UNITED NATIONS
ECONOMIC COMMISSION FOR AFRICA

Geographic Information Management

the nature of resource and environmental information, data sources and their organization

A FRAMEWORK FOR THE ESTABLISHMENT AND UTILIZATION OF NATIONAL GEOGRAPHIC INFORMATION SYSTEMS

December 1993 - *Revision June 1994*

Provisional document

12289

ECA/NRD/CRSU/93-7
Distr: LIMITED
Original: English

Geographic Information Management

the nature of resource and environmental information, data sources and their organization

A FRAMEWORK FOR THE ESTABLISHMENT AND UTILIZATION OF NATIONAL GEOGRAPHIC INFORMATION SYSTEMS

UNITED NATIONS
ECONOMIC COMMISSION FOR AFRICA
December 1993 - Revision June 1994

Provisional document

CONTENT

INTRODUCTION

Scope

This document attempts two things: (a) to provide those who are responsible of the management of the land and the protection of the environment (usually non-specialists on computer spatial data manipulation) with a simplified vista on the capabilities and potentials of Geographic Information Systems (GIS) and to familiarize them with the means the computer uses to represent and manipulate the real world, and (b) to provide a set of general conditions and rules, exclusive of a particular sector or user's view, on how to proceed for the establishment of a national GIS composed by a network of databases and individual institutional GIS.

Although implementing GIS entails, in practice, highly specialized computer hardware and software, and expertise, this document avoids technicalities normally of interest to the specialist but not to the decision-maker who must take decisions concerning the implementation and management of a GIS.

The first section outlines a number of basic conditions at government and institutional level that are thought critical to the establishment of a national spatial information infrastructure. It lists the major steps towards the establishment of an institutional GIS, within the scope of a nationwide information system. Appendix 1 supplements this section by giving an outline of how geographic data is stored, structured and handled by the system.

The second section focuses on general spatial information management and includes issues such as the nature and importance of geographic information, and its accessibility and manipulation.

Appendix 2: Chapter 40 of Agenda 21: "Information for decision-making", should be a helpful addition.

Justification for these guidelines

GIS is becoming the new creed for those who deal, in one way or another, with land resource development, environment protection and other activities related to land occurrences. This phenomena is, to a great extent, a result of the power of these systems to cope with increasing amounts of varying spatial information, and its ability and flexibility to process and display data according to the different purposes and planning scenarios. It can also be misleading to those who want a GIS to solve all their problems. Some planners may feel that GIS must be introduced at any cost and without delay in their operations, without fully understanding all the implications of establishing a GIS, much less the kind of GIS that is suitable to them. Surrounded by an overwhelming amount of GIS-promoted software available, ranging from the complex and expensive to the simple and cheap, many organizations are tempted to purchase one particular software and begin to collecting digital data, often on ad-hoc basis, in the belief that they are establishing the GIS they need. Within a region, a country, a municipality or even within an institution, a proliferation of ill-suited GISs will most likely result in nothing more than a lamentable waste of resources.

On the other hand, many may be reluctant to enter GIS for fear of getting involved in something that is too technical, too costly, implies too much work and is not worth it.

When these two attitudes --the positive-but-naive and the negative-- persist, the high front-end costs coupled with confusion and frustration will by far outweigh any gains, and what initially were dreams and expectations turn into nightmares. The results, rather than beneficial, will prove damaging and may delay or destroy any further attempts to use the technology.

In order to avert these situations, GIS developers, producers, scholars and users, usually from the private sector, have generated various models providing rules and guidelines on how to construct a GIS. However, they are part of the scientific literature and are "promulgated" in specialized publications, dispersed among or entwined with technical articles of all sort, and do not the reach the normal planner. This document compiles, reconciles and integrates many schemes proposed by renowned scientist and specialists as well as published technical material.

SECTION ONE

FRAMEWORK OF BASIC CONDITIONS AND GUIDELINES FOR THE ESTABLISHMENT OF A NATIONAL RESOURCE AND ENVIRONMENTAL GEOGRAPHIC INFORMATION SYSTEM.

There is no general model that suits every country for the establishment of a national geographic information system, as there is no general set of procedures or steps to implement such a system. These is a large range of conceptual models and implementing procedures among which to choose to better satisfy the needs and peculiarities of each individual country. However, a number of basic enabling conditions and guidelines are identifiable as being crucial to the success of building up a useful geographic information infrastructure in any country. These exist at both national level and institutional level. An attempt is made here to provide those that are thought of major relevance.

At national level.

Understanding of modern information technologies and their potentials by those who have the keys to development.

It is essential a full understanding of the potential of modern information technologies, in particular geomatics, and of the ***need*** to use then in the inventorying, assessment and management of natural resources and the environment, and in setting up and steering sustainable development paths that will solve socio-economic problems of African countries, all urgent and overwhelming.

It must be realized that the possession of a reliable and efficient spatial information infrastructure is as important as other national infrastructures upon which the governments concentrate their efforts, such as health, education, transport, energy, etc. Making a parallel, what expectations could have an international airline without possessing a reservation database server or that is not linked to the global reservation network?. How could a modern bank presently operate without possessing a client/account information system networked with its branches and subsidiaries?.

Such an understanding must be done at the highest levels among planners, decision and policy makers, in the government (heads of state, ministers, etc.), in the political scenario (chiefs of political parties, senators, in some cases governors and majors, local community leaders, etc.), in the production and industry sectors, etc.

The appreciation for base-line spatial data and information.

Base-line data and information, in particular topographic and other land information maps, is normally the least appreciated tool. National mapping programmes, which in the past were at the forefront of development, are nowadays given little preeminence in national development programmes. As a consequence, the cartographic coverage of the majority of African countries, with some notable exceptions, is deficient and outmoded, in some areas nonexistent or the mapping is only planimetric, and in others little attention was given to geodetic control. Although valuable efforts have been made by some leading countries, the situation for many of them remains unchanged and may worsen as time elapses. The governments must grant to this activity a high priority, allocating the resources that it needs to complete, improve and maintain their national cartographic coverage. Otherwise, their efforts to build any national geoinformation infrastructure will be futile, and whatever is done will lead only to a lamentable waste of resources and an anthology of frustrations. Simply, a building without foundations cannot be erected.

The will and commitment to change.

Understanding potentials and realizing needs is not enough. It is necessary to have the commitment to change, by conviction. Only then we can expect to practically materialize this change.

This commitment has to be a long-term commitment. It has to last at least until the development of the system is well in place, and its benefits appear. Then the users (among which the government will be the major one) will be the driving force to maintain and continuously expand the system that may become self-sustainable.

***The materialization of the will to change: A Plan of Action** (or the implementation of Agenda 21).*

A plan of action, issued at the highest level, that would lead to a national programme for the creation and management of a geographic information system, would be the first step to materialize the will to change. Such a Plan of Action would be a great leap in the implementation of chapter 40 of Agenda 21. A copy of this chapter appears as Appendix 2. Such a plan of action can (should) initially be simple and should be inscribed within a broader Plan (Agenda 21). It would, at further stages and as results of its own implementation, be amended and improved, going in deeper detail, changing strategies and procedures, adding new components, fixing new responsibilities, etc.

This initial plan of action would:

(i) provide principles on flow, access and supply of geographic information.

(ii) define the goals at short, medium and long-term.

(iii) identify the sectors that would be addressed by the system: principal, secondary, tertiary.

(iv) identify the actors including those of the private sector: national and sub-national agencies, research and educational institutions, agricultural and livestock associations, scientific and professional associations, etc.

(v) create a national steering body, constituted by a core set of actors, including the private sector, with the task of defining the nature, characteristics of the system as the methodology, procedures and time-frames of implementation. When necessary, any other members can be coopted by the steering body. This body would be accountable to the highest levels of government, to which it would report regularly at specific intervals.

(vi) Define provisional budgets and budget lines.

A National steering body on geomatics:

It can be an ad-hoc Committee or a permanent Council. Sometimes, a leading Institution may be selected to coordinate the work of the Committee or Council. Its work would mainly be conducted by working groups, whose initial functions would be:

(i) Identify the spatial datasets that are required for each sector or aspect of development. This exercise will necessarily comprise the identification of data users, data needs and data sources, required data accuracy, as well as the selection and prioritization of scales.

In general, the following datasets have been accepted as an appropriate basic set for national planning.

Fundamental data sets:	(a) Topographic map:	geodetic control elevation (DTM) water areas, drainage and shorelines terrestrial communication lines cultural elements general vegetation: forests, rangeland, cultivated areas and pastures. principal administrative boundaries
Other data sets	(b) baseline satellite image (c) cadastre, land tenure, detailed administrative boundaries. (d) geology/mining (e) land use/land cover (f) soils and land vocation (g) energy (h) restrictive sites (public lands, special tenures) (i) climate (j) fauna species distribution	

It is very important to bear in mind that the topographic map is the basis upon which the other datasets are geometrically fit. *Hence, it must have the highest priority.*

(ii) Compile a classified catalogue of all existing data and information that is deemed relevant, assessing the attributes of each piece of information. The committee will also be responsible for assuring that such metadatabase is properly maintained.

(iii) Identify datagaps, and provide concrete recommendations for new data collection.

(iv) Select the model of the national geographic information system (central database, distributed database, combined central and distributed) and, accordingly, how the different datasets are organized.

(v) Compile an inventory of all existing hardware and software, within participating parties, qualifying its relevance to the project: operating platforms, capacity, compatibility, risk of out-datedness, flexibility, facility of maintenance, possibility to expand, etc.

(vi) Analyze the current mandates of the different participating institutions. The Committee would agree on new limits to the mandates for the purposes of the establishment of the national geographic information system, in order to avoid the normal conflicts due to overlapping and duplication, or to fill gaps, that inevitably will always be found. Precise responsibilities and roles of each institution would be clearly defined within the system, so that every actor knows what to do and what to expect without ambiguities.

(vii) In the light of the above, the Committee would propose or set up the rules, as applicable, concerning legal and technical aspects of mandates, proprietary rights, security and confidentiality, flow of information, access and supply, pricing of the information, data quality standards and standards for data collection, data up-dating, data conversion, integrity and integration of datasets (geometric, datatransfer, etc.)

(viii) At least a written master agreement is necessary to protect the interest of all parties to the extent possible, where the issues in (iv) and (v) above are clearly defined.

(ix) Provide advice on and revise the datamodels and datastructures proposed by each institution for the datasets of which it is responsible for producing and maintaining, verifying that they satisfy the needs of all the development sectors that will use those datasets.

(x) Revise the necessities and procurement of new equipment and software, assuring that there is full compatibility with the equipment and software of the other participants of the systems, and that appropriate existing equipment and software has maximum, but reasonable, possible utilization.

(xi) Design and coordinate the execution of pilot studies and pilot projects, where tests of small integrated systems would be carried out as a means of obtaining practical experience with the technology and its possibilities, and test system conditionalities such as networking and communication among GIS sub-systems, standards, database export-import, data integrity, data flow, etc.

(xii) Propose, set up and phase implementation plans.

At institutional level

Variety of databases and GIS within the system:

It is clear than the datamodels and GIS type and characteristics, as well as the strategies of implementation within the individual institutions will depend on each institution concerned, and will vary from one to the other. Among other factors, they will have to take into account and will be conditioned by: (a) whether it is a source producer, a user, or a combination of the two, (b) the type and nature of the data, (c) the volume and complexity of the data, (d) the analysis required on the data to satisfy the needs of the users, (d) the physical, financial and human resources available, (d) the institutional structures, (e) etc.

So far, it would be unrealistic to conceive uniform individual systems within the general national geographic information system network. Each institution will be responsible for selecting the GIS it will use, but the choice must not be isolated from the rest of the system. What is important is the conditionality that they communicate freely and that the databases are integrated together.

Building up and maintenance of the databases:

In a distributed system, with is being adopted more and more widely, the producer (owner) of the data and information is also their custodian. In such context, he defines his own datatamodel and data structure, but will assure that they satisfy the needs of the users within the national information system, and in such a way as to facilitate the up-date and expansion of the content. Data conversion, data maintenance and data up-dating will be the responsibility of the producer. He will also be responsible of making his database fit to the geometry of the fundamental topographic database. In an ideal case, common elements from the two databases would be adjusted to share the same set of primitives, assuring full integrity.

Hence, the conversion of the fundamental database should be done with priority, as the other datasets are just layers to it.

At any rate, although it implies a greater rate of investment, the conversion of data should be done with great care and within the shortest possible period. The reason of preferring a short conversion period is that the effects and benefits will not be obtained until a complete database exists for at least one application theme covering a comprehensive geographic area.

Organizational and cultural conditions:

The introduction of GIS technologies leads to changes not only in existing routines for information exchanges between and within national authorities and agencies, but also entails changes in old-time conceptions on the nature itself of spatial information, conceptions considered immutable in the minds of those who have been producing and managing that information. These changes imply organizational changes within the institutions and also challenges the cultural attitude of the staff, *at all levels*, as the new technology threatens their system of values. A certain amount of the personnel will oppose to them, either deliberately or not.

Institutions switching to GIS must be aware, from the very beginning, that these organizational and cultural problems are more intricate to solve than technical ones, and at the same time are crucial to the degree of success achieved. A great deal of time and attention has to be devoted these matters. Clear and open information on the new methods/goals of the organization to all concerned staff (operational, supervisory and managerial) should be a priority, where an internal convincing marketing of the new working procedures, new products and overall expectations would be earnestly conducted. Awareness and education through periodic seminars and workshops must be organized, encouraging and entertaining open discussions and creating a sense of solidarity and interdependence, whereby the units of the organization would feel genuinely responsible for developing and maintaining the new functions. If an important part of the organization does not accept them, then it should be removed from any role/activity related to the new system.

Training:

Qualified staff is a sinequanom in geomatics. The success or failure of individual GISes will directly depend on the availability of competent staff, who can understand the processes behind the technology. Adequate and continuous training at all levels is required: (i) *Operators* for data conversion (digitizing, editing), up-dating and display. Sometimes experienced but enthusiastic cartographers and draughtmen can be successfully (and easily) trained, as they understand, better than others, the mechanics of cartographic data representation; (ii) *Supervisors*: They must be individuals with a full understanding of the mandates of the organization and of the processes utilized to carry out those mandates, in particular concerning the generation and the purpose of the institutional data that will be stored and used by the system. These individuals must also possess a thorough understanding of the system's datamodel and datastructure, with good expertise in the manipulation of the GIS dataconversion modules, and must understand how the database is to be queried; (iii) *GIS applications staff*. These are professionals (engineers, surveyors, foresters, soil scientists, environmentalists, urban/rural planners, etc.) specialized in spatial data analysis with GIS, with an intense and dedicated training in the use and characteristics of the particular GIS package being implemented by the organization. Data analysis and manipulation, modelling, planning scenarios, etc., would be the responsibility of this group. It would also be responsible in assuring that the GIS meets the goals that have been set up and that the (growing) users' needs are met; (iv) *Institution managers*: A good understanding of the components and of hardware and software, database models and structure, functionalities of the GIS and how the system can be queried and analyzed, is essential, and can be attained through the organization of seminars and workshops; (v) *Decision-makers*: who are the real end-users of the system. Training would focus on knowing the content and structure of the database and in the use of the GIS modules to manipulate and analyze the database. As a rule, they are assisted in their task by the applications staff.

