Updated: 10 Nov 2003

International Conference
SUSTAINABILITY INDICATORS
Valletta, Malta: 6-8 November 2003

ANALYTICAL TOOLS FOR ENVIRONMENTAL MANAGEMENT GEOGRAPHICAL INFORMATION SYSTEMS

SAVIOUR FORMOSA
Information Resources Manager,
Malta Environment & Planning Authority

Paper

Introduction

The issues pertaining to environmental management can be tackled from different perspectives, ranging from policymaking, through impact assessments to decision making exercises. Each sector can be taken as a research topic in isolation, however the main fulcrum of the process revolves around the data-management processes aimed at setting-up and maintaining information resources structures. Not exclusively anchored to the environmental management disciplines, this process nonetheless finds broad scope in this field particularly due to availability of specialized tools and user-friendly environments.

This paper reviews the data management process, gives an overview of the basic data definitions, looks at the tools available and delves into the spatial scenario. The main topic of discussion revolves around Geographical Information technologies and issues related therein. Data models are investigated with particular reference to the components of spatial analysis and how datasets in any sector can be analysed through a robust querying facility. The final section looks at innovations in the area and posits examples related to the final product, that of data dissemination. Case studies on urban development, coastal inundation and air quality data interpolation are reviewed.

The SI-MO Experience

Managing information resources entails the primary need to know one's data structure. This ambiguous terminology entails the knowledge of the contents of data repositories, ongoing data gathering exercises, the tools needed to process the data and the final dissemination procedure.

Taking the SI-MO (Sustainability Indicators for Malta) exercise as a base case, one concludes that data is available in a variety of formats from text documents to simple tabular data in flat-tables, complex relational databases, through to fully-fledged spatial information systems. Whilst the legal documentation is mostly stored in textual format, socio-economic data is mainly tabular-based whilst data in such areas as the environmental sectors of air, waste, water and biodiversity are stored in spatial format.

This type of data-storage variation is normal in multi-disciplinary projects and reflects experiences within environment and planning organisations where business processes range from policy making to implementation of management projects. This environment creates the potential to review the availability of information systems that would integrate all the datasets into an interactive medium. The technology currently exists where this is possible with value-added results: Geographical Information Systems (GIS) provide the launching pad for the analysis of data from a wide range of sources.

Data Processes

To achieve excellence in data management a series of steps need to be followed in order to achieve reliable and repeatable results. This can only be sustained through the recognition of adequate processes that would enhance the main pillars of data management and the use of adequate tools.

The processes involved in any data gathering exercise as in the case of sustainability indicators can be summarised as follows:

- Data **input** and **verification**

  This initial step looks at the data acquisition phase and includes a package of different technologies such as scanning, digitisation, manual data capture and use of AI to capture such data from remotely sensed imagery. Data acquisition also requires users to verify sources, remove errors, and carrying out essential quality control exercises.

- Data **storage** and database **management**

  Essentially concerned with hardware constraints and the need for more storage space, this step sees users going beyond the physical issues and looks at ways to store data in reliable and easily accessible formats. This major process involves the building up of entities such as are datasets where users can access data in a variety of forms and designs. However, care needs to be taken to ensure that the correct structures are used. Data models, database management systems (DBMS) and metadata issues are tools enabling structural integration. Data managers need to decide whether to use such systems based on a set of Rule's such as those designed by Codd in the 1980s, with systems based on a relational model (RDBMS), object-oriented (OODBMS) or object-relational (ORDBMS).

  The main advantage with current databases is the ability to access attributes within different databases situated in remote sites, facilitating the access to data across networks such as the global Internet.

  Another issue related to this step relates to metadata where information on the contents of datasets is made available with information on users, creators, time stamps, attributes list, projections and other related information.

