GIS project management

A case study: A GIS implementation in South Africa

Rosan van Wilgenburg,

2004

Submitted in fulfilment of the requirements for the degree of Master of Science in
Geographical Information Systems (UNIGIS)
Faculty of Earth and Life Sciences
Vrije Universiteit Amsterdam

Contents

1 Introduction

The introduction, use and management of GIS entail a number of specific management issues. The various stages of system development require careful planning and management. Good project design and management are essential to produce a useful and effective GIS application (Heywood et al, 2001).

In this paper the theory of GIS project management will be investigated. Several project management theories will be described.
An extensive description of a project carried out in South Africa will serve to examine the theory. The case study will be used to showhow a project can be managed , by practical experience. Before starting the project not much was known about the project, so no particular framework was followed. After the project the theoretical framework was written and it was investigated how it fitted in into the theory, The case study is derived from a project that was carried out for the Eastern Cape Manufacturing Advisory Centre (ECMAC), an organization that asked for the introduction of a GIS. Later in this paper the organisation structure of ECMAC is going to be explained.

This paper will answer the question:
How does the ECMAC project fit into the theoretical framework of project management?

2 Project management

This chapter describes critical success factors for an implementation and summarizes different visions on how a project should be undertaken in order for it to become a success.

The introduction of GIS in an organization often starts with one enthusiast that sees possibilities for GIS.
The introduction, use and management of GIS entail a number of specific management issues. The various stages of system development require careful planning and management. Good project design and management are essential to produce a useful and effective GIS application (Heywood et al, 2001).

When an introduction of GIS isn't managed properly there is a big chance it fails.

2.1 Implementation strategies

This paragraph summarises some implementation strategies as an introduction to project management methodologies.

Tai On Chan that GIS researchers have suggested many GIS implementation strategies.

Nolan has a general theory about stages that an IT growth cycle goes through from introduction to a mature enterprise system.

Somers (1994) suggests a dual-track development strategy. After a preliminary requirement analysis and core design, the strategy follows two tracks simultaneously. By the first track, one collects short-term data and implements immediate applications, while by the second track; one continues the detailed analysis and design.

Peuquet and Bacastow (1991) suggest an iterative prototyping strategy. After defining a preliminary set of requirements, a working model is developed and implemented, provoking experimental organizational changes. Both the functional and the organizational requirements are derived and tested through iterations.

Anderson suggests a proactive approach in which GIS is implemented in five phases, namely, participation, context evaluation, vision creation, change and implementation. The five phases are non-linear and proceed concurrently, and if necessary repeatedly.

A project management methodology depends on the implementation strategy used. Some project management methodology are based on IT project management, while others are derived from general management theories.

The implementation of a GIS is often preceded by a business case. A business case for investment in a GIS serves as a transition from a potential project to an actual (approved) project. The making of a business case can be seen as a project in itself.

2.2 The need for a Business Case

A successful introduction of GIS in an organization depends on understanding the decision-making processes that are part of a business strategy, but also on understanding the role that geographic information can play in the wider information strategy of an organization. Understanding this combination is a necessity for a sound business case for investment. (Grimshaw, 1999)

GIS used to be technology and application driven. In the eighties it was moreover looked at as an information system. GIS is used to turn data into information with meaning. The meaning depends on the context, or the decision for which the information is required. Information systems rarely fail because of technical factors: rather they fail for organizational reasons (Grimshaw, 1988).

A business case serves as a document on which the decision to invest in the implementation of a GIS is based. The approval process involves assessing cost, value and risk. An important aspect in this process is to create managerial consensus. A broad backing from the stakeholders (beneficiaries) is essential in the implementation phase of the project to ensure that resources are allocated to the GIS project.

The process of making a business case for investment in a GIS is often seen as a bureaucratic hassle, but it is a useful exercise to maximize the benefits of a GIS. A thorough understanding of the business requirements and the opportunities offered by GIS technology, gains also insight in the best use of resources.

A business case is often used to justify the technology to the top management. An important part of a business case is traditionally a cost-benefit analysis. A danger of putting too much emphasize on cost reduction is that managers tend to exaggerate the benefits in order to get expenditure approval (Ward, Taylor, and Bond, 1996). Changing the focus from cost to value added is needed to gain maximum value from a GIS (Grimshaw, 1999). The need for a business case depends on the size of a project. The bigger, the more usefull a business case is.

A business case can be part of the methodologies as described below.

2.3 Incremental approach

K. E. McDaniel and B. E. Sherman (AM/FM Conference, 1992) propose the incremental approach. This methodology has a general character and can by employed on small and bigger GIS projects. It uses a phased approach in order to reach the final goal: the implementation of a GIS.

The implementation strategy for a typical GIS project follows the following steps:

1. Project evaluation, definition and plan
2. Analyse, feasibility, select system
3. Customize and test system
4. Pilot, conversion and operation
5. Implement System, conversion and operation

Ad 1) Project Evaluation is the process of gathering information on the technology available in the market. This information helps to identify potential benefits and application for GIS within the organization. The outcome of the project evaluation is the beginning of a project plan.