The training scheme will depend on the organization's structure and its resources, on the type and complexity of the data to be converted and used, as well as in the schedule and time-frame for the system's implementation. Some of the staff may be identified within the organization and some will have to be recruited. Some training can be done on-the-job (operators, supervisors) by the system's vendor during the initial installation phase and pilot projects, or, if available, by contracting with some experienced group in the country. If the data conversion phase is done by an external firm, provisions can be made whereby this process is utilized to provide the required on-the-job training to operators. A number of professionals will necessarily have to be sent for formal training and education to universities and specialized training centres, within or outside the country. The Regional Centre for Training in Aerospace Surveys (RECTAS) at Ile-Ife, Nigeria, established under the auspices of the Economic Commission for Africa, constitutes an excellent alternative for training and education in Spatial Information Systems in Africa at the three levels of technician, technologist and postgraduate. It has recently initiated the implementation of a new and full fledged course in Geomatics. The Regional Centre for Services in Surveying, Mapping and Remote Sensing (RCSSMRS), also under the aegide of the commission, although it does not offer regular courses as does RECTAS, organizes short and medium-term training courses in GIS to African nationals. Further, both centres have experience in organizing customized seminars and workshops, which are apt to provide the required know-how for those in (iv) and (v) above.

Whatever scheme is applied, any public organization will quickly realize that finding and retaining staff with adequate skills may be a serious problem, as it is the adoption of measures and strategies to solve this problem, such as the introduction of special salary scales and effective incentive strategies.

Database Administrator (DBA):

The person or a group of persons, normally constituted by computer scientist(s) and specialized in GIS design and management, responsible for the overall control of the database. Among other important tasks, he will decide how the data will be defined and stored (conceptual and internal data definition languages), and structured, he will liaise with partners and users ensuring that the data they require is available and accessed, will define authorization checks and validation procedures, will ensure that the data meets the established standards and that the data and information that is generated flow smoothly both internally and externally. Finally, he will organize the system so as to get the performance that is "best for the organization", and will make the necessary adjustments to respond to changes in requirements. This important and complicated task, which until not long ago was a serious ordeal, has nowadays been eased as most of the major GIS packages of the market have incorporated the necessary utilities of the DBA within the system's DBMS, many of which are performed automatically and efficiently in a way that is invisible to the user.

A GIS has been rightfully compared to a car (Konecny, 1993), where the hardware and software supplied by vendors is the car itself, the data is the fuel and the administrator (manager) is the driver, without which the car could not go anywhere (and would serve no purpose).

If funds are available, the recruitment of a specialized consultancy firm to perform the duties of the Database/System Administrator, can be a simple yet satisfactory solution, at least during the entire period of implementation, until the whole system is working successfully and when experts within the organization are highly knowledgeable and familiar with the tasks of the Administrator.

Database design and conversion:

Database design: During this stage, the content of the database, which is conditioned by the requirements of the users' needs, is defined and documented. A datamodel, which is an abstraction of the "reality" of a particular application [1], is first developed, representing, in a simplified manner, the entities of interest, the attributes that have to be recorded about those entities, and their releshionships. It is then followed by the development of a database dictionary, which classifies, lists and codifies every theme (and sub-themes), object and entity, down to the last identifiable and meaningful element. The database design should include as well descriptions of the specifications and standards, sources of data, and of the processes for data input and conversion, updates, maintenance and archival.

At the final stage of the database design, the actual structure of the database is developed and documented against the software platform that, by then, must have necessarily been already selected. However, the original concept of the database should not be confined or restrained to a particular hardware and software, as the lifetime and cost of the database surpasses by far the lifetime and cost of hardware and software. In this regard, data independence is a major objective of database systems.

Another important conditionality is that the database design takes into account interfaces and communication with existing and planned computerized datasets, as it is the case within a national geographic information system, more so when the datasets are spread across a network of distributed datasets.

Database conversion: Several options can be used to populate the database. The process can be done by an external contractor, it can be done internally, or part externally and part internally. Each one possesses advantages and disadvantages depending on the singularities of the organization and the database.

There are, however, good reasons that favour the first one, that is external conversion, at least for the bulk of the task:

(i) The database can be very large and therefore, it requires a large number of conversion units (digitizers, scanners, editing stations), all of which are costly and which the organization won't use after the database is completed. *It should be borne in mind that maintaining the database will only need a minimum of units, and that data analysis within the organization can nowadays be performed with inexpensive but powerful desktop PC's and software.* A large number of experienced operators is also required, who at a later stage may just sit idle, entailing all the problems related to staff redeployment, laying-off, etc. These operators can, of course, be recruited only to perform the task of data conversion, but again they have to be adequately trained, and once the process is over, the institution will lose the investment made. Finally, if the organization can not invest in an appropriate number of conversion units or cannot find and train sufficient operators, then the time to populate the database may increase substantially, postponing results (and benefits), and augmenting the risks of failure.

[1] *Nevertheless, the database system should be independent from the application. In fact, different applications will need different views of the same data, in particular within the concept of a national distributed geographic information systems, where the different datasets will be use to satisfy the needs of different sectors.*

(ii) The process of data conversion is new to the organization. If not enough care is taken and quality control is deficient, too many errors are bound to be committed affecting the quality of the database, which sooner or later will be detected and will have to be corrected, either via re-digitizing and editing and re-editing, large portions of the database. In addition to unexpected delays, internal and external criticism and pressures, compounded with lack of confidence and frustration, may show very damaging to the implementation of the GIS.

(iii) Data conversion, as other routine tasks, are often carried out more efficiently by the private sector than by the public sector (see below the role of the private sector).

At any rate, a close monitoring of the data conversion process must be done, assessing the quality of the quality of results, whereby the data converted is accepted or rejected following a clear set of rules, standards and specifications. The responsibility for such control can be assigned to a unit of the organization, if the required know-how is available in house, or can be entrusted to a specialized external firm.

Pilot projects.

Pilot projects are essential elements of any GIS implementation plan. These pilot projects are necessary to test that all the components and functionalities of the system meet its objectives.

The realization of a pilot project consists of the data conversion of a small geographic area of the dataset, which is loaded into the hardware/software selected. Data content and structure, data storage and access, data analysis and queries are verified against the original specifications. The pilot project will also test data conversion and acceptance procedures.

Within the context of the national geographic information system, it is clear that an integrated pilot project, comprising data from distinct datasets from different sources, must be designed and conducted, to test, inter alia, data export/import and integration, compatibility among the individual systems of the network, and pilot applications.

Benchmarks should also be designed and carried out before the acquisition of the particular equipment that the system(s) will use is decided. The primary role of a benchmark is to provide an unbiased mechanism to measure the suitability and efficiency of a supplier's proposed GIS hardware-software solution within the context of the institutions' requirements and environment. The benchmark content should closely replicate the user's application and dataset characteristics, but should be concise and focussed on key aspects that can quickly exposed any potential inadequacies.

The role of the private sector vs. the public sector.

Many tasks are cheaper and more rapidly performed by the private sector than by the public sector. This fact, until recently considered an anathema by public officials at all levels, in both developed and developing countries, is now being slowly recognized and accepted. With notable exceptions, major public institutions have traditionally, by their nature, a strong inertia with stiff hierarchical structures, inflexible norms and rigid procedures, coupled with a "civil servant mentality", all of which make them difficult to perform fast and efficiently. "Social service", and not rentability, is the driving force. The sense of urgency is replaced by a sense of "gravity", where in order to do things properly, these must be done carefully at a determined pace that can not be accelerated. Taking risks is something that is totally out of question. When, in a given case, a decision is made to move faster, the organizational inertia will apply the brakes.

The private sector, on its side, is not limited by traditional cultures or procedures. It is therefore more flexible, and not only can adapt easily to changes, but these are welcome. The driving force is cost-effectiveness, and private firms will swiftly test and incorporate innovative technologies and methodologies to speed up production rates while at the same time lowering costs, taking any necessary risks. The staff, in expertise and number, is also elastic and is easily accommodated to fit the needs of the firm. The danger lies on the "responsibility" of the firm, as quality may be sacrificed on behalf of speed and cost.

To a greater or lesser extent, many States still consider that the private sector: (a) has no business in national mapping or in the collection and management of other types of spatial resource information, or: (b) are competitive bodies and that such competition should not be encouraged or allowed, or: (c) both of the above.

Such attitudes can not be any longer sustained in a modern market economy, where complementarity and partnership between the two sectors are main conditions for development. The private sector can lend/transfer to the national institutions the expertise and know how they may be lacking, and add to and complement their production capacities, enabling them to absorb new technologies as well as to cope with increasing demands for products and services. On its side, public entities, by contracting part of their activities with private firms, will encourage the development of this sector, not only creating new jobs and opportunities, but will also ensure the acquisition of appropriate endogenous technical capacities within the country.

Successful examples of such partnership are not difficult to find as it is the case in many countries with cadastre systems regulated and controlled by the State whereas the surveys are carried out by private surveyors.

Typical components and steps of an institutional GIS implementation process

* INITIATIVE TO INVEST IN A GIS

↓

* EDUCATION

Orientation to national planners, decision-makers and staff through seminars, workshops, formal and informal meetings.

- understanding GIS
- relational databases, topology, queries
- implications to the agency's role
- performance expectations
- demostrations

↓

CONCEPTUAL DESIGN

Establishes the feasibility of GIS, establishes a concept for the system and provides an overall implementation strategy.

- organizational assessment: establishes a starting point: assets and facilities
  gaps and deficiencies
  staff knowledge and motivation
- * outline of responsibilitites and relashioships among all institutions involved
- * identification of users' needs and applications
- * identification and analysis of datasources
- * determing datasets, formats, scales and media
- determing training needs at all levels
- feasibility study: estimation of required resources: invest agressively or within normal budget
  cost/benfit analysis
- * study and appraisal by national steering committee

↓

* APPROVAL

↓

* SYSTEM AND DATABASE DESIGN

- furhter development and streamlining of system concept
- datamodel
- database
- benchmark(s)

↓

* In cooperation and agreement with the national steering body on geomatics

↓

* IMPLEMENTATION PLAN

Provides a (normally multi-year) programme of tasks for establishing the GIS

- revision of resources
- strategies for: training
  - selection and design of pilot project(s)
  - procurement and installation of equipment and software
  - data conversion:
  - database and system administration
  - data up-dating
  - data analysis and manipulation
  - overall system appraisal
  - phasing
- schedule of tasks

↓

IMPLEMENTATION

It is the materialization of the implementation plan

- * training programme launched:
- procurement of equipment: bid/tender documents
  - * analysis of responses, benchmark carried out
  - * selection & award of contract
  - site preparation
  - installation
  - acceptance test
- specific operator's training completed
- data conversion process:
  - external: conversion purchase process
    - pilot conversion: test data base
      - requirements met
    - full conversion
  - internal: recruitment of GIS data conversion consultant
    - pilot conversion
    - full conversion
- pilot project: test overall system performance
  - * test data transfer and communication within the network
  - * test existent/immediate applications
  - * develop and test new applications
  - test GIS management structure
  - Reassess budget requirements - cost/benefit analysis

↓

DATABASE ROUTINE MAINTENANCE AND MANAGEMENT

↓

SYSTEM REVIEW AND EXPANSION

* In cooperation and agreement with the national steering body on geomatics

SECTION TWO

GEOGRAPHIC INFORMATION MANAGEMENT

1. The nature of geographic information

In our case, the real world is formed by physical objects/ features/elements (what we will call **entities**) located on the surface of the earth or below it, having homogenous or heterogenous properties and characteristics (what we will call **attributes**). These entities can be very simple or very complex, and are combined to form other entities. They all have a location within the space that defines our three dimensional world, they occupy a portion of this space (length/ volume/ surface), although it can be punctual, and maintain a spatial relationship with the other elements (distance, adjacency, intersection, connectivity, overlapping, inclusion, etc). These elements, when considered with their spatial characteristics, are referred as *spatial data, georeferenced data or geographic data*. Geographical data are referenced to locations on the earth's surface by using a standard system of coordinates. The coordinate system may be purely local, as in the case of a study of limited area, or it may be that of a national grid or internationally accepted projection such as the Universal Traverse Mercator Coordinate System (UTM).

Geographical data are very often described in terms of well-established geographical "objects", or phenomena. All geographical studies have used phenomenological concepts such as "town", "river", "soil association", "district" as fundamental building blocks for analyzing and synthesizing complex information. They are often grouped or divided in hierarchically defined taxonomies, for example the hierarchy of country-province-town-district, or the hierarchy of soil classification systems or of plants and animals. They are also associated in themes and sub-themes (classes and sub-classes) such as roads and its types, soils and its types, urban blocks, etc.

Data and information that are relevant to resource and environmental management are normally collected and produced by a large number of local, national and regional sources, both public and private, in order to fulfill each one's objectives. This leads to a diversity in its existence, reliability, accuracy, scale, storage, representation, etc, within a given area. Further, in developing countries, it is not an easy task to know what information is available, who produces what and where it can be accessed. Too frequently, studies are conducted to meet a particular need or user, but the results are not made available or disseminated to other users that may require it. As a result, existing information, when obtained, will resemble a patchwork with pieces of rich and adequate information, adjacent to pieces of poor information or lacking it. Rational integrated analysis is then very difficult, if possible at all.

Cataloguing of Information - The meta-data-base

In order to avoid confusion, losses of time and resources, duplication and triplication of efforts, misuse of adequate information as well as use of inadequate information, and all the frustration that may come from the scheme just described, it is necessary, as a first and essential step, to undertake a thorough inventory of all existing information, including actual and potential producers of information that is relevant to natural resources, land planning and the environment. Such a catalogue should not be a simple listing, but should include the attributes of each piece of information: The sources (producer, client, depositary, availability), positional accuracy (for spatial data), up-to-datedness (year of production) and other ancillary information such as scale, projection, completeness, source material from which the data was produced, etc, should be normally part of the attributes. If possible, an atlas depicting graphically the inventory of all available information should be constructed.

