large.

Disadvantages of the traditional census

- High costs
- Logistics, HR and management intensive
- The infrequency of enumeration
- The problems of accurately enumerating certain subgroups of the population such as the young mobile, people living in squatter camps and some other marginalised sections of the community, although modern methods of census mapping through the use of remote sensing and aerial imaging is alleviating this problem

Because of the high cost of traditional census taking, many countries carry out a sample enumeration in conjunction with the census to collect more detailed information on a separate longer questionnaire. Separate from decennial sample enumeration, some countries employ sample household surveys as an integral part of their census operations. Household surveys are the most flexible of census data sources. In principle, almost any subject can be investigated though this method.

There is a move in especially developed countries to collect census information through the use of administrative registers. Many social statistics are produced as a by product of administrative processes. Data from different registers can be merged (provided they have a common georeference, such as a postcode, place name or administrative area) thus reducing the burden on the respondent by making use of whatever data is already in the system.

Other benefits include the avoidance of costly field data collection and a greater degree of control over the timing and frequency of statistical reporting.

There are however also disadvantages of using administrative registers. The content and detail may substantially be less than that available from questionnaire based surveys and there may be data quality and error propagation problems, which make the merging of registers problematic. There is also the significant challenge of convincing the public about the confidentiality of the merging and reporting of processes – issues of data privacy and protection.

In less wealthy nations where the cost of providing a census or other types of data infrastructure is prohibitively expensive, the use of remotely sensed imagery has been suggested.

As for the future, it is clear that the nature of the census and the manner in which it is conducted is changing in many countries. There are several factors involved. First is the impact of technology on traditional census. This has an impact on all aspects of the census process, from planning and management, through data capture, processing and storage to data dissemination. It is fuelled by tools such as GIS, intelligent character recognition, the internet as a data collection and dissemination medium and by new analysis and reporting software.

Second is the improving range of data sources for a register based census. An increasing number of countries are adopting the register based approach to supplement or even replace the traditional census. Pressures to implement population registers where these are not currently available together with the availability of new and highly innovative data sources will tend to fuel this trend.

Thirdly, in a world with accelerating population and socio-economic change, the traditional decennial census is becoming less and less relevant after the first few years following a census. With cost the major barrier to a more frequent traditional census, there will be greater pressure for the use of register based data and frequent sample surveys to supplement or supersede the census.
2.2 Census data sources - differences in what is asked where

The census questions that are asked in different national censuses are broadly similar in scope, covering: age, ethnicity and household characteristics, housing and various measures of affluence and deprivation including education, occupation, industry and car ownership. Despite these broad similarities there are interesting and important differences in each market, often reflecting differences in the level of economic development. These differences often have a significant effect on the shape of the eventual classification.

In Australia and the USA, for instance, the census provides unusually detailed information on rents, house values and personal and household incomes, making it correspondingly easier to differentiate neighbourhoods characterised by high income and high housing cost - which predominantly appear in larger cities - from less well off neighbourhoods in smaller towns and rural locations. The Brazilian, Namibian and South African censuses are, by contrast, much more focused on topics relating to access to utilities such as water and electricity that differentiate shanty towns from mainstream development. These censuses cover whether or not households have access to running water, electricity or mains sewerage. In these countries, the censuses also have quite detailed information on literacy.

2.3 Basic African census methodology

Most African countries use the traditional census method. In general, this method is applied successfully, although it is costly. Due to the critical nature of census information, development and donor agencies are usually eager to provide the necessary funds to conduct these censuses. The main issue in the African context is not how the data is collected or processed, but rather how it is used.

Typically, a census consists out of the following phases:

**Demarcation phase:** Where the national geographic frame is defined and EA boundaries are demarcated. Typically this is done though extensive fieldwork operations covering every inch of a country. However, recently the use of satellite and aerial imagery has come to the fore as a means to minimize fieldwork activities. EA demarcation has a definite GIS component, since GIS is used to digitize the census geography onto a spatial database with the resulting base data (1:50000 topographic map sheets, infrastructural spatial data, satellite imagery). However, most African statistical organizations do not have functional GIS units and use is made of sketch mapping techniques to create EA maps. While this technique can be used successfully without GIS, it has reduced accuracy which compromises enumeration quality, while the impact of the lack of GIS is also felt during the analysis and dissemination phase, since no spatial analysis can be done and the statistical data cannot be mapped effectively.

**Enumeration phase:** Maps (sketch maps or digital hardcopy maps) are created for every EA and an enumerator is typically sent to such an EA with the map and all the necessary questionnaires. The enumerator therefore goes from house to house to deliver and administer (in some cases) the questionnaire.

**Processing phase:** All the questionnaires are collected and processed, meaning the information on the questionnaires are entered digitally into a database, either manually, automatically through scanning software or through a combination of both.

**Analysis phase:** The census attribute data captured in the database is analyzed by demographers and other specialists to develop core census indicators and summary statistical information. As noted, the lack of a functional GIS unit means that the information is analyzed only in the statistical context and not in the spatial context, meaning that underlying causal trends or occurrences which are spatial in nature will probably be missed.

**Dissemination phase:** The resultant statistical information is disseminated via reports, census atlases and the web to government and private agencies in order to be used for socio-economic
decision and policy making. Again, the lack of a GIS Unit means that proper thematic mapping cannot be done which limits dissemination possibilities and products.

The problem in Africa is that the census data is being collected and processed, but the analysis and dissemination phases especially, leave much to be desired. Information can only have an impact if it is accessible and used effectively.

2.4 Challenges affecting the use of statistics and GIT in Africa

2.4.1 Accessibility of data
Unfortunately, there is not a prevalent culture of data sharing in Africa. Government departments especially, are very reticent to make data available to other government departments, let alone private agencies. Due to a constant and acute lack of funding to especially statistical and survey and mapping departments, these departments sell their data to other departments and private agencies at high costs, thereby inhibiting the use of the data. Even in statistical agencies themselves, it will be difficult for one section to obtain data from another section. To be the custodian of an important dataset is seen as a position of power. To give the data away would be to give the power away. This is not unique to Africa and is also a problem in many developed countries. South Africa recently implemented a Spatial Data Infrastructure act which states that data collected by public sector institutions (such as census data) is in the public domain and should therefore be available for free to the public and private institutions, as long as privacy laws and ethical codes are not violated.

The fact remains that if census data is not provided to the public domain, the value of the data is diminished since it is not being used. It is like having a fertile piece of agricultural land with a healthy crop on it, but never actually harvesting the crop.

2.4.2 Use of data
Even in statistical agencies, census data is sometimes not analyzed and used to its full potential due to the lack of further downstream funding. When it comes to a census, financial impetus is usually lost by the time the preliminary census results is published. Funding dries up and the statistical agencies have again to make due with their normal budget allocation. As a result, many statistical agencies have non-existent or barely operational GIS units which mean spatial analysis of the census data almost never takes place. Again, the power of the census data is under utilized, which is a shame considering the amount of cost and effort it takes to obtain it in the first place.

2.4.3 Lack of funding and base data
Lack of funding and base data will continue to play a major role in how census methodology is conceptualised in Africa. As noted, the base data (topographic data, infrastructural data, settlement data) in most African countries are out of date which makes accurate census demarcation very difficult. The reason that this data becomes out of date it that the agencies mandated with the maintenance of this data are constantly under funded. Thus, lack of funds leads to bad data which means more funds are needed during census time to make up for the lack of data by acquiring satellite or aerial imagery, for example. It is therefore a vicious circle which will only be resolved once statistical and survey and mapping agencies get the financial assistance from their governments that they deserve.

2.5 The future of GIT in Africa
GIT based classification and analysis techniques can be used to address both development issues and the encouragement of private sector investment. It is only through an understanding of where there are very poor communities from a human development, socio-economic and service access perspective that African countries will be able to encourage international agencies to provide funding for development and begins to address socio-economic disparities on the continent.

Ideally, other information that makes Africa unique in terms of its natural resources, different cultures, mineral wealth, wilderness landscapes, scenic beauty and economic opportunities must be made available to encourage economic investment. International businesses want to grow into Africa, but too often they are afraid to do so as they are unable to get sufficient information,
especially in a spatial context at a localized level, to enable them to decide whether sufficiently large markets exists and whether it is a viable investment. Information that is required in this regard is often related to demographics, socio-economics and infrastructure of the country, such as the per capita income, disposable income, education skill base and road and airport infrastructure.

It is also important to recognize that there are other secondary data sets that might not directly influence development or foreign investment, but are needed to get a holistic picture of what is happening throughout the continent. Prime examples of secondary data sets are crime, environment, education, health, labour forces and social welfare. What is therefore needed is a total spatial information system, which is a challenge because a lot of data in African countries are disparate, in different formats, of differing accuracies and currencies and are sometimes difficult to access, as previously mentioned.

When examining countries in Africa one can see that, the largest and most common use of GIT is still in the environmental field. Only recently, has GIT been used in the public utilities environment and for the capture of socio-economic information in some African countries. The true use of GIT for socioeconomic analysis for development or its use to encourage private sector investment is still to happen. This will occur when census information at an enumerator area level becomes available in African countries. It will also occur when African leaders accept the fact that information, especially spatial information, is required to bring about sustained socio-economic development.

Furthermore, it is necessary that they place sufficient emphasis on their countries allocating enough resources to integrate this information into GIS. However, it must be emphasized that this should be done within a spatial information framework that is founded on policies and strategies of the continent and its individual countries.

Not only is the collection of relevant spatial information for countries in Africa a challenge but so to is the effective use of the information and technology. Research has recently shown that very few people are effectively able to use spatial information for problem solving. Therefore, the conclusion that has been drawn from this is that for spatial information to be used for decision-making will require facilitation by specialists with knowledge of the relevant spatial information and GIT technology. This requires the development of this expertise in countries throughout Africa, especially in the use of socio-economic information in GIT and geodemographics. Decision support systems, especially using interactive web mapping technology, needs to be further developed and the challenge of making spatial information readily available and easy to use needs to be addressed.

### 2.6 A socio-economic perspective of Africa

The socio-economic, political and environmental character of Africa is both unique and very complex. These unique features are a consequence of the many historical (e.g. colonization), political and natural factors that have molded the African landscape.

Of all the social problems faced in Africa, poverty is suggested to be the overriding priority. However, it is felt that unemployment and HIV/AIDS are probably the more important issues on the African continent. Poverty does not happen on its own, but is caused by several factors including environmental degradation and the inability of people to secure employment. This is often related to the economic situation, education of the labour force and the availability of jobs.

Unemployed people are more at risk of being infected with HIV and becoming impoverished and, thereby, continuing the vicious circle of human deprivation. Africa is also the only continent where poverty is expected to increase in the next century. Therefore, economic growth, job creation and poverty reduction remain the primary challenges for Africa.

Another issue that characterizes many African countries is the under provision of basic services and infrastructure, such as water, sanitation, electricity, telephones, education, health and roads. To improve access and the quality of services and infrastructure requires an understanding of what is the status quo at a national and sub-national level. This requires information on the size of the population as well as the availability of these basic services and infrastructure. Migration across
national borders and within countries is a factor that is now being recognized as having a major impact on governments planning for the provision of basic services and infrastructure. By looking at the per capita access to basic services gives a better understanding of the under provision within and between African countries.

A previous initiative conducted in the SADC region was to map the access to services and infrastructure at a sub-national level (i.e. district). Another part of the project looked at whether SADC countries have population statistics from their census available at a localized level. This is required to provide information for decision making in terms of both development and encourage private sector investment in the region. Again, geodemographic based GIS analysis is tailor made to provide the analytical answers to many of these developmental questions.

Taking into account all of the above, the following priority development sectors can be determined:

- Provision and maintenance of services and infrastructure (e.g. roads, electricity, water, sanitation, police, formal housing);
- Providing information and communications technology (e.g. radio, television, telephones, cellular networks, internet connections and hubs);
- Provision of education and health services and the development of skills;
- Effective agricultural programs; and
- Develop of local and international export markets.

GIT based data collection, analysis and dissemination techniques, on one or another level, stands central to the addressing of these priorities. The need has therefore been identified. It is however up to the government and private sectors within African countries to create the necessary will to implement national data frameworks from which data can be accessed and new data created in order to build national geostatistical systems.

2.7 More reasons why statistical organizations need GIT

It is evident that the increased demand for developmental socio-demographic statistics with regard to issues such as poverty and famine monitoring is exposing existing statistics data collection, integration, analysis and dissemination techniques. Developmental issues all have a distinct geographic or spatial component which must be part and parcel of the whole statistics collection and creation process for the statistics to be fully relevant and meaningful. Many existing geographic or GIS units within statistical organizations faces a real challenge with regard to educating line divisions within these organizations to make use of the potential of GIS. The fact that GIS and the use thereof with regard to statistical agencies in developing countries is a fairly new concept leads to inadequate understanding on how it must best be implemented and utilized.

The following are typical challenges faced by GIS units within statistical organizations:

- Inadequate awareness within the organization as a whole regarding the potential and use of GIS in statistical agencies
- Lack of institutional support regarding the ongoing maintenance and sustainability of GIS
- Inadequately integrated and properly designed data warehouse and database
- Inadequate geographic base, primary and secondary data
- Available data lacking in accuracy and currency
- Inadequate equipment
- Lack of skilled staff
- Inadequate funding and operational assistance

A fully functional and supported GIS can provide statistical agencies with the following benefits:
• Greater data analysis and dissemination possibilities, in turn advocating statistics and raising the profile of statistics
• The creation of an accurate administrative and statistical boundary spatial database
• Accurate, cost effective Census Mapping operations and with the end result being accurate EA demarcation and map creation.
• Facilitating the design and implementation of a Master Sample frame and maintaining this frame for effective survey implementation and operations
• Increasing the scope of spatial statistics analysis and census product creation
• Further developing IT capacity within the organization
• Fast-tracking relevant staff skills development
• Acting as a corporate service provider to the organization
• Increasing the accuracy and effectiveness of census and survey data collection
• Saving cost and time in the long run due to decreased fieldwork activities because of better survey and census implementation methods.

2.7.1 Obtaining the most from the GIS Unit

An institutional shift will have to take place with regard to the use of GIS and its functions within statistical agencies. The role of GIS must be re-defined as a corporate service provider to statistical agencies with the following services in mind:

• Spatial data and digital map provider to all divisions and external users
• Spatial database custodian responsible for new data creation and maintenance
• Custodian of the Master Sample frame responsible for its maintenance, Secondary Sampling Unit creation according to relevant survey needs and updating
• Responsible for all spatial analysis according to internal and external user needs (for economic and socio-demographic data). Analysis to be done in conjunction with relevant unit expertise.
• Responsible for Census Mapping revision and maintenance
• Responsible for GIS attribute data integration, updating and maintenance
• Responsible for spatial analysis and graphic dissemination of the Census Atlas
• Responsible to host the spatially enabled web application spatial database.
• Responsible for the acquisition and use of remote sensing imagery
• Responsible for internal and external ad hoc map creation
• Responsible for the integration, geocoding and mapping of GPS derived data

A properly implemented and integrated GIS will assist a great deal in meeting some of the challenges facing statistical agencies, especially concerning the following:

- The creation of an accurate geographic base to plan and implement census and survey operations from will increase the accuracy of data collection, thereby improving the accuracy of the resultant statistics, thereby increasing the demand and profile of statistics in the user community.
- The added analysis and dissemination possibilities (thematic maps, etc) will enhance decision-making. Accurate and relevant maps will go a long way to advocate the use of statistics and further increase the profile of statistics.
- Institutional capacity will be increased due to the added dimension GIS brings to all the core processes within statistical organizations. Statistical products can be expanded and the GIS Unit can act as a corporate service provider for the whole organization.
- Master sampling frames and updated survey methodology can now be created and implemented, further increasing the accuracy of surveys and census data collection as well as being able to speak directly to user-specific needs.
- The timeliness of statistics delivery will increase due to time saved on planning, sample drawing, fieldwork activities, logistics management and the like.
- Socio-demographic data especially, can now be collected more cost and time effectively with fewer resources.
The IT infrastructure and knowledge base will be strengthened with the strengthening of the GIS infrastructure.
A cartographic mind-set concerning data collection, analysis and dissemination will eventually filter through the whole GSS as well as other data producers and users.

3. We are GIS

Currently, our everyday lives are affected by GIS in more ways than one. Everyday, our actions, behaviour and decisions are guided or influenced directly or indirectly by factors which are inherently linked to GIS. To prove this statement, let's look at a typical day in the life of Person A.

3.1 A day in the life of GIS

Table 1: A day in the life of GIS (adapted from Longley et al, 2005)

<table>
<thead>
<tr>
<th>TIME</th>
<th>ACTION</th>
<th>RELEVANT TO GIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>7:00</td>
<td>My alarm goes off</td>
<td>The electricity for the alarm comes from the local power company, which uses GIS to manage its electricity network and do future demand modelling</td>
</tr>
<tr>
<td>7:05</td>
<td>I jump in the shower</td>
<td>The water comes from the local water company, which uses GIS to maintain its pipe and valve network and predict future water usage</td>
</tr>
<tr>
<td>7:35</td>
<td>I open the mail</td>
<td>I get a property tax bill from my local municipality which uses GIS to store property data and automatically produce tax bills. A second item is a special offer for property insurance, from a firm that uses its geodemographic techniques to target neighbourhoods with low past claim histories. We receive less junk mail than we used to, because geodemographic GIS methods is used to target mailings more precisely, thus reducing waste.</td>
</tr>
<tr>
<td>8:05</td>
<td>I walk my kids to the bus stop</td>
<td>The school district administrators use GIS to optimize the routing of school buses</td>
</tr>
<tr>
<td>8:15</td>
<td>I drive to work</td>
<td>The roads I drive on are maintained by the municipality and the local roads agency. They use an extensive GIS to map the type and condition of roads in order to maintain them properly and plan future roads to alleviate traffic problems.</td>
</tr>
<tr>
<td>8:40</td>
<td>I arrive at work and read the newspaper</td>
<td>The paper for the newspaper comes from sustainable forests managed by a GIS. Geodemographic profiling enables the newspaper to determine which kind of sport and economic articles to focus on, and where to distribute the newspaper.</td>
</tr>
<tr>
<td>12:00</td>
<td>I drive to the local deli to grab some lunch</td>
<td>The owner of the deli appointed a GIS specialist to determine the best location for the deli within the neighbourhood using profiling and drive time analysis techniques</td>
</tr>
<tr>
<td>6:30</td>
<td>I go shopping</td>
<td>I received a booklet of discount coupons in</td>
</tr>
</tbody>
</table>
We are affected by GIS on a daily basis without even realising it, such is the power of the spatial context in which information is collected and analysed and decisions are being made. Much of the data used to implement the GIS techniques described above are demographic and business data which originates from statistical agencies. Thus, to make any decision accurately information or data is needed. Where possible, information should be analysed in both the spatial and statistical context in order to make the decision process more accurate. We will now explore the basic principles and uses of GIT.

### 4. Geographic Information Systems: Basic principles

The maturing of the computer age has greatly changed the manner in which we can merge, compare, and manipulate multiple maps and other data sets. Computers and their software have significantly enhanced data handling capacity and flexibility. A powerful new tool, known as the Geographic Information System (GIS) emerged in the 1970s.