Once the preliminary study has been completed the next stage is to carry out a functional evaluation of the current manual processes. The current workflow and the specific problems that occur should be analysed in order to create a Functional Requirements Document (FRD). This document defines the overall project scope and it can change again during the course of the project. The FRD will serve as a basis for cost benefit analyses and to indicate the feasibility of achieving the project goals.

Ad 2) System selection is a timely process that should be undertaken with care. It is best done through a benchmark study in which the required functionality is tested with real data. Also aspects like security, required hardware and speed of the system have to be determined. This outcome of this study results in a Go or No Go or a revalidation of the project goals. This document can be seen as a business case.

The detail design document defines the database graphics specifications interfaces, calculations and any specific GIS applications. Interfaces with other current systems should be described and a database has to be designed. The application has most likely involves custom-made aspects as every organization is different and has different requirements.
At this stage the acceptance criteria and the delivery schedule have to be determined.

A prototype test serves to determine how the actual performance of the system compares with the requirements as defined in the project scope. The prototype test should take place with a data set the same size as the data sets that are going to be used. The documented test results can be used to correct the system before the pilot test starts.

The pilot is the last stage before the implementation. In the pilot the daily operations of the system, like response time and connectivity to other systems are evaluated. The pilot is the last decision point in the technical and financial analysis. If the pilot is successful a written approval leads to the implementation of the project.

The method described above is efficient and offers clear goals and objective as also guidelines for the GIS project. Many GIS projects have failed due to the missing of stated objectives and the missing of a defined project scope.

An incremental approach offers the advantage of reduced risks for an organization, while with a total data conversion approach more sophisticated systems can be obtained.

2.4 Systems life cycle and Prototyping Project Management Methodologies

There are two approaches that are often used to manage GIS projects: systems life cycle (SLC) approach and prototyping approach. The SLC approach is also being called the waterfall model as the outputs of the first phase of the project are being used as input for the following phase and so on.

The SLC approach starts with a feasibility study, followed by a system investigation and analyses, which leads to a system design. During the last phase, the implementation, review and maintenance phase the users get the first opportunity to comment on the application. One of the advantages of this system is that it provides a structured framework for the management of a GIS project. One of the many disadvantages is that it does not put the user at the centre of the design and therefore there is no possibility to change the scope and character of the problem.

The prototyping approach enables continuous feedback from the user to the developer. First the basic requirements are defined on which the first prototype is based. Then the users experiment with the system and sharpen their first requirements. The main advantage to this system is the involvement of users, because of which the chance of a successful implementation increases. The main disadvantage is that time management is difficult, because clear objectives are not being set. This can also been seen as something positive as it can bee argued that few GIS systems are ever finished.

The incremental approach discussed before suggests it is a prototyping approach, because it makes use of a prototype. It is, however not stated whether or not users are involved in testing the prototype. When the prototype test only considers technical aspects it can be a SLC approach too.

Regardless the methodology followed some common implementation problems can occur:

- data is in the wrong format for the GIS software
- lack of GIS knowledge imposing technical and conceptual constraints
- users of the GIS chance their mind about what they want the GIS to do

2.5 Summary

The aim of project that involves the introduction of (new) GIS technology in an organization is a successful implementation in which all the criticall succes factors are met. Several different implementation strategies exist. A business case can be used to convince the top management of the need for a GIS, but serves in the mean time to gain a better understanding in the business process and the role geographic information can play.

A project management methodology depends on the implementation strategy used. Systems life cycle approach and prototyping approach are 2 of them. Different approaches have been described shortly.

This chapter served to put the case study in perspective.

3 Case study

This chapter describes a project that was carried out for the Eastern Cape Manufacturing Advisory Centre (ECMAC) from November 2003 till February 2004 by the author of this paper. It starts with explaining what ECMAC is and how it relates to NAMAC, the mother organisation, which is followed by the a description of the course of the project. Some of the original correspondence is included in the appendices.

3.1 ECMAC

Eastern Cape, Manufacturing Advisory Centre (ECMAC) is, together with 17 other centres, part of the National Manufacturing Advisory Centre (NAMAC) Programme. NAMAC is a Small, Medium and Micro Enterprise support agency within the DTI group (the South African Department of Trade and Industry), and is widely recognised as one of the most successful SMME development and support agencies in South Africa.

NAMAC has developed an extensive structure across South Africa that serves as a channel for the application of new tools, information, products and projects, thus enabling the effective delivery of solutions aimed at Small, Medium and Micro Enterprise. NAMAC can be compared to the chamber of commerce in the Netherlands but with the difference that it focuses on manufacturing and on Enterprises with less than 100 employees.