Data Representation

Land related information is stored and represented by a number of ways. It can be constituted by texts in the form of books, scientific magazines or journals. It can also be constituted by descriptive characteristics containing the geographic location, the extension and the limits of the elements of interest, in graphic form, where the attributes of each element are either in tabular form, or they are incorporated in the graphics in the form of symbols, or as a combination of the two.

```
                          real world
                              ↓
                         (Formats of)
                 geographic data representation
                   /                       \
                   ↓                       ↓
            non-graphic                  graphic
          (tables, reports)             /       \
                                        ↓        ↓
                                 line maps     images
                          (topographic series, (aerial photos,
                             thematic maps,     satellite images)
                                     sketches)
```

The graphic format, in particular maps, has been used since early and simpler times, probably spontaneously, as the best means to represent data on land resources. It still remains unchallenged and unsurpassed by any other format, noting that in recent times analog graphic representation is yielding a long time reign in favour of digital graphic storage. The importance of mapping can never be overemphasized. The great majority of researches of natural resources and other land occurrences require the positioning of features and observations, and plotting them in map. A map is the most rational kind of spatial representation of data of earth resources and phenomena, either static or dynamic. A map will always yield the quickest and fullest perception of the information, provide the correlation between objects and phenomena, and enable the measurements of spatial-time characteristics. Geologic exploitation, land management, water resource management, cadastre, infrastructure design and implementation, etc, are practically impossible without cartographic support.

Graphic representation goes from simple sketches without scale nor geographic reference, to full fledged cartographic representations complying with the highest mapping specifications of accuracy and content.

A means that has proved very effective and valuable in the last decades to handle spatial data is the use of aerial photographs. Not only the photographic image serves as the base on which results and phenomena are delimitated and consigned, but through interpretative processes the task of the technician and scientist in procuring raw data is highly facilitated. Although precise measurements of the elements on the photographs are not possible without resorting the sophisticated equipment and highly technological expertise, a fair assessment of distances, surfaces, altitudes, elevation differences and slopes, can be made with very simple means and procedures. When the resource scientist or resource manager requires higher accuracy, then the interpreted elements are readily transferred to a precise map by visual correlation or other methods, and finally the thematic map is produced. However, this requires that a map be available at a suitable scale, which in Africa is not always the case.

Orthophotographs and orthophotomaps became popular in the seventies and early eighties. Orthophotos are photographic images, brought to a predefined scale and where the distortions produced by ground relief and by the non orthogonality of the camera axes have been removed. They conserve all the richness of the photographic image, while at the same time have the standard precision of the map. An orthophoto and a line-map differ in that the in first one the identification of the details are left to the user, while in the latter, the details have been interpreted and symbolized by the map maker. An orthophotomap is an orthophoto or mosaics of Orthophotos, cut to fit the map sheet format, where the map grid and some linear features have been superimposed. Rural cadastre and land use mapping became the most attractive applications of these products in developing countries.

Another product that has come with great force is the satellite image and its corresponding orthophotomap called the space image map or space map. Satellite images are less distorted that photographic images and covers a broader area. There is a broad constellation of observation satellites collect data of the earth and the atmosphere with a wide number of sensors of different spatial, spectral, temporal and radiometric resolution. These constellation will be continuously improved with satellites possessing better resolving power and other characteristics. Satellite images can be obtained and manipulated in both analog form or digital form.

Structures of geographic data storage: analog and digital

Analog form: Until very recently, this "analog" form has been the only way of representation since the very first map was produced and a different "mode" was not even conceived. Thus, practically the totality of the spatial geographic information that has been produced exist in the form of analog maps and documents.

Structures of data storage

real world
↓
graphic representation
/ \
↓ ↓
analog **digital**
↓ ↓
maps drawn on paper (symbols drawn over a reference grid + legend) | computer data files (coordinates representing points, lines and polygons + attributes)

Digital mode: The developments in computer science has led to a different mode of storage and representation. The geographic data is stored in digital form in computer data files which represent the real world.

2. Major Data sources

Lets try to visualize the most important data sources:

	Type	Scale	Positional Accuracy	up-datedness/ repetitivity	Supplier
GRAPHICS					
Maps	**Topographic** series	1:500 k/1.000 k 1:200 k/ 250 k 1: 50 k/ 100 k	150m-300m * 60m - 75m * 15m - 30m *	low (5 - 20 + years)	National/Federal Mapping Institutions
	Thematic maps				
	geological	1:250k-1.000k	**	adequate	Geol. Dpts. Min. of Mines and Energy, Oil companies
	hydrogeological	1:250k-1.000k	**	adequate to low	Geol., Nat. Resources and Energy Dpts.
	rural cadastre, land tenure	1: 50k on remote areas to 1: 5K on semiurban areas	rel.acc.: high abs.acc.: high to very low	medium 5 to 10 years	Cadastre, Land Tenure Depts.
	vegetation cover; land use; soils	1:200k/250k 1: 50k/100k	**	low, nat.coverage is normally partial	Forest, Nat. Res., Agriculture authorities. Research institutes,

	oceanographic & hydrographic	1:500k/1000k 1:200k/ 250k 1: 50k/ 100k	variable	very low, partial coverage	Oceanog. & hydrographic agencies, mapping institutions, Navy forces
	rainfall & veget. index	1:4.000k-10.000k	**	very high -weekly-	Meteorological authorities
	other (census, population, nat. parks, livestock, fishery, etc	variable	**	low, some produced on ad-hoc basis	National, federal, municipal authorities, research instit., IGOs, NGOs

* The national topographic series offer the highest positional accuracy attainable at the scale of the map. This accuracy is not normally surpassed by other analog maps.

** The uncertainty of boundaries of categorical maps, makes the issue of positional accuracy difficult to appraise. What may be more important is the quality itself of the sources and the consistency and methodology used in building the map.

Images	**aerial photographs**				
	normal	1:30k - 60k	variable ***	low 5 - 10 years	
	high definition- high altitude	1: 100k +	variable ***	medium 2 - 5 years	National/ Federal Mapping Authorities. Military forces
	orthophotos, orthophomaps	1: 30k - 100k	same of the topographic map	low to medium	
	satellite images	***Swath***	***Resolution/average accuracy***		
	Meteosat	5 km	7.5km	30 min	Meteo. Depts.
	NOAA		1.5/6km	12 hrs	Meteo. Depts.
	LANDSAT	185x170 m	120m	17 days	
	LANDSAT TM	185x170 m	45m	17 days	Mapping, natural resources, environmental, authorities, Remote sensing centres
	SPOT-XS/P	60km	10-20/15-30m	28 days	
	JERS-1	75km	18m/25m	44 days	
	ERS -1 (slar)	100km	30m/45m		
	IRS-1	150km	36-72m/50-110m	22 days	

*** The scale within a photographic image is not constant. The positional accuracy depends mainly on flight height, ground relief and the inclination of the camera axis.

NON-GRAPHIC

studies, reports, publications, statistical data, socio-economic data, etc.	Everybody

Further characteristics of geographic data and information sources

The **topographic map** is probably the most important data source. It is the regular representation of earth features and phenomena, normally constructed following rigorous procedures and specifications, and constitutes the standard reference upon which other databases are build. However, (i) Offers "limited" environmental information: while general areas of forest, rangeland, deserts, cultivated and populated land are shown, it doesn't provide major detail of the characteristics within those areas, unless for very large scales; (ii) it is costly and time consuming to establish and maintain, hence, it is usually out-of-date. Even considering the "ideal scheme" whereby the national mapping can be kept up within a 5 and a 10 year revision cycle for dynamic and less dynamic areas respectively, rapid changes, whether gradual of sharp, will not be detected timely.

Concerning **thematic maps**, with some exceptions, these are not produced and revised in developing countries in a systematic manner, and in many cases only offer a partial coverage, and again, if they are not up-to-date their utility will be reduced or disappear.

Cadastre and land tenure maps, if produced in the context of a multi-disciplinary system, are invaluable source of environmental information. These maps will not only portray the boundaries of the parcels, but also how the land is possessed and used. A sound cadastre system will equally indicate the type of soils and their aptitude or vocation. Maps depicting zones of homogenous characteristics, such as slope, erosion, road accessibility, water availability, value of the land, etc, are natural by-products of such systems.

Aerial photographs have been for many years the basic and most appreciated source of data for studies of earth related phenomena, and still constitutes the major input for topographic and thematic mapping. It offers a richness of detail unsurpassed by any other source, and can be obtained in standard black and white (B&W), in natural colours, in B&W infrared -where the infrared reflection of the earth surface is captured-, or in (false) colour infrared, providing a vast range of possibilities for the detection and interpretation of the imaged elements. However, each photography has a reduced coverage and for some applications, such as geomorphological studies, and when the territory to be analyzed is vast, it will only give a limited vision, and often too many photographs will have to be treated. It has the inconvenience that it requires of cloud free weather and therefore the obtention of a good aerial coverage becomes very difficult in tropical areas, sometimes impossible. The cost (around $4/km2 for small scales) is also a limiting factor to embark in periodic and systematic aerial photo coverage. The recent development and production of aerial cameras with forward image motion compensation (FIMC), have marked one of the important revolutions in mapping technology in the last decade. They have allowed the use of high resolution film with which we obtain images with the same detail and accuracy as those from photographs having half the scale, being specially suitable for high altitude photography. Scales of 1:100.000 and above can now be used. This advancement allows speeding up of production while cutting down costs notably.

Radar imagery: From active sensors installed in airborne or space platforms, radar images of the earth surface are obtained independently of weather or light conditions, through cloud coverage and at any time of the day or night. However, major inconveniences of airborne radar images are the heterogenous scale within the image, and the "shadows" in areas of high relief caused by ground slopes. Although in these images the geomorphology and the water areas show very well, their interpretative capacities for land use and land cover still have to be proved.

Space technology has brought a new dimension to resource mapping. Satellite images and their corresponding orthophotos or spacemaps, provide general views of any area of the globe, at low cost and at regular intervals in time with a high repetitivity. It has pros and cons, detractors and defenders, but it is proving to be the best option so far available that can effectively solve the major land and environmental information gaps of developing countries, including the completion and revision of their national topographic mapping. In Asia and America they are being introduced and accepted as standard operations in their cartographic and natural resources institutions and programmes. In this regard, it is gratifying to note that some countries in the region have already incorporated the use of satellite images in their regular programmes ; others are in the verge of doing so.

It has to be said that in some regions of Africa, cloud coverage imposes serious restrictions to obtain good satellite images, but these restrictions are even more serious for conventional aerial photography. In such areas, the high repeatability of the satellite image is a factor favoring the possibility of obtaining workable images with reasonable delays. In this respect, the radar capabilities of ERS-1 satellite should prove of particular interest. The high resolving power and other characteristics of the system make it very well-suited for territorial applications in the tropical belt in Africa because it is not affected by meteorological problems. Radar, moreover, has a special sensitivity for the morphological characteristics of the surface and the topography, and could prove to be a particularly effective tool in the production of forest and geological maps.

From the rigorous point of view of planimetric accuracy, LANDSAT TM imagery proves to be perfectly apt for the revision and up-dating of maps at scales up to 1:150.000. Strictly speaking, it should be the system selected with priority over other possible systems, both satellite and conventional. Furthermore, for vast areas of the African continent, where the landscape is less developed, a more relaxed accuracy can easily be accepted and the use of LANDSAT imagery be extended without major restrictions or hesitations to the scale of 1:100.000.

Similarly, SPOT imagery should be the system selected with priority over conventional systems for planimetric mapping and revision at scales ranging from 1:150.000 up to 1:66.000, and for remote and less developed areas, up to 1:50.000 (SPOT P).

Of all the operational satellites, SPOT is the only one offering altimetric determination capabilities. The standard accuracy attained has been assessed at 5 to 10+ meters, depending on the conditions (identification of control points, temporal separation and the B/H ratio). In general, the altimetric accuracy is recognized to be compatible with topographic mapping with minimum contour intervals of about 30 to 40 meters.

Precision in positioning is not the only factor that determines the applicability of satellite imagery in mapping. Another equally important factor is the capacity to interpret from the image the elements of interest that must be consigned in the map. Such capacity is directly related to the spacial resolution or pixel size of the image, which conditions the level of detail detection and identification. It is beyond any doubt the factor of major controversy regarding satellite mapping suitability at medium and small scales, ranging from those who totally ostracize it and those who accept it.

Tables 2, 3 and 4 in Annex 1 shows us how the conditions of accuracy and interpretability are ordinarily met with LANDSAT TM and SPOT satellite images.

3. Data organization

When studying one particular environmental or development problem, the expert has to identify first what type of data is needed to solve his problem. Then he has to locate data sources, find it is existent and available and if its quality is OK. This location and assessment of data may be, as it is often, a tall order. The expert may not be able to identify all needed sources, and much less the quality of the data. It may also happen that, initially, he can not identify all his needs, regardless how keen or specialist he is in the matter, as there may be valuable documents that can assist in solving the problem but of which I am not aware of.

If the documented catalogue mentioned at the beginning of this presentation is available and properly maintained and accessed, then the puzzle is easily solved. If it is not, the expert can be in deep trouble. Indeed, the data and information required for any environmental aspect is a set constituted by numerous subsets of data pertaining to different sectors. On the one hand, modern technological development is leading to a convergence of what previously were discrete disciplines. On the other, we can not handle any more one resource or one environmental aspect in isolation from others resources or aspects, as it was done in simpler times. In addition to sound spatial and temporal data, information and knowledge on discrete aspects of the environment, which eventually can be obtained independently for each of them, their actual understanding and management cannot be properly conducted without considering their interaction. As an example, water resources is dependent on climate, rain fall, soil structure, geomorphology, vegetation cover, weather behavior, hydraulic and hydrological infrastructure, etc. Food security will be influenced by land use, land potentiality, natural vegetation, techniques in use, human influence, industrial development, population growth and migrations, industrial development, and many others.

The identification of a suitable location for a refuge camp, will require, e.g. to locate land with water accessibility, benevolent climate conditions, a particular soil structure, adequate slopes, proper drainage, etc, which are all find in different documents of different quality.

In conclusion, the data and information set of any aspect of development is constituted by numerous subsets of data and information pertaining to different sectors.

climate rain-fall
soil structure geomorphology
vegetation cover
weather behavior hydraulic and hydrological infrastructure
etc etc

the vegetation set

trade hard currency avail.
exports vs. imports
land use land potentiality natural vegetation
techniques in use human influence
industrial development population growth and land tenure policies
migrations and many others.

the food security set

In a nutshell, the task of the planner is to identify and access the required subsets of information that are relevant to solve his problem, and discard redundant and inadequate data. Then, what is relevant has to be simplified, and synthesized, adding intelligence to it so that appropriate decisions are made. In other words, two are the main issues to resolve: firstly, the provision of means for accessing accurate, reliable and timely data and information; and secondly, the provision of means for the integrated analysis of the different types of data that is made available, comprising both internal and external factors.