- Data **analysis** and **modelling**

  Data models help to create a system of information processes such as layering, cross-dataset linkages and integration of internal and external datasets. Analysis can take the form of querying functions through languages supporting the data. Structured Query language (SQL) is a useful tool, though even more rudimentary tools such as functions within Excel and other base software can help achieve good results. Other software such as Statistical Package for the Social Sciences (SPSS) help in the analysis of socio-economic data, though SQL and spatial options within SQL. These tools aid environmental scientists to carry out multi-dimensional analysis: such as in the case of Spatio-temporal analysis of habitat change, inundation predication, urban sprawl, and flood prediction. Modelling is important at this stage as it aids the researcher to build up a functional model that could also be dynamic and deliver automated analysis for eventual report development.

Platform-independent modern tools are aiding the global-spread of small but powerful technologies using the Internet medium. Java applets, java scripts, and the potential of emerging technologies as *.net* have proven that small applications can give real-time results. Tools can be programmed to access local clients or remote servers that carry out analytical functions.

- Data **display** and **outputs**

  The final aspect of the data process concerns the issue of data display and output. This can take a variety of forms: histograms, tables, maps, amongst others. On the dissemination side there are a variety of technologies that help users to publish their research results, ranging from on-line html reports to dynamic web-mapping services.

Tools Availability

Tools necessary for the development of data services can span a variety of methodologies that are made available on a two-pronged scenario: Push **-** Pull and Pull – Push. Entities who were pushed to develop their systems based on the availability of software and provision of base maps were experiencing a **push-pull** though this caused problems though to non-functional use of the same software. Users of environmental data had to cope by creating their data around the software itself, which was generally broader than their requirements and may not cover their specific needs; a case in point is dispersion modelling in air quality analysis where GIS analysts had to use interpolation models that were too generic to deliver excellent results. This trend has seen major developments and currently it is more a situation of **pull-push** where private organisations finally broke through by creating their own data and then going for the software that would aid their work. This is becoming more the norm as software packages are becoming specialised or else take the form of componentware and users can acquire modules that fit their needs.

Tools can be said to fall within these following categories:

- **Manual**

  Applications that do not have access to any automated functions or else users do not need the functions. It also includes those exercises where data is inputted manually and digitisation processes.

- **Semi-Automated**

  Applications that aid users to acquire data such as software used in remote sensing projects, automated querying functions, digitisation, etc. This category includes such software such as object-oriented programs that scan imagery and extract land-cover categories, as an example.

- **Automated with limited access**

  Fully automated applications that use dynamic links to store and disseminate data. Use of these tools is restricted within an organisation through intranet access. Projects may include air quality data capture modules that store real-time data, store it in a database and generate graphs or maps in an intranet website.

- **Automated and Distributed**

  Applications that serve users across the globe through a series of client and server hardware collectively known as the Internet or other network that have developed through access to protocols such as HTTP, ASP and FTP. Projects include the distribution of real-time geographic data, planning permitting schemes, through the use of mapservers and other on-line programs.

The Case for GIS

During recent decades a new range of applications became more common and have gained widespread use, particularly in the case of Geographical Information Systems (GIS). This technology, also known as spatial information systems and geomatics has come at an opportune time to aid data management and though originally developed for the physical sciences has been expanded to include planning, economic and social sciences.

GIS has come a long way from the first attempts at manual analysis of spatial patterns. These include the good use of spatial analysis in 1854 by Dr. John Snow who identified the source of a cholera outbreak that centered around the use of a contaminated water pump in Soho London. At the other end of spatial analysis use, a case of abuse was experienced by Massachusetts in 1812 when Governor Elbridge Gerry changed the boundaries of his districts to enforce his opponents' voters into a smaller area resulting in a salamander shape, thus the name *gerrymandering*, a term used to signify corrupt election purposes.

GIS has since developed into a series of tools, both inter-related and proprietary through of an amalgamation of four main systems. It can be stated that computer-aided design, computer cartography, remote-sensing and database management tools have helped to further the development of GIS as it has elements of all four.