NAMAC's programmes are aligned with the following national priorities:

- SMME development
- Black economic empowerment
- Job creation
- Competitiveness
- Geographic outreach (rural focus)

As one of two Pilot MAC's in South Africa, ECMAC (originally PERMAC) was established in December of 1996, when the PE Manufacturing Advisory Centre Task Team was established. (www.namac.co.za)

3.2 The course of the project

In October 2003 ECMAC heard about the possibilities of GIS and hired me as a GIS consultant through the Port Elizabeth Technikon and the international UNIGIS programme. The introduction meeting took place on 19 November 2003 (see appendix 1 for the full meeting report):

At that moment NAMAC was looking for a centralized information system. That time a proposal wass being written to integrate all 18 databases into one central database that is maintained at one place and that was planned to be available over the web.

As part of the information system ECMAC would like to have a GIS solution that meets the same demands: available over the web, maintained at one place. The idea was that end users can query the database and have the results shown on a map. Combining this with census and other socio-economic or physical data would be interesting. Wayne (the director of the ECMAC programme, mentioned the GIS application of the Agricultural Research Council as an example.

However it was clear that to implement an application like this will take much more time than the three and a half months that Rosan has available.

A new objective was set: to write a proposal containing the following information:

- What will be possible,
- What problems might be encountered,
- How much time it will take,
- An indication of the costs involved.

The proposal will possibly come with maps in jpeg format that can be shown on the Internet or can be used in a sample application that supports the proposal.

In order to get a feel for the data and possibilities it was decided that Rosan is going to set up a desktop GIS environment. ECMAC needs to define what kind of information has to be available in the GIS system. Grant (the person responsible for the databases at ECMAC) is going to look at this.

After ascertaining the data and functional requirements, it was decided to have a look at the software. The development of the central MS Access database can continue in the mean time.

A meeting on 21 November followed this meeting. By that time it was all ready expected that geocoding at street level or house number level is not easily possible in South Africa.

The database that was used when Namac first started was still being used. The database developed as the programme developed. The offices are all using the same database structure. The database is managed locally and only the reports are aggregated to a national level.

Recently most regions in South Africa are covered so growth has stabilized. Therefore a final centralized version of the database was being designed at that moment. The database was planned to be operational at the first of April.

The development of the database will continue regardless of the GIS. The GIS will be designed based on a copy of the database. The two systems will be merged at a later stage.

The database consists of two types of clients:

- inactive contacted clients
- active clients

The inactive clients only have contact details, the active client can have a whole history of projects attached. There are 5 types of projects.

ECMAC indicated that it will be very interesting for them to see the geographical spread of the project types and of the size of the company.

The expectations of the GIS are:

- On a national level monthly reporting per office. (Office regions are defined by magistrate areas, for example, the East Cape consists of 4 regions)
- Per office, to start with ECMAC, an information system in which SMME information is shown per suburb.

The GIS has to have a dynamic link to the database so it is updated automatically. It is going to be a desktop application at first, in later instance it can be migrated into an Internet GIS. The data does not have to be combined with socio-economic or other data.

Priority will be given to the ECMAC information system.

- It was decided that Rosan will start setting up the system with a copy of the most recent SMME table .

It became clear that the table was insufficient for detailed use in a GIS,

In a following meeting (5-12-03) with Grant, Ryno an Rosan it was decided that for now it is sufficient to start designing a GIS with the three fields available. Ryno is responsible for the database design and will implement the desired fields. In a later stage, when all data is available, the other preferred fields can be added.

Rosan was given the full ECMAC database, in order to obtain the necessary data. Rosan also got an example of the MAC reports in Excel, showing statistics per province. This report can be used as a basis for designing a GIS for monthly reporting per office on national level.

During the course of the benchmark study (Appendix 3, Benchmark study) it became clear that Microsoft Maps for Excel (Appendix 4, Microsoft Maps for Excel) is the best solution for the above mentioned reporting functionality.

After preparation of the attribute data in Microsoft Access and manipulation of the map data in MapInfo Professional, Rosan was able to show the possibilities with the existing data.
In a meeting before the summer holiday (15-12-03), Rosan showed results so far and some new wishes regarding the visualisation of data were defined.

After the summer holiday Rosan showed results to Wayne, Ryno and Grant. A discussion followed on what to do next. At ECMAC/NAMAC there is a good understanding now of what is and what isn't possible in a GIS in combination with the current database.

Before new database fields are going to be defined in the new database ECMAC and NAMAC should decide what they precisely want to do with a GIS. Possibilities are:

- To use a desktop GIS to analyse data on office level. First starting with ECMAC and in a later instance it can be extended it to all offices,
- To use a desktop GIS at NAMAC for monthly reporting and monitoring,
- To use a GIS (possible via the Internet) to sell data to other parties interested.

All agreed that the quality of the data should be up to date and correct before it can be used. Next step for Rosan is to prepare a presentation on how GIS can be used for NAMAC/ECMAC, using as much real data as possible. This presentation is going to be used to show the stakeholders the profitability of GIS. Within the presentation hyperlinks will be used so that one can migrate form a national reporting level, to a "micro level" displaying raw material inputs, customer linkages and a hybrid model whereby a number of layers can determine the best place for a joint venture, given a set of criteria, much like the example Wayne used for the CSIR model for determine the best place for farming certain crop.