Many of those who developed GIS were inspired by the 1969 publication of the classic *Design with Nature* by Ian McHarg (Doubleday/Natural History Press), a leading landscape architect then at the University of Pennsylvania. This document pointed the way to planning and decision-making through comparative, integrated maps and related data types. Since its inception, GIS has become a major growth industry, now conducted worldwide at the multi-billion-dollar level. It has blossomed into the main way for using maps (novel and practical) in most endeavors that focus and rely on geographically-based data of many kinds.

Because remote sensing has routinely provided new images of the Earth's surface, it too has become intertwined with GIS as a means to constantly and inexpensively update some of the GIS data (such as land use and cover).

**The Association for Geographic Information defines GIS as:**

A system for capturing, storing, checking, integrating, manipulating, analyzing, and displaying data which are spatially referenced to the Earth' (usually) land surface.

A simpler working definition is: A computer-based approach to interpreting maps and images and applying them to problem-solving. The inclusion of computers to store, process, manipulate, interpret, and display GIS information is the critical ingredient that separate modern GIS from the more conventional (traditional) methods of using maps and correlative data prior to the 1970s.

GIS technology integrates common database operations such as query and statistical analysis with the unique visualization and geographic analysis benefits offered by maps. These abilities distinguish GIS from other information systems and make it valuable to a wide range of public and private enterprises for explaining events, predicting outcomes, and planning strategies.

Hence GIS is looked upon as a tool to assist in decision-making and management of attributes that needs to be analysed spatially.
4.1 Factors aiding the rise of GIS

- Revolution in Information Technology.
- Computer Technology.
- Remote Sensing.
- Global Positioning System.
- Communication Technology.
- Rapidly declining cost of Computer Hardware, and at the same time, exponential growth of operational speed of computers.
- Enhanced functionality of software and their user-friendliness.
- Visualizing impact of GIS corroborating the Chinese proverb "a picture is worth a thousand words."
- Geographical feature and data describing it are part of our everyday lives & most of our everyday decisions are influenced by some facet of Geography.

4.2 Philosophy of GIS

The proliferation of GIS is explained by its unique ability to assimilate data from widely divergent sources, to analyse trends over time, and to spatially evaluate impacts caused by development.

For an experienced analyst, GIS is an extension of one's own analytical thinking. The system has no built in solutions for any spatial problems, problem solving depends on the analyst.

GIS involves a complete understanding about patterns, space, and processes or methodology needed to approach a problem. It is a tool acting as a means to attain certain objective quickly and efficiently. Its applicability is realized when the user fully understands the overall spatial concept under which a particular GIS is established and analyses his specific application in the light of those established parameters.

Before the GIS implementation is considered the objectives, both immediate and long term, have to be considered. Since the effectiveness and efficiency (i.e. benefit against cost) of the GIS will depend largely on the quality of initial field data captured, organizational design has to be decided upon to maintain this data continuously. This initial data capture is most important.

4.3 Advantages of GIS

The use of GIS has been in vogue primarily due to the advantages mentioned below:

- Planning of projects
- Make better decisions
- Visual Analysis
- Improve Organizational Integration

4.3.1 Planning of projects

The advantage of GIS is often found in detailed planning of projects having a large spatial component, where analysis of the problem is a pre requisite at the start of the project. Thematic maps generation is possible on one or more than one base maps, example: the generation of a land use map on the basis of a soil composition, vegetation and topography. The unique combination of certain features facilitates the creation of such thematic maps. With the various modules within GIS it is possible to calculate surface, length, width and distance.

4.3.2 Making decisions

The adage "better information leads to better decisions" is as true for GIS as it is for other information systems. A GIS, however, is not an automated decision making system but a tool to query, analyze, and map data in support of the decision making process. GIS technology has been used to assist in tasks such as presenting information at planning inquiries, helping resolve territorial disputes, and siting pylons in such a way as to minimize visual intrusion.

4.3.3 Visual analysis

Digital Terrain Modelling (DTM) is an important utility of GIS. Using DTM/3D modelling, landscape can be better visualized, leading to a better understanding of certain relations in the landscape.
Many relevant calculations, such as (potential) lakes and water volumes, soil erosion volume (Example: landslides), quantities of earth to be moved (channels, dams, roads, embankments, land levelling) and hydrological modelling becomes easier. Not only in the previously mentioned fields but also in the social sciences GIS can prove extremely useful. Besides the process of formulating scenarios for an Environmental Impact Assessment, GIS can be a valuable tool for sociologists and demographers to analyze administrative data such as population distribution, market localization and other related features.

### 4.3.4 Improving organizational integration

Many organizations that have implemented a GIS have found that one of its main benefits is improved management of their own organization and resources. Because GIS has the ability to link data sets together by geography, it facilitates interdepartmental information sharing and communication. By creating a shared database one department can benefit from the work of another, thus data can be collected once and used many times.

As communication increases among individuals and departments, redundancy is reduced, productivity is enhanced, and overall organizational efficiency is improved. Thus, in a utility company the customer and infrastructure databases can be integrated so that when there is planned maintenance, affected people can be informed by computer-generated letters.

### 4.4 How does this relate to us?

As noted, we are all GIS, since we use and make decisions based on spatial data all the time. For example, the locations of your dwelling, work place, school, nearby stores, banks, and local landmarks are all included in your personal spatial database and are normally what you would think of when asked about spatial data. However, don't forget the less obvious things, like computer keyboards, remote controls, locations of items in a store, and the location of your furniture (important for the 3 a.m. bathroom run).

We pose questions, called queries in the jargon, to our spatial databases, like where is the nearest grocery store, how do I get there, or perhaps in idle speculation like what is the average income in a certain place or area? When we move to a new part of town (or even a new town), our queries often come up blank and we have to update our neighborhood databases with the locations of stores, bus stops, parks, and so on.

We also make decisions using spatial data, some of which are quite complex, on a daily basis. Perhaps the most common is route planning, usually from your home to some other place. This can be made more complex by your significant other calling and asking that you stop by a grocery store on the way home and pick up some broccoli for dinner. If the store is significantly out of your way, you may have to adjust the route for your trip home. Others that you might not immediately consider include how to pack stuff in boxes and where to put the boxes in the truck, designing a flower garden, and even interior decorating.

The point is that a GIS is a tool we use to help us to store and manipulate large datasets and to perform complex operations that would take a human a long time (with plenty of opportunity for errors) to do. However, the algorithms and storage techniques that it uses are usually analogous to human thought processes.

### 4.5 Components of GIS

GIS constitutes of six key components:

- Hardware
- Software
- Data
- People
- Method and workflow
- Maintenance
There is a common misconception that GIS is software only, thus, if you have the software and you know how to use the software you can “do GIS”.

GIS is, in fact, DATA.

Consider the following analogy – a sustainable and functional GIS can be compared to a vehicle.

- **Hardware**: This is the body of the vehicle (PCs, printers, IT Networking, server, GPS Units)

- **Software**: This is the engine of the vehicle and can consist of GIS proprietor software, open source software and peripheral software, such as CAD, image processing etc. GIS software provides the functions and tools needed to store, analyze, and display geographic information. It is important to choose the correct software for your needs.

- **Data**: This is the fuel for the vehicle. Without fuel, you can go nowhere and do nothing. You need clean and good quality fuel to keep the vehicle running optimally, just as you need good quality, current and accurate data to keep your GIS sustainable. Thus, without data GIS is useless. Spatial and attribute data can be collected in the field though GPS data collection or with the use of remote sensing products and traditional fieldwork.

  Data can also be commercially acquired and various different data sets can be merged and analyzed to create new datasets. Data is simply the most important and expensive commodity any statistical organization will ever invest in. Therefore, special consideration must always be given to data accuracy, quality and currency.

  Moreover, data standards regarding projection and datum systems and accuracy must be observed. The accuracy of analysis or data display is only as current or accurate as the data loaded into the GIS in the first place. Care must be taken with the storage and maintenance of data.

- **People**: You need someone with a license to drive this vehicle. Similarly, you need skilled and experienced staff to manage your GIS. GIS users range from technical specialists who design and maintain the system to those who use it to help them perform their everyday work. GIS users have to be multi-skilled, not only understanding GIS principles and their specific software, but also be conversant in their own subject domains and related fields such as statistics, geography, environmental studies, demography and economics. To be a GIS specialist essentially means to be a jack of all trades. The most important skill to have is to know which questions to ask of the GIS (and which data to have) in order to obtain answers for a specific subject field.

- **Method and workflow**: One cannot just drive the vehicle anywhere without a plan. You need to know where you want to go and what you need to get there. With GIS therefore, you need a common vision about the purpose of the GIS which needs to culminate into a definitive methodology and operational workflow in order for the GIS to function optimally. Therefore, above all a successful GIS operates according to a well-designed plan and business rules, which are the models and operating practices unique to each organization.

- **Maintenance**: This vehicle needs to be serviced and maintained, otherwise it will break down. The maintenance of the GIS must be budgeted for in the annual line budget of the relevant department. This must include not only consumables such as printer and digital media but also software and hardware upgrades and most importantly, staff skill enhancement and additional training. GIS will always die a slow death if it is not maintained properly.

Although each of these components is vital, DATA, the fuel of your GIS, is the most important.
4.6 Maps and Spatial Information
To understand the specific requirements of Geographic Information Technology, one needs to understand its underlying principles. The following section will deal with the basic principles regarding maps and spatial information.

The main method of identifying and representing the location of geographic features on the landscape is a map. A map is a graphic representation of where features are, explicitly and relative to one another. A map is composed of different geographic features represented as either points, lines, and/or areas. Each feature is defined both by its location in space (with reference to a coordinate system), and by its characteristics (typically referred to as attributes). Quite simply, a map is a model of the real world.

The map legend is the key linking the attributes to the geographic features. Attributes, e.g. such as the species for a forest stand, are typically represented graphically by use of different symbology and/or color. For GIS, attributes need to be coded in a form in which they can be used for data analysis. This implies loading the attribute data into a database system and linking it to the graphic features.

Figure 1: What is a map?
Maps are simply models of the real world. They represent snapshots of the land at a specific map scale. The map legend is the key identifying which features are represented on a map.

For geographic data, often referred to as spatial data, features are usually referenced in a coordinate system that models a location on the earth's surface. The coordinate system may be of a variety of types. The most common are:

- **geographic coordinates** such as latitude and longitude, e.g. 56°27'40" and 116°11'25". These are usually referred by degrees, minutes, and seconds. Geographic coordinates can also be identified as decimal degrees, e.g. 54.65°.
- **a map projection**, e.g. Universe Transverse Mercator (UTM) where coordinates are measured in metres, e.g. 545,000.000 and 6,453,254.000 normally reference to a central meridian. Eastings refer to X coordinates while Northings refer to Y coordinates.

Geographic or spatial data is distinguished from attribute data in that it is referenced spatially by a coordinate system, e.g. it has a spatial extent. Maps are the traditional method of storing and displaying geographic information.

A map portrays 3 kinds of information about geographic features.
- Location and extent of the feature
- Attributes (characteristics) of the feature
- Relationship of the feature to other features.
Geography has often been described as the study of why what is where. This description is quite appropriate when considering the three kinds of information that are portrayed by the traditional map:

- the location and extent of a feature is identified explicitly by reference to a coordinate system representing the earth's surface. This is where a feature is.
- the attributes of a feature describe or characterize the feature. This is what the feature is.
- The relationship of a feature to other features is implied from the location and attributes of all features. Relationships can be defined explicitly, e.g. roads connecting towns, regions adjacent to one another, or implicitly, e.g. close to, far from, similar to, etc. Implicit relationships are interpreted according to the knowledge that we have about the natural world. Relationships are described as how or why a feature is.

The geographic information system distinguishes between the spatial and attribute aspect of geographic features. Therefore, the identification of relationships between features, within a common theme or across different themes, is the primary function of a GIS.

### 4.7 Characterizing Geographic Features

All geographic features on the earth's surface can be characterized and defined as one of three basic feature types. These are **points**, **lines**, and **areas**.

**Point** data exists when a feature is associated with a single location in space. Examples of point features include a fire lookout tower, an oil well or gas activity site, and a weather station.

**Linear** data exists when a feature's location is described by a string of spatial coordinates. Examples of linear data include rivers, roads, pipelines, etc.

**Areal** data exists when a feature is described by a closed string of spatial coordinates. An area feature is commonly referred to as a **polygon**. Polygonal data is the most common type of data in natural resource applications. Examples of polygonal data include forest stands, soil classification areas, administrative boundaries, and climate zones. Most polygon data is considered to be homogeneous in nature and thus is consistent throughout.

---

**Figure 2: GIS Data Structure**

(Adapted from Berry)

Commonly, an identifier accompanies all types of geographic features. This description or identifier is referred to as a label. Labels distinguish geographic features of the same type, e.g. forest stands, from one another. Labels can be in the form of a name, e.g. "Lake Victoria", a description, e.g.
"WELL" or a unique number, e.g. "123". Forest stand numbers are examples of polygon labels. Each label is unique and provides the mechanism for linking the feature to a set of descriptive characteristics, referred to as attribute data.

It is important to note that geographic features and the symbology used to represent them, e.g. point, line, or polygon, is dependant on the graphic scale (map scale) of the data. Some features can be represented by point symbology at a small scale, e.g. villages on a 1:1,000,000 map, and by areal symbology at a larger scale, e.g. villages on a 1:10,000 map. Accordingly, the accuracy of the feature's location is often fuzzier at a smaller scale than a larger scale. The generalization of features is an inherent characteristic of data presented at a smaller scale.

It is important to note that data can always be generalised to a smaller scale, but detail cannot be created!

Remember, as the scale of a map increases, e.g. 1:15,000 to 1:100,000, the relative size of the features decrease and the following may occur:

- Some features may disappear, e.g. features such as ponds, hamlets, and lakes, become indistinguishable as a feature and are eliminated.
- Features change from areas to lines or to points, e.g. a village or town represented by a polygon at 1:15,000 may change to point symbology at a 1:100,000 scale.
- Features change in shape, e.g. boundaries become less detailed and more generalized.
- Some features may appear, e.g. features such as climate zones may be indistinguishable at a large scale (1:15,000) but the full extent of the zone becomes evident at a smaller scale (1:1,000,000).

Accordingly, the use of data from vastly different scales will result in many inconsistencies between the number of features and their type. Therefore, the use and comparison of geographic data from vastly different source scales is totally inappropriate and can lead to significant error in geographic data processing.

Figure 3: Map Scale

<table>
<thead>
<tr>
<th>Scale in</th>
<th>Example of Features</th>
<th>Example of Symbology</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:1,000</td>
<td>Features disappear</td>
<td>Changes to point</td>
</tr>
<tr>
<td>1:5,000</td>
<td>Features change in shape</td>
<td>More detailed and less generalized</td>
</tr>
<tr>
<td>1:20,000</td>
<td>Some features appear</td>
<td>Changes in climate zones</td>
</tr>
<tr>
<td>1:50,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1:250,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1:1,000,000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5. Data Accuracy and Quality

The quality of data sources for GIS processing is becoming an ever increasing concern among GIS application specialists. With the influx of GIS software on the commercial market and the accelerating application of GIS technology to problem solving and decision making roles, the quality and reliability of GIS products is coming under closer scrutiny. Much concern has been raised as to the relative error that may be inherent in GIS processing methodologies.

Moreover, in the African context many data sets are often used without relevant and updated metadata to accompany it, leading in many cases to inappropriate datasets used together for analysis and dissemination.

Several practical recommendations have been identified which help to locate possible error sources, and define the quality of data. The following review of data quality focuses on three distinct components, data accuracy, quality, and error.

5.1 Accuracy

The fundamental issue with respect to data is accuracy. **Accuracy is the closeness of results of observations to the true values or values accepted as being true.** This implies that observations of most spatial phenomena are usually only considered to be estimates of the true value. The difference between observed and true (or accepted as being true) values indicates the accuracy of the observations.

Basically two types of accuracy exist. These are positional and attribute accuracy. Positional accuracy is the expected deviance in the geographic location of an object from its true ground position. This is what we commonly think of when the term accuracy is discussed. There are two components to positional accuracy. These are **relative** and **absolute** accuracy. Absolute accuracy concerns the accuracy of data elements with respect to a coordinate scheme, e.g. UTM. Relative accuracy concerns the positioning of map features relative to one another.

Often relative accuracy is of greater concern than absolute accuracy. Specifically for census and survey purposes, relative accuracy is of paramount concern, since, in many cases, GIS created field maps are used for fieldwork.

Attribute accuracy is equally as important as positional accuracy. It also reflects estimates of the truth. Interpreting and depicting boundaries and characteristics for forest stands or soil polygons can be exceedingly difficult and subjective. Most resource specialists will attest to this fact. Accordingly, the degree of homogeneity found within such mapped boundaries is not nearly as high in reality as it would appear to be on most maps.

5.2 Quality

**Quality can simply be defined as the fitness for use for a specific data set.** Data that is appropriate for use with one application may not be fit for use with another. It is fully dependant on the scale, accuracy, and extent of the data set, as well as the quality of other data sets to be used. The recent U.S. Spatial Data Transfer Standard (SDTS) identifies five components to data quality definitions.

- Lineage
- Positional Accuracy
- Attribute Accuracy
- Logical Consistency
- Completeness

5.2.1 Lineage (Metadata)

The lineage of data is concerned with historical and compilation aspects of the data such as the:

- source of the data;
• content of the data;
• data capture specifications;
• geographic coverage of the data;
• compilation method of the data, e.g. digitizing versus scanned;
• transformation methods applied to the data; and
• the use of an pertinent algorithms during compilation, e.g. linear simplification, feature generalization.

5.2.2 Positional Accuracy
The identification of positional accuracy is important. This includes consideration of inherent error (source error) and operational error (introduced error).

5.2.3 Attribute Accuracy
Consideration of the accuracy of attributes also helps to define the quality of the data. This quality component concerns the identification of the reliability, or level of purity (homogeneity), in a data set.

5.2.4 Logical Consistency
This component is concerned with determining the faithfulness of the data structure for a data set. This typically involves spatial data inconsistencies such as incorrect line intersections, duplicate lines or boundaries, or gaps in lines. These are referred to as spatial or topological errors.

5.2.5 Completeness
The final quality component involves a statement about the completeness of the data set. This includes consideration of holes in the data, unclassified areas, and any compilation procedures that may have caused data to be eliminated. The ease with which geographic data in a GIS can be used at any scale highlights the importance of detailed data quality information. Although a data set may not have a specific scale once it is loaded into the GIS database, it was produced with levels of accuracy and resolution that make it appropriate for use only at certain scales, and in combination with data of similar scales.

5.3 Error
Two sources of error, inherent and operational, contribute to the reduction in quality of the products that are generated by geographic information systems. Inherent error is the error present in source documents and data.

Operational error is the amount of error produced through the data capture and manipulation functions of a GIS. Possible sources of operational errors include:

- mislabelling of areas on thematic maps;
- misplacement of horizontal (positional) boundaries;
- human error in digitizing
- classification error;
- GIS algorithm inaccuracies; and
- human bias.