The following section lists the different aspects of each technology:

Computer-Aided Design (CAD)

- CAD automated manual drafting techniques
- Design and automatic drafting of new objects and not to reflect reality
- Creation of a "blueprint" of a proposed new world
- CAD use is also made of arcs and polygons with few vertices but no fractal techniques
- CAD has rudimentary links to databases
- CAD works with precise geometric techniques and needs to be perfect – GIS techniques allow correction by means of software rectification processes, a process which is not acceptable to CAD tools
- Modern CAD taking on board some of the functions of other tools such as GIS
- Vendors have bundled packages together to try and sell CAD systems as GIS systems
- "Darwin vs Creationists" dilemma - just added further functions to an old or rebuild the whole software

Computer Cartography

- The main strength of these systems is the display of maps and high quality vector format outputs - single focus in the discipline is the map
- Lacks attribute functions
- Limited linkage to database management systems
- Computer cartography must be seen as a means to an end rather than as an end in itself

Remote Sensing Information Systems

- Tools designed to collect, store, manipulate and display raster data
- Production of images taken from a sensor on a satellite or a plane
- Allows the enhancement and classification of data
- Cannot handle vector data and limited in handling attribute data

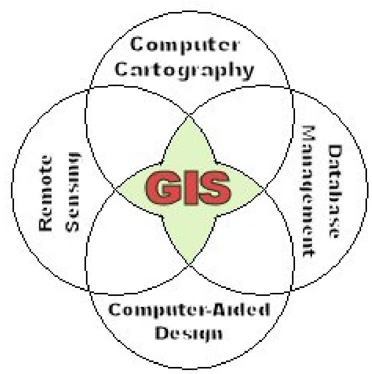
- Difference between raster imagery and rectified orthoimagery

Database Management

- Databases are mainly systems optimized for storing and retrieving non-graphic attribute data
- There are different types of DBMSs – RDBMS, OODMBS, ORDBMS
- Use of centralized control, data sharing, control of redundancy of data, direct user access and user views
- Databases are composed of records, fields and keys
- Limited in retrieving graphical data and in anything other than simple analytical functions
- Allows use of Version Managed Data Storage (VMDS)

Geographical Information Systems

GIS caters for most of these functions and has major added value through the use of spatial querying functions. The inter-relationships of the different types of software is a major attribute to the success of GIS, since it can work together with other systems apart from having its own unique characteristics that are not available in the other packages.



The main issue in GIS relates to the locational aspect of data. Since relatively few real-life cases exist where data does not have a geographical (spatial) component, the availability of tools catering for this field has grown with extensive development. The fields are vast and include: analysis of demographic trends, installation of a new client for utility services, tracking of vehicles and air emissions, real-time on-line maintenance, trouble hotspot localization, crime analysis and rerouting of services, and development planning services.

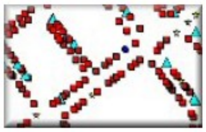
GIS allows data managers to depict their data in spatial format that is essentially based on the following categories: **Position**: (l*ocation),* **Topology** (study of the properties of a geometric figure which are *not dependent on position* – connectivity, containment and adjacency of features), and **Attributes (**the *data attached to a spatial construct*). Each has its own important function particularly position and topology whilst the third aspect, that of attributes, is primarily what differentiates GIS from cartographic tools and also aids its linkages to other databases in any format it is stored in.

GIS have the ability to generates maps, stores and manipulates map data as a model of the real world, produces new information by combining existing databases and outputs results in different formats such as maps, databases, 3D models and VRML outputs.

Through the use of data models such as Vector (points, lines, areas encoded in **xy** coordinates) and Raster (grid cells) GIS manages to depict all spatial analysis that help reflect the real world. Each has its own structure and data storage issues such as huge sizes for raster as against smaller size and more topology in vector formats.

Spatial Data Components

Data in GIS is stored in special entities called spatial data components. These are basically composed of three types of entities: points, lines and polygons (areas) with the another two types of these entities called networks and surfaces being a combination of lines and polygons.