The presentation is made in Microsoft PowerPoint combined with movies in Flash in order for it to be self-explaining. The presentation was burned on a CD-Rom and handed to ECMAC at February 2, 2004.

Rosan proposed to set up a user environment and to prepare a training in Geomedia Professional while waiting for further GIS needs to be discussed.

The project ended with a final meeting on 25 February 2004 in which a 2 CD-ROMs were handed over. The first contains a 1 year evaluation copy of Geomedia Porfessional, the second contains:

- All correspondence, including the appendices of this paper,
- All tables in Mapinfo and Shape format tables,
- A database in Microsoft Access with tables and queries used,
- Maps prepared for Microsoft Maps,
- Installation guide for Microsoft Maps written for ECMAC,
- An environment, geoworkspace, in Geomedia Professional using ECMAC data,
- A manual/course to set up an environment in Geomedia Professional

3.3 Data related issues

A further investigation to the way data geographical regions are defined in South Africa was necessary. In Appendix 2, Data, the results of this investigation together with specific difficulties with the ECMAC data is explained.

While Rosan prepared the data, ECMAC provided her with the following lists of fields that they want to have shown in a GIS.

Here are some of the preferred fields that need to be included in the GIS database.
This information would typically reflect Y/Y (year on year) results and trends.

- Industry type / sector
- Technology / machinery type (hardware)
- ICT (software)
- No of employees (Direct and indirect)
- Skills / qualifications of staff
- Primary raw material (plastic, wood, metal etc.)
- Product type (corporate clothing, furniture, more specific definition than sector type)

- Training (manhours)
- Average capacity utilization
- Customer location/s
- Mode of distribution (air, road, rail, sea)
- Warehouse space utilization
- Number of project interventions
- Type of project interventions
- Annual sales / turnover
- Annual raw material usage / expenses
- Raw material source/s
- Gross value added (depreciation, NPBT and wages and salary bill)
- Annual scrap value tonnage (wooden off-cuts etc, sawdust etc.)
- Annual waste tonnage (through poor workmanship)

Only three of these fields exist in the current database: Industry type/sector, number of employees and annual turnover. All other fields have to be included in the database design that is currently developed. Thought has to be given to the following:

Since the data of all the SMMEs will be combined at a higher level, that of suburbs, it will not be possible to aggregate all data fields. For example, the count of the total number of employees per suburb will not be a problem. But it will not be possible to give an average or a count of raw material sources per suburb.

The field size and type is not clear for every field. For most data field a numeric field will be sufficient and for some it will be a fixed entry ("Mode of distribution" can be either: air, road, rail or sea). But for others, like "Customer locations", it is not clear whether it should be a text field that can be filled in freely or that it also is a fixed list. This is also related to the question, how to present it in GIS?

4 Discussion and conclusion

4.1 Evaluation

Before the project started ECMAC/NAMACs expectations for GIS in their organisation where high; they were thinking of a complex Internet GIS application. ECMAC hired a consultant for 3 months to advice them on GIS and to possibly implement a GIS application.

During the introduction meeting it all ready became clear that there would not be enough time to do more than to write a proposal containing the following information:

- What will be possible,
- What problems might be encountered,
- How much time it will take,
- An indication of the costs involved.

Instead of writing a proposal for a complete enterprise GIS, it was decided to start smaller by setting the following short term objectives:

- A desktop solution for monthly reporting per office on a national level,
- A desktop information system in which SMME information is shown per suburb.

The underlying aim of these objectives was twofold:

- ECMAC/NAMAC would get more insight in the possibilities of GIS, and would so be better able to define their wishes,
- The consultant would get more insight in the ECMAC/NAMAC data.

Decreasing the project scope by breaking the project down in smaller objectives, in this case to start with setting up a desktop GIS, does not mean that the incremental project approach was used, although the name would suggest that. The incremental approach as described in paragraph 2.4 is called incremental because it splits a project in different functional steps, and not in different objectives.

The short term objectives can be seen as a project in itself, apart from the long term aim: the introduction of Internet GIS.

During the course of the project the underlying aims were met. In various meetings in which temporary results were discussed, and in correspondence in which functionality was defined by ECMAC/NAMAC, both parties were forced to think about GIS and the possibilities it can offer in practise to ECMAC/NAMAC. Also shortcomings in the ECMAC/NAMAC data and in the desired GIS functionality became clear. GIS at ECMAC/NAMAC evolved from a abstract idea to a more realistic view.

The objectives itself are met, but a full implementation, including a pilot or a prototype never took place. Software and data were handed over on CD-Rom, including an installation guide.

One of the reasons for the fact that a implementation never really took place was that the new NAMAC database, which will have the desired fields to base a GIS on is not yet ready. At the end of February 2004 it was still in the design phase. Expected implementation is April 2004.

An other reason is that the underlying aim, to gain insight in the possibilities of GIS in the ECMAC/NAMAC organisation, was met during the course of the project. This aim is more important at the end than the implementation of a desktop GIS.