While error will always exist in any scientific process, the aim within GIS processing should be to identify existing error in data sources and minimize the amount of error added during processing. Because of cost constraints it is often more appropriate to manage error than attempt to eliminate it. There is a trade-off between reducing the level of error in a database and the cost to create and maintain the database.

The validity of any decisions based on a GIS product is directly related to the quality and reliability rating of the product.

Often because GIS data is in digital form and can be represented with a high precision it is considered to be totally accurate. In reality, a buffer exists around each feature which represents
the actual positional location of the feature. For example, data captured at the 1:20,000 scale commonly has a positional accuracy of +/- 20 metres. This means the actual location of features may vary 20 metres in either direction from the identified position of the feature on the map. Considering that the use of GIS commonly involves the integration of several data sets, usually at different scales and quality, one can easily see how errors can be propagated during processing.

5.3.1 Example of areas of uncertainty for overlaying data.
Several comments and guidelines on the recognition and assessment of error in GIS processing have been promoted in papers on the subject. These are summarized below:

- There is a need for developing error statements for data contained within geographic information systems
- The integration of data from different sources and in different original formats (e.g. points, lines, and areas), at different original scales, and possessing inherent errors can yield a product of questionable accuracy
- The accuracy of a GIS-derived product is dependent on characteristics inherent in the source products, and on user requirements, such as scale of the desired output products and the method and resolution of data encoding
- The highest accuracy of any GIS output product can only be as accurate as the least accurate data theme of information involved in the analysis
- Accuracy of the data decreases as spatial resolution becomes more coarse
- As the number of layers in an analysis increases, the number of possible opportunities for error increases

6. Sources of Data
As previously identified, two types of data are input into a GIS, spatial and attribute. The data input process is the operation of encoding both types of data into the GIS database formats.

The creation of a clean digital database is the most important and time consuming task upon which the usefulness of the GIS depends. The establishment and maintenance of a robust spatial database is the cornerstone of a successful GIS implementation.

As well, the digital data is the most expensive part of the GIS. Yet often, not enough attention is given to the quality of the data or the processes by which they are prepared for automation.

The general consensus among the GIS community is that 60 to 80 % of the cost incurred during implementation of GIS technology lies in data acquisition, data compilation and database development.

A wide variety of data sources exist for both spatial and attribute data. The most common general sources for spatial data are:

- hard copy maps;
- aerial photographs;
- remotely-sensed imagery;
- point data samples from surveys and GPS captured data;
- existing digital data files.

Existing hard copy maps, e.g. sometimes referred to as analogue maps, provide the most popular source for any GIS project.

Because of the large costs associated with data capture and input, government departments are often the only agencies with financial resources and manpower funding to invest in data compilation. However, as mentioned, data in many African countries are out of date, disparate, of poor quality, difficult to access and very often very expensive. The scourge of one government department selling data at a premium cost to another government department also still exists in
many countries. In the African context, donor agencies are in many cases valuable sources of data, however, data accuracy is often not what is necessary and thematic coverages are sometimes not updated since a lot of data creation is project based.

Attribute data has an even wider variety of data sources. Any textual or tabular data than can be referenced to a geographic feature, e.g. a point, line, or area, can be input into a GIS. Attribute data is usually input by manual keying or via a bulk loading utility of the DBMS software.

The following figure describes the basic data types that are used and created by a GIS.

Figure 4: Basic data types

<table>
<thead>
<tr>
<th>BASE DATA</th>
<th>DERIVED</th>
<th>INTERPRETED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elevation</td>
<td>Slope</td>
<td>Habitat</td>
</tr>
<tr>
<td>Roads</td>
<td>Aspect</td>
<td>Aesthetics</td>
</tr>
<tr>
<td>Water</td>
<td>Watersheds</td>
<td>Best Route</td>
</tr>
<tr>
<td>Vegetation</td>
<td>Stream Buffer</td>
<td>Harvesting Schedule</td>
</tr>
<tr>
<td>Structures</td>
<td>Viewsheds</td>
<td></td>
</tr>
<tr>
<td>Political</td>
<td>Accessibility</td>
<td></td>
</tr>
<tr>
<td>Features</td>
<td>Covertype Diversity</td>
<td></td>
</tr>
<tr>
<td>Ownership</td>
<td>Housing Density</td>
<td></td>
</tr>
<tr>
<td>Soils</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent Rock</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depth to DR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mtv. Units 100 yr. Flood</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Base data represents information which is collected and encoded into the GIS database. Standards in classification, geographic registration and data exchange formats are in development. Derived maps are created through algorithms, such as the derivation of slope from an elevation map. Interpreted maps involve the calibration and modeling of base and derived maps in the expression of an GIS application, such as Elk habitat. Interpretation standards for maps created within the GIS are needed.

(after Berry)

6.1 The identification of Fundamental Geospatial Datasets within the African context

Recently, the South African Human Sciences Research Council (HSRC) and EIS Africa, under contract from the UN Economic Commission for Africa (ECA), completed a study to determine the key fundamental geospatial datasets for Africa. The project was funded by the Department of Land Affairs, South Africa.

Very good work was done in determining the fundamental geospatial datasets applicable to all countries in Africa, specifically pertaining to socio-economic development.

Fundamental geospatial datasets implies implicitly that these datasets must provide critical base information for development planning. As stated in the ECA report, it must provide complete coverage of the relevant area of interest, there should be a consistent need for it, and it must provide sufficient detail and accuracy so that users from many different sectors can gain value by using the dataset. Moreover, the datasets needs to adhere to certain minimum data standards and quality assurance processes so that it can be perceived as consistent, reliable, useful and accurate.

In short, it should be the base data always needed when designing a decent GIS database for broad use by various different institutions dealing with governance, social, infrastructural, service delivery and economic issues.

The report goes on to define a fundamental geospatial dataset as the following:

“Fundamental geospatial data sets are the minimum primary sets of data that cannot be derived from other data sets, and that are required to spatially represent phenomena, objects, or themes
important for the realisation of economic, social, and environmental benefits consistently across Africa at the local, national, sub-regional and regional levels”

As a result, the report went on to determine the following fundamental data themes regarding geospatial data for Africa:

- Geodetic Control Network
- Imagery
- Hypsography
- Hydrography
- Boundaries
- Geographic names
- Land management units/areas
- Transportation
- Utilities and Services
- Natural Environment

These have been further classified into different categories and themes to provide the specific data detail. The study furthermore divided data into different levels to determine whether it could be classified as primary or secondary data. In general terms, primary data can be seen as base data created without any analysis or interpretation (for example aerial photography). Secondary data will be data created by analysing or thematic mapping of primary data and additional data derived by field data collection, data manipulation or imagery interpretation.

The study determines the following levels of data. Primary data is denoted as levels 0 – II while fundamental secondary data was classified as level III. Non fundamental secondary data was classified as level IV.

Consider the following table:

Table 2: Structure of Geospatial Data (As adapted from “Determination of Fundamental Datasets for Africa” – ECA, 2005.

<table>
<thead>
<tr>
<th>PRIMARY DATA – FUNDAMENTAL GEOSPATIAL DATASETS</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEVEL 0</td>
</tr>
<tr>
<td>LEVEL I</td>
</tr>
<tr>
<td>LEVEL II</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SECONDARY DATA – FUNDAMENTAL GEOSPATIAL DATASETS</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEVEL III</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SECONDARY DATA – NON FUNDAMENTAL GEOSPATIAL DATASETS</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEVEL IV</td>
</tr>
</tbody>
</table>
The data can be qualitative or quantitative, as long as they can be spatially referenced.

The end result of the study provides us with a detailed list of perceived fundamental geospatial datasets which are necessary for development planning. The datasets consequently defined were:

- Transportation
- Administrative Boundaries
- Hydrography
- Hypsography
- Land Cover
- Cadastre and Tenure
- Imagery

The following table further details the specific data sets related to each of the previously identified data themes.

### Table 3: Fundamental geospatial datasets, spatial features and their attributes
(Adapted from “Determination of Fundamental Datasets for Africa” - ECA, 2005.)

<table>
<thead>
<tr>
<th>DATA THEME</th>
<th>DATASET</th>
<th>SPATIAL FEATURES</th>
<th>ATTRIBUTES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geodetic control network</td>
<td>Geodetic control points</td>
<td>Trigonometric points and coordinates</td>
<td>Coordinates; history of the network; network design</td>
</tr>
<tr>
<td></td>
<td>Height datum</td>
<td>Points</td>
<td>Primary height values</td>
</tr>
<tr>
<td></td>
<td>Geoid model</td>
<td>Reference ellipsoid</td>
<td>Name of reference ellipsoid; origin; numerical values of ellipsoid parameters</td>
</tr>
<tr>
<td>Rectified imagery</td>
<td>Aerial Photography</td>
<td>Ortho Photos</td>
<td>Date, time, scale or resolution, format, projection, level of rectification</td>
</tr>
<tr>
<td></td>
<td>Satellite Imagery</td>
<td>Orthorectified images</td>
<td>Date, resolution, bands, format, projection, level of rectification</td>
</tr>
<tr>
<td>Hypsography</td>
<td>Digital elevation model</td>
<td>Contour lines, Bathymetry lines</td>
<td>Height values</td>
</tr>
<tr>
<td></td>
<td>Spot Heights</td>
<td>Spot Heights</td>
<td>Heights of peaks</td>
</tr>
<tr>
<td></td>
<td>Bathymetry</td>
<td>Contours, point ocean features and grid</td>
<td>Type, depth</td>
</tr>
<tr>
<td>Hydrography</td>
<td>Coastline</td>
<td>Coastline</td>
<td>Scale, source, date</td>
</tr>
<tr>
<td></td>
<td>Natural water bodies</td>
<td>Streams and rivers (perennial, intermittent, seasonal), canals, ponds, lakes, wetlands, wells</td>
<td>Unique code, name, length, surface area</td>
</tr>
<tr>
<td>Boundaries</td>
<td>Governmental units</td>
<td>International, national, provincial, regional, district, local government, traditional authority, ward, township, tribal</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Populated places</td>
<td>Capitals, urban areas, towns, villages, suburbs, localities, rural settlements</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Enumeration areas</td>
<td>EA units</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Census indicators aggregated to small</td>
<td>Small area boundaries</td>
<td>Aggregated demographic, economic and social census</td>
</tr>
<tr>
<td>Area Statistics</td>
<td>Indicators</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geographic names</td>
<td>Place Names</td>
<td>Village, town, suburb, city</td>
<td>Unique code, names, synonyms, type, source, date</td>
</tr>
<tr>
<td>Feature names</td>
<td>River, mountain, farms, landforms, etc.</td>
<td>Unique code, names, synonyms, type, source, date</td>
<td></td>
</tr>
<tr>
<td>Land management units/areas</td>
<td>Land parcels/Cadastre</td>
<td>Land parcels</td>
<td>Parcel number, owner, size, date acquired</td>
</tr>
<tr>
<td>Land tenure</td>
<td>Address point</td>
<td>Street number, street name, street type, postal code, place name, province</td>
<td></td>
</tr>
<tr>
<td>Street address</td>
<td>Postal or zip code</td>
<td>Post code areas or zones</td>
<td>Unique code</td>
</tr>
<tr>
<td>Land use planning zones</td>
<td>Conservation areas, heritage sites and other restricted sites (state protected areas)</td>
<td>Name, area</td>
<td></td>
</tr>
<tr>
<td>Transportation</td>
<td>Roads</td>
<td>Main, trunk, national road, secondary road, tertiary road, minor road, street, tract</td>
<td>Unique code, name, surface, length, number of lanes</td>
</tr>
<tr>
<td>Road centerlines</td>
<td>National, main, major, minor, trails, secondary, other</td>
<td>Unique code, name, surface, length</td>
<td></td>
</tr>
<tr>
<td>Railways</td>
<td>Railway line, station</td>
<td>Unique code, name, type, length</td>
<td></td>
</tr>
<tr>
<td>Airports and ports</td>
<td>Airports, airfields, landing strips, harbour</td>
<td>Unique code, name</td>
<td></td>
</tr>
<tr>
<td>Structures</td>
<td>Structures</td>
<td>Bridges, tunnels, ferries, towers, stadiums</td>
<td>Unique code, name, type, latitude/longitude</td>
</tr>
<tr>
<td>Utilities and services</td>
<td>Power</td>
<td>Power stations, power lines</td>
<td>Unique code, name, capacity, type</td>
</tr>
<tr>
<td>Telecommunications</td>
<td>Telecommunication towers, telephone network</td>
<td>Unique code</td>
<td></td>
</tr>
<tr>
<td>Natural environment</td>
<td>Land cover</td>
<td>Rangelands, forests, woodland, scrub, urban or built up areas, wetlands</td>
<td>Name, surface and area</td>
</tr>
<tr>
<td>Soils</td>
<td>Soil types</td>
<td>Name, code, area, depth, land capability, clay, content, agricultural constraints, etc.</td>
<td></td>
</tr>
<tr>
<td>Geology</td>
<td>Lithogical, units/contacts</td>
<td>Unique code, name, age, Stratigraphy</td>
<td></td>
</tr>
<tr>
<td>Structure</td>
<td>Unique code, name, type, age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regional Boundary</td>
<td>Regional structure features</td>
<td>Type of features (e.g. faults, joint)</td>
<td></td>
</tr>
<tr>
<td>Major ore deposit</td>
<td>Type, name, commodity</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It should be noted that one dataset was added to the table which is the aggregated census indicators at small area level (if available). This data is invaluable to development planners, the
private sector and local councils since it allows them to accurately integrate key social, demographic and economic indicators into their decision making.

Now that we have an idea of what the fundamental geospatial datasets are, we need to know where to find these datasets.

The following table provides us with the information on which entity or entities are the most appropriate to contact regarding these datasets. It might be that more than one entity or institution per dataset is listed because in certain cases entities will cooperate to create a specific dataset.

<table>
<thead>
<tr>
<th>DATA THEME</th>
<th>DATASET</th>
<th>APPROPRIATE DEPARTMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geodetic control network</td>
<td>Geodetic control points</td>
<td>Survey Department</td>
</tr>
<tr>
<td></td>
<td>Height datum</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Geoid model</td>
<td></td>
</tr>
<tr>
<td>Rectified imagery</td>
<td>Aerial Photography</td>
<td>Survey Department Statistics Agency Electricity provider Telecommunications provider Various other local councils</td>
</tr>
<tr>
<td></td>
<td>Satellite Imagery</td>
<td>Survey Department Statistics Agencies Department of Defense Commercial vendors, such as SPOT, Geoeye, Quickbird</td>
</tr>
<tr>
<td>Hypsography</td>
<td>Digital elevation model</td>
<td>Survey Department</td>
</tr>
<tr>
<td></td>
<td>Spot Heights</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bathymetry</td>
<td></td>
</tr>
<tr>
<td>Hydrography</td>
<td>Coastline</td>
<td>Survey Department Water Affairs Forestry Agriculture</td>
</tr>
<tr>
<td></td>
<td>Natural water bodies</td>
<td></td>
</tr>
<tr>
<td>Boundaries</td>
<td>Governmental units</td>
<td>Survey Department Election Commission Statistics Agencies</td>
</tr>
<tr>
<td></td>
<td>Populated places</td>
<td>Survey Department Election Commission Statistics Agencies</td>
</tr>
<tr>
<td></td>
<td>Enumeration areas</td>
<td>Statistics Agencies</td>
</tr>
<tr>
<td></td>
<td>Census indicators aggregated to small area statistics</td>
<td>Statistics Agencies</td>
</tr>
<tr>
<td>Geographic names</td>
<td>Place Names</td>
<td>Survey Department Election Commission Statistics Agencies</td>
</tr>
<tr>
<td></td>
<td>Feature names</td>
<td>Survey Department</td>
</tr>
<tr>
<td>Land management units/areas</td>
<td>Land parcels/Cadastre</td>
<td>Survey Department</td>
</tr>
<tr>
<td></td>
<td>Land tenure</td>
<td>Survey Department Registrar of Deeds</td>
</tr>
<tr>
<td></td>
<td>Street address</td>
<td>Survey Department Registrar of Deeds Local Councils</td>
</tr>
<tr>
<td></td>
<td>Postal or zip code</td>
<td>Post Office</td>
</tr>
<tr>
<td></td>
<td>Land use planning zones</td>
<td>Survey Department Local Councils</td>
</tr>
<tr>
<td>Transportation</td>
<td>Roads</td>
<td>Transport Department</td>
</tr>
<tr>
<td></td>
<td>Road centerlines</td>
<td></td>
</tr>
</tbody>
</table>
The information above is by no means definitive but provides a good basis of data and suppliers of that data.

7. Data Input Techniques

Since the input of attribute data is usually quite simple, the discussion of data input techniques will be limited to spatial data only. There is no single method of entering the spatial data into a GIS. Rather, there are several, mutually compatible methods that can be used singly or in combination.

The choice of data input method is governed largely by the application, the available budget, and the type and the complexity of data being input.

There are at least four basic procedures for inputting spatial data into a GIS.

- Heads-up digitizing;
- Automatic scanning;
- Entry of coordinates using coordinate geometry; and the
- Conversion of existing digital data.

7.1 Digitizing

This method entails the GIS user digitizing point, line or area features directly in the GIS using appropriate backdrop, such as aerial photography or satellite imagery.

Digitizing has many advantages.

- Quickest way to create data;
- Flexibility and adaptability to different data types and sources;
- Easily taught in a short amount of time - an easily mastered skill
- Generally the quality of data is high;
- Ability to easily register and update existing data.
- Topology is built on-the-fly as features are digitized. Cleaning of data after digitizing is however still necessary.

7.2 Automatic Scanning

A variety of scanning devices exist for the automatic capture of spatial data. While several different technical approaches exist in scanning technology, all have the advantage of being able to capture spatial features from a map at a rapid rate of speed. However, as of yet, scanning has not proven to be a viable alternative for most GIS implementation.

Scanners are generally expensive to acquire and operate. As well, most scanning devices have limitations with respect to the capture of selected features, e.g. text and symbol recognition. Experience has shown that most scanned data requires a substantial amount of manual editing to create a clean data layer. Given these basic constraints some other practical limitations of scanners should be identified. These include:
• hard copy maps are often unable to be removed to where a scanning device is available, e.g. most companies or agencies cannot afford their own scanning device and therefore must send their maps to a private firm for scanning;
• hard copy data may not be in a form that is viable for effective scanning, e.g. maps are of poor quality, or are in poor condition;
• geographic features may be too few on a single map to make it practical, cost-justifiable, to scan;
• often on busy maps a scanner may be unable to distinguish the features to be captured from the surrounding graphic information, e.g. dense contours with labels;
• with raster scanning there it is difficult to read unique labels (text) for a geographic feature effectively; and
• scanning is much more expensive than digitizing, considering all the cost/performance issues.

The sheer cost of scanning usually eliminates the possibility of using scanning methods for data capture in most GIS implementations. Large data capture shops and government agencies are those most likely to be using scanning technology.

Currently, general consensus is that the quality of data captured from scanning devices is not substantial enough to justify the cost of using scanning technology. However, major breakthroughs are being made in the field, with scanning techniques and with capabilities to automatically clean and prepare scanned data for topological encoding. These include a variety of line following and text recognition techniques. Users should be aware that this technology has great potential in the years to come, particularly for larger GIS installations.