Points: points are single locations in two or three-dimensional space based on an **xy** coordinate structure.
Example: The dot representing a tree on a map.

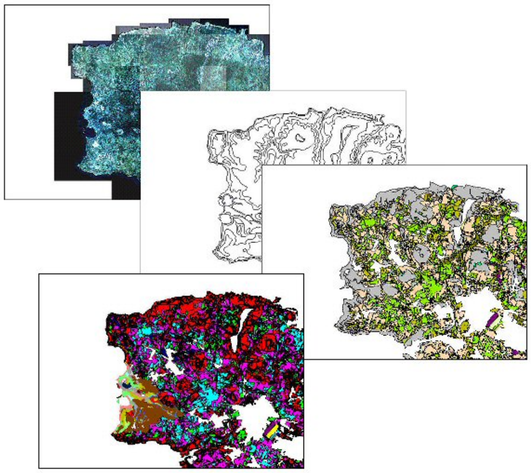
Lines: lines can be isolated, within a tree-structure, or elements of a network structure.
Example: River or road systems

Polygons: polygons can be isolated, adjacent, or nested.
Example: State boundaries or contour lines in a map.

Layering and Querying Functions

GIS operates within a layering environment that allows users to analyse data spread across a number of datasets in a comparative environment to analogue transparency overlays. In such a scenario, GIS helps allows users to view data right through the complete workspace and query the different datasets across the whole workspace such as an analysis of Landcover identification based on topography, agriculture and habitats. Through the use of such tools both attribute and spatial queries can occur within the same data request, facilitating the development of thematic imagery and other data outputs to fit the need of the researcher.

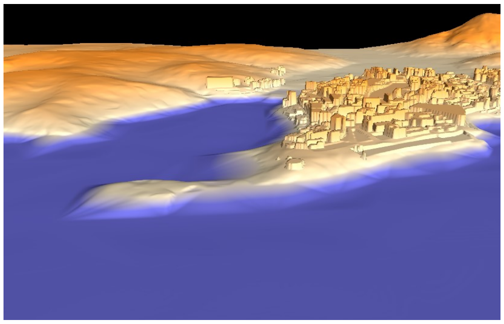
Another tool function within GIS is the use of buffering techniques that allow users to query the spatial parameters and attributes falling within a specified boundary around a site being analysed. This is especially important for such analysis as environmental impact assessments, contaminated sites, and areas of ecological importance. Buffers are created around areas of interest and querying functions acquire the information contained within that buffer. This type of querying together with the spatial parameters to spatial SQL commands (as are within, contain, intersect) distinguishes a spatial query as against a tabular query.



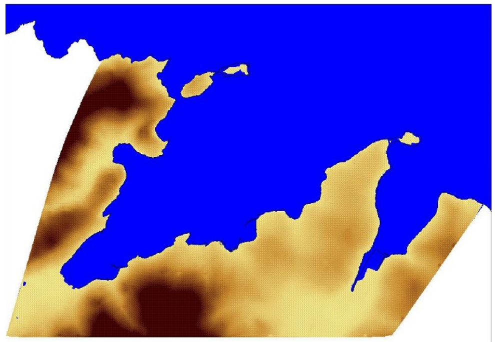
Case Examples of GIS use in spatial and environmental mapping

The following examples depict 3 case studies on the utility of GIS in land-use mapping, change mapping and interpolation techniques. In all three cases MapInfo and Vertical Mapper software were used.

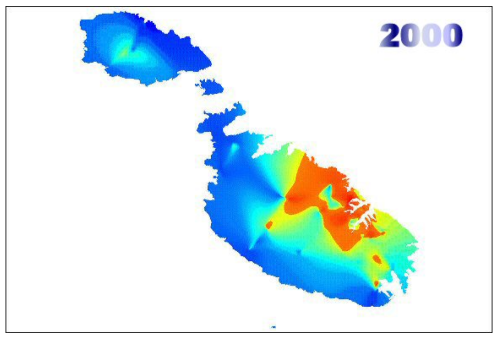
Case I: **Urban Land-**Use - Overlaying and Thematic mapping: This land-use planning project aimed to verify the availability of vacant land in the town of St. Paul's Bay in Malta as well as to analyse the potential of this technology for enforcement purposes. Buildings that exceeded the height limitations were identified through the use of z-heights attributes that are interpolated into a 3d grid file with drape overlays. Colour schema help identify the higher buildings and perspective views aid visualisation.