The main conclusions/insights is that the current database of ECMAC/NAMAC is unsuitable for interesting GIS analysis. This is because the geocoding of the addresses of the SMMEs involves a lot of handwork and because there are not enough variables available to do meaningful analyses with. Thought has to be given to which fields have to be added in the new database. Also, before data can be sold to other interested parties, data should be correct and up to date, which is doubtful at the moment.

A conclusion that is left unmentioned, or unsaid, is that there might not be enough reason to make an investment in GIS profitable. But that can be an outcome of a business case. The slideshow, produced to illustrate the profitability of GIS to stakeholders, may be presented at a later stage, when a good idea for a GIS implementation arises, that would justify the investment.

4.2 Project management

In this paragraph the project is being compared with the theory of project management as described in paragraph 2.5: Systems life cycle and Prototyping Project Management Methodologies.

In case of the ECMAC project, a rich picture has never been drawn; instead we immediately used the root definition approach. This was merely due to the fact that only one party's perspective (that of ECMAC) was involved. Two meetings with people involved resulted on written project objectives. (see meeting reports in the case study)

ECMAC's first objective was to develop a complex Internet GIS in which their clients are able to upload and download data, and in which analysis related to logistical issues can be undertaken. Their clients should for example be able to combine logistical means in order to reduce costs.
As the consultant was the only resource available for a time period of only three months and because there was a lot on certainty about data quality and required functionality in general, it was decided soon that this objective could not be met and it was agreed to break up the project into smaller parts. During the time span of 3 months a desktop GIS environment would be created, in order to give the people of ECMAC a feel of what is possible in GIS with the data they have. The development of a complex Internet GIS was extracted from the first objective and brought outside the scope of this project.

The data used in the project consists of the attribute data available in the ECMAC database and geographical data to visualize the data in the database. Data commonly used in business GIS, like demographic, infrastructure and competitor data, are not being used in the project. Therefore the geographical data model is fairly simple. Maps consist of 1 map layer shown attribute information in a thematically way. The attribute database however has a complex data model. Since the ECMAC database is in the process of redesign and is not available yet in its future format and with its future content, it is agreed upon that the GIS project will start with defined fields, independent of the underlying query structure. The database design is the responsibility of NAMAC.

During meetings in which progress was being shown wishes changed continually based on new insight in the possibilities. From that point it can be said the prototype approach was being used. The four steps for the design of a conceptual model are taken more than once, although they are not explicitly written down every time.

During the phase in which the data was prepared different map algebra equations were used: overlaying, extracting, combining, etc. These operations are documented, but a written data model is not being created on beforehand. During the design for the desktop GIS itself cartographic modelling did not take place, due to the fact that the data model is fairly simple and that complex geographic data processes do not taken place.

The ECMAC project can be said to use the prototype approach. Due to the unfamiliarity of the people at ECMAC with GIS and due to the consultants lack of knowledge on the data availability in South Africa, the project had the character of trail and error. After showing the first results, the geo referencing of their database on suburb level, new ideas were being raised by ECMAC. Decided was to train one employee in GIS because it was foreseen that new ideas would go to be developed even after the three months period. The prototype approach is being used in an ever-developing GIS environment and this project can be seen as a beginning of corporate GIS implementation.

4.3 Conclusion

This paper aimed to answer the question:

How does the ECMAC project fit into the theoretical framework of project management?

The project can be seen as a business case, or a project evaluation after which and Go or No Go for an approved project can be given. A business case for investment in a GIS serves as a transition from a potential project to an actual (approved) project.

The ECMAC project can be said to use the prototype approach, although an implementation never took place. The project had the character of trail and error.

5 Appendices

Appendix 1, Table SMMEs contacted

The SMME table from ECMAC consists of 752 records. The following fields are defined:

Name	Type	Size
ID	Long Integer	4
CompanyID	Text	255
CompanyRef	Double	8
CompanyName	Text	255
HDI	Yes/No	1
FemaleOwned	Yes/No	1
Co-Owned	Yes/No	1
DateContacted	Double	8
DTI Sector	Text	255
NatureOfBusiness	Text	255
Type	Text	255
Process Category	Text	255
Process	Text	255
IndAdvisorId	Text	255
Current IA	Text	255
Courtesy Title	Text	255
CoFirstName	Text	255
CoSurname	Text	255
Province	Text	255
Centre	Text	255
Satellite	Text	255
CoTelephone	Text	255
CoFax	Text	255
CoCell	Text	255
CoAltTelNo	Text	255
AltCoFirstName	Text	255
AltCoSurname	Text	255
AltCoTelephone	Text	255
CoEmail	Text	255
CoWebsite	Text	255
Co Postal Address1	Text	255
Co Postal Address 2	Text	255
Co Postal Address 3	Text	255
CoSuburbP	Text	255
CityP	Text	255
Postal CodeP	Text	255
Co Physical Address 1	Text	255
Co Physical Address 2	Text	255
Co Physical Address 3	Text	255
CoSuburbPH	Text	255
CityPH	Text	255
Postal CodePH	Text	255
DateOfGMVisit	Text	255
Working With Company	Yes/No	1
Date Converted to Working With	Text	255
ConfidentialityAgreementCompleted	Yes/No	1