7.3 Coordinate Geometry
A third technique for the input of spatial data involves the calculation and entry of coordinates using coordinate geometry procedures or geocoding. This can involve the downloading and integration of GPS captured coordinates digitally or the integration of analogue coordinates or addresses.

7.4 Conversion of Existing Digital Data
A fourth technique that is becoming increasingly popular for data input is the conversion of existing digital data. A variety of spatial data, including digital maps, are openly available from a wide range of government and private sources. Several ad hoc standards for data exchange have been established in the market place. These are supplemented by a number of government distribution formats that have been developed. Given the wide variety of data formats that exist, most GIS vendors have developed and provide data exchange/conversion software to go from their format to those considered common in the market place.

As digital data becomes more readily available this capability becomes a necessity for any GIS. Potential purchasers of commercial GIS packages should determine and clearly identify their data conversion needs, prior to purchase, to the software vendor.

8. Data Editing and Quality Assurance
Data editing and verification is in response to the errors that arise during the encoding of spatial and non-spatial data. The editing of spatial data is a time consuming, interactive process that can take as long, if not longer, than the data input process itself.

Several kinds of errors can occur during data input. They can be classified as:

• Incompleteness of the spatial data.
  This includes missing points, line segments, and/or polygons.

• Locational placement errors of spatial data.
These types of errors usually are the result of careless digitizing or poor quality of the original data source.

- **Distortion of the spatial data.**
  This kind of error is usually caused by base maps that are not scale-correct over the whole image, e.g. aerial photographs, or from material stretch, e.g. paper documents.

- **Incorrect linkages between spatial and attribute data.**
  This type of error is commonly the result of incorrect unique identifiers (labels) being assigned during manual key in or digitizing. This may involve the assigning of an entirely wrong label to a feature, or more than one label being assigned to a feature.

- **Attribute data is wrong or incomplete.**
  Often the attribute data does not match exactly with the spatial data. This is because they are frequently from independent sources and often different time periods. Missing data records or too many data records are the most common problems.

The identification of errors in spatial and attribute data is often difficult. Most spatial errors become evident during the topological building process. Most topological building functions in GIS software clearly identify the geographic location of the error and indicate the nature of the problem. Comprehensive GIS software allows users to graphically walk through and edit the spatial errors. Others merely identify the type and coordinates of the error. Since this is often a labour intensive and time consuming process, users should consider the error correction capabilities very important during the evaluation of GIS software offerings.

### 8.1 Spatial Data Errors

A variety of common data problems occur in converting data into a topological structure. These stem from the original quality of the source data and the characteristics of the data capture process. Usually data is input by digitizing. Most GIS software has utilities to clean the data and build a topologic structure. If the data is unclean to start with, for whatever reason, the cleaning process can be very lengthy. Interactive editing of data is a distinct reality in the data input process.

**Experience indicates that in the course of any GIS project 60 to 80 % of the time required to complete the project is involved in the input, cleaning, linking, and verification of the data.**

The most common problems that occur in converting data into a topological structure include:

- slivers and gaps in the line work;
- dead ends, e.g. also called dangling arcs, resulting from overshoots and undershoots in the line work; and
- bow ties or weird polygons from inappropriate closing of connecting features.

Of course, topological errors only exist with linear and areal features. They become most evident with polygonal features, such as Enumeration Areas.

Slivers are the most common problem when cleaning data. Slivers frequently occur when coincident boundaries are digitized separately, e.g. once each for adjacent Enumerator Areas, or after polygon overlay. Slivers often appear when combining data from different sources. It is advisable to digitize data layers with respect to an existing data layer, e.g. hydrography, rather than attempting to match data layers later. A proper plan and definition of priorities for inputting data layers will save many hours of interactive editing and cleaning.

Dead ends usually occur when data has been digitized in a spaghetti mode, or without snapping to existing nodes. Most GIS software will clean up undershoots and overshoots based on a user defined tolerance, e.g. distance. The definition of an inappropriate distance often leads to the formation of bow ties or weird polygons during topological building. Tolerances that are too large
will force arcs to snap one another that should not be connected. The result is small polygons called bow ties. The definition of a proper tolerance for cleaning requires an understanding of the scale and accuracy of the data set.

The other problem that commonly occurs when building a topologic data structure is duplicate lines. These usually occur when data has been digitized or converted from a CAD system. The lack of topology in these type of drafting systems permits the inadvertent creation of elements that are exactly duplicate. However, most GIS packages afford automatic elimination of duplicate elements during the topological building process.

Most GIS software will provide the capability to eliminate bow ties and slivers by means of a feature elimination command based on area, e.g. polygons less than 100 square metres. The ability to define custom topological error scenarios and provide for semi-automated correction is a desirable capability for GIS software.

### 8.2 Attribute Data Errors
The identification of attribute data errors is usually not as simple as spatial errors. This is especially true if these errors are attributed to the quality or reliability of the data. Errors as such usually do not surface until later on in the GIS processing. Solutions to these type of problems are much more complex and often do not exist entirely. It is much more difficult to spot errors in attribute data when the values are syntactically good, but incorrect.

Simple errors of linkage, e.g. missing or duplicate records, become evident during the linking operation between spatial and attribute data. Again, most GIS software contains functions that check for and clearly identify problems of linkage during attempted operations. This is also an area of consideration when evaluating GIS software.

### 8.3 Data Verification
Six clear steps stand out in the data editing and verification process for spatial data. These are:

- **Visual review.**
  On-screen and printing maps

- **Cleanup of lines and junctions.**
  This process is usually done by software first and interactive editing second.

- **Weeding of excess coordinates.**
  This process involves the removal of redundant vertices by the software for linear and/or polygonal features.

- **Correction for distortion and warping.**
  Most GIS software has functions for scale correction and rubber sheeting. However, the distinct rubber sheet algorithm used will vary depending on the spatial data model, vector or raster, employed by the GIS. Some raster techniques may be more intensive than vector based algorithms.

- **Construction of polygons.**
  Since the majority of data used in GIS is polygonal, the construction of polygon features from lines/arcs is necessary in some cases. Usually this is done in conjunction with the topological building process.

- **The addition of unique identifiers or labels.**
  Often this process is manual. However, most systems do provide the capability to automatically build labels for a data layer.

These data verification steps occur after the data input stage and prior to or during the linkage of the spatial data to the attributes. Data verification ensures the integrity between the spatial and
attribute data. Verification should include some brief querying of attributes and cross checking against known values.

9. Data Organization and Storage

The second necessary component for a GIS is the data storage and retrieval subsystem. This subsystem organizes the data, both spatial and attribute, in a form which permits it to be quickly retrieved for updating, querying, and analysis. Most GIS software utilizes proprietary software for their spatial editing and retrieval system, and a database management system (DBMS) for their attribute storage. Typically, an internal data model is used to store primary attribute data associated with the topological definition of the spatial data. Most often these internal database tables contain primary columns such as area, perimeter, length, and internal feature id number. Often thematic attribute data is maintained in an external DBMS that is linked to the spatial data via the internal database table.

9.1 Organizing Data for Analysis

Most GIS software organizes spatial data in a thematic approach that categorizes data in vertical layers. The definition of layers is fully dependent on the organization's requirements. Spatial data layers are commonly input one at a time. Accordingly, attribute data is entered one layer at a time. Depending on the attribute data model used by the data storage subsystem data must be organized in a format that will facilitate the manipulation and analysis tasks that will be required. Most often, the spatial and attribute data may be entered at different times and linked together later. However, this is fully dependent on the source of data. The clear identification of the requirements for any GIS project is necessary before any data input procedures, and/or layer definitions, should occur.

It is mandatory that GIS users fully understand their needs before undertaking a GIS project. Experience has shown that a less than complete understanding of the needs and processing tasks required for a specific project, greatly increases the time required to complete the project, and ultimately affects the quality and reliability of the derived GIS product(s).

9.1.1 Spatial Data Layers - Vertical Data Organization

In most GIS software data is organized in themes as data layers. This approach allows data to be input as separate themes and overlaid based on analysis requirements. This can be conceptualized as vertical layering the characteristics of the earth's surface. A variety of terms are used to define data layers in commercial GIS software. These include themes, coverages, layers, levels, objects, and feature classes. Data layer and theme are the most common and the least proprietary to any particular GIS.

In any GIS project a variety of data layers will be required. These must be identified before the project is started and a priority given to the input or digitizing of the spatial data layers. This is mandatory, as often one data layer contains features that are coincident with another. Data layers are commonly defined based on the needs of the user and the availability of data. They are completely user definable.
The definition of data layers is fully dependent on the area of interest and the priority needs of the GIS. Layer definitions can vary greatly depending on the intended needs of the GIS.

When considering the physical requirements of the GIS software it is important to understand that two types of data are required for each layer, attribute and spatial data. Commonly, data layers are input into the GIS one layer at a time. As well, often a data layer is completely loaded, e.g. graphic conversion, editing, topological building, attribute conversion, linking, and verification, before the next data layer is started. Because there are several steps involved in completely loading a data layer it can become very confusing if many layers are loaded at once.

The proper identification of layers prior to starting data input is critical. The identification of data layers is often achieved through a user needs analysis. The user needs analysis performs several functions including:

- identifying the users;
- educating users with respect to GIS needs;
- identifying information products;
- identifying data requirements for information products;
- prioritizing data requirements and products;
- determining data gaps and needs;
- determining GIS functional requirements.

Often a user needs assessment will include a review of existing operations, e.g. sometimes called a situational assessment, and a cost-benefit analysis. The cost-benefit process is well established in conventional data processing and serves as the mechanism to justify acquisition of hardware and software. It defines and compares costs against potential benefits. Most institutions will require this step before a GIS acquisition can be undertaken.

Most GIS projects integrate data layers to create derived themes or layers that represent the result of some calculation or geographic model, e.g. forest merchantability, land use suitability, etc. Derived data layers are completely dependant on the aim of the project. Each data layer would be input individually and topologically integrated to create combined data layers. Based on the data model, e.g. vector or raster, and the topological structure, selected data analysis functions could be undertaken. It is important to note that in vector based GIS software the topological structure defined can only be traversed by means of unique labels to every feature.
9.2 Editing and Updating of Data

Perhaps the primary function in the data storage and retrieval subsystem involves the editing and updating of data. Frequently, the following data editing capabilities are required:

- interactive editing of spatial data;
- interactive editing of attribute data;
- the ability to add, manipulate, modify, and delete both spatial features and attributes (independently or simultaneously);
- the ability to edit selected features in a batch processing mode.

Updating involves more than the simple editing of features. Updating implies the resurvey and processing of new information. The updating function is of great importance during any GIS project. The life span of most digital data can range anywhere from 1 to 10 years. Commonly, digital data is valid for 5 to 10 years. The lengthy time span is due to the intensive task of data capture and input. However, often periodic data updates are required. These frequently involve an increased accuracy and/or detail of the data layer. Changes in classification standards and procedures may necessitate such updates. Updates to a forest cover data layer to reflect changes from a forest fire burn or updates to primary sampling units or EAs due to settlement development in peri-urban areas are typical examples.

Many times data updates are required based on the results of a derived GIS product. The generation of a derived product may identify blatant errors or inappropriate classes for a particular layer. When this occurs updating of the data is required. In this situation the GIS operator usually has some previous experience or knowledge of the study area.

9.3 Data Retrieval and Querying

The ability to query and retrieve data based on some user defined criteria is a necessary feature of the data storage and retrieval subsystem. Data retrieval involves the capability to easily select data for graphic or attribute editing, updating, querying, analysis and/or display.

The ability to retrieve data is based on the unique structure of the DBMS and command interfaces are commonly provided with the software. Most GIS software also provides a programming subroutine library, or macro language, so the user can write their own specific data retrieval routines if required.

Querying is the capability to retrieve data, usually a data subset, based on some user defined formula. These data subsets are often referred to as logical views. Often the querying is closely linked to the data manipulation and analysis subsystem. Many GIS software offerings have attempted to standardize their querying capability by use of a Standard Query Language (SQL). This is especially true with systems that make use of an external relational DBMS. Through the use of SQL, GIS software can interface to a variety of different DBMS packages. This approach provides the user with the flexibility to select their own DBMS. This has direct implications if the organization has an existing DBMS that is being used for to satisfy other business requirements. Often it is desirable for the same DBMS to be utilized in the GIS applications.

This notion of integrating the GIS software to utilize an existing DBMS through standards is referred to as corporate or enterprise GIS. With the migration of GIS technology from being a research tool to being a decision support tool there is a requirement for it to be totally integrated with existing corporate activities, including accounting, reporting, and business functions.

There is a definite trend in the GIS marketplace towards a generic interface with external relational DBMS’s. The use of an external DBMS, linked via a SQL interface, is becoming the norm. A flexibility as such is a strong selling point for any GIS. SQL is quickly becoming a standard in the GIS software marketplace.
## 10. Data Analysis

The major difference between GIS software and CAD mapping software is the provision of capabilities for transforming the original spatial data in order to be able to answer particular queries. Some transformation capabilities are common to both GIS and CAD systems, however, GIS software provides a larger range of analysis capabilities that will be able to operate on the topology or spatial aspects of the geographic data, on the non-spatial attributes of these data, or on both. **The main criteria used to define a GIS is its capability to transform and integrate spatial data.**

### 10.1 Manipulation and Transformation of Spatial Data

The maintenance and transformation of spatial data concerns the ability to input, manipulate, and transform data once it has been created. While many different interpretations exist with respect to what constitutes these capabilities some specific functions can be identified. These are reviewed below.

#### 10.1.1 Coordinate Thinning

Coordinate thinning involves the weeding or reduction of coordinate pairs, e.g. X and Y, from arcs. This function is often required when data has been captured with too many vertices for the linear features. This can result in redundant data and large data volumes. The weeding of coordinates is required to reduce this redundancy.

The thinning of coordinates is also required in the map generalization process of linear simplification. Linear simplification is one component of generalization that is required when data from one scale, e.g. 1:20,000, is to be used and integrated with data from another scale, e.g. 1:100,000. Coordinate thinning is often done on features such as contours, hydrography, and forest stand boundaries.

#### 10.1.2 Geometric Transformations

This function is concerned with the registering of a data layer to a common coordinate scheme. This usually involves registering selected data layers to a standard data layer already registered. The term rubber sheeting is often used to describe this function. Rubber sheeting involves stretching one data layer to meet another based on predefined control points of known locations. Two other functions may be categorized under geometric transformations. These involve warping a data layer stored in one data model, either raster or vector, to another data layer stored in the opposite data model. For example, often classified satellite imagery may require warping to fit an existing forest inventory layer, or a poor quality vector layer may require warping to match a more accurate raster layer.

#### 10.1.3 Map Projection Transformations

This functionality concerns the transformation of data in geographic coordinates for an existing map projection to another map projection. Most GIS software requires that data layers must be in the same map projection for analysis. Accordingly, if data is acquired in a different projection than the other data layers it must be transformed.

#### 10.1.4 Conflation - Sliver Removal

Conflation is formally defined as the procedure of reconciling the positions of corresponding features in different data layers. More commonly this is referred to as sliver removal. Often two layers that contain the same feature, e.g. EA and district boundaries both with a specific lake, do not have exactly the same boundaries for that feature, e.g. the lake. This may be caused by a lack of coordination or data prioritization during digitizing or by a number of different manipulation and analysis techniques. When the two layers are combined, e.g. normally in polygon overlay, they will not match precisely and small sliver polygons will be created. Conflation is concerned with the process for removing these slivers and reconciling the common boundary.

There are several approaches for sliver removal. Perhaps the most common is allowing the user to define a priority for data layers in combination with a tolerance value. After polygon overlay if a polygon is below the size tolerance it is classified a sliver. To reconcile the situation the arcs of the...
Another approach is to simply divide the sliver down the centre and collapse the arcs making up the boundary. The important point is that all GIS software must have the capability to resolve slivers. Remember that it is generally much less expensive to reconcile maps manually in the map preparation and digitizing stage than afterwards.

**10.1.5 Edge Matching**

Edge matching is simply the procedure to adjust the position of features that extend across typical map sheet boundaries. Theoretically data from adjacent map sheets should meet precisely at map edges. However, in practice this rarely occurs. Misalignment of features can be caused by several factors including digitizing error, paper shrinkage of source maps, and errors in the original mapping. Edge matching always requires some interactive editing. Accordingly, GIS software differs considerably in the degree of automation provided.

**10.1.6 Interactive Graphic Editing**

Interactive graphic editing functions involve the addition, deletion, moving, and changing of the geographic position of features. Editing should be possible at any time. Most graphic editing occurs during the data compilation phase of any project. Remember typically 60 to 70 % of the time required to complete any project involves data compilation.

Accordingly, the level of sophistication and ease of use of this capability is vitally important and should be rated highly by those evaluating GIS software. Many of the editing that is undertaken involves the cleaning up of topological errors identified earlier. The capability to *snap* to existing elements, e.g. nodes and arcs, is critical.

The functionality of graphic editing does not differ greatly across GIS software offerings. However, the user interface and ease of use of the editing functions usually does. Editing within a GIS software package should be as easy as using a CAD system. A cumbersome or incomplete graphic editing capability will lead to much frustration by the users of the software.

**10.2 Integration and Modelling of Spatial Data**

The integration of data provides the ability to ask complex spatial questions that could not be answered otherwise. Often, these are inventory or locational questions such as how much? or where?. Answers to locational and quantitative questions require the combination of several different data layers to be able to provide a more complete and realistic answer. *The ability to combine and integrate data is the backbone of GIS.*

Often, applications do require a more sophisticated approach to answer complex spatial queries and what if? scenarios. The technique used to solve these questions is called spatial modelling. Spatial modelling infers the use of spatial characteristics and methods in manipulating data. Methods exist to create an almost unlimited range of capabilities for data analysis by stringing together sets of primitive analysis functions. While some explicit analytical models do exist, especially in natural resource applications, most modelling formulae (models) are determined based on the needs of a particular project. The capability to undertake complex modelling of spatial data, on an ad hoc basis, has helped to further the resource specialists understanding of the natural environment, and the relationship between selected characteristics of that environment.

The use of GIS spatial modelling tools in several traditional resource activities has helped to quantify processes and define models for deriving analysis products. This is particularly true in the area of resource planning and inventory compilation. Most GIS users are able to better organize their applications because of their interaction with, and use of, GIS technology. The utilization of spatial modelling techniques requires a comprehensive understanding of the data sets involved, and the analysis requirements. *The critical function for any GIS is the integration of data.* Consider the following figure:
10.3. Integrated Analytical Functions in a GIS

Most GIS's provide the capability to build complex models by combining primitive analytical functions. Systems vary as to the complexity provided for spatial modelling, and the specific functions that are available. However, most systems provide a standard set of primitive analytical functions that are accessible to the user in some logical manner. Aronoff identifies four categories of GIS analysis functions. These are:

- Retrieval, Reclassification, and Generalization;
- Topological Overlay Techniques;
- Neighbourhood Operations; and
- Connectivity Functions.

The range of analysis techniques in these categories is very large. Accordingly, this section of the guide focuses on providing an overview of the fundamental primitive functions that are most often utilized in spatial analyses.