How can we achieve the above objective?

The current analog mode of the documents containing the spatial and non-spatial data and information, that is drawings and text on paper form, does not facilitate the extraction of the selected elements, neither the integration with other analog data, much less their analysis. First, it is not easy to obtain copies of the documents, and to physically manipulate them. Further, the data contained in analog form is of rigid nature, making it difficult to separate what is relevant from the rest of the information that is disturbing and adds confusion to the work. But what is more problematic are the difficulties inherent in integrating heterogenous data coming from an different sources, with variable resolutions, scales and inaccuracies. We can use projectors to blow up (or reduce) and plot a map at a different scale or we can use photographic enlargers. We can draw a series of overlays, one for each selected element, trying to overcome the geometric inconsistencies by visual interpolation, or forcing the features to fit the geometry of a particular map. Then we can, more or less visually, interpret the results and make synthesis-maps portraying those results. You will agree with me that when the amount of data sources is high and the area subject to analysis covers numerous map sheets, then the task of extracting and integrating the data is simply formidable. Another inconvenient of this form of presentation is that normally the results, including resulting maps, are normally consigned in a report, and the new maps are lost for other potential users.

Fortunately, today we have powerful tools hat have been perfected to do such a task. These are the Geographic Information Systems (GIS).

The nature and importance of GIS

These systems have emerged from the above need to evaluate processes and evaluate geographic data in an integrated, multidisciplinary way, as a realization that it does not function independently of each other (Geeta Varadan et al, 1992). Although GIS evolved from previous Computer Aided Design and Mapping systems -CAD/CAM-, which initially were developed to accelerate and improve manual map production operations, the advances in computing and information sciences have evidenced a continued increasing potential of the real value of digital spatial data, surpassing any expectations and changing the conception itself of mapping science and what it represents. Geographic Information Systems are more than just coding, storing and retrieving geographic data. The data in a Geographic Information System represent a model of the real world, and because these data can be accessed, transformed and manipulated interactively, is it possible for planners and decision-makers to explore a range of possible scenarios and to obtain an idea of the consequences of a course of action before the mistakes have been irrevocably made (P. A. Burrough, 1986). It has moved from graphic inventory to modeling potential uses. It has been argued that GIS is a change in manual manipulation of spatial data like the cocoon's change to a caterpillar and butterfly (Joseph K. Berry, 1993). "Ugly, but effective".. "To those on the outside, the cocoon appears to just sit there. But on the inside, there is a total up-heaval and complete restructuring. Such is the metamorphosis brought by the digital map"... " The butterfly is obviously superior to the worm. But more importantly, it is radically different".

There are many definitions of GIS, such as:

> "a number of computer software, hardware and peripherals that transform geographically referenced spatial data into information on locations, spatial interactions, and geographic relationships of the fixed and dynamics entities that occupy space in the natural and built environments. (US Federal Interagency Coordinating Committee on Digital Cartography)

or

> "a system of computer hardware, software, and peripherals designed to support the capture, management, manipulation, analysis, ..., and display of spatially referenced data for solving complex planning and management problems" (Environmental Systems Research Institute)

The last definition gives a good indication of what modern GISs are, indicating explicitly that the end goal of GIS is the solution and management of problems.

Synthesis of major examples of GIS application classes

Sector	applications	general characteristics
rural and urban data management	land use programme cadastre networks and services construction planning ...	large scale cartography large data volume repetitive operations (daily, weekly, monthly) good precision static nature

Impact studies	road layout industrial plant commercial centre tourist resort ...	medium scale 1: 5000 → local level 1:50000 → regional level punctual, but long term multi-source required (maps, photos, census data) dynamic nature : "what if"
Transport assistance	fleet management navigation assistance road network ...	variable scale no precision dynamic nature
Environmental studies	natural resources inventories pollution prevention risk areas (earthquake, floods, ...) intervention or evacuation plans ...	multi-source (satellite, aerial data, ground surveys...) variable scale large volume static or dynamic complex if dynamic
Cartographic presentation	economy politics population promotion ...	no precision required punctual studies interactivity statistical data simple
		.....
Military applications	...	

GIS functionalities

A GIS manipulates the geographic data in its three dimensions, spatial (shape and location), thematic and temporal, and organizes it in such a way that the user sees and handles it as a series of layers (overlays) each one corresponding to a particular theme. In turn, the subthemes of each theme can also be subdivided and organized in layers.

The user can retrieve any particular theme or group of themes (and sub-themes), he can retrieve individual elements and combine them according to different associations of graphic and non-graphic attributes. Relational algebra, in particular Bolean operators can be applied to the data. Let us keep in mind that traditional mathematical capabilities, plus an extensive set of advance nap processing operations, are available in modern GIS packages. You can add, subtract, multiply, divide, exponentiate, root, log, cosine, differentiate, and even integrate maps. After all, maps in a GIS is just an organized set of numbers. Queries such as:

"In "noname" municipality, show me plots where surface area is greater than 50 hectares (cadastral graphic database), slope is less than 5% (topo database) and average annual rain is above 100mm (rainfall database). Indicate plot No., owner and cadastral value (cadastral alphanumeric database)",

"Within 50 miles around "locoloco" locality, show areas where soil-type is abcd or wxyz (soil database). Indicate land use (land use database) and compute individual and total area",
"Show coastal and riparian areas affected if an ten hour oil-spill of 100m3 occurs in well No 1234, at 15:00 (topographic data base + tidal database + risk model). Predict damage in fisheries, wildlife, farms (fisheries, wildlife, cadastre database)",

"Show possible paths connecting points x1,y1 and x2,y2 having a slope less than 5%, 7% and 10% (topodatabase). Show geologic formations and rock and sand deposits within a corridor of 15 miles along each path (geologic data base)",

"In "unique" state, show areas within national parks and other protected areas (protected areas database) where severe deforestation has taken place in the last decade (vegetation/forest past and current databases), compute rate of degradation, show new settlements and population indexes (census database), show current land use of deforested areas (land use database)",

which were just expectations a few years ago, are now a full reality. The current technology allows us to make them and obtain the answers.

However, two paramount conditions must exist:

(i) Existence of data in computer form: The digital spatial database

(ii) Integration of data of different nature coming from different sources.

Creating the database

A GIS can only handle data that is in computer readable form, organized in such a way that the system can perform its functions, stored in a database (see appendix 1). It is clear that without a database the system is entirely useless, just like a car without fuel.

The process of populating a database is referred as <u>database conversion</u>, and includes the transformation of the existing analog data and information, such as paper maps, into the digital format required by the GIS.

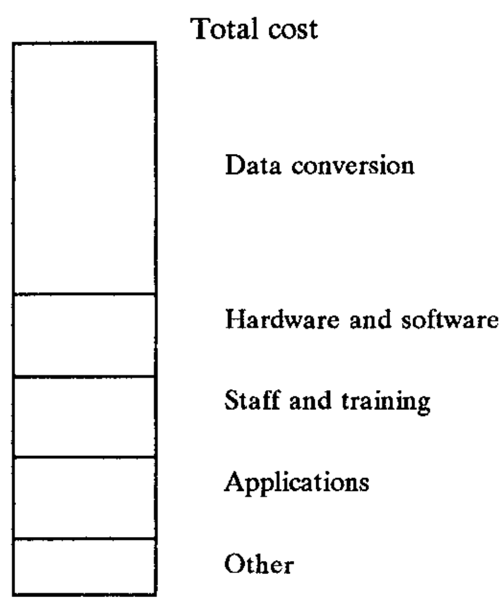
In Africa, like in most developing countries, few efforts have been done to create, in a systematic way, the fundamental databases that are required for modern spatial analysis. It has been argued, with reason, that the contribution of modern cartographic science to the development processes of these countries, has been, to say the least, marginal. Geographic Information Systems are not used by planners and decision-makers mainly because the data needed by such systems is not available. A recent study [W. Ottichilo et al, 'Establishing GIS in Africa: problems and challenges', 1993] showed that one of the impediments for the development of GIS in Africa, was the lack of adequate data, not only already converted data, but also the lack or inadequacy of source data from which the digital databases can be obtained.

Data conversion process

sources: existing analog data
(maps, graphs and images + associated records and narratives)
(field surveys data, datafiles, etc)
↓
database design
(to fit the requirements of the GIS)
↓
conversion technologies and approach
(digitizing, scanning, editing)
↓
conversion process
(internal, external)
↓
database

The creation and maintenance of a core integrated set of fundamental datasets would be a major step in the establishment and use of GIS technology in the continent. In particular, the topographic map, the most important source of base-line information, and constituting the skeleton around which the other datasets are built, should be tackled with priority (see next section: data integrity).

With the current techniques, the total cost of conversion typically exceeds 50% of the overall cost of implementing a GIS. This is due primarily to the significant amount of tasks that have to be done manually. Although some automated techniques have been successfully developed, such as scanning and line-tracking, the substantial amount of post processing and editing, in order to add intelligence to the database and ensure topology, will not allow any major decrease in the cost.



GIS implementation costs

The cost (and the time) spent on the data conversion process is highly dependent on the scale/accuracy of the original data and is estimated at $ 5,000-10,000 per square meter of analog graphic document digitized. The data conversion cost normally varies as the second power of the scale ratio: e.g. it would cost 100 times more to digitize a map at the scale of 1:20.000 than for a map at the scale of 1:200.000, with the same density of detail.

It would be unrealistic to conceive the initiation of a national GIS based on the conversion of nation-wide datasets at large scales from the beginning. Even if the original analog datasets are available — which is not the usual case unless it is a small country — it would take too long before it produces effects, too many resources would be devoted on something that is relatively new to the organizations, and the huge databases obtained would be complex to archive and manage. If the datamodels are not carefully conceptualized and structured (e.g. datadictionary, topology, formats and standards, etc) their utility to serve the needs of the national GIS may be substantially limited, representing a large waste of resources until the models and the processes are readjusted.

On the other hand, national institutions, sources and custodians of resource and environmental data, should not hesitate to start the conversion of their datasets, in a progressive way, from small to large. It is more and more accepted within the community dealing with land resources and environmental information that scales of 1:250k/200k should be adopted for the basic national database reference system. In further steps, databases at larger scales would be build up.

In particular, as a first and essential step for the creation of national resource information infrastructure, mapping institutions should embark without delay in the design of their database systems, and start converting the analog maps to digital form as soon as possible, as the resulting digital topographic databases are urgently needed by the other institutions in order adjust and integrate their own data.

For a large average country like Ethiopia, Nigeria, Tanzania or Niger, etc, the estimated cost[2] of converting their national map series would be:

First phase:	**1:250,000 scale:**	**$ 125,000 for the entire country**
	can be achieved in a very short time: aprox. one year with 4 conversion units working two shifts per day.	
Second phase:	**1:100,000 scale:**	**$ 800,000, initially for selected areas, then for the rest of the country**
Third phase:	**1:50,000 scale:**	**$ 3.2-$4 per square km.** Normally selected areas would be treated, as not all the territory would justify, for the time being, the construction of a nation-wide database at this scale. In this case the cost would add up to $ 3,500,000

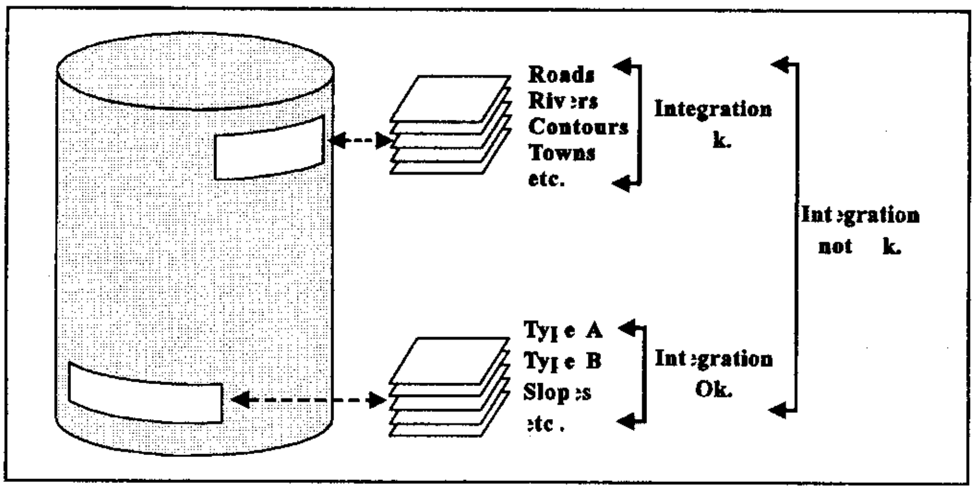
These (mapping) institutions should not be reluctant to start the process because they do not have the required expertise at all or some levels, nor the appropriate equipment. The private sector and specialized regional centres, such as RECTAS and RCSSMRS, are ready to assist in the process. GIS contractors can be used to perform the required data conversion and provide the training to the

2 Only data conversion costs at an average value of $7.500/sq.m.

organization in maintaining and handling the database. On their side, the mapping organizations only require relatively simple and inexpensive equipment & software to handle the small-scale database created during the first phase. Further, with a modest investment and with the assistance of GIS consultants, any organization can make effective use of their existing physical and human resources to populate the initial small-scale database.

Data Integrity

Let's look at the national GIS database (described in appendix 1).



The data in the databases is organized in separated "overlays", each one containing the data pertaining to a different theme.

Different overlays, originating from a unique source, will fit perfectly together. This is logical as they were derived from a unique map, where all elements are integrated.

But, when different sets of data come from different sources, then their integration is not evident, and may not be possible at all.

- <u>Data integration</u> relates to:

  Geometry
  Content (classification systems/specifications)
  Data format
  Structure (raster and vector)
  Socio- economic data (non-spatially structured)

Geometric Integration

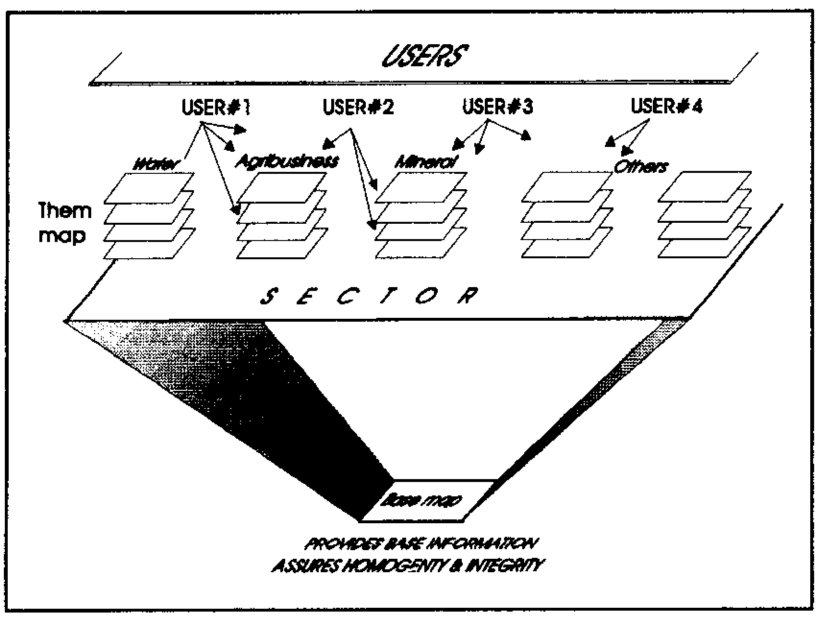
Geometric integration is possible when the geometry of different sets of data is based on the same reference system and the elements have been positioned with "good accuracy" with respect to that reference system.