Case II: **Inundation** Analysis - 3d Change Mapping: The use of topographic data such as height maps helped generate as grid file that allowed further analysis for inundation mapping. Thematic maps based an increase in the z-value through a colour schema helped generate maps depicted sea-level rise at current 0m, 0.5 and 1m levels. The main change-mapping tool used was Vertical Mapper where parameters were changed to simulate the changing sea level. Such a study assumes that no geological shifts would occur in which case more spatial parameters would have to be interpolated.



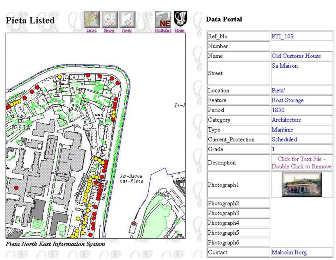
Case III: **Interpolation** Techniques - Air Quality mapping: Though ideally carried out through the use of a specialised dispersion analysis tool, this project employed Nearest Neighbour Analysis (NNA) interpolation techniques to analyse Benzene monitoring. Data sources from diffusion-tubes spread around the Maltese Islands indicate change in the levels of benzene in air at quarter-hourly averages and data is aggregated in hourly, daily, monthly and annual datasets. Interpolation helps to depict a particular time capture data or an average annual aggregation as depicted in the image. Sequential image outputs enables researchers to analyse change at all data capture levels where further multimedia techniques aid the presentation of results. The use of animation technologies helps to visualise change and aid in the identification of sites requiring monitoring and policy regulation and implementation.

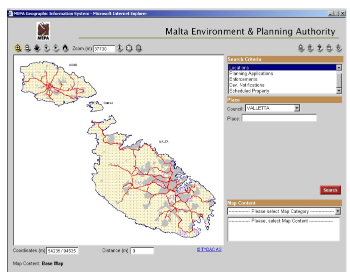


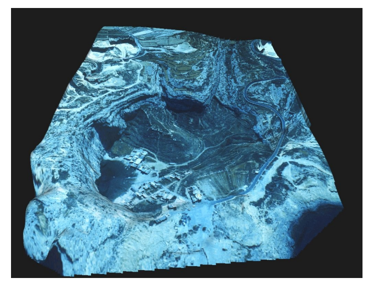
Future developments within the technologies

The main future developments envisaged in the spatial areas include development around web technologies such as that employed by the Malta Environment & Planning Authority (MEPA) technologies. The site (www.mepa.org.mt) has a number of on-line information systems such as an EcoExplore (database of animals, plants and habitats), National Protective Inventory heritage data (ImageMapping) and Mapserver (full web-GIS). Future developments would include querying functions and the integration of environmental geosets. The use of Open-GIS technologies though XML, GML and VRML is being explored.

The main aim of these technologies is to present users viewing the data from remote sites in an organisational intranet or across the internet to view, query and output the data needed for their requirement. A number of technologies exist to help distribute the data. These include simple textual and tabular dissemination, through an imagemap that helps users to view information should a mouse be pointed at specific areas to the state-of-the art technology as webmapping were users can interact with spatial data and query the information. Examples of interactive webmappers can be viewed at the following sites: MEPA mapserver and IG GIS (http://www.geoportal.ch/Internet_2.htm).







ImageMapping – This technology gives users the possibility to view data that does not need a server to generate data requests. The systems work on hotspot maps generation that are linked to data layers. Each data layer would however need manual updating on each change. Users can access data with a common web browser and save the results to their hardware. Users may not be able to query data.