10.3.1 Retrieval, Reclassification and Generalization

Perhaps the initial GIS analysis that any user undertakes is the retrieval and/or reclassification of data. Retrieval operations occur on both spatial and attribute data. Often data is selected by an attribute subset and viewed graphically. Retrieval involves the selective search, manipulation, and output of data without the requirement to modify the geographic location of the features involved.

Reclassification involves the selection and presentation of a selected layer of data based on the classes or values of a specific attribute, e.g. cover group. It involves looking at an attribute, or a series of attributes, for a single data layer and classifying the data layer based on the range of values of the attribute. Accordingly, features adjacent to one another that have a common value, e.g. settlement type, but differ in other characteristics, e.g. place name or EA type, will be treated and appear as one class.

Reclassification is an attribute generalization technique. Typically this function makes use of polygon patterning techniques such as crosshatching and/or color shading for graphic representation. In a vector based GIS, boundaries between polygons of common reclassified values should be dissolved to create a cleaner map of homogeneous continuity. Raster reclassification intrinsically involves boundary dissolving. The dissolving of map boundaries based on a specific attribute value often results in a new data layer being created. This is often done for visual clarity in the creation of derived maps. Almost all GIS software provides the capability to easily dissolve boundaries based on
the results of a reclassification. Some systems allow the user to create a new data layer for the reclassification while others simply dissolve the boundaries during data output.

One can see how the querying capability of the DBMS is a necessity in the reclassification process. The ability and process for displaying the results of reclassification, a map or report, will vary depending on the GIS. In some systems the querying process is independent from data display functions, while in others they are integrated and querying is done in a graphics mode. The exact process for undertaking a reclassification varies greatly from GIS to GIS. Some will store results of the query in query sets independent from the DBMS, while others store the results in a newly created attribute column in the DBMS. The approach varies drastically depending on the architecture of the GIS software.

10.3.2 Topological Overlay
The capability to overlay multiple data layers in a vertical fashion is the most required and common technique in geographic data processing. In fact, the use of a topological data structure can be traced back to the need for overlaying vector data layers. With the advent of the concepts of mathematical topology polygon overlay has become the most popular geoprocessing tool, and the basis of any functional GIS software package.

Topological overlay is predominantly concerned with overlaying polygon data with polygon data, e.g. District polygons and EA polygons. However, there are requirements for overlaying point, linear, and polygon data in selected combinations, e.g. point in polygon, line in polygon, and polygon on polygon are the most common. Vector and raster based software differ considerably in their approach to topological overlay.

Most GIS software makes use of a consistent logic for the overlay of multiple data layers. The rules of Boolean logic are used to operate on the attributes and spatial properties of geographic features. Boolean algebra uses the operators AND, OR, XOR, NOT to see whether a particular condition is true or false. Boolean logic represents all possible combinations of spatial interaction between different features. The implementation of Boolean operators is often transparent to the user.

To date the primary analysis technique used in GIS applications, vector and raster, is the topological overlay of selected data layers.

Generally, GIS software implements the overlay of different vector data layers by combining the spatial and attribute data files of the layers to create a new data layer. Again, different GIS software utilize varying approaches for the display and reporting of overlay results. Some systems require that topological overlay occur on only two data layers at a time, creating a third layer. This pairwise approach requires the nesting of multiple overlays to generate a final overlay product, if more than two data layers are involved. This can result in numerous intermediate or temporary data layers. Other systems allow the user to overlay multiple data layers at one time. Each approach has its drawbacks depending on the application and the nature of the implementation. Determining the most appropriate method is based on the type of application, practical considerations such as data volumes and CPU power, and other considerations such personnel and time requirements. Overall, the flexibility provided to the operator and the level of performance varies widely among GIS software offerings.

The following diagram illustrates a typical overlay requirements where several different layers are spatially joined to created a new topological layer. By combining multiple layers in a topological fashion complex queries can be answered concerning attributes of any layer.
10.3.3 Neighbourhood Operations

Neighbourhood operations evaluate the characteristics of an area surrounding a specific location. Virtually all GIS software provides some form of neighbourhood analysis. A range of different neighbourhood functions exist. The analysis of topographic features, e.g. the relief of the landscape, is normally categorized as being a neighbourhood operation. This involves a variety of point interpolation techniques including slope and aspect calculations, contour generation, and Thiessen polygons.

Interpolation is defined as the method of predicting unknown values using known values of neighbouring locations. Interpolation is utilized most often with point based elevation data.

This example illustrates a continuous surface that has been created by interpolating sample data points.

Elevation data usually takes the form of irregular or regular spaced points. Irregularly spaced points are stored in a Triangular Irregular Network (TIN). A TIN is a vector topological network of triangular facets generated by joining the irregular points with straight line segments. The TIN structure is utilized when irregular data is available, predominantly in vector based systems. TIN is a vector data model for 3-D data.

An alternative in storing elevation data is the regular point Digital Elevation Model (DEM). The term DEM usually refers to a grid of regularly spaced elevation points. These points are usually stored with a raster data model. Most GIS software offerings provide three dimensional analysis capabilities in a separate module of the software. Again, they vary considerably with respect to their functionality and the level of integration between the 3-D module and the other more typical analysis functions.
Without doubt the most common neighbourhood function is buffering. Buffering involves the ability to create distance buffers around selected features, be it points, lines, or areas. Buffers are created as polygons because they represent an area around a feature. Buffering is also referred to as corridor or zone generation with the raster data model.

Buffering is typically used with point or linear features. The generation of buffers for selected features is frequently based on a distance from that feature, or on a specific attribute of that feature. For example, some features may have a greater zone of influence due to specific characteristics, e.g. a primary highway would generally have a greater influence than a gravel road. Accordingly, different size buffers can be generated for features within a data layer based on selected attribute values or feature types.

10.3.4 Connectivity Analysis

The distinguishing feature of connectivity operations is that they use functions that accumulate values over an area being traversed. Most often these include the analysis of surfaces and networks. Connectivity functions include proximity analysis, network analysis, spread functions, and three dimensional surface analysis such as visibility and perspective viewing. This category of analysis techniques is the least developed in commercial GIS software. Consequently, there is often a great difference in the functionality offered between GIS software offerings.

Proximity analysis techniques are primarily concerned with the proximity of one feature to another. Usually proximity is defined as the ability to identify any feature that is near any other feature based on location, attribute value, or a specific distance. A simple example is identifying all the forest stands that are within 100 metres of a gravel road, but not necessarily adjacent to it. It is important to note that neighbourhood buffering is often categorized as being a proximity analysis capability. Depending on the particular GIS software package, the data model employed, and the operational architecture of the software it may be difficult to distinguish proximity analysis and buffering.

Figure 9: Proximity Analysis

Proximity analysis is often used in urban based applications to consider areas of influence, and ownership queries. Proximity to roads and engineering infrastructure is typically important for development planning, tax calculations, and utility billing.

The identification of adjacency is another proximity analysis function. Adjacency is defined as the ability to identify any feature having certain attributes that exhibit adjacency with other selected features having certain attributes. A typical example is the ability to identify all EAs of a specific type, adjacent to a specific road or river.

Network analysis is a widely used analysis technique. Network analysis techniques can be characterized by their use of feature networks. Feature networks are almost entirely comprised of
linear features. Hydrographic hierarchies and transportation networks are prime examples. Two example network analysis techniques are the allocation of values to selected features within the network to determine capacity zones, and the determination of shortest path between connected points or nodes within the network based on attribute values. This is often referred to as route optimization. Attribute values may be as simple as minimal distance, or more complex involving a model using several attributes defining rate of flow, impedance, and cost.

**Three dimensional analysis** involves a range of different capabilities. The most utilized is the generation of perspective surfaces. Perspective surfaces are usually represented by a wire frame diagram reflecting profiles of the landscape, e.g. every 100 metres. These profiles viewed together, with the removal of hidden lines, provide a three dimensional view. As previously identified, most GIS software packages offer 3-D capabilities in a separate module. Several other functions are normally available.

These include the following functions:

- user definable vertical exaggeration, viewing azimuth, and elevation angle;
- identification of viewsheds, e.g. seen versus unseen areas;
- the draping of features, e.g. point, lines, and shaded polygons onto the perspective surface;
- generation of shaded relief models simulating illumination;
- generation of cross section profiles;
- presentation of symbology on the 3-D surface; and
- line of sight perspective views from user defined viewpoints.

While the primitive analytical functions have been presented the reader should be aware that a wide range of more specific and detailed capabilities do exist.

**The overriding theme of all GIS software is that the analytical functions are totally integrated with the DBMS component. This integration provides the necessary foundation for all analysis techniques.**

## 11. Implementation Issues and Strategies

### 11.1 Current Options and Software Assessment

Perhaps the first question asked by anyone when discovering GIS is *what are the current options available?* This question is often asked as directly as *what is the best GIS?* Quite simply, there are no best GIS. A wide variety of GIS software offerings exist in the commercial market place. Commercial surveys often are a good starting point in the assessment of GIS software.

One of the problems in evaluating the functionality of GIS software is the bias one gets from using one system or another. Comparing similar functions between systems is often confusing. Like any software, ultimately some do particular tasks better than others, and also some lack functionality compared to others. Due mostly to this diverse range of different architectures and the complex nature of spatial analysis no standard evaluation technique or method has been established to date.

Any GIS should be evaluated strictly in terms of the potential user's needs and requirements in consideration of their work procedures, production requirements, and organizational context! The experienced GIS consultant can play a large and valuable role in the assessment process.

A current accepted approach to selecting the appropriate GIS involves establishing a benchmark utilizing real data that best represents the normal workflow and processes employed in your organization. The identification of potential needs and requirements is essential in developing a proper benchmark with which to evaluate GIS software packages. A formalized user need analysis is absolutely critical to the successful implementation of GIS technology.

Development of the benchmark should include a consideration of other roles within your organization that may require integration with the GIS technology. A logical and systematic
approach as such is consistent with existing information systems (IS) planning methodologies and will ultimately provide a mechanism for a successful evaluation process.

11.2 Justification and Expectations

GIS is a long term investment that matures over time. The turnaround for results may be longer term than initially expected. Quite simply, GIS has a steep learning curve. The realization of positive results and benefits will be not achieved overnight. Both initial investment funding and continued financial support are major determinants in the success or failure of a GIS.

Most often the justification and acquisition of a GIS centres on technical issues of computer hardware and software, functional requirements, and performance standards. But experience has shown that, as important as these issues may be, they are not the ones that in the end determine whether a GIS implementation will succeed or not.

Even though the proper assessment of an appropriate GIS product requires a good understanding of user's needs, most often systems are acquired based on less than complete and biased evaluations. Nonetheless, even with the GIS in hand a properly structured and systematic implementation plan is required for a successful operation (the method component of a GIS). Generally, a GIS implementation plan must address the following technical, financial, and institutional considerations:

- system acquisition tactics and costs;
- data requirements and costs;
- database design;
- initial data loading requirements and costs;
- system installation tactics, timetable, and costs;
- system life cycle and replacement costs;
- day-to-day operating procedures and costs;
- staffing requirements and costs;
- user training and costs; and
- application development and costs.

Potential GIS buyers should be aware of the necessary investment required in hardware, software, training, supplies, and staffing. The cost of establishing a successful GIS operation is substantial. However, with realistic expectations and support the development of GIS within an organization that manipulates geographic data will almost certainly prove beneficial. Certain considerations of data longevity, data capture, personnel hiring, etc. are the practical concerns of GIS implementation. The longer term implications, such as hardware/software maintenance and replacement, should also be considered.

The acquisition of GIS technology should not be done without seriously considering the way in which GIS will interact with the rest of the organization. Without adequate financial and institutional support from the statistical organization, the GIS implementation will fail and it will never be sustainable. The problem with many African statistics organizations is that GIS is implemented as part of a large project, such as a population census, for which money is generally made available. However, after the completion of the census, the maintenance and additional development costs needed for the GIS to be sustainable and flourish are never allocated as part of an annual budget, meaning that the unit dies a slow death.

Therefore, it absolutely needs buy in and support from senior management and the organization as a whole in order for it to benefit the organization as a whole.

It is simply not enough to purchase a computer, a plotter, a display device, and some software and to put it into a corner with some enthusiastic persons and then expect immediate returns. A serious commitment to GIS implies a major impact on the whole organization.
11.3 Implementation Issues

The mere presence of an implementation plan does not guarantee success. Most organizations do not have sufficient staff to cope with the commitment and extra work required when introducing a GIS to existing operations. GIS implementation must also consider all technology transfer processes.

11.3.1 Common Pitfalls and Lessons Learnt

Several pitfalls exist that most often contribute to the failure of a GIS implementation strategy. These are identified below:

- **Failure to identify and involve all users**
  Users in an operational GIS environment consist of operations, management, and policy levels of the organization. All three levels should be considered when identifying the needs of your users. Moreover, the data product needs of external users must also be identified.

- **Failure to match GIS capability and needs**
  A wide spectrum of GIS hardware and software choices currently exist. The buyer is presented with a significant challenge making the right choice. Remember, the right choice will be the GIS that provide the needed performance no more, no less for the minimum investment. The success of a GIS implementation is particularly sensitive to the right hardware and software choices!

- **Failure to identify total costs**
  The GIS acquisition cost is relatively easy to identify. However, it will represent a very small fraction of the total cost of implementing a GIS. Ongoing costs are substantial and include hardware and software maintenance, staffing, system administration, initial data loading, data updating, custom programming, and consulting fees. It is therefore essential that the GIS Unit annual costs are part of the annual budget.

- **Failure to conduct a pilot study**
  The GIS implementation plan concerns itself with the many technical and administrative issues and their related cost impacts. Three of the most crucial issues, are database design, data loading and maintenance, and day-to-day operations. The pilot study will allow you to gather detailed observations, provided it is properly designed, to allow you to effectively estimate the operational requirements.

- **Lack of a basic implementation plan**
  An implementation plan needs to be created where the relevant issues are documented and implementation processes identified per GIS component. Additional logistical issues such as procurement, office space etc. must also be taken into account.

- **Giving the GIS implementation responsibility to only existing Geography Unit staff**
  Because of the distinct differences of the GIS from conventional geography and field mapping operations, there is a need to add additional GIS specialist staff which has the necessary background. The specialized skills of the ‘GIS analyst’ are required at this stage. Reliance on conventional geography personnel who lack these skills will ensure failure. However, care must be taken to do a skills analysis amongst existing staff to determine which staff members has the potential to be trained as GIS technicians.

- **Failure to consider technology transfer**
  Training and support for on-going learning, for in-house staff as well as new personnel, is essential for a successful implementation. Staff at the three levels should be educated with respect to the role of the GIS in the organization. Education and knowledge of the GIS can only be obtained through on-going learning exercises. Nothing can replace the investment of hands on time with a GIS!
11.4 The Learning Curve

Contrary to information provided by commercial vendors of GIS software, there is a substantial learning curve associated with GIS. It is normally not a technology that one becomes proficient in overnight. It requires an understanding of geographical relationships accompanied by committed hands-on time to fully apply the technology in a responsible and cost effective manner. Proficiency and productivity are only obtained through applied hands-on time with the system! GIS is an applied science. Nothing can replace the investment of hands-on time with GIS. The following figure presents the typical learning curve for GIS installations.

Figure 10: The GIS Learning Curve

The learning curve is dependent on a variety of factors including:

- the amount of time spent by the individual with hands-on access;
- the skills, aptitude and motivation of the individual;
- the commitment and priority attached to GIS technology dictated by the organization and management;
- the availability of data; and
- the choice of software and hardware platforms.

A critical requirement for all GIS implementations is that adequate education and training is provided for operational staff, as well as realistic priorities are defined with which to learn and apply the technology. This is where a formal training curriculum is required to ensure that time is dedicated to learning the technology properly. Adding GIS activities to a staff member's responsibilities without establishing well defined milestones and providing adequate time and training mechanisms is prone to failure. A focused and properly trained operations staff that has consistent training will result in greatly reduced turnaround times for operations, and ensure consistency in quality of product.

The threshold point of the learning curve is typically around the two year time frame. However, this is dependent on the ability of the organization to establish a well defined and structured
implementation plan that affords appropriate training and resources for technical staff. The flat part of the learning curve can be shortened if proper training is provided, data is available for use, the right software and hardware is acquired.

The typical learning curve reflects a long initial period for understanding spatial data compilation requirements and database architecture. However, after data models are well understood and sufficient data compilation has been completed the learning curve accelerates. Once a formal application development environment is established and user needs are well defined an infrastructure exists for effective application of the technology. Building operational applications based on formal functional specifications will result in continued accelerated learning. The data hurdle is often a stumbling block for many GIS users, meaning that relevant data is unavailable to accelerate the learning process.

11.4.1 GIS Productivity
GIS is a long term investment that matures over time. The turnaround for results may be longer than initially expected. The establishment of a formal implementation strategy will help to ensure that realistic expectations are met. Data is the framework for successful application of GIS technology. In this respect, the investment in establishing a solid data platform will reap rewards in a short term timeframe for establishing a cost-effective and productive GIS operation. The availability of quality data supplemented by a planned implementation strategy is the cornerstones of achieving a productive and successful GIS operation. A robust database should be considered an asset.

However, even with a well defined and systematic implementation strategy GIS technology will not provide immediate benefits. Benefits and increased productivity are not achieved overnight. GIS technology is complex in nature, has a generally steep learning curve, and requires a complement of skills for it to be applied successfully. In fact, most organizations realize a loss in overall operational productivity over the short term while the GIS platforms are being installed, staff is trained, the learning curve is initiated, and data is being captured. This is common of all office automation activities.

A GIS acquisition based on well defined user needs and priorities is more likely to succeed than without. A major pitfall of most installations with GIS technology, particularly government agencies, is the lack of well defined user needs on which to base the GIS acquisition and implementation.

12. Implementing a GIS
Implementing a GIS within a statistical organization might not be a straight forward process as we would like to think. The following section provides some details as to which implementation processes can be followed to ensure a sustainable GIS and use of spatial information. Implementation can be seen as a six phase process.

12.1 Creating an awareness
GIS needs to be sold within an organization. The education of staff is very important. Depending on the way in which GIS technology is being introduced to the organization the process for creating awareness may differ. Technical workshops are often appropriate when a top-down approach exists; while management workshops are often more relevant when a bottoms-up approach exists. Education of the new technology should focus on identifying existing problems within an organization. These often help justify a GIS acquisition.

They include:

- spatial information is poorly maintained or out of date;
- spatial data is not recorded or stored in a standard way;
- spatial data may not be defined in a consistent manner,
- data is not shared between departments within an organization;
- data retrieval and manipulation capabilities are inadequate to meet existing needs; and
new demands are made on the organization that cannot be met with existing information systems (specifically with regard to the analysis and dissemination of census data, small area statistics creation and poverty mapping.

12.2 Identifying System Requirements

The definition of system requirements is usually done in a user needs analysis. A user needs analysis identifies users of a system and all information products required by those users. Often a prioritization of the information products and the data requirements of those products is also undertaken. A proper user needs analysis is crucial to the successful evaluation of GIS software alternatives.