Two sets of data, with different reference systems, can be brought together if the these systems are mathematically known and, therefore, one can be transformed into the other. This is the usual case of satellite data and topographic maps.

In order to integrate the two disparate sets which are not geometrically homogeneous, one set has to be forced into the other until common features to coincide, operation that not always is feasible, and requires a substantive amount of editing of one or both of the two of the datasets. (GIS packages have special functions, such as "rubbering", to facilitate this task).

The ideal situation would be, for any type of geographic information system, is that all data sets are geometrically homogenous, where all the elements have been adjusted to one common reference system, which is of high quality, reliable, and is a correct mathematical model of the earth.

It is realized, immediately, that the rational way of achieving that condition is by using the national topographic map as the reference system upon which the other datasets are adjusted. It is the only base that can assure the condition of geometric homogeneity and integrity.



Further, the ideal situation is that all data sets share a unique set of primitives (points, lines, polygons — see appendix 1). Thus, any up-date (addition, correction, deletion) in any of the datasets will be automatically reflected in the rest of the sets that contain the element up-dated.

The idea is simple in theory, however its practical realization pose many problems.

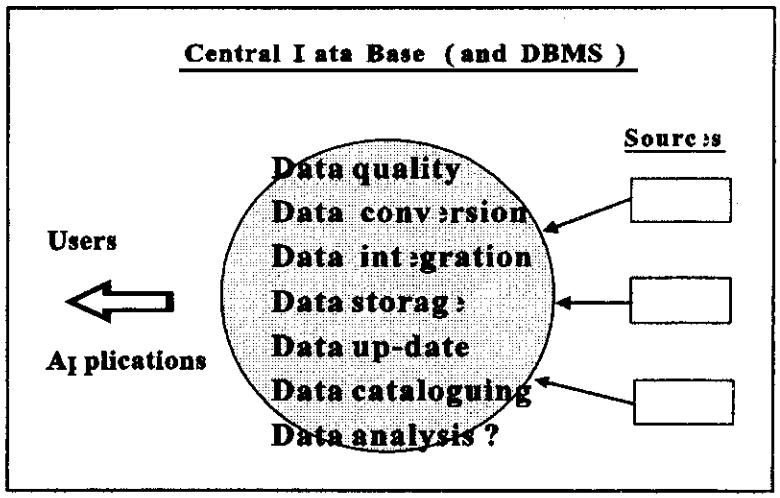
There are two main schemes of implementation: one is axed around a central database, where the different subsets of data are stored and integrated in one physical location; the other is based upon "distributed databases" which is, typically, a database that is not stored in its entirety in one physical location, but is rather spread across a network of computers that are geographically dispersed and connected via communication links.

Central data base (and DBMS management)

All data is converted, handled and stored in one central data base. It assures the highest degree of integrity.

Problems:

size
coordination
resources
priorities
flow of data
mandates and responsibilities
protection
ownership

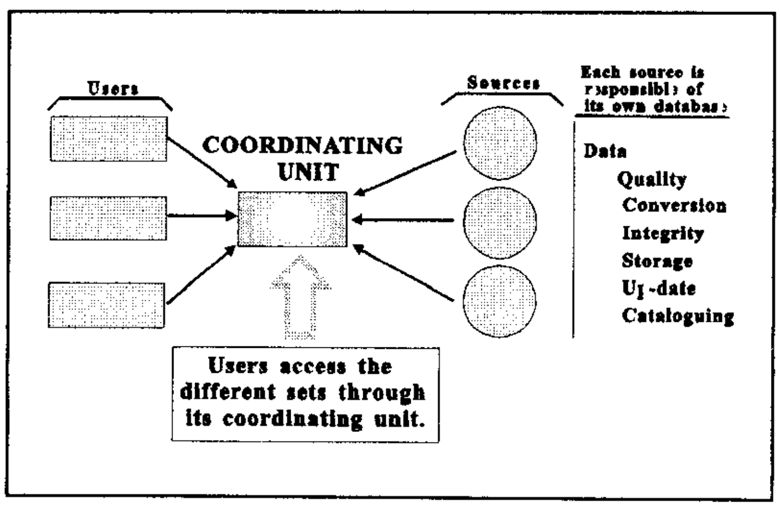


Distributed Database

Each dataset is converted, handled and stored by the producers of the data

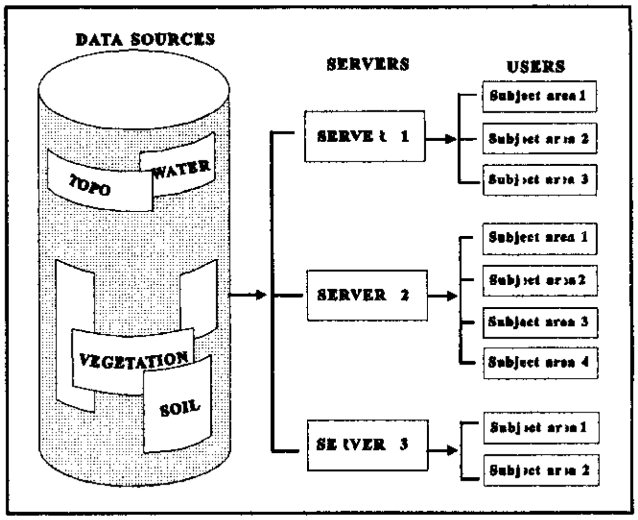
Problems:

standards
data integration
flow of information
transmission of up-dates
heterogeneity of hardware/ software

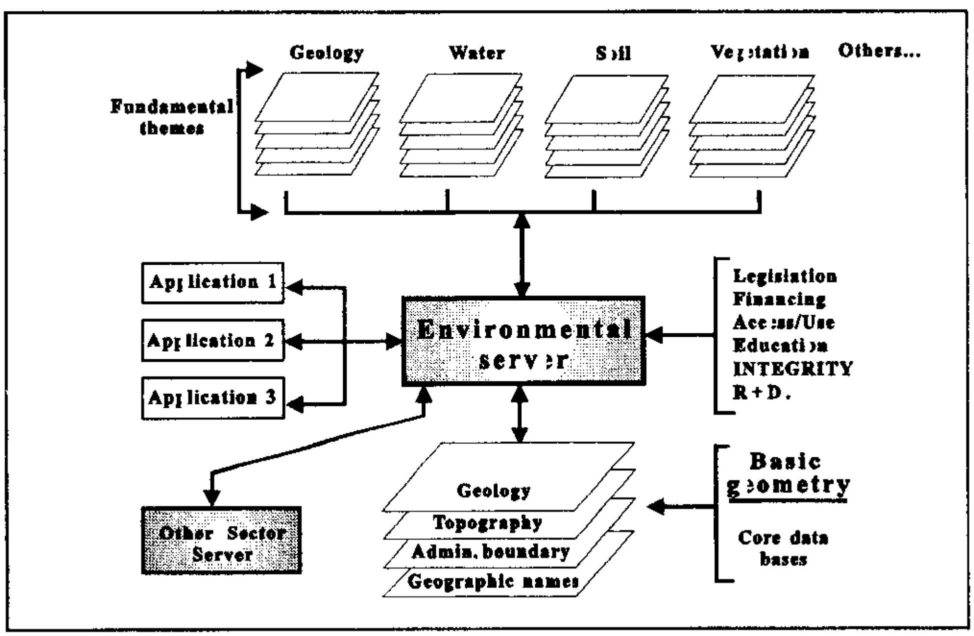


Dedicated servers

One scheme, which is a sort of a combination of the two schemes above, is proving to be very promising. This approach uses the concept of application (dedicated) servers that banking and credit companies use, or the flight information and booking servers utilized by airlines. Each server handles the database that is relevant to a particular application.

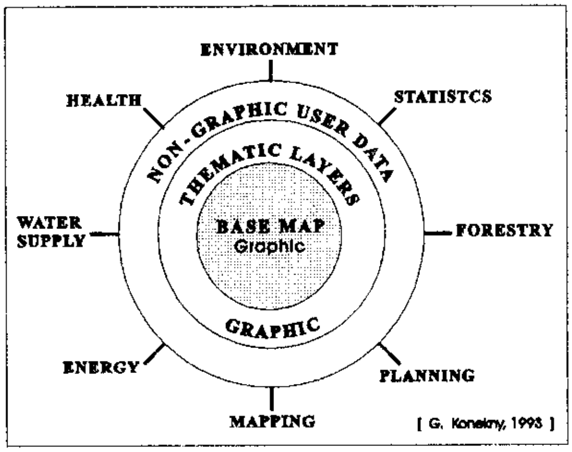


Scheme of *a dedicated server*



The importance of the topographic map and the additional great responsibility that mapping institutions have in providing the required data base can never be overemphasized. If mapping institutions don't do it, the users will start converting the maps on an ad-hoc basis, many times without knowledge of the cartographic science, using different criteria, formats and structures, and at the end they will have added chaos instead of solutions.

The following scheme (G. Konecny, 1991) is clear, showing the topographic map as the core of the system where all other thematic maps appear as layers to it.



What happens in those areas where the base-map is highly outmoded or does not exist ?

This problem, very serious some years ago, can be overcome, relatively expeditiously and cheaply, with the use of satellite images such as LANDSAT and SPOT (and ERS-1 and INS-1). These images will allow the rapid up-date of obsolete maps at scales up to 1:100.000 and in some cases up to 1:50.000, And , combined with GPS measurements, replace the map where it does not exist, providing the necessary geometric support to other databases. In this context, some of the leading advantages of satellite technology are:

as background to line maps.

to up-date maps. There is no need to use expensive equipment. It is enough to have geometrically corrected images ($0.18 - 0.90/km2) and to follow simple cartographic techniques requiring modest equipment.

to replace the base- map where it does not exist

to establish a nation-wide environmental resource satellite image data base for multiple use, that can expedite the preparation of thematic maps for quick decisions on the environment and natural resources management.

It is interesting to note the *FAO AFRICOVER PROJECT* [3] will cater for all of the above.

Other aspects of Integrity

Content: One of the problems of data integrity, independent from geometric considerations, raises when trying to combine thematic layers representing the same theme coming from different data bases, and the qualification of the phenomena is different due to different standards, specifications, and classification schemes, set up for each map in order to satisfy individual needs.

It may be the case when studying soil series aiming at establishing a suitability class map for reforestation of an area overlapping two or more administrative boundaries, and each one has produced a map with different classification systems.

Date format and standards. Until recently, conflict in data exchange format were frequently faced when importing data from different datasets. In many cases it was not just not possible, and when it could be achieved, important losses in the topological structure were experienced. Currently, the systems in use are becoming more open and, practically, no major difficulties exist in exporting/importing spatial data. Another improvement is that, due to the high number of users of some successful GIS, these are becoming standards themselves. It is the case, particularly, with packages like ARC-INFO and MAPINFO, among the most important.

Further, major mapping US organizations, such as the USGS and the DMA, have developed data transfer formats (Digital Line Graphics, DLG and Spatial Data Transfer Standard, SDTS) which most likely will soon be university adopted.

Concerning data semantics and quality, sets of norms have also been developed, such as the ISO 900. The Digital Geographic Information Working Group (DGIWG), which is constituted by 11 North American and European countries, has been developing DIGEST (Digital Information Geographic Exchange Standard), which specifies the exchange structure and feature attribute coding scheme for vector, raster and matrix data.

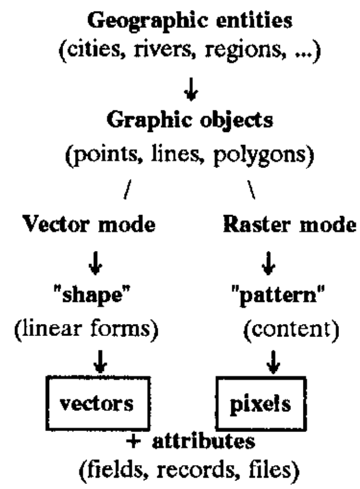
---

[3] *The overall objective of the AFRICOVER PROJECT is to strengthen the capacities of African countries to manage their natural resources sustainability with particular emphasis on food security and environmental protection.*

The specific objectives are to: (a) establish a land cover digital database for, and by African countries, based on a harmonized methodology and advanced geographic information technologies and existing data; (b) produce a land cover map at scales 1:250 000 and 1:1 000 000 from the digital database; (c) strengthen national and regional capacities in the application of advanced geographic information technologies to land cover mapping, natural resources assessment and environmental monitoring.

The implementation of the project will count with the full involvement of the relevant national institutions in member States, under the overall coordination and monitoring of the corresponding regional remote sensing centre in each sub-region concerned.

Integration of vector and raster data structures



There are two data structures for the storage of geographic data in the computer: *raster* and *vector*. Each one of them is suited for a particular type of data. The vector mode is used for linear elements, such as roads, fences, contour lines, pipe lines, etc, and whenever a high positional accuracy is desired. A topographic map will be stored in vector form. The data conversion from analog to digital is done through linear digitization (or through scanning and then automatic vectorization plus manual editing). The raster mode is appropriate for the storage and representation of regions, such as soil types, hypsometric values, etc. Precision is lower than in the vector mode and depends on the pixel size. Data conversion is done through scanning procedures. In addition, images, such as satellite images or photographs are stored in raster form.

It can be easily seen that many applications will require the analysis of sets of data with the two different structures. However, until now, there is no GIS that allows the full integration of raster and vector data-files. Queries combining attributes from the two different structures are not possible. What can be done is two superimpose the vector file over the raster file and viceversa, provided that they are geometrically homogenous, and make the spatial analysis by visual correlation and interpretation. It is also possible to convert vector data into raster data, but it must be kept in mind that loss of precision of the vector file will occur. Raster data can equally be converted into vector data, but the process is very complicated, requiring in many cases powerful workstations, and the results are not always satisfactory.