Mapserver – The state-of the art technology is bases on a fat-server thin-client approach. Users query the interface which send the request to a remote server that does all the work and send back small packets of images/data for client consumption. Advanced querying tools guide users to view large scale spatial entities such as building parcels, habitat areas as well as small scale maps such as national orthophotos.

VRML development – Technology that will be further developed over a number of years to help enable users to interact in 3D real-time with their datasets. Examples of this technology could help users of predictive analysis to build watercourse modelling systems, flood mapping and inundation. Free plugins such as Cartona aid the visualisation aspect of these data formats. (http://www.parallelgraphics.com/products/cortona/)

Concluding points

Tools employed in the processes of data capture, storage, analysis and output provide a basis for the management and dissemination of information. Whilst textual and tabular tools provide users with a basis for analytical approaches, spatial tools bring in a new dimension to the research domain. Outputs such as change-mapping, 3D mapping and virtual reality help policy makers with the potential effects of their decision-making as well as the general public through the user-friendly dissemination technologies associated with spatial tools.

GIS has helped to bring together a large number of disciplines and allowed for the analysis of impacts on vulnerable locations through its integration of various datasets and structures. The use of such related technologies as digitization, scanning, remote sensing, querying functions, mapping output, web mapping and distributed access has helped analysts to enhance their research in both social and physical sciences research.

In summary, there are a number of benefits pertaining to the spatial analysis tools:

- The integration of a number of technologies such as cartography, remote sensing, computer-aided design (CAD) and database management
- The ease of use through helpful windows-based environments giving access to visual interpretation
- The development of spatial data models allowing better data storage and pushing forward the analysis of different data layers and datasets
- The integration of multi-sourced spatial and non-spatial data across one project that was primarily difficult to access within other technologies
- The introduction of manual, visual and automated change-pattern analysis that enables researchers to monitor change
- The integration of other data outputs such as population density, land-use monitoring, and impact analysis into a simple map form
- A querying function that looks beyond the *what* but more importantly at the *where*
- The dissemination issue being enabled by tools such as Web-mapping that evolved into a state-of-the-art tool giving any user with any specific or generic request the ability to query the data through a system of clients and servers

Definitions

Data: data are information coded and structured for subsequent processing, generally by a computer system (British Computer Society, 1989)

Information: information is the meaning given to data by the way in which it is interpreted (British Computer Society, 1989)

Geographic Information: information which can be related to specific locations on the earth (UK Department of the Environment, 1987)

Spatial Referencing: the means by which information can be related to a specific position or location (Shand and Moore, 1989)

MetaData: Data about Data: a list of datasets in a repository

References

Harder Christian, (1998), Serving Maps on the Internet: Geographic Information on the World Wide Web, Environmental Systems Research Institute (ESRI)

Grimshaw D.J., (1996), Brining Geographic Information Systems into Business, GeoInformation International, Cambridge

Longley P., Goodchild M., Maguire D., and Rhind D., (2001), Geographical Information Systems and Science, John Wiley and Sons, Chicester West Sussex

Tanier Richard, (1997), The Intranet Data Warehouse: Tools and Techniques for Connecting Data Warehouses to Intranets, John Wiley and Sons, Incorporated

Skidmore A., (2002), Environmental modelling with GIS and remote sensing, Taylor and Francis, London

Worboys M.F., (1995), GIS A Computing Perspective Taylor and Francis Ltd., London, UK

Sites (accessed November 2003)**:**

Cartona: http://www.parallelgraphics.com/products/cortona/

ESRI: www.esri.com

IG GIS: http://www.geoportal.ch/Internet_2.htm

ITworld.com, (5th July 2001), Data Management Strategies – Codd's 12 Rules
http://www.itworld.com/nl/db_mgr/05072001/

MapInfo Corporation: www.mapinfo.com

Malta Environment & Planning Authority (MEPA): www.mepa.org.mt

MetaData and WWW Mapping Home Page: www.blm.gov/gis/html