After user needs have been identified and prioritized they must be translated into functional requirements. Ideally, the functional requirements definition will result in a set of processing functions, system capabilities, and hardware requirements, e.g. data storage, performance. Experienced GIS consultants often play a major role in this phase. Moreover, the following issues must be carefully assessed with findings documented and recommendations made:

- Hardware required
- Software required
- IT requirements
- Data and database required
- Staff and training (skills sets) required
- Methodology to be implemented
- Institutional arrangement (where will the GIS Unit fit into the organizational structure)
- Logistical considerations
  - Office space
  - Furniture and additional equipment
  - Procurement procedures
  - Funding
- Maintenance and sustainability issues

12.3 System Evaluations

Evaluating alternative hardware and software solutions is normally conducted in several stages. Initially a number of candidate systems are identified. Information to support this process is acquired through demonstrations, vendor literature, etc. A short listing of candidates normally occurs based on a low level assessment. This followed by a high level assessment based on the functional requirements identified in the previous phase. This often results in a rating matrix or template. The assessment should take into account production priorities and their appropriate functional translation. After systems have been evaluated based on functional requirements a short list is prepared for those vendors deemed suitable. A standard benchmark is then used to determine the system of choice.

It should be noted that currently the 2 major GIS software used by statistical agencies in Africa are the ESRI ArcGis suite and the Intergraph Geomedia Professional suite. These two are the most prominent, but also the most expensive. Various free GIS software packages exist but cannot be compared with vendor specific software regarding functionality and maintenance. Open source GIS options, such as GRASS also exist but require skilled staff to drive further application development. Open source GIS software options also have some way to go to compare favourably with vendor driven applications regarding functionality and maintenance, but ongoing development means it will always be an option for cash strapped departments.

12.4 Justifying the System Acquisition

The proper justification of the chosen system requires consideration of several factors. Typically a cost-benefit analysis is undertaken to analyze the expected costs and benefits of acquiring a system. To proceed further with acquisition the GIS should provide considerable benefits over expected costs. It is important that the identification of intangible benefits also be considered.
The justification process should also include an evaluation of other requirements. These include data base development requirements, e.g. existing data versus new data needs and associated costs; technological needs, e.g. maintenance, training, and organizational requirements, e.g. new staff, reclassification of existing job descriptions for those staff who will use the GIS.

12.5 System Acquisition and Start Up
After the system, e.g. hardware, software, and data, is acquired the start up phase begins. This phase should include pilot projects. Pilot projects are a valuable means of assessing progress and identifying problems early, before significant resources have been wasted. Also, because of the costs associated with implementing a GIS it is often appropriate to generate some results quickly to appease management. First impressions are often long remembered.

12.6 Operational Phase
The operational phase of a GIS implementation involves the on-going maintenance, application, and development of the GIS. The issue of responsibility for the system and liability is critical. It is important that appropriate security and transaction control mechanisms be in place to support the system. A systematic approach to system management, e.g. hardware, software, and data, is essential.

12.7 Implementation plan – a backbone structure
The following can be seen as a backbone structure which can be followed to design your implementation plan.

12.7.1 Phase 1 - Awareness
Selling the GIS to your organization
- Obtaining a Champion for your GIS proposal
  - Management as Champion
  - Workers as Champions
  - Consultant as Champion
- Reasons for considering a GIS?
  - Data management issues
  - Data retrieval & manipulation issues

12.7.2 Phase 2 - System Requirements
- Identify components of your organization
- Data Input and Output
- Procedures and Policies
- Applications
- Users
- Functional Requirements Grid
  - Identify Current Technology
  - Identify Current Users
  - Identify Current Tasks
- Perform a formal Needs Analysis
  - Goals, Objectives, and Purposes
    1. Detail Organizational Requirements
    2. Data Source
    3. Primary Applications
    4. Identify Users
    5. Deadlines
- Current System Analysis
  - Data quality and completeness
  - Hardware scalable?
  - Recommendations

12.7.3 Phase 3 - System Evaluation
Evaluate alternative systems -- comparative analysis of competing systems
- Use Needs Analysis of Phase 2
  - State Goals and Objectives
    - Organization Policy
    - Primary Applications
    - Data Sources
  - System Integration with current system
  - System Cost / Expense Analysis
    - Hardware Conversion
    - Software Conversion & Training
    - Data Conversion
- System Benchmarking
  - Evaluate systems with user data.
    - Perform analysis and manipulation of data as indicated in Goals and Objectives
    - Develop Pilot study with
      - Costly but informative
- Compare performances
  - Did systems work as advertised?
  - How did systems compare to system requirements?
- Generate Report
  - Recommendations for system
  - Identify funding source
  - Identify Staff requirements
  - Identify organizational changes

### 12.7.4 Phase 4 – Develop Implementation Plan

Multiple approaches to implementation
- Acquisition Issues
  - When to Purchase?
  - What to Purchase?
    - Hardware, Software, Data, and Training
  - Needs for Purchase
    - Support Structure, Vendor Information, Vendor Contracts
    - Cost of Purchase
- Organizational Changes
  - Management of GIS
  - Prepare for Changes
- Funding Strategies
  - What are the costs?
  - Where is the funding?
- Application Development
  - Purpose
  - Goals
  - Methods
  - Procedures
  - Instructions
- Data Conversion and Development
  - QA/QC
  - Accuracy Issues
  - Metadata

### 12.7.5 Phase 5 – System Start-up

- Contracting Issues
  - Request for Proposal (RFP)
    - Detail Goals and Objectives
      - Required Applications
      - Required Data


- **Organization Policy**
  - Detail Services
  - Detail Requirements
  - Perform Test
  - Specify Maintenance Contract
    - Tech Support
    - Product Upgrades
    - Application trouble-shooting

- **Apply the Implementation Policy**
  - Dealing with unexpected problems
  - Pilot Project
    - Develop Pilot Application
    - Develop complete dataset for small study area.
      - Pursue an easily attainable project with high probability of success
  - Plan and execute as if full-scale application
    - Participation of all staff/users
    - Use your resources
      - Available staff
      - Consultants

- **Report Results**
  - Detail failures and corrections necessary

### 12.7.6 Phase 6 - Operational Phase

- Define & Implement Procedures
  - Updates/Upgrades
    - Hardware Upgrades
    - Software Updates
    - Training
    - Data Management

- Promote GIS
  - Actively Promote GIS to remind all of benefits
    - Ensures budgetary needs

- Establish Security & Accountability
  - Develop Information Policies
    - Distribution Policy
      - Technical, Legal, Economic, and Political implications of distribution.

- Reporting

### 12.8 Country experiences regarding GIS implementation

Relevant country experiences through the GDDS initiative has brought to the fore many lessons learnt regarding GIS implementation. It seems that a lot of confusion exists regarding the equipment, software, training, staffing and data needed to implement a GIS. The following section attempts to shed some light on the issue by way of a short implementation summary. This summary will be structured according to the following headings:

- Hardware
- Software
- Human Resources and Training
- Data and Database
- Institutional Arrangement
- Administrative issues

#### 12.8.1 Hardware

**Computers**
The number of PCs needed for a GIS Unit usually depends on the size of the organization as well as the amount of data available. Typically, 3-5 high end personal computers are sufficient to kick start
GIS operations. Note that these should not be hand-me-down PCs or PCs normally used for word processing and the like. GIS software and data demands are such that computers with sufficient CPU processing power, storage capacity, memory capacity and graphics interface are necessary.

**PC screens**
It is desirable to work on high definition large format screens, at least 17 inch wide since it greatly increases the ease of GIS panning and digitising operations while also enhances on-screen image interpretation.

**Printers and Plotters**
Since large numbers of fieldwork maps need to be printed for survey and census operations, it is desirable to have at least one high end A3 format colour printer, coupled with one high end large format A0 colour plotter to cater for ad hoc map dissemination tasks. These printers and plotters should be networked through a Local Area Network (LAN) to the GIS computers and server.

**Data server**
Due to the large amount of spatial and attribute information being generated by a successful GIS, it is necessary to have all your GIS data stored on one server to which GIS users have access to. Too often GIS data is stored and duplicated on individual PCs with no backups made. Pretty soon various versions of the same dataset is everywhere and confusion reigns. A data server is therefore essential, especially when digital imagery comes into play. This data server should also be networked to the GIS user's PCs via a LAN.

**Internet Connectivity**
GIS users are heavily dependent on internet access for issues such as product licensing, online maintenance and support as well as online data mining and data sharing. Fast, broadband internet access is therefore essential and should be supported by an effective IT network.

**GPS Units**
GPS units are an indispensable field data collection tool and every statistics organization should use them. High end data loggers which are able to capture line, point and polygon features and sufficient attribute data with data dictionary functionality are desirable (such as the Trimble Juno). GPS units able to capture only waypoints and basic attribute information are NOT desirable.

### 12.8.2 Software

**GIS Software**
Many GIS software options exist and the choices can become quite confusing. GIS Units in statistical agencies do a lot of data collection, data creation, digitising and data integration from various sources, in various formats and various coordinate systems. Moreover, spatial analysis and thematic mapping is essential to analyse and disseminate census and survey results, not to mention the demands of poverty mapping. Therefore, your first question should be if the GIS you ultimately choose have this functionality? To provide some guidance, there are currently two commercial GIS software options available that are widely used in Africa by statistical agencies. These are the ArcGis suite from ESRI as well as the Geomedia Professional suite from Intergraph. Both these are safe bets. MapInfo is also a good choice. All three these options are however expensive and have many add on modules that can be acquired at an additional expense. It is always a good idea to consult a GIS specialist regarding your specific needs. Also, always ensure that maintenance and upgrades are part of the pricing package before signing on the dotted line.

**GPS Software**
Typically, a high end GPS unit will be accompanied by software which enables one to download, view and edit GPS data, such as the Trimble pathfinder software. When buying GPS units, make sure that if there are any specific software in addition to the hardware, that it is also acquired.

**Database software**
Most high end GIS software interfaces with relational or object orientated database management software, such as Oracle, MS Access or SQL Server. Due to the amount and types of data and statistics collected by statistical agencies, it is imperative that data is stored within a stable database management system. Spatial data can also be stored within these database management systems. The two systems most commonly used are Oracle and SQL Server and the software would typically run on your dedicated data server which is networked to relevant users.

**12.8.3 Human Resources and training**

It is imperative that skilled staff is appointed to drive the GIS. Skills and education gaps that might exist within the current cartographic or geographic staffing structures needs to be assessed and skills gaps identified. Moreover, it is important that the trainability of staff is also assessed. GIS is an IT intensive technology and therefore basic IT and computer skills need to exist. If these basic skills do not exist, then additional, appropriately qualified persons must be appointed. Experience has also shown that a tertiary qualification in GIS or applied geography does not necessarily mean that the relevant person can do GIS in a practical environment. Additional training will always be necessary, GIS is a dynamic and multi disciplinary application field with new functionality, uses and operations being developed constantly, which means there should be a budget allocation for refresher or specialised training at least every two years for GIS staff.

Again, it would be prudent to appoint a qualified consultant to assess current staff skills and training needs. While formal training is necessary in the long run, in the short term, to get your system up and running, intensive practically base short courses of two to three weeks in various disciplines have proven very effective. GIS is, after all, a practical application. Typical subjects for training might be:

- Basic GIS principles and use
- Remote sensing use and interpretation for statistical agencies
- Specific GIS use for statistical agencies regarding:
  - Census Mapping
  - Survey design, mapping and implementation
  - Map creation and printing
  - Fieldwork management and monitoring
  - Data collection
  - Data management, storage, manipulation and analysis
  - Spatial analysis
  - Thematic mapping
  - GIS based census and survey data products
  - Spatial statistics dissemination
  - Web based GIS options
  - Database options and design
  - Database administration
  - Infrastructure set up and GIS implementation
  - Organizational requirements
- GPS use and field operations
- GPS data integration
- Specific GIS software training

**12.8.4 Data and database**

As mentioned before, GIS is data. There are certain base datasets that every GIS Unit in a statistical agency needs access to and these datasets should preferably be stored in a proper database management system on a dedicated data server. It would be necessary to do a data audit on current available data and identify the necessary data gaps that might exist. The data audit should take the following into account:

- Is it primary or secondary data?
- Who is the owner of the data?
- What is the currency of the data?
• Are there any restrictions on the use of the data?
• How was the data generated?
• What is the accuracy of the data?
• Are there any metadata available for a particular dataset?

**The following types of data should be investigated:**

- **Geographic data**
  - Primary base data, including satellite imagery, digital aerial photography, topographic 1:50000 map sheets, cadastral data, infrastructure data
  - Secondary data, including points of interests, landmarks, statistical boundaries such as health districts, current census related boundaries, such as EA and SA boundaries, administrative boundaries, zoning schemes, current land use, utilities (electricity network, water articulation network), waste management sites, environmental data

- **Attribute data linked to spatial data**
  - Survey specific demographic data, Census data, data gathered from socio-economic surveys, community survey data, infrastructure and utilities attribute data, service delivery attribute data, land information and ownership data, address data, municipal accounts data, land valuation data.

- **Attribute data not necessarily linked to spatial data**
  - Previous census and survey data not digitally linked to spatial features
  - Economic and business data
  - Donor generated data

A detailed gap analysis should follow which will investigate which data sets are lacking in which areas. Moreover, available datasets which does not have the desirable accuracy or currency should be flagged for replacement or alternative datasets should be identified.

The current database management system (if any) should be assessed as well to determine whether the planned GIS software you want to acquire can seamlessly link with the relevant DBMS. One would also need to look at the type and amount of data that needs to be stored and managed within the DBMS. Whether a DBMS exist or not, it will be necessary to plan, design and build a relevant spatial database to which the GIS can link. If in-house expertise does not exist, a qualified GIS database consultant needs to be appointed to assess your needs and design and populate the database, with the appropriate training needed.

**12.8.5 Institutional Arrangement**

The statistics organization as a whole needs to be ready for the implementation of GIS. Often, due to ignorance, senior managers might be sceptical about the potential use of GIS and how it can benefit them. Therefore, awareness needs to be created within the organization about the potential and use of GIS. Moreover, awareness needs to be created within all government structures as to how GIS can increase better service delivery on all levels. Too often the focus is only on map creation and not on enabling GIS to be a corporate service provider to the entire organization. The issue of sufficient budget allocation and support has also been discussed and is vital to the sustained success of any GIS unit.

**12.8.6 Administrative issues**

Experience has shown that tedious and ineffective financial allocation, tender and procurement procedures has delayed and negatively affected GIS implementation in many organizations. This is an unfortunate circumstance that needs to be attended to.

GIS offices, due to their equipment intensive nature, needs to be properly air-conditioned and sufficient space needs to be available to accommodate not only staff, but large format plotters and hardcopy map archives.
13. Developments and Trends

The development and application of geographic information systems is vibrant and exciting. The term GIS remains one of the most popular buzz words in the computer industry today and even more so in the socio-economic survey industry. While GIS has been in use at statistical agencies in developing countries for many years (and in fact is a crucial part in the organizational structure), developing countries, specifically in Africa, lag behind. Very few statistics agencies in Africa has GIS Units which perform to their true potential and forms an integral part of the survey, analysis and dissemination operations of these organizations.

GIS is very much a multi-disciplinary tool for the management of spatial data. It is inherently complex because of the need to integrate data from a variety of sources. Functions must accommodate several application areas in a detailed and efficient manner. A variety of important developments are occurring which will have profound effects on the use of GIS. They are identified in the following sections.

13.1 New Data Sources

The generation of data from new sources is an ongoing development. Application specialists have traditionally attempted to research and implement new data sources into their work. Most of these new data sources are based strictly on scientific technological developments.

13.1.1 Remote Sensing

Remote sensing will become, if it is not already, the primary source for new data. Due to recent technological developments in hardware most GIS software can now accommodate remotely sensed imagery at high resolutions, and in varying formats. Remote sensing data can include aerial photographs, satellite imagery, radar imagery, etc. Some of the past problems with using remotely sensed imagery have been the inability to integrate it with other data layers, particularly vector encoded data. Remote sensing specialists stress that their data is of most value when combined with, and substantiated by, other data sources. Several commercial GIS products are now offering their software bundled with an image processing software package. Many of these packages allow you to interactively view data from both systems simultaneously, and also afford the conversion of data between systems.

Remote sensing is a technology used for obtaining information about a target through the analysis of data acquired from the target at a distance. It is composed of three parts, the targets - objects or phenomena in an area; the data acquisition - through certain instruments; and the data analysis - again by some devices. This definition is so broad that the vision system of human eyes, sonar sounding of the sea floor, ultrasound and x-rays used in medical sciences, laser probing of atmospheric particles, are all included. The target can be as big as the earth, the moon and other planets, or as small as biological cells that can only be seen through microscopes.

Remote sensing data acquisition can be conducted on such platforms as aircraft, satellites, balloons, rockets, space shuttles, etc. Inside or on-board these platforms, we use sensors to collect data. Sensors include aerial photographic cameras and non-photographic instruments, such as radiometers, electro-optical scanners, radar systems, etc.

Electro-magnetic energy is reflected, transmitted or emitted by the target and recorded by the sensor. Because energy travels through the medium of the earth's atmosphere, it is modified such that the signal between the target and the sensor will differ.
Digital imagery can greatly enhance a GIS. Imagery is a powerful visual aid and serves as a source of derivative information such as planimetrics and classification schemes to derive such information as land use or vegetation. If your GIS cover a particularly large area, aerial imagery may not be a practical or economical choice. Satellite imagery is often the preferred choice of imagery for larger regions. More and more choices of satellite imagery are becoming available and the cost associated with its purchase is dropping.

The integration of GIS and image processing capabilities offers a great potential for resource specialists and has not gone unnoticed in the field of demographic surveys. Geodemography is essentially the study of where people are and remote sensing provides an interesting alternative to traditionally based fieldwork.

Imagery provides a statistical agency with an accurate and current geographic base from which to plan, implement and monitor socio-demographic surveys from. Due to the accuracy of imagery, the geographic frame of the country, with all its administrative boundaries, can be re-demarcated to fit the imagery. Where, previously, fieldworkers had to use sketch maps and outdated toposheet maps to facilitate field operations, it is now possible to provide a fieldworker with a photo map, enabling them to navigate correctly and eliminate demarcation and enumeration errors due to outdated or inappropriate base data. Imagery also provides an ideal base from which to implement spatial analysis and dissemination operations from. Especially in developing areas, where current and accurate base information rarely exists, imagery can provide the answer and provides the following advantages:

- **User needs**
  - User needs have become more sophisticated and many users demand the visual impact and accuracy of remotely sensed imagery with traditional thematic mapping products
- **Time**
  - The use of imagery decreases the amount of time spent in the field since fieldworkers operate more effectively with current imagery maps in stead of maps with inappropriate or outdated base data. Moreover, the use of imagery with GIS allows one to plan better and enables a GIS Unit to immediately convert statistical and administrative boundaries to digital format through heads-up digitising.
- **Advances in technology**
  - Advances in CPU speed and storage capacity now allows users to easily store, manipulate and view images on PC workstations and laptops
- **The cost factor**
Since fieldwork time is greatly minimised, fieldwork cost are also minimised and traditionally, demarcation fieldwork is the most expensive part of any census mapping effort.

- Capacity building
  - Introducing the use of imagery in your GIS forces staff and users alike to become familiar with the various products and its uses as well as how to interpret imagery correctly.