Integration of Socio-economic data

It has been reported (AFRICA-GIS'95) that the majority of accessible socio-economic data is collected on ad-hoc basis, not systematically to serve useful databases, and without clear objectives and targets. The georeference of this data is normally administrative boundaries. It has been further noted that no standard procedures exist for the integration of socio economic data in spatial databases, and for each work a particular approach has to be devised.

Socio economic theories, at the moment, are not using the benefits or potentialities of GIS. Some experts are just starting to understand it and use it in their studies and theories. Similarly, GIS specialists are not familiarized with socio-economic data. They are related to environment, land use, etc.

APPENDIX 1

FROM THE REAL WORLD TO COMPUTER DATA FILES: THE GEOGRAPHIC DATABASE

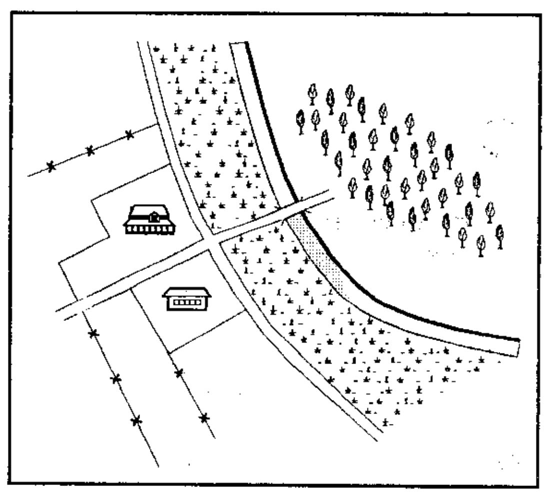
1. Geographic data files

A GIS handles geographic data stored in digital form in computer files which represent the real world, following the general process indicated below.

real world
↓
graphical representation
↓
geometry + topology
↓
geographical data

The real world

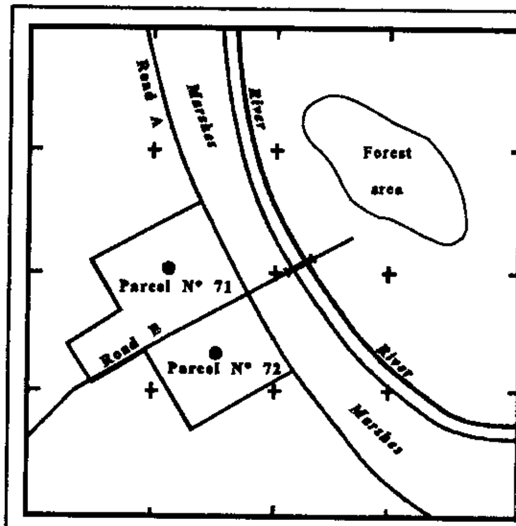
The physical world is composed by elements and objects that occupy a certain space and have a particular location on the earth (and by extension on the universe). These objects and elements have their own identity represented by their physical characteristics, normally referred as attributes. These elements and objects constitute the geographic data[1] whose study, together with their relation with men, has been the realm of the geographer.



The real world

[1] *Also referred as georeferenced data or spatial data.*

Graphic representation: Cartography is the science in charge of determining and depicting the shape of the earth and the location and space that geographic entities (data and phenomena) occupy on it. Graphics and more particularly maps, have been historically the means to represent their location, space and the geometric relationships that exist among these entities.



Graphic representation
A map, a drawing made with points, lines and symbols

In a map, the different elements (entities) are drawn on a flat surface, over a grid representing the reference system that defines the space (two dimensional or three-dimensional), usually in plane coordinates (x,y,(z)), although some representations use the geographic coordinates of longitude and latitude (λ,ϕ): navigation charts, small scale maps, etc. The geometric characteristics (shapes) are represented by points and lines. The (non-graphic) attributes are identified by a set of symbols, whose meaning is given in a legend. Graphic representation goes from simple sketches without scale or geographic reference, to full fledged cartographic representations complying with the highest mapping specifications of accuracy and content. A different and recent (since the beginning of this century) form of representation of earth data and phenomena is the aerial photograph, taken from airplanes or space platforms, and even more recently (just a few decades) by space images obtained from specialized satellites. Maps and photographs, are basically distinct in nature. In the map the elements are drawn with lines and symbols following the interpretation made by the cartographer. The photograph shows an image of the earth, constituted by a 'texture" of shades of grays or colours, just like the human eye sees it, and the interpretation of the elements are left to the user.

Geometry :

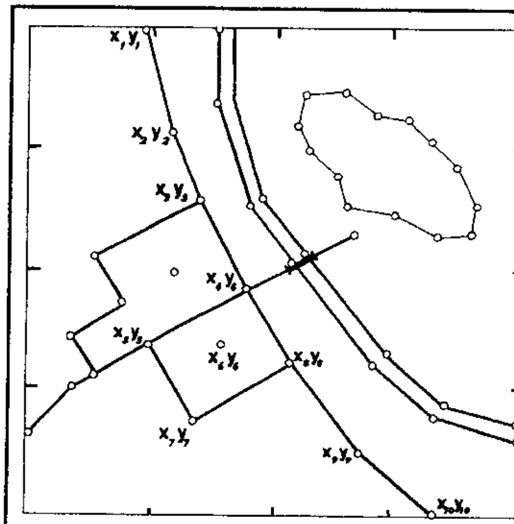
Two are the basic structures used to organize the geographic data in the computer: the vector and the raster structure.

Vector structure: The geometry of the line-map is structured having the vector as the basic logic unit. It uses the following elements called "primitives":

<u>Points,</u> constituted by a pair x,y (2D) or a triplet x,y,z (3D) of coordinates.

<u>Lines,</u> a sequence of connected points.

<u>Polygons,</u> a closed line, whose starting and ending points are the same.



o point of coordinates x_i, y_i
Vector geometry
Stores coordinates of points

<u>Points</u> will represent all geographic entities that are punctual, such as a power tower, a ground monument, a corner of a block of houses, a city in a small scale map, etc. The lower house of the example in the box will be represented by point (x_6, y_6).

<u>Lines</u> will represent linear entities, such as roads, rivers, contour lines, parcel boundaries, etc. The road A will be the string of points (x_1, y_1), (x_2, y_2), ... ,(x_8, y_8),(x_9, y_9),(x_{10}, y_{10})

<u>Polygons</u> will represent "regions", such as soil types, parcels, lakes, etc. The lower property by the closed string (x_4, y_4), (x_5, y_5), (x_7, y_7), (x_8, y_8), (x_4, y_4)

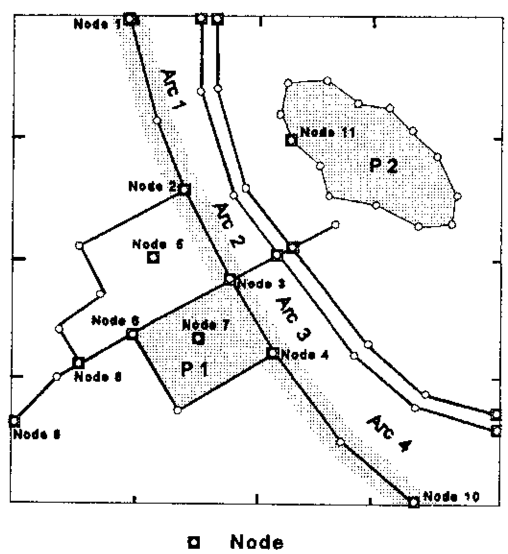
The attributes of each entity will be provided by alpha-numerical codes attached to the above "primitives".

Topologic primitives: Alternatively, the following similar basic elements are used. However, they belong to a higher level or data organization as they facilitate the building and recognition of the spatial relationships existing among the entities.

<u>Nodes,</u> a point of intersection of two or more lines.
a starting or an ending point.
a point entity.

<u>Arcs,</u> two nodes and the line connecting them.

<u>Polygons,</u> formed by a closed loop of successive arcs.



Topologic vector structure

The two houses would be represented by NODES 9 and 10, respectively.

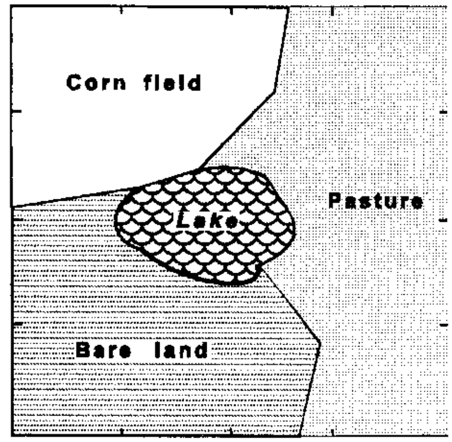
Road A would be represented by the successive following ARCs: Arc1 (N1-N2), arc2 (N2-N3), arc3 (N3-N4), arc4 (N4-N10)

Parcel 72 would be represented by POLYGON P1 constituted by: Arc3 (N3-N4), arc5 (N4-N6), arc6 (N6-N3)

The forest would be represented by polygon P2 formed by the closed Arc7 (N11,N11)

Raster structure: The basic unit of representation is a tiny unit of area, called cell or pixel (picture element) The geographic entities are represented by sets of cells. A point will be one isolated cell, a line a succession of tangent cells, a polygon a group of cells. The position of the cell is defined by the coordinates of its centre (centroid) or one corner. Considering that the cells form a regular grid, the position can be deducted from the sequential number of the cell within the array, or by the row and column to which it belongs.

The attributes will be provided by codes and a value given to each particular cell or groups of cells.



Thematic map representing surface areas of distinct nature

Raster structures are considerably more simpler than vector structures, However storage requirements are larger,[2] and the accuracy is poorer, limited to pixel size. Phenomena which have no sharp boundaries, such as those represented in categorical maps (soils, geologic structures, land cover, etc) are better represented as rasters.

Some thematic spatial analysis operations, such as Bolean operations and polygon overlay, are more easily performed with raster mode than vector mode.

1	1	1	1	1	1	2	2	2
1	1	1	1	1	1	2	2	2
1	1	1	1	1	2	2	2	2
1	1	1	1	2	2	2	2	2
1	1	1	3	3	2	2	2	2
4	4	3	3	3	3	2	2	2
4	4	4	4	3	2	2	2	2
4	4	4	4	4	4	2	2	2
4	4	4	4	4	4	2	2	2
4	4	4	4	4	4	4	4	4

Values identify poligons:
1: corn fields 2: pastures 3: lake 4: bare land

Raster structure

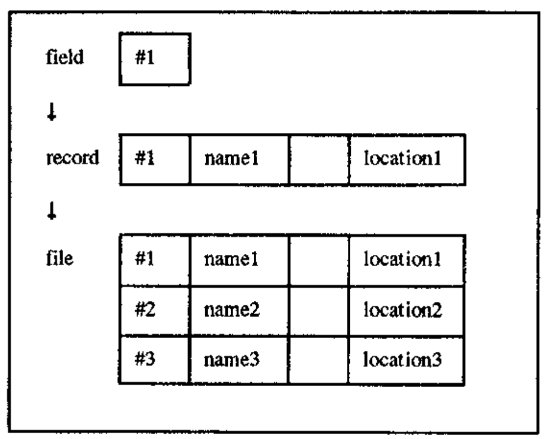
[2] *Presently, some efficient techniques , such as quadtrees , have been devised to store raster data without using too much storage space. Further, the growing capacities of storage media are reducing this disadvantage.*

The **physical organization of geographic data in the computer** can be synthesized as follows.

Geographic entities
(cities, rivers, regions, ...)
↓
Graphic objects
(points, lines, polygons)
/ \
Vector mode ***Raster mode***
↓ ↓
"shape" ***"pattern"***
(linear forms) *(content)*
↓ ↓
vectors ***pixels***
+ attributes
(fields, records, files)

Geographic data files

The geographic data for each geographic entity consists in a set of alphanumerical values registering the non-spatial attributes plus coordinates registering the geometry, which in turn are also attributes of the entity, stored as **fields**, **records** and **files**. The smallest unit of data is called a **field**. An attribute will be represented (and stored) by one field or several fields. For instance, a database containing cadastral information would probably include a field called "parcel number", a field called "owner", a field for the "corner" nodes, one or several fields for the "boundaries" or the parcel, one field for its area, etc. The collection of associated fields relating to a geographic entity is a record. A **file** is a collection of associated records.



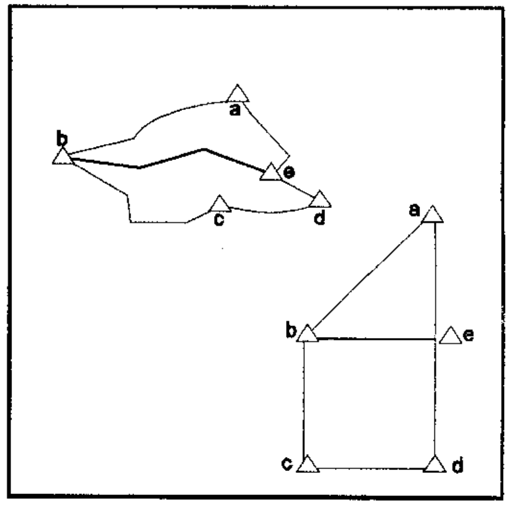
Geographic data files

Relational databases

In practically all modern GIS databases the geographic data is located in bidimensional tables. They are particularly suited to ensure the identification and conservation of the relationships existing among the different entities, by indexes and pointers defining the links between one particular table and another. The most common database structure, referred as relational database structure, obeys to the above scheme.

*Spatial relationships: **Topology***

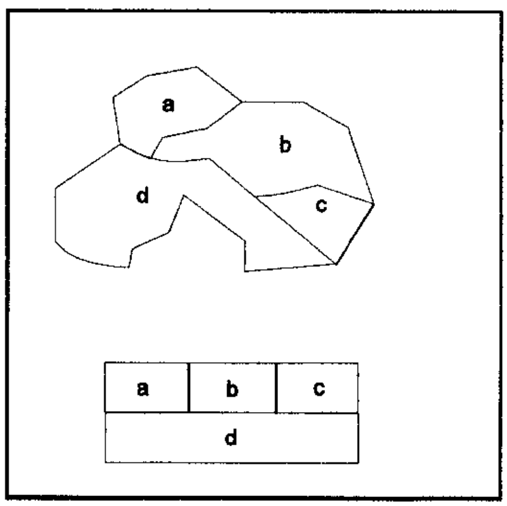
The study of the spatial relationships among spatial entities is known as topology. Rigorously is the analysis of geometric properties not tied to distance and coordinate measure. It is a coordinate-free geometry. Today, it is the mathematical field that studies the characteristics of geometry that are invariant under continuous transformations. In simpler terms it indicates the way in which the geographic elements (e.g. arcs, nodes, polygons and points) are linked together. For example, the topology of an arc includes its from-and to-nodes, and its left and right polygons. Topological structures are built from simple elements into complex elements.