DateConfidentialityAgreementSigned	Text	255
ServiceContractCompleted	Yes/No	1
DateServiceContractSigned	Text	255
Dormant	Yes/No	1
Notes	Text	255
Financial Year End	Text	255
Datevalidated	Double	8
SMMENotes	Text	255
Date Created	Double	8
Date Appended	Double	8

Physical locations of SMMEs are in the SMME table are described by using:

- CO Physical address 1,
- CO Physical address 2,
- CO Physical address 3,
- CoSuburbPH,
- CityPH,
- Postal CodePH

Appendix 2, Data

The search for information on geographical data and on the definition of geographical regions in South Africa took place by means of an Internet research, phone calls to different geographical data suppliers and an interview with a GIS specialist of Urban Dynamics.

Available datasets

Urban Dynamics is a town and regional planning company. They use cadastral data for planning purposes. The Metro supplies this data to them. The Metro consists of 3 departments: GIS, Treasury and Town planning. Cadastral data is provided in Shape files.

From a cadastral perspective, Port Elizabeth is divided in allotments. Allotments have a 2 digit ID. Each allotment is divided in erfs. Erfs have a 5-digit erfnumber. Erfs can be divided into portions, having a 4 digit ID. Every polygon has therefore an 11-digit tag. To each erfnumber an owner name is attached. The ECMAC does not contain erfnumbers. ECMAC confirmed that the SMMEs are not always the owners of the erf and therefore owner names cannot be used to relate the SMME table to cadastral information. At Urban Dynamics they do not have streets addresses related to erfnumbers. The validity of that kind of information is doubtful, if it exists.

Allotments and erfs might have a relationship with postal codes and suburbs. Wards do not bear any relation to allotments, erfs, postal codes and suburbs. Wards are not always logical from a town planner’s perspective, as they are based on population numbers.

The coordinate system used in South Africa is WGS 84, zone 25, Hartebeesthoek datum 1994.

Requests for information to different GIS companies, software and data suppliers, in South Africa resulted in the possibility to obtain a dataset of the whole of South Africa. House number data for the whole of South Africa costs 800.000 Rand and a set with street data for the whole of South Africa costs 8.500 Rand. These datasets are said to be accurate for the bigger cities.

Addresses

A problem that occurs in topographical data in South Africa is the fact that geographical objects, like streets, cities, suburbs can be spelled in several different ways depended on the language being used. For example, the Admirality Road is also called the Admirality

Way or the Admiraliteiten weg. This makes it harder to find a table with street names that cover all spelling possibilities used.

At ECMAC the industrial advisors manually enter the SMMEs address in the database. Street names are not being selected form a fixed pop-up list. Often the name of a building is used instead of or in addition to a street name. Three fields can be used to enter the physical address of a SMME. As a result addresses are not being entered in a uniform way. A selection of the SMME table shows the physical addresses of SMMEs in the suburb Kwazakhele.

Co Physical Address 1	Co Physical Address 2	Co Physical Address 3
10652 Stofile Street		
11517 Ghana Street		
11520 Site & Service		
11520 Site & Service		
12931 Ebazi Street		
1644 Mandela Street		
204 Hostel		
2535 Mavuso Street /2435		
2539 Sali Street		
302 Ekuphumleni Street		
3721 Site & Service	Kwazakhele	
3770 Site & Service		
3957 Sali Street		
4484 Tubali Street	Site and Service	
50079 Daku Road		
508 Site & Service	Struandale	
5625 Jakavula Street		
5726 Kulati	Njoli	
6598 Salamntu Street		
740 Nduma Street		
82015/5 Site and Service		Salamntu Street
8351 Daku Road	Kwa Zakhele	
9293 Site & Service		
9927 Site & Service		
Block 260		Kwa Zakhele Hostel
Daku Road		
Embizweni & Njoli Square		
Embizweni and Njoli Squares	New Brighton & Kwazakhele 2991	
Enkuthazweni Urban	Stofile Street	
Masakhisizwe Building	Daku Road	
Masakisizwe Resource Centre	Daku Street	
Red Cross Building	Qgunda Street	
Red Cross Building,	Qgunda Street	
Room 5 , SBDC Building	Struanway	
Stofile Street		

Geocoding this table, if a valid geo-referenced table exists, will involve a tremendous amount of handwork. It is decided that geocoding on street or house number level is not going to take place due to the high costs and the presumed unavailability of valid data, combined with the quality of the content of the data fields that indicate addresses.

In South Africa the geographical area covering a postal code is bigger than a suburb and often even bigger than a city or village. The smallest geographical area to georeference the SMME table is suburb. It is agreed upon with ECMAC that suburbs are going to be used to show the SMMEs on a map. A consequence of this is that information on SMMEs has to be aggregated on suburb level. It was also decided that only suburbs in Port Elizabeth were going to be used, as the majority of the SMMEs is located there and a first look at the data reviewed that it would consume a lot of time to match all the not automatic matched records manually.