- The issue of quality
  - In Africa, the base information used for most urban and rural areas are in most cases topographic map information of a scale 1:50000 and in some cases, outdated town plan maps of dense urban areas in particular. Consequently, the use of high resolution imagery greatly enhances the accuracy of census mapping and therefore enumeration and therefore the census results in general.

Individual choice of imagery will depends on many factors such as resolution, image type and cost. The most useful imagery for census applications however, is typically high resolution natural color imagery (2.5m or less) on which dwellings, infrastructure and structures can be clearly identified. (SPOT 5 Supermode, Quickbird, Ikonos or Geoeye, for example)

Many of the imagery can be purchased directly from the associated agencies. In addition, there are online sites that also specialize in satellite imagery sales.

Here are some basic satellite imagery related terms that are useful:

**Hyperspectral imagery**
Useful for classifying material types on the Earth’s surface - beneficial in agriculture and forestry management, mineral exploration, environmental monitoring and national security activities.

**Interval**
Amount of time that passes before the satellite scans the same point of the globe.

**Multispectral**
Two or more images taken simultaneously, but each image taken in a different part of the electromagnetic spectrum.

**Panchromatic**
Imagery taken of all wavelengths within the visible spectrum, though not uniformly.

**Orbit**
The act of circling the earth. The type of orbit describes the path the satellite takes as it circles the earth.

**Resolution**
The amount of ground covered in one pixel of the image. For example an image with one meter resolution means that each pixel in the image represents one square meter on the ground. Click here for examples of imagery at different resolutions.

**Sun-synchronous**
An orbit that always passes over the earth at the same local sun time.

**Swath**
Amount of ground covered lengthwise in the passing of the satellite.
13.1.2 Global Positioning System (GPS) technology

Global Positioning System (GPS) devices can be found everywhere - they're used in cars, boats, airplanes, and even in cellular phones. Handheld GPS receivers are carried by hikers, surveyors, map makers, and others who need to know where they are.

The GPS is a technical marvel made possible by a group of satellites in earth orbit that transmit precise signals, allowing GPS receivers to calculate and display accurate location, speed, and time information to the user.

It consists of the control segment that monitors the satellites, the space segment including the satellite constellation, and us, the user segment. These three segments enable a GPS position (Latitude, Longitude and Height) to be calculated anywhere in the world at any time, in any weather by determining the distance to at least four satellites and knowing where the satellites are in space. Thus, by capturing the signals from four or more satellites (among a constellation of 31 satellites available), GPS receivers are able to triangulate data and pinpoint your location. With the addition of computing power, and data stored in memory such as road maps, points of interest, topographic information, and much more, GPS receivers are able to convert location, speed, and time information into a useful display format.

GPS was originally created by the United States Department of Defense (DOD) as a military application. The system has been active since the early 1980s, but began to become useful to civilians in the late 1990s. Consumer GPS has since become a multi-billion dollar industry with a wide array of products, services, and Internet-based utilities.

GPS works accurately in all weather conditions, day or night, around the clock, and around the globe. There is no subscription fee for use of GPS signals. GPS signals may be blocked by dense forest, canyon walls, or skyscrapers, and they don't penetrate indoor spaces well, so some locations may not permit accurate GPS navigation.

GPS receivers are generally accurate within 15 meters, and newer models that use Wide Area Augmentation System (WAAS) signals are accurate within three meters. The accuracy can be improved to the meter or even cm level by using different combinations of technique, receivers and software.

While the U.S. owned and operated GPS is currently the only active system, five other satellite-based global navigation systems are being developed by individual nations and by multi-nation consortiums.

Applications that use the Global Positioning System continue to grow. The GPS hardware and software is becoming more powerful. A good basic knowledge of both the GPS strengths and weaknesses will be useful for those willing to learn. The advances in the GPS technology during the last few years have been truly remarkable. While today's GPS equipment manufacturers try to make their products easier to use, they are also just as easy to misuse. Proper care must be taken to ensure the correct use of positions from GPS receivers.

With regard to statistical agencies and census taking, GPS use initially was introduced during the 2000 round of censuses in Africa and many agencies have since been using it on a regular basis to collect point, line and polygon field data. However, most agencies own basic handheld devices which can only capture points and has very limited attribute data capturing capability. New field specific GPS data loggers are continually being developed, culminating in devices called UMPC (Ultimate Mobile Personal Computers) which are ruggedized mobile field computers with high end GPS capability, real time and post processing positioning, geotagging functionality and mobile GIS functionality. These units are however very expensive.

Most GIS software packages also have the capability to directly integrate GPS captured information into the GIS data warehouse but that depends on the type of software and type of GPS used.
Be that as it may, as long as fieldwork and surveys are implemented, GPS use will have its place. For statistical agencies, GPS use can also greatly improve the monitoring of fieldworkers, while it provides a valuable tool for data collection, specifically if current digital imagery is not available for a specific area. Moreover, improvements in attribute data logging capabilities now allows fieldworkers to capture a lot of digital attribute data on these devices themselves which mean whole questionnaires can be captured digitally directly on the unit which negates the use of hardcopy questionnaires and data capturing.

14. The use of population census data in small area statistics

Small area statistics is the coupling of information recording location with other demographic, attribute data. This coupling defines a geographical dataset and since GIS are systems to capture, store, transform, analyse and display geographical data, it should not be surprising that GIS and small area statistics creation are stable mates.

It seems logical that census information should be the cornerstone of small areas statistics and classification since census datasets, for the most part, contains the core indicators needed to build such classifications, such as population, ethnicity, age, income, employment, access to amenities and communication and general living conditions. Moreover, with the use of GIS in statistical organizations, it now becomes possible to electronically link these datasets to a small geographical area, such as a census EA, zone or tract. By using spatial aggregation functionality resident within most GIS, it is therefore possible to use small census areas as the building blocks for neighbourhood classification and clustering, enabling the analyst to incorporate the spatial context of information within the demographic classification.

The term census originally referred to the enumeration and registration of people and property, often for the purpose of taxation. The knowledge gained from a census about the numbers of people, their characteristics, and their housing is of critical importance to governments, businesses, researchers and service providers.

Geospatial analysis can create new information. In the past, the reporting of census information was typically limited to standard administrative units. The power of geospatial analysis allows data to be retrieved for custom geographic areas (for example, a one km radius around a store, or a buffer along a transportation corridor.) In this new era of electronic data, the internet, and GIS, geospatial tools have removed the limitations of printed reports and standard administrative boundaries. The result is a fundamental change in how Census data can be utilized.

Demographics about populations can be collected in a variety of ways. The most comprehensive collection of demographics is via a Census. Typically, a Census is conducted with the goal of counting every member of a population. A critical aspect of Census data is that the data is typically reported as geographic totals. Data about individuals is usually not publicly available (at least not in a form that can identify specific individuals). Rather, it is summarized to a variety of geographic levels.

A powerful use of small area statistical analysis is to link a customer’s address to the census geography where that customer lives (geocoding). Once the census geography of the customer is known, the demographic characteristics of that census geography are attached to the customer. The assumption is made that people tend to live in areas with other people similar to themselves. This process provides a way to “estimate” the demographic characteristics of the customer. From this process, measures and conclusions about customers can be calculated.

The type of census data collected varies from country to country. Some countries, for example, include questions about income while others do not. Moreover, there is always a debate as to how many socio-economic indicators should be included in the census questionnaire, which has bearing on the length of enumeration and the length and cost of data processing. Developed countries also tend to focus more on detailed consumer statistics, such as house rent, car ownership, and access
to internet while developing countries focus more on developmental issues, such as access to water, electricity, clinics and the like.

It therefore means that the geodemographic classification for each country will be different, as one can only classify that which has been collected. In some countries the level at which census data is disseminated (district level or constituency level, for example) might be too coarse to be used outright for small area classification, which means that privately collected survey data might be more useful. Moreover, most censuses are conducted once every ten years, which means by the end of the cycle the data might be outdated and more useful data such as longitudinal survey data (labour and depravity sample surveys) might be more useful. In the public sector, for example, it might be that the local municipality has recent socio-economic and demographic area based information which is more current and at a smaller scale that that of the National Census.

Small area statistical mapping is aligned to vector GIS which have their origins in cartography. Under the vector model, suburbs, places or neighbourhoods are usually represented as two-dimensional areas - polygons. Mapping geodemographic data usually requires a boundary file to be obtained or created to represent the neighbourhood and to which the attribute information is joined. Other GIS functions, such as overlay, aggregation or point-in-polygon (spatial intersection) analysis may be required.

### 14.1 Small area statistics and poverty mapping

The quest to eliminate poverty is one of the great human endeavors of our time. Never before has humankind had the economic resources and technological tools to end human deprivation. We do now.

**One essential tool is information.** Information is key to enable people to lift themselves out of poverty—a condition characterized not only by lack of income and material goods, but also by lack of opportunity. Foremost among these are the opportunity to learn and the opportunity to participate in decision-making.

Poverty maps can be used by international, national, and local decision-makers to direct investments in human development. They demonstrate the importance of the spatial dimensions of poverty. Poverty mapping pinpoints places where development lags. It can highlight the location and condition of infrastructure and natural resource assets that are critical to poverty reduction.

Armed with this knowledge, policymakers can deploy highly targeted antipoverty expenditures and interventions to reach the neediest people by the most effective and affordable means. **Specifically for statistical agencies, poverty mapping must be viewed as an essential, long-term capacity development and institutional strengthening exercise.**

#### 14.1.1 What are poverty maps?

**Poverty maps can be defined as:**

- The spatial representation and analysis of indicators of human wellbeing and poverty
- Poverty maps are spatial representations of poverty assessments.

Many developing country policymakers use poverty maps when planning public investments in education, health, sanitation, water, transport, and other sectors. Social funds or education and health ministries often use poverty maps because geographically targeted investments are thought to reach many poor citizens and to have wide ranging spillover effects in depressed areas. Decision-makers need information tools such as poverty maps to help them identify areas where development lags and where investments in infrastructure and services could have the greatest impact.

The assessment information comes from a variety of sources and can be presented at various levels (global, national and local). Indicators of income poverty (such as GDP per capita or daily subsistence levels), or of well-being (such as life expectancy, child mortality, or literacy) are most
frequently used in poverty maps, and are derived from national census data or household surveys. Sometimes various indicators are combined to give an index of poverty or human development (such as the Human Development Index, a composite of life expectancy, literacy and income).

14.1.2 The uses of poverty maps

Poverty maps also allow easy comparison of indicators of poverty or well being with data from other assessments, such as access to infrastructure or services, availability and condition of natural resources, and distribution of transport and communications facilities. Specifically;

- Poverty maps can quickly provide information on the spatial distribution of poverty that in turn proves the targeting of intervention or development projects.
- GIS based poverty analysis makes it easier to integrate poverty data from various sources.
- Geo-referenced information can free analysis from the restrictions of fixed geographical boundaries. For instance, data can be converted from administrative to ecological boundaries which are often more meaningful in a natural resources management context.
- Mapped information on the levels and distribution of poverty make the results of analysis more easily understandable to a non-specialist audience.
- This greatly assists in the targeting and implementation of development projects, and the communication of information to a wide range of stakeholders.

Poverty mapping therefore supports more effective priority creation and intervention design, by better situating development challenges within their salient geographic contexts.

14.1.3 Poverty mapping information sources

**Surveys and censuses**

Surveys typically provide very comprehensive information on a broad range of dimensions of living standards and their determinants or correlations. For instance, multi-topic household surveys typically report information on income and/or consumption, while Demographic and Health Surveys present information on anthropometric measures and other indicators of health status. On the downside, surveys only cover a relatively small subset of households or individuals.

Censuses provide information on all individuals and households in a country, thereby allowing for the finest geographic disaggregation. Census data can be compiled for small administrative areas, for communities, villages, and towns. On the downside, censuses are typically not carried out very frequently and they only collect information on a limited set of indicators (in particular, income or consumption are typically not available).

These two sources can also sometimes be combined to build on their respective strengths and obtain detailed information that is representative at a very low level of disaggregation. A poverty map is most useful if it can be constructed at a fine level of geographical disaggregation.

**Administrative data**

Another critical source of information to analyze the link between welfare indicators, their determinants and interventions is administrative data (e.g. information on schools, health facilities, markets, roads, and the administrative hierarchy of the country etc.). For instance, information on the transport network and its quality can be used to estimate the distance or travel time that communities face to reach essential services or to access inputs or outputs markets.

Other data sources can also be central to the measurement of poverty, its understanding and policy design. These include information on rainfall and agro-climatic conditions which can be used to indicate communities’ susceptibility to food shortages. Several major initiatives have developed monitoring systems to assess food security and coordinate drought relief operations.
14.1.4 The need for GIS in poverty mapping

Consider the following figure:

**Figure 12: The need for GIS in Poverty Mapping**

We have established that GIS is a technology that enables us to geographically reference space within which natural and human induced processes and structures occur. Therefore, we can geographically overlay many of the natural and man made elements and indicators which influences poverty and analyze possible correlations and causality within a spatial context, which in turn allows us to integrate specific poverty related data within the spatial context. This proves invaluable when making planning, policy and implementation decisions regarding poverty relief programs and interventions. Consider the following figure;

**Figure 13: Integrating Geo-referenced Databases**

In order to present disaggregated information on maps, one needs to have some kind of geographic location coordinate for each observation. GIS permits the analysis of spatial association between different dimensions. In particular, it permits the simultaneous analysis of variables which are observed at different levels.

For instance, poverty status might be observed at the district level while climate is recorded at the level of agro-climatic zones. Or some infrastructure might serve broad areas (hospitals, major roads) while others serve smaller zones (primary school or health post). The GIS allow the
simultaneous analysis of information from heterogeneous sources, as long as they have geographic location coordinates. For instance, for each village, a GIS can generate the distance to the nearest market town, the average rainfall within a 20 kilometer radius, demographic indicators, and village-level estimates of income poverty.

The key to the use of poverty mapping is to present information that is sufficiently disaggregated to capture heterogeneity. Small area estimation maps combine the depth of information in a survey (information on consumption and/or income) with the complete spatial coverage available from a census (without detailed information on welfare).

### 14.1.5 Issues to consider for successful poverty mapping initiatives

The following issues should be considered:

- Close linkage of the map production aspect with user demand promotes the use and impact of poverty maps.
- The use of a credible, transparent method is an important factor in the eventual acceptance and impact of a poverty map.
- Proper access to data.
- Producing poverty maps at high levels of resolution using up-to-date information also promotes extensive use and impacts.
- Collaborative interagency approaches supported by senior officials greatly enhances map production and use.
- Active, strategic dissemination can have a powerful impact on awareness and use of poverty maps.
- Poverty mapping has a positive effect on institutional strengthening in developing countries. In addition to the expected rise in the technical skills of participating researchers, case studies contained numerous reports of increased institutional credibility and standing in policymaking circles in the wake of a poverty mapping exercise.

There is as yet no standard methodology for producing high resolution poverty maps. Various methods have been used and refinements of techniques continue to be developed. Each method has its own particular strengths and weaknesses. Data needs differ depending on the analytical methods chosen, and various methods have different implications for the timeframe and costs involved in conducting the analysis. Moreover, some methods require a higher level of statistical and econometric expertise than do others.

The choice of methods and data sources for poverty mapping should be determined according to the purpose for which the resulting map will be used, which often dictates the appropriate level of precision and resolution. In developing countries, it is also important to take into account the prevailing level of technical and human capacity development.

### 14.1.6 Generic steps for Poverty Mapping Implementation

8 Generic steps have been identified. These steps highlight key decision points faced by researchers and map producers. Not every poverty-mapping exercise will include all eight steps or follow them sequentially.

**Define purpose and expected use of mapping**

In an ideal world, all poverty mapping would start here. Maps may be needed to show that certain regions are disadvantaged, to rapidly assess options for food emergency interventions, to target public investment to areas of greatest need, or to investigate specific causes of poverty. The purpose and intended use of poverty maps determine the scope and the required precision of the mapping exercise and should shape methodological choices described below.

**Select measure(s) of poverty and human wellbeing**

Choosing an indicator or indicators of poverty is a pivotal step in map production. Poverty is a multidimensional phenomenon, including economic, social, and other aspects of human wellbeing. The selected indicator may be a monetary or non-monetary variable—for example, the proportion of
households below a certain income level or the proportion of households without access to sanitation. Researchers sometimes distinguish between status and outcome variables—e.g., access to safe drinking water (status) versus incidence of waterborne diseases (outcome) — but because indicators of poverty are interdependent, the distinction between status and outcome measures is not always clear.

A poverty indicator may measure a single important dimension of human wellbeing, such as household expenditure compared to a minimum necessary level or poverty line. Alternatively, the indicator may be multidimensional, for instance, a composite index that depicts deficits in basic human needs, such as education, health care, and sanitation. Each type of poverty indicator has its own strengths and weaknesses, and the choice of indicator will certainly influence who is classified as “poor.”

**Select input data**

Data used to construct a poverty map typically are drawn from population or agricultural censuses, household surveys, or spatial (GIS) databases in which values are fixed to specific locations on a grid. Increasingly, poverty mapping relies on data from many sources. Data used in poverty mapping may vary in coverage, collection method, and level of resolution, all of which may have Methodological implications.

Data coverage may be comprehensive—such as a national census or a detailed map covering the entire geographic area under consideration—or it may be partial, for example, a survey of household expenditures covering a representative sample of the population. Researchers may face choices with respect to data collection methods, including qualitative versus quantitative approaches, or top-down versus participatory methods. The level of resolution of input data used in poverty mapping may be high (e.g., household level) or relatively coarse (e.g., averages for Easy or administrative units).

**Select method of estimating or calculating poverty indicator**

Researchers may choose to estimate a single variable, such as per capita household expenditures compared to a specific standard of living (i.e., poverty line). Alternatively, they could use a composite index, which may be calculated by simple aggregation (i.e., equal weighting) of a few variables or by multivariate analysis, such as principal components or factor analysis.

**Select a method to calculate, estimate, or display poverty indicator for geographic area**

Depending on the chosen poverty indicator, input data, and method of estimation/calculation, researchers will have different options for calculating or estimating the poverty indicator across a geographic area. For instance, if map producers are using census-level data made available at the household level, then simple aggregation of the data for the selected geographic unit may suffice.

However, researchers often need techniques that are more sophisticated. Poverty maps often combine census data (featuring complete country coverage) with household survey data (encompassing a representative sample of the selected population). This is accomplished by means of advanced statistical methods based on econometric techniques, sometimes referred to as small area estimation.

Combining data from these two sources enables a poverty mapping study to benefit from both the complete spatial coverage of the census and from a relevant poverty indicator in the household survey. Such statistical techniques help overcome the survey’s insufficient sample size, which could not be aggregated to small administrative units, and the census’ lack of an appropriate poverty measure.

**Decide on number of units for final map (resolution) to present poverty data**

For many poverty-mapping methods, this step is often combined with the previous one. In the case of small area estimation relying on household-unit data, researchers cannot map an individual household; they must aggregate household-level data to larger units to reduce the statistical error in their prediction model.
Sensitivity tests conducted by researchers suggest that a minimum of 5,000 households is needed to reduce statistical error to an acceptable level. The number of households required may be significantly higher in other cases, especially if the statistical model is not as strong in its predictive power.