Topologically equivalent networks

Topology is essential to GIS because many of the spatial functions and operations required for the analysis of spatial data, in particular modelling, do not require coordinates, but topological information. For instance, to find an optimal path between two points requires a "list" of which arcs connect to each other and the cost of traversing along each arc in each direction. Coordinates are only needed to draw the path after it is calculated. Queries such as intersection, adjacency, such as intersection, adjacency, inclusion, polygon overlay, are done and answered using algebraic topology.

Topology is also important to reduce redundancy and assure integrity, checking the integrity and the consistency of the data base. Inconsistency between two entries representing the same "fact" is an example of lack of integrity. For instance if a point belonging to a line and a polygon are assigned different coordinates in each case, the system will detect a contradiction as two different points can not represent a unique node. Topology will assure that common elements will share the same primitives (points, lines). A sound database structure can therefore be aware of possible redundancies and assume the responsibility of propagating up-dates: creation, modification, deletion.



Topologically equivalent surface areas

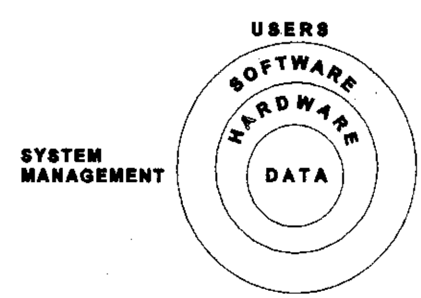
In a modern GIS much or all of the topological encoding should be carried out by the system, if possible in a way transparent to the system. In limited GIS the topology has to be created manually by the operator during the process of digitization of the data, increasing tremendously the time and cost for data conversion. In a system were the object is spatial data analysis and manipulation, this limitation should be avoided. If the purpose is just graphic representation, the construction of topology can at all be spared (but in this case we do not have -do not need- a GIS).

2. The components of a GIS

The basic scheme of a GIS do not differ from other information systems. The major difference lies on the geographic nature of the data base, and, by extension, on the spatial processing and queries made to it. The nature and structure of the geographic data base is discussed later in this document.

It involves four major components:

- ***data***
- ***hardware***
- ***software***
- ***users***



GIS components

They will be considered briefly below:

Data

The data stored in the system is positioned in one, two or more **databases.**

A database, then, is a repository for stored data. In general, it is both **integrated** and **shared**. By "integrated" we mean that the data is not independently duplicated in different files, and that any redundancy among files is eliminated. By "shared" we mean that individual pieces of data in the database may be shared among several different users, in the sense that each of those users may have access to the same piece of data (and may use it for different purposes). Such sharing is a consequence of the fact that the database is integrated. Another consequence of this integration is that any given user will normally be concerned only with some subset of the total database; moreover, different users 's subsets will overlap in many different ways. In other ways, a given database may be perceived by different users in a variety of different ways. (Even if two users share the same subset of the database).

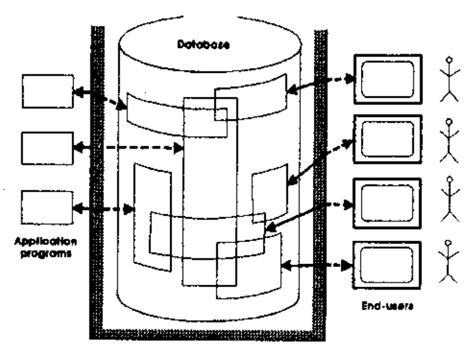


Fig. 1.1 Simplified picture of a database system. [Date C.J.]

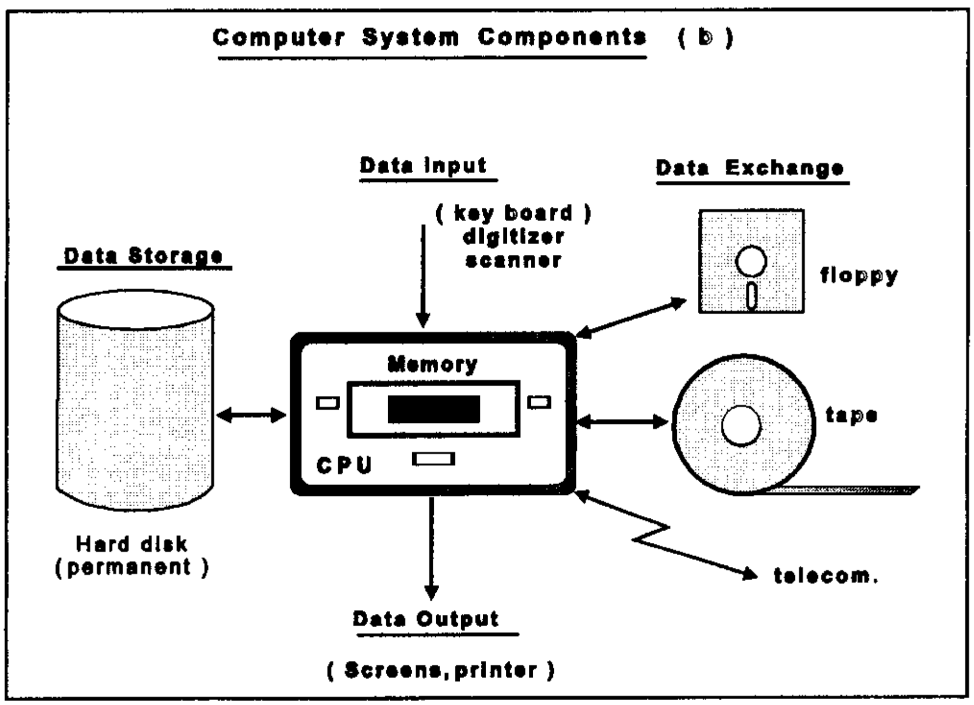
Simplified picture of a Database

The term "shared" is frequently used to cover, not only sharing as just described, but also concurrent sharing: that is the ability for different users to be actually accessing the database - at the same time - (such a system is referred also as a multiuser system)

"Distributed" data bases are, typically, a data base that is not stored in its entirety at a single physical location, but is rather spread across a network of computers that are geographically dispersed and connected vis communication links [3]. Consider the case of a national geographic data base containing data on different sectors: water, land use, base-map, geology, etc, and the data pertaining to each sector is stored in a sub-database located at the ministry or institution concerned with the sector: Water Development Authority, Forest Department, Mapping Institute, Minister of Mines and Energy, etc.

Hardware:

It consists of all physical computer devices, control units, and peripherals used by the system, such as CPU, RAM and Disk memory, elements for data storage, data conversion (input, editing), data exchange, data output etc.



GIS hardware components

[3] *Data, C.J. -An introduction to data bases*

Software

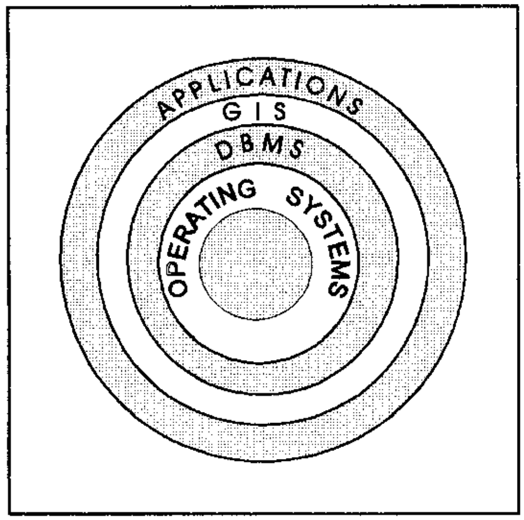
The software is normally constituted by:

(1) the Operation System (OS) of the computer platform in use,

(2) the Database Management System (DBMS)

(3) the GIS software itself

(4) the application programmes.



Software Components

Operating System:

It is the heart of the system, without which no application can run. Its basic functions are:

user management	accounting access protection	
process management	"process": elementary system/user task	create hold/suspend terminate
memory management	allocation read (load) write (store) protection	
Input/Output (I/O) management	allocate peripheral send data (write) receive data (read) interruptions handling	

file management	allocate file open close read (record) write (record) control access
communications management	send message receive message send file receive file

Database management system (DBMS)

It is the software that handles all access to the database. Conceptually: (i) a user issues a request using a data manipulation language, normally SQL (or squel) (Structured Query Language), (ii) the DBMS intercepts the message and interprets it, (iii) the DBMS inspects, the external and internal schema, conceptual schema, and the storage structure definition and, (iv) the DBMS performs all the necessary operations on the stored database.

Query: "show me plot No. 335 and give me the area"
↓
compile query
↓
find/retrieve data
↓
extract or
compute area (if not stored)
↓
give answer

Example of a query

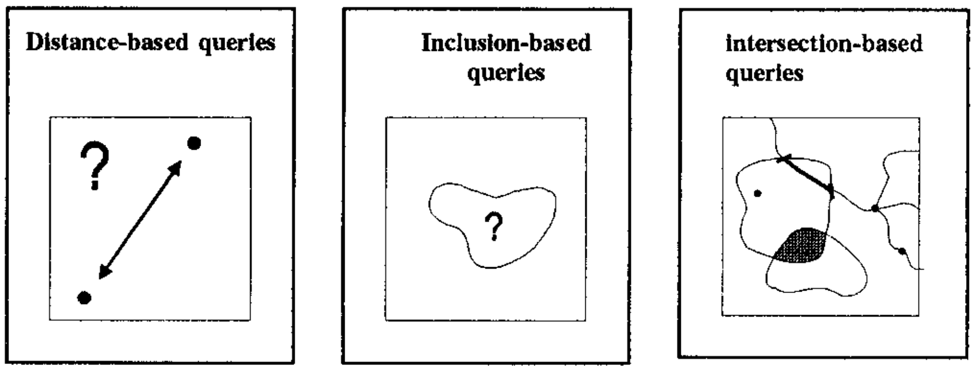
DBMS functions:

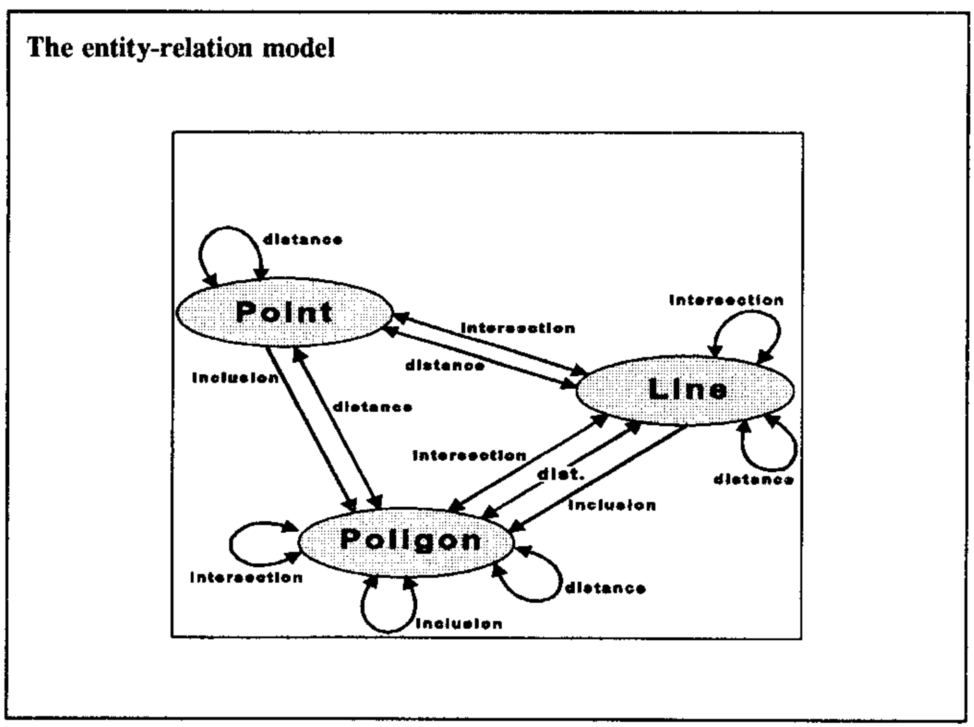
Schema creation/modification → *datadictionary*	protection → *unauthorized access*
User interface → *query language*	data sharing → *multiple users*
query processing → *decomposition, optimization, access*	Consistency/ integrity checking → *validation of inputs and outputs*

3. GIS functionalities:

The GIS software allows the capture and structuring of the geographic data base, its query and analysis, according to combinations of geometric and non-geometric attributes, normally interactively, and the output of results. The GIS software may have its own proprietary DBMS and query language. In particular, the spatial relationships among the entities are preserved in the database allowing distance-based queries, inclusion-based queries and intersection based-queries using relational algebra operators.

Spatial relationships:





Major GIS functionalities

Basic Geographic Data Processing	automated cartography	for storage and display only of graphic and maps	
	thematic cartography	cartographic presentation of (tabular) statistical data: superimposition of layers": topography, land use, road network, cadastre,	distribution of population, political results, agriculture, industry, ...
	geographic analysis	non-spatial queries SQL queries to a relational database	distance, intersection, inclusion
	modeling and simulation	what if ?	

Query processing	statistical operations	histograms, classification, graphics
	geometric operations	distance, surface, centroid, buffer zone generation, polygon generation, digital terrain model generation, contour line generation
	set operations	intersection, union, inclusion, merge, (polygon overlay)
	conversion operations	scale, projection, internal representation (vector vs. raster)
	3D representations	perspective, digital terrain modeling
	multimedia processing	integration of numeric images linkage with analog photos and video sequences
	graphic techniques	windowing, zooming, panning, browsing
	network operations	optimum path

APPENDIX 2

AGENDA 21, CHAPTER 40

INFORMATION FOR DECISION-MAKING

INTRODUCTION

40.1 In sustainable development, everyone is a user and provider of information considered in the broad sense. That includes data, information, appropriately packaged experience and knowledge. The need for information arises at all levels, from that of senior decision makers at the national and international levels to the grass-roots and individual levels. The following two programme areas need to be implemented to ensure that decisions are based increasingly on sound information:

(a) Bridging the data gap;

(b) Improving information availability

PROGRAMME AREAS

A. Bridging the data gap

Basis for action

40.2. While considerable data already exist, as the various sectoral chapters of Agenda 21 indicate, more and different types of data need to be collected, at the local, provincial, national and international levels, indicating the status and trends of the planet's ecosystem, natural resource, pollution and socio-economic variables. The gap in the availability, quality, coherence, standardization and accessibility of data between the developed and the developing world has been increasing, seriously impairing the capacities of countries to make informed decisions concerning environment and development.