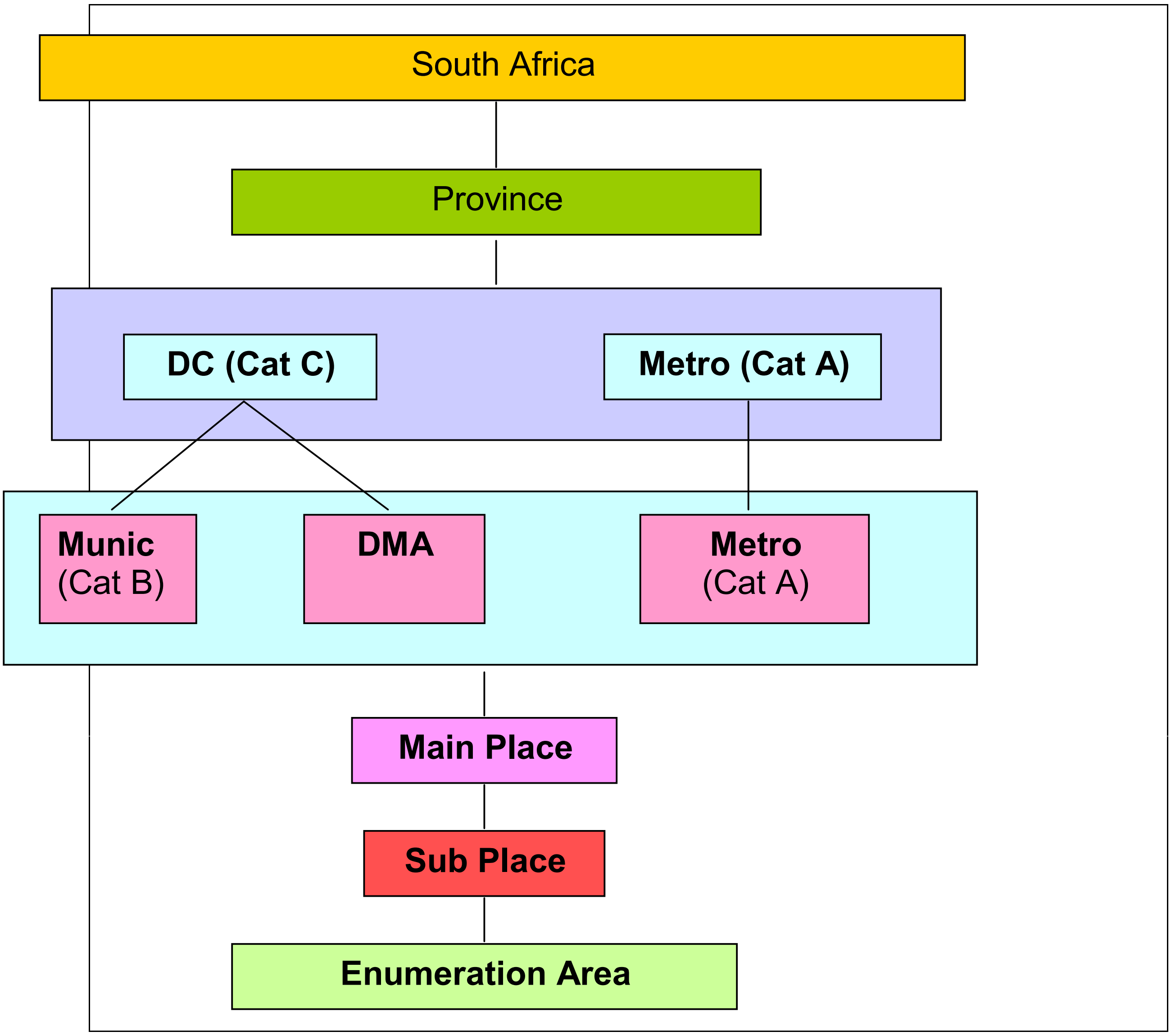
Suburbs

11 out of the 752 records of the ECMAC table have no suburb filled in. Only 64 of the 752 records in the ECMAC table contain a city name, for the rest the field CityPh is left empty. An automatic selection of suburbs in PE can therefore not be done. Grouping SMMEs per suburb resulted in a list of 125 suburbs for the whole of the Eastern Cape.

Suburb names in the ECMAC table are derived form an Excel sheet that can be downloaded from the official website of the South African Post Offices (www.sapo.co.za). The table contains suburb name, postcode and city name. For cities and suburbs both the African and English names are defined if they are different, for example Parsons hill is the same as Parsons heuwel. In total there are 14.295 suburb names in the table. Querying the table for suburbs in PE resulted in 268 selected records. Due to different spelling of names there is some overlap. It seems however, that the industrial advisors at ECMAC preferred the English spelling above the Afrikaans spelling when a suburb was selected from the list.

From Statistics South Africa (www.statssa.gov.za) a free copy of a CD-Rom with the GIS spatial data used in the 2001 census dissemination was obtained. The CD-Rom contains Shape files of different geographical areas in South Africa. South Africa's geography has some inconsistent entities at the same level of reporting, which do not fit perfectly one onto the other. This results in cross-boundary geographical entities at all levels of the hierarchy structure.

The hierarchy structure looks as follows:



Statistics South Africa, census 2001, Metadata, Geography hierarchy and attributes, report 03-02-25 (2001)

Sub place is the next spatial level up from the Enumeration Area and one below the main place in the place name hierarchy. There are 15.966 unique sub place names in South Africa according to the Statssa data.

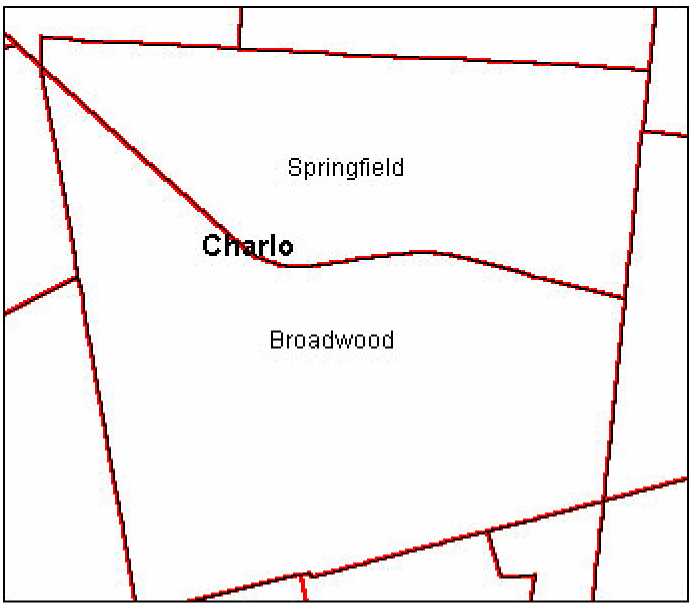
After converting the shape files to MapInfo .tab files, a geographical selection was done to select only Port Elizabeth's suburbs. Sub places were selected by selecting all sub places (Sp-SA) that are within or intersect the municipality (Mn_SA) of Nelson Mandela. Nelson Mandela is the new name for Port Elizabeth. This selection resulted in a table of 222 records (15 of which have the name "NONE").

Comparing the names of the 222 PE suburbs of Statssa with the 268 PE suburbs of Sapo resulted in only 90 suburbs having the same name. Matching the 125 Eastern Cape suburbs from the ECMAC table with 222 PE suburbs of Statssa resulted in 55 similar names. In other words, 70 suburbs did not match. These 70 are either outside PE or their names are spelled differently.

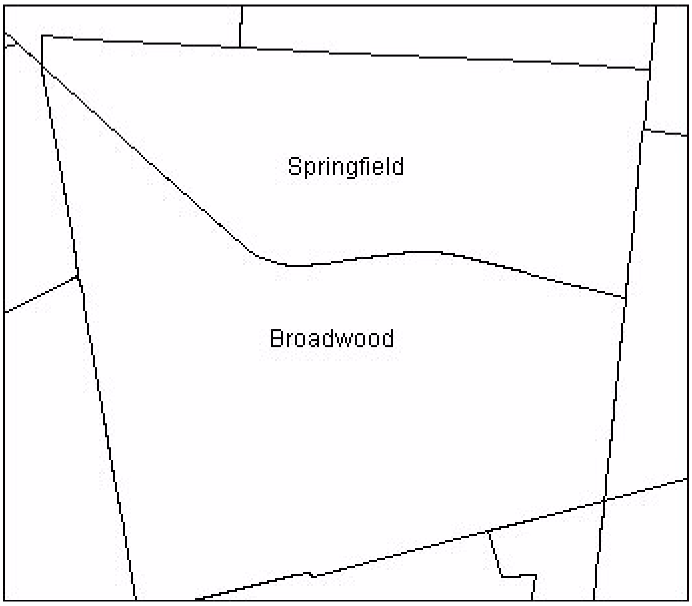
A geographical selection of all sub places in the Eastern Cape (EC) province resulted in 7.962 sub places. Matching the 7.962 EC suburbs with the 125 suburbs of the ECMAC table resulted in 74 matched suburbs. Which means that at least 19 of the suburbs of the ECMAC table are outside PE. The remaining 51 suburbs had to be matched manually, in such a way that the names in the ECMAC table stay the same. Consequently names but also physical regions had to be changed in the Statssa map and table.

Some names could easily be match by changing the spelling: Booysen Park to Booysens Park. The process of changing suburbs in the Statssa table involved local knowledge that was derived from a paper map of PE, added with help form people that life in PE. Subjective decisions had to be made, as suburbs are not always clearly defined regions. They tend to overlap and people have different opinions about the actually suburb region and name. This can also be seen in the way the physical locations, address and suburb, are being entered in the ECMAC. For example: 1 SMME in the ECMAC table is located on 303 Kempston road in Korsten, while an other is located on 303 Kempston road in Sidwell.