**Produce and distribute maps**
Mapping software is used to produce a spatial representation of the geographic distribution of calculated/estimated poverty indicators. Maps and supporting analyses are distributed to the targeted decision-makers. Increasingly, map producers are supplementing hardcopy maps with other products, such as interactive decision-support tools and/or datasets on compact discs (CDs), aimed at various audiences (technical, general, or mixed).

**Monitor usage and feedback**
Poverty maps are used for various purposes, ranging from identifying and understanding the causes of poverty, to assisting in program development and policy formulation, to guiding allocation of anti-poverty investments and expenditures. Map producers should monitor and evaluate the various ways in which their maps are being used by decision-makers and/or researchers, and users should provide feedback on the impact and limitations of poverty maps to map developers.

### 14.1.7 Limitations of poverty mapping
Although poverty mapping can be a powerful tool for analyzing poverty and communicating the results to technical and non-technical audiences, experts hasten to point out the limitations of these techniques.

Poverty maps are not a panacea for understanding or solving poverty problems; they are only one tool among many for investigating the complex phenomenon of poverty. They should be used in conjunction with other information and analysis that provide context and ground truthing within communities. Poverty maps can be used to explore the spatial aspects of various components of human poverty. However, indirect estimation of poverty, as opposed to direct observation in the field, introduces some degree of uncertainty. Careful additional analyses are needed before conclusions are drawn on any meaningful correlation, much less causal relationships, between these variables.

In addition, it is important that poverty mapping is always seen in the overall context of a country’s decision-making processes. Technical tools like poverty maps run the risk of being abandoned once initial donor support has waned. To ensure a path of sustained use and support for poverty maps, fundamental questions need to be addressed, such as how to retain skilled analysts in the public sector, overcome limited or lacking demand and funding from policymakers, and convince decision-makers that continued investment in poverty maps is worthwhile in an environment that does not follow a purely technical approach to decision-making.

### 14.1.8 Methodological issues

**The importance of a credible, transparent method**
Case study results indicate that small area estimation provides a consistent and replicable method of poverty mapping. Acceptance and use of such maps has been significantly aided by the institutional clout of the World Bank, along with the commitment of the Bank's Research Group to providing technical assistance, refining statistical tools, and building developing-country poverty mapping capacity.

Despite the methodological rigor of the small area estimation technique, not all intended uses of poverty maps require detailed household survey data and sophisticated econometric analyses. In some countries, highly complex analytical methods are impractical and too expensive. It is important to recognize that debating the use of many different, competing methodologies may discourage the use of poverty maps. In Peru, for example, discussion of poverty mapping is increasingly bogged down in debates about which mapping technique to use rather which programs would benefit most from the application of poverty maps. In some instances, methodological
controversies are thought to have provided a means of sidestepping the development and use of poverty maps altogether.

**Overcoming problems with data availability and access**

Map producers typically face two major challenges: do the necessary data exist and can access to them be secured at a reasonable cost in an acceptable timeframe? These issues are most prominent for researchers using the small-area estimation methodology at a household-unit level, which requires comprehensive data from the national census as well as household surveys. Many governments are extremely reluctant to release highly disaggregated census data to independent parties, citing legitimate confidentiality concerns.

One means of addressing such constraints has been the forging of collaborative arrangements through which researchers at government institutions first manipulate the raw data directly and then work jointly with analysts from other institutions on model development and distribution of results. Similarly, the direct involvement of the national statistics departments in poverty map production in Guatemala, Madagascar, Malawi, Nicaragua, Panama, and South Africa is thought to have helped map developers obtain full access to census data.

Another approach has been to make use of coarser data that governments are willing to release.

**Issues of data quality**

Case study results point to two general data-quality problems: quality of input data coming from household surveys or censuses, and quality of geographic attribution of input data. For example, in Malawi, researchers had to painstakingly remove half of the household sample, due to unreliable consumption information in the 1997 household survey.

South African map producers faced geo-coding issues involving the time-consuming task of reconciling incompatible datasets from the apartheid and post-apartheid eras. Burkina Faso’s reliance on GIS for additional information required the investment of significant resources to check spatial data for errors, reconcile boundary information, and integrate various data in a “seamless” GIS.

Commonly agreed-upon minimum data standards (e.g., phrasing of questions on household characteristics or locating boundaries of enumeration areas) for household surveys, national censuses, and geospatial information like administrative boundaries are one way to address these issues. High-quality, documented baseline data reduces costs and increases accuracy of poverty map results.

**Significance of up-to-date information and high-resolution maps**

The use and impact of poverty maps can be considerably reduced if the map is based on data (typically census) that are several years old. Recent data are especially important for applications requiring a great deal of accuracy, such as geographic targeting. Outdated information can also lead to serious misunderstandings; a Brazilian municipality attempted to sue UNDP over its HDI ranking, not realizing that the indicator was based on 1991 census data rather than more recent information.

Whether data are “current” is not an absolute, but rather depends on the context in which the information will be used. Such factors as the speed of change in a country or the sensitivity of poverty measure used greatly influence data requirements.

Packaging data at various spatial and temporal scales also appears to encourage broad usage.

**Need for comparative analyses and more sophisticated spatial analyses**

With the proliferation of poverty maps, it will become even more imperative for mapping experts to encourage systematic, comparative analyses that highlight the strengths and weaknesses of different methods. There is also a need for more sophisticated spatial analysis.
Dimensions that could be explored by such analyses include basic environmental services (solid waste collection and disposal, drinking water supply, wastewater treatment and management, mass transit), identification and prevention of environmental health risks, rights of access to natural resource use and development, siting and management of protected areas, and fiscal expenditures and subsidies for environmental goods and services at municipal, county, or state levels.

### 14.1.9 Institutional issues related to poverty mapping

#### Importance of senior-level support

The support of senior-level officials is crucial to the success of poverty map production and use, especially when there are issues of data access and testing of new statistical methods. One example drawn from our case studies is that of South Africa, where the leadership of Statistics SA supported the idea of conducting a formal evaluation of data for income and poverty estimation. This evaluation demonstrated how census data underestimated income, especially for rural households with significant earnings in non-cash income (Alderman et al. 2000). Initial senior-level endorsement was subsequently extended to support for production of a poverty map using small area estimation to combine census income data with household expenditure data from a national survey. The involvement of senior management also helped researchers gain full access to census and survey data.

#### Need for collaborative approaches

Collaborative interagency approaches greatly facilitate map production and use.

#### Importance of an active dissemination strategy

Extensive and strategic dissemination of poverty maps is crucial if poverty maps are to be widely used to influence decision-making in government and among other users in society. Distribution of data products, such as CDs containing data and user-friendly mapping and statistical software, has also been an effective way of promoting familiarity with and adoption of poverty maps by a broad spectrum of users.

#### Institutional strengthening

Among the most significant impacts of poverty mapping have been the effects—both planned and unanticipated—on institutional strengthening in developing countries.

In addition to the expected rise in the technical skills of participating researchers, case studies contained several reports of increased institutional credibility and standing in senior policymaking circles in the wake of a poverty mapping exercise. Experiences with poverty mapping has led to greater institutional awareness of the need for better coordination of efforts, especially with respect to survey and census design, data issues (collection, processing, analysis, data standards, and data sharing), and project development.

Many of the institutions participating in poverty mapping had poor track records concerning interagency collaboration. In particular, poor coordination has led to the development of incompatible datasets. Long-term data standards development requires better coordination among institutions responsible for baseline data, such as surveys and statistics departments and mapping agencies.

#### Strategy for building a cadre of skilled poverty map producers and analysts

Such a strategy would include developing training programs and guidelines for using poverty maps and incorporating poverty map techniques and results in university curricula.

#### Tools and skills for sustainability

Equipment and tools used to develop and disseminate poverty maps need to account for local conditions. Another potentially important tool is the customized poverty mapping software developed by the World Bank. This software will not only facilitate complex econometric modeling, but will also obviate the need for expensive purchases and high maintenance fees associated with commercial statistical software (e.g., SAS).
Data comparisons are greatly facilitated by the use of GIS technology allowing for data overlays and comparisons. With anticipated rising demand for comparing poverty map results with other spatial data (e.g., in environmental applications), adequate support for developing and maintaining GIS skills and technologies in developing countries will become even more important. Likewise, support will be needed for the capacity to use GIS to generate additional variables for econometric models.

**15. Thematic mapping**

Since poverty mapping is essentially the thematic mapping of poverty indicators, it will be useful to conclude this guide with some thoughts on thematic mapping, specifically since the ease of use of creating thematic maps on GIS has led to ignorant users misleading themselves and other users through the misuse of thematic maps.

A thematic map displays the spatial pattern of a theme or series of attributes. In contrast to reference maps which show many geographic features (forests, roads, political boundaries), thematic maps emphasize spatial variation of one or a small number of geographic distributions. These distributions may be physical phenomena such as climate or human characteristics such as population density and health issues. These types of maps are sometimes referred to as graphic essays that portray spatial variations and interrelationships of geographical distributions. Location, of course, is also important to provide a reference base of where selected phenomena are occurring. While general reference maps show where something is in space, thematic maps tell a story about that place.

**15.1 Why do we use thematic maps?**

Thematic maps serve three primary purposes. First, they provide specific information about particular locations. Second, they provide general information about spatial patterns. Third, they can be used to compare patterns on two or more maps. Common examples are maps of demographic data such as population density. When designing a thematic map, cartographers must balance a number of factors in order to effectively represent the data. Besides spatial accuracy, and aesthetics, quirks of human visual perception and the presentation format must be taken into account.

Of equal importance is audience. Who will “read” the thematic map and for what purpose helps define how it should be designed. A political scientist might prefer having information mapped within clearly delineated county boundaries (choropleth maps). A state biologist could certainly benefit from county boundaries being on a map, but nature seldom falls into such smooth, man-made delineations. In which case, a dasymetric map charts the desired information underneath a transparent county boundary map for easy location referencing.

In constructing any type of thematic map (or any map for that matter) it is understood that location is a key feature. After selecting the physical area to examine, the next step is collecting data sets.

Data dealing with one subject is called univariate, which examines occurrences of a single type of event. The distribution of population, cancer rates, and rainfall are all examples of univariate data.

Bivariate mapping shows the distribution of two sets of data to explore possibilities of correlations. For example, we can examine population density in relation to textile manufacturing. Other examples could be cancer rates and population density, or rainfall and elevation.

More than two sets of data leads to multivariate mapping. Taking three or more data sets and displaying the result on a map helps determine possible correlations between different phenomena. Map makers must be careful in designing thematic maps that display too much information or suggest phenomenon have a correlation when in fact they do not.
15.1.1 Types of thematic maps

**Choropleth maps**
The most commonly used method of thematic mapping. Choropleth maps are particularly suited for charting phenomena that are evenly distributed within each set area.

**Proportional symbol maps**
Also known as graduated symbols, these maps represent data associated with point locations (i.e., cities or villages). The data is displayed with proportionally sized symbols to graphically represent a realistic difference in occurrence.

**Isarithmic maps**
These maps, also known as contour maps, depict smooth continuous phenomena such as precipitation. They are also well-suited to displaying three-dimensional values such as elevation i.e; on topographic maps.

**Dot maps**
A map using dots to show the presence of a feature or occurrence and display a spatial pattern. Note, though, that a dot is not required to represent a single unit and may indicate any number of entities; 1000 persons, 50 cattle, 100 voters, etc.

Some Namibian examples of thematic maps: (All examples sourced from the Namibian Ministry of Environment and Tourism website; http://www.met.gov.na )

**Figure 14:** Choropleth map with buffer analysis for the North Central region in Namibia describing population access to health facilities.
Figure 15: Dot map for the North Central region in Namibia describing health facility locations

Figure 16: Graduated symbol map for the North Central region in Namibia describing the number of people and cattle per household
15.1.2 The dangers of thematic mapping
The mapping functionality of GIS greatly enhances the dissemination possibilities of geodemographic classification, specifically thematic mapping. Consequently, it is prudent to explain some of the pitfalls of thematic mapping.

Thematic mapping, especially choropleth mapping, can sometimes ‘lie’, revealing patterns that might not exist in real life. The visual characteristics of a map can be extremely persuasive, so much so that they can lead to a sometimes false sense of security in the objectiveness, neutrality and truthfulness of the map and its designer. In practice, maps, like statistics, can be manipulated to paint very particular pictures – a property that is useful but open to abuse. For example, changing the ranges or classes for average household expenditure, even though using the same dataset can paint a very different picture of household expenditure within a specific area. Similarly, there are many different ways one could classify, symbolise or map income data. Each can give a different impression of the distribution of incomes across a country. Monmonier (1999) comments “a single map is but one of an indefinitely large number of maps that might be produced for the same situation or from the same data”. He warns that because of personal computers, GIS and electronic publishing, map users can now easily lie to themselves – and be unaware of it!
There are three main limitations to the use of thematic GIS maps for census data small area classification:

- It is implied that populations are distributed uniformly within each area and that all zones are fully occupied by residents, therefore, the implication is that there are no non-residential spaces within each area.
- It is also implied that populations are social-economically uniform within each area division with all residents within a neighbourhood assigned the same map label, whether it be average income, expenditure, ethnicity, newspaper preference etc.
- It is implied that all change in geodemographic condition occurs only at the borders between zones and thus never within a zone or neighbourhood.

The limitations of a choropleth map as a tool to visualise neighbourhood data helps explain why point and dot density maps are often used to better explain specific geodemographic conditions.

### 16. Determining GI T ethical codes of conduct

It is important that GIS practitioners remember that codes of conduct, statements of ethical principles and codes of practice are found in many other information handling professions, many with long histories, not just in law or medicine, engineering or government service. Information is information, after all, and geographic information, misused or abused, can cause similar or worse harm to persons or society as can, for example, unethical journalism.

Neither data nor information is inherently 'unethical'. Rather, the uses of data and information can be potentially unethical, and this is where computer software and applications, dissemination networks and various data policies enter the ethical debate. It is therefore necessary that information producers, users and governments to promote ‘ethical behaviour and respect for community standards and values in respect of the activities and content of information producers, users and service providers.

With regard to spatial information, i.e. information with an important location attribute, several types of ethical issues arise in relation to collection, dissemination, use and abuse of such information:

- ethical application of intellectual property law, especially for copyright, e.g. the right to reuse spatial information for one’s own purposes and how or where is intellectual property law being challenged on what can be construed as ‘ethical’ grounds
- ethics regarding rights of citizens for access to information gathered at public expense
- moral obligations for governments to create frameworks and infrastructure, both legal and physical, permitting universal access to information, especially that in the public domain or for public sector information not already in the public domain
- protecting personal privacy by preventing unethical use of publicly available information, especially spatial information that can be used to track individuals.

The basic ‘right’ versus ‘wrong’ or ‘good’ versus ‘bad’ interpretations of our actions, which constitute ‘acting ethically’ within the accepted norms of our particular society, may not have changed intrinsically in the past decades. Yet the situations and environments commercial, social, personal - in which such decisions are made, have altered dramatically. As members of an active GIS/spatial information community, both as creators of data, technology and applications, and as users of the same, we ignore considering the ethical issues at our peril, both in the near term and long term.

Given the generally promoted view that GIS is everywhere in information space and underpins so much of our society, economy and government, we can also argue that all ethical issues impinge on the spatial information community. Indeed, if GIS is really that important to society, then the GIS community should have the strongest ethical statements and practices of any sector in the information society, globally.
There are general moral and professional code attributes which can be considered as pertinent in the search for a code of ethics for GIS.

**General moral code issues:**
- Contribute to society and human well-being.
- Avoid harm to others (health, safety, welfare, etc.)
- Be honest and trustworthy, maintain high personal integrity
- Be fair and take action **not** to discriminate
- Honour property rights including all intellectual property rights
- Respect privacy of others
- Honour confidentiality of information
- Be actively aware of consequences of your actions
- Speak out about public issues within your competency area

**Professional code issues:**
- Maintain and update professional competence
- Do not misrepresent professional qualifications or expertise
- Respect the work of colleagues and other professionals
- Uphold the reputation of the profession in your own conduct
- Treat all information with respect, e.g. accuracy, quality, etc.
- Maintain objectivity, independence and openness at all times
- Be aware of environmental issues and concerns

### 16.1 The ten commandments of GIS ethics

By taking all of the above to heart with regard to data and privacy protection and moral and professional ethical behaviour within the information community, we might take a stab at creating a “Ten Commandments” for GIS practitioners”. The following is a suggestion of how these commandments might look like.

**The Ten Commandments of GIS Ethics**

1. You will not use a GIS technology to harm other people.
2. You will not interfere with other people's computer work.
3. You will not snoop around in other people's computer files, nor misuse distributed access to geospatial databases or location-based technology in such a way as to infringe of the rights of your fellow humans.
4. You will not use a computer to steal or use geospatial data or information or GIS applications to aid in theft, whether directly or indirectly.
5. You will not use a GIS database or GIS technology to bear false witness.
6. You will not copy or use proprietary GIS, visualisation, GI database management or remote sensing analysis software for which you have not paid.
7. You will not use other people's computer resources or geospatial databases without authorization or proper compensation, unless such resources are openly advertised as being freely available and in the public domain.
8. You will not appropriate other people's intellectual output.
9. You **will** think about the social consequences of the GIS or other spatial analysis program(s) you are writing or the GI application system you are designing, including its potential for misuse or abuse by those less principled than yourself.
10. You will always use a computer, information network and spatial applications in ways that insure consideration and respect for your fellow humans.

### 16.2 Conclusion

Ultimately, enforcing issues of data protection and privacy as well as ethical behavior begins with the individual. By our individual and collective example it is possible to inspire confidence in the GIS industry regarding ethical issues.
17. References


17.1 Internet references and links

17.1.1 Geographic Information Systems
http://gislounge.com
http://lagic.lsu.edu/gisprimer/strategies.asp
http://lagic.lsu.edu/gisprimer/
http://www.colorado.edu/geography/
http://www.gis.com/
http://www.gisdepot.com
http://www.innovativegis.com
http://www.met.gov.na
http://www.physicalgeography.net/fundamentals/2f.html

17.1.2 Remote Sensing
http://ccrs.nrcan.gc.ca/resource/tutor/fundam/index_e.php
http://gislounge.com/remote-sensing-principles/
http://rst.gsfc.nasa.gov/start.html
http://rst.gsfc.nasa.gov/Front/overview.html
17.1.3 **Poverty Mapping**  
http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY  
http://www.ciesin.columbia.edu/povmap/  
http://www.povertymap.net

17.1.4 **Global Positioning Systems**  
http://gislounge.com/gps-resources/  
http://www.eaa1000.av.org/technicl/gps/gps.htm  
http://www.colorado.edu/geography/gcraft/notes/coordsys/coordsys_f.html  
http://www.colorado.edu/geography/gcraft/notes/datum/datum_f.html  
http://www.gisdevelopment.net/tutorials/tuman004.htm

17.1.5 **Spatial Data Infrastructure**  
http://geoinfo.uneca.org/sdiafrica/default1.htm  
http://www.gsdi.org

18. **Suggested reading**


*Note, this very useful handbook can be downloaded online from http://unstats.un.org/unsd/Demographic/sources/census/2010_PHC/Publications/Series_F103_more.htm*