40.3. There is a general lack of capacity, particularly in developing countries, and in many areas at the international level, for the collection and assessment of data, for their transformation into useful information and for their dissemination. There is also need for improved coordination among environmental, demographic, social and developmental data and information activities.

40.4. Commonly used indicators such as the gross national product (GNP) and measurements of individual resource or pollution flows do not provide adequate indications of sustainability. Methods for assessing interactions between different sectoral environmental, demographic, social and developmental parameters are not sufficiently developed or applied. Indicators of sustainable development need to be developed to provide solid bases for decision-making at all levels and to contribute to a self-regulating sustainability of integrated environment and development systems.

Objectives

40.5. The following objectives are important:

(a) To achieve more cost-effective and relevant data collection and assessment by better identification of users, in both the public and private sectors, and of their information needs at the local, provincial, national and international levels;

(b) To strengthen local, provincial, national and international capacity to collect and use multisectoral information in decision-making processes and to enhance capacities to collect and analyze data and information for decision-making, particularly in developing countries;

(c) To develop or strengthen local, provincial, national and international means of ensuring that planning for sustainable development in all sectors is based on timely, reliable and usable information;

(d) To make relevant information accessible in the form and at the time required to facilitate its use.

Activities

(a) Development of indicators of sustainable development

40.6. Countries at the national level and international governmental and non-governmental organizations at the international level should develop the concept of indicators of sustainable development in order to identify such indicators. In order to promote the increasing use of some of those indicators in satellite accounts, and eventually in national accounts, the development of indicators needs to be pursued by the Statistical Office of the United National Secretariat, as it draws upon evolving experience in this regard.

(b) Promotion of global use of indicators of sustainable development

40.7. Relevant organs and organizations of the United Nations system, in cooperation with other international governmental, intergovernmental and non-governmental organizations, should use a suitable set of sustainable development indicators and indicators related to areas outside of national jurisdiction, such as the high seas, the upper atmosphere and outer space. The organs and organizations of the United Nations system, in coordination with other relevant international organizations, could provide recommendations for harmonized development of indicators at the national, regional and global levels, and for incorporation of a suitable set of these indicators in common, regularly updated, and widely accessible reports and databases, for use at the international level, subject to national sovereignty considerations.

(c) Improvement of data collection and use

40.8. Countries and, upon request, international organizations should carry out inventories of environmental, resource and developmental data, based on national/global priorities for the management of sustainable development. They should determine the gaps and organize activities to fill those gaps. Within the organs and organizations of the United Nations system and relevant international organizations, data-collection activities, including those of Earthwatch and World Weather Watch, need to be strengthened, especially in the areas of urban air, freshwater, land resources (including forests and rangelands), desertification, other habitats, soil degradation, biodiversity, the high seas and the upper atmosphere. Countries and international organizations should make use of new techniques of data collection, including satellite-based remote sensing. In addition to the strengthening of existing development-related data collection, special attention needs to be paid to such areas as demographic factors, urbanization, poverty, health and rights of access to resource, as well as special groups, including women, indigenous peoples, youth, children and the disabled, and their relationships with environment issues.

(d) Improvement of methods of data assessment and analysis

40.9. Relevant international organizations should develop practical recommendations for coordinated, harmonized collection and assessment of data at the national and international levels. National and international data and information centres should set up continuous and accurate data-collection systems and make use of geographic information systems, expert systems, models and a variety of other techniques for the assessment and analysis of data. These steps will be particularly relevant, as large quantities of data from satellite sources will need to be processed in the future. Developed countries and international organizations, as well as the private sector, should cooperate, in particular with developing countries, upon request, to facilitate their acquiring these technologies and this know-how.

(e) Establishment of a comprehensive information framework

40.10. Governments should consider undertaking the necessary institutional changes at the national level to achieve the integration of environmental and developmental information. At the international level, environmental assessment activities need to be strengthened and coordinated with efforts to assess development trends.

(f) Strengthening of the capacity for traditional information

40.11. Countries, with the cooperation of international organizations, should establish supporting mechanisms to provide local communities and resource users with the information and know-how they need to manage their environment and resources sustainably, applying traditional and indigenous knowledge and approaches when appropriate. This is particularly relevant for rural and urban populations and indigenous, women's and youth groups.

Means of implementation

(a) Finance and cost evaluation

40.12. The secretariat of the Conference has estimated the average total annual cost (1993-2000) of implementing the activities of this programme to be about $1.9 billion from the international community on grant or concessional terms. These are indicative and order of magnitude estimates only and have not been reviewed by Governments. Actual costs and financial terms, including any that are non-concessional, will depend upon, inter alia, the specific strategies and programmes Governments decide upon for implementation.

(b) Institutional means

40.13. Institutional capacity to integrate environment and development and to develop relevant indicators is lacking at both the national and international levels. Existing institutions an programmes such as the Global Environmental Monitoring System (gems) and the Global Resource Information Database (GRID) within UNEP and different entities within the systemwide Earthwatch will need to be considerably strengthened. Earthwatch has been an essential element for environment-related data. While programmes related to development data exist in a number of agencies, there is insufficient coordination between them. The activities related to development data of agencies and institutions of the United Nations system should be more effectively coordinated, perhaps through an equivalent and complementary @Development Watch@, which with the existing Earthwatch should be coordinated through an appropriate office within the United Nations to ensure the full integration of environment and development concerns.

(c) Scientific and technological means

40.14. Regarding transfer of technology, with the rapid evolution of data-collection and information technologies it is necessary to develop guidelines and mechanisms for the rapid and continuous transfer of those technologies, particularly to developing countries in conformity with chapter 34 and for the training of personnel in their utilization.

(d) Human resources development

40.15. International cooperation for training in all areas and at all levels will be required, particularly in developing countries. That training will have to include technical training of those involved in data collection, assessment and transformation, as well as assistance to decision makers concerning how to use such information.

(e) Capacity-building

40.16. All countries, particularly developing countries, with the support of international cooperation, should strengthen their capacity to collect, store, organize, assess and use data in decision-making more effectively.

B. Improving availability of information

Basis for action

40.17. there already exists a wealth of data and information that could be used for the management of sustainable development. Finding the appropriate information at the required time and at the relevant scale of aggregation is a difficult task.

40.18. Information within many countries is not adequately managed, because of shortages of financial resources and trained manpower, lack of awareness of the value and availability of such information and other immediate or pressing problems, especially in developing countries. Even where information is available, it may not be easily accessible, either because of the lack of technology for effective access or because of associated costs, especially for information held outside the country and available commercially.

Objectives

40.19. Existing national and international mechanisms of information processing and exchange, and of related technical assistance, should be strengthened to ensure effective and equitable availability of information generated at the local, provincial, national and international levels, subject to national sovereignty and relevant intellectual property rights.

40.20. National capacities should be strengthened, as should capacities within Governments, non-governmental organizations and the private sector, in information handling and communication, particularly within developing countries.

40.21. Full participation of, in particular, developing countries should be ensured in any international scheme under the organs and organizations of the United Nations system for the collection, analysis and use of data and information.

Activities

(a) Production of information usable for decision-making

40.22. Countries and international organizations should review and strengthen information systems and services in sectors related to sustainable development, at the local, provincial, national and international levels. Special Emphasis should be placed on the transformation of existing information into forms more useful for decision-making and on targeting information at different user groups. Mechanisms should be strengthened or established for transforming scientific and socio-economic assessments into information suitable for both planning and public information. Electronic and non-electronic formats should be used.

(b) Establishment of standards and methods for handling information

40.23. Governments should consider supporting the efforts of governmental as well as non-governmental organizations to develop mechanisms for efficient and harmonized exchange of information at the local, national, provincial and international levels, including revision and establishment of data, access and dissemination formats, and communication interfaces.

(c) Development of documentation about information

40.24. The organs and organizations of the United Nations system, as well as other governmental and non-governmental organizations, should document and share information about the sources of available information in their respective organizations. Existing programmes, such as those of the Advisory Committee for the Coordination of Information Systems (ACCIS) and the International Environmental Information System (INFOTERRA), should be reviewed and strengthened as required. Networking and coordinating mechanisms should be encouraged between the wide variety of other actors, including arrangements with non-governmental organizations for information sharing and donor activities for sharing information on sustainable development projects. The private sector should be encouraged to strengthen the mechanisms of sharing its experience an information on sustainable development.

(d) Establishment and strengthening of electronic networking capabilities

40.25. Countries, international organization, including organs and organizations of the United Nations system, and non-governmental organizations should exploit various initiatives for electronic links to support information sharing, to provide access to databases and other information sources, to facilitate communication for meeting broader objectives, such as the implementation of Agenda 21, to facilitate intergovernmental negotiations, to monitor conventions and efforts for sustainable development to transmit environmental alerts, and to transfer technical data. These organizations should also facilitate the linkage of different electronic networks and the use of appropriate standards and communication protocols for the transparent interchange of electronic communications. Where necessary, new technology should be developed and its use encouraged to permit participation of those not served at present by existing infrastructure and methods. Mechanisms should also be established to carry out the necessary transfer of information to and from non-electronic systems to ensure the involvement of those not able to participate in this way.

(e) Making use of commercial information sources

40.26. Countries and international organizations should consider undertaking surveys of information available in the private sector on sustainable development and of present dissemination arrangements to determine gaps and how those gaps could be filled by commercial or quasi-commercial activity, particularly activities in and/or involving developing countries where feasible. Whenever economic or other constraints on supplying and accessing information arise, particularly in developing countries, innovative schemes for subsidizing such information-related access or removing the non-economic constraints should be considered.

Means of implementation

(a) Finance and cost evaluation

40.27. The secretariat of the Conference has estimated the average total annual cost (1993-2000) of implementing the activities of this programme to be about $165 million from the international community on grant or concessional terms. These are indicative and order of magnitude estimates only and have not been reviewed by Governments. Actual costs and financial terms, including any that are non-concessional, will depend upon, inter alia, the specific strategies and programmes Governments decide upon for implementation.

(b) Institutional means

40.28. The institutional implications of this programme concern mostly the strengthening of already existing institutions, as well as the strengthening of cooperation with non-governmental organizations, and need to be consistent with the overall decisions on institutions made by the United Nations Conference on Environment and Development.

(c) Capacity-building

40.29. Developed countries and relevant international organizations should cooperate, in particular with developing countries, to expand their capacity to receive, store and retrieve, contribute, disseminate, use and provide appropriate public access to relevant environmental and developmental information, by providing technology and training to establish local information services and by supporting partnership and cooperative arrangements between countries and on the regional or subregional level.

(d) Scientific and technological means

40.30. Developed countries and relevant international organizations should support research and development in hardware, software and other aspects of information technology, in particular in developing countries, appropriate to their operations, national needs and environmental contexts.

Planimetric accuracy:

ANNEX 1

Table 1

m_{pl} = 1 pixel size

With good geometric models and well defined sufficient control points, the standard planimetric accuracy m_{pl} of LANDSAT and SPOT satellite images have been normally assessed at 1 times the pixel size:

Sensor	m_{pl} (m)	Largest mapping scale based on m_{pl} = .3 mm at mapping scale
Landsat TM	30	1:100.000
Spot XS	20	1:65.000
Spot P	10	1:33.000

Table 2

m_{pl} = 1.5 to 2 times pixel size

In Africa, elaborate and rigorous accuracy tests carried out in Sudan and Yemen (G. Petrie and A.E.H. El Niweiri, 1992) shows that when there is a deficiency of well defined points the results are less satisfactory: m_{pl} = 1.5 to 2. times the pixel size:

Sensor	m_{pl} (m)	Largest mapping scale based on m_{pl} = .3 mm at mapping scale
Landsat TM	45 to 60	1:150.000 to 1:200.000
Spot XS	30 to 40	1:100.000 to 1:113.000
Spot P	15 to 20	1:50.000 to 1:66.000

Table 3

Interpretability of LANSAT and SPOT imagery

LAND USE CLASSES	LANDSAT	SPOT 1/
Communication lines	Hard surface roads are easily seen, but there is almost total failure to detect and identify single track railways, unsurfaced roads and car roads.	The interpretability of roads is excellent. In general, even small single lane roads (3-5 m) are interpreted (90% +) and at least half of the car roads are seen.
Cultural features	Large and medium large cities are readily identified, but not so their internal pattern. Villages and small settlements remain undetected, although some can be inferred but not identified.	Large and medium cities and built up areas are seen quite well, including the street pattern. Villages and small settlements can generally be detected, although classification is more difficult
Vegetation and land cover	Woodlands and cultivated land can generally be well detected and identified.	Vegetation areas and agricultural land can be delimited and identified with high accuracy(85%+).
Water areas and drainage	Open water areas, large canals, rivers and natural drainage show up very well and with great complexity. Small rivers and irrigation canals are very difficult to detect	Water areas and rivers are detected and classified with high accuracy. Interpretation of rivers narrower than 5 m. is less accurate. Results improve with stereoscopic interpretation.

Analysis of LANDSAT are based on results on image interpretability (U. of Glasgow & U. of Khartoum) published in ISPRS journal, 47 (1992) Analysis of SPOT are based on the OEEPE test (1990) and on the test carried out by the author with IGN (France) (1989)

[1] *The interpretability of SPOT imagery increases with stereoscopic analysis, and particularly when examined in an analytical plotter, but in turn, it depends largely, among other factors, on the temporal and seasonal separation between the two images.*

Bibliography

Adeniyi Peter O. *Issues in, and strategies for developing and managing resource information in Africa, 1994*

Boursier Patrick *Class notes, Université de Paris Sud, 1986*

Date, I. C. *An introduction to Database Systems*

Montgomeery G.E & Schuch C.H. *GIS Data conversion Handbook, 1993*

Castle G.H., III, Editor *Profiting from a geographic Information System, 1993*

Wellar Barry *Profiting from a Geographic Information System, 1993*

Landis John D. *GIS capabilities, Uses, and Organizational Issues, 1993*

Nino-Fluck O. *Class notes, Universided de America, Bogota, 1990*

Brassel Kurt *Overview of Spatial Information System, 1990*

Uhlenbruck M. & Barteleme N. *Spatial Data: Storage and access, 1990*

Spooner R. *Advantages and problems in the creation of a topologically structured database, 1990*

Nino-Fluck O. *Satellite imagery for African mapping requirements, 1993*

Hutchinson C. & Hassan H.M. (Editors) *Natural Resource & Environmental Information for decision-making, 1992*

Burrough P. A., 1986 *Profiting of Geographic Information Systems for Land Resources Assessment*

Varadan Geeta et al, 1992

Konekny G. *Mapping and Remote Sensing, the Backbone of Environmental Information Systems in Africa, 1991*

Harrison & Dagermond, ESRI *Five tracks to GIS, 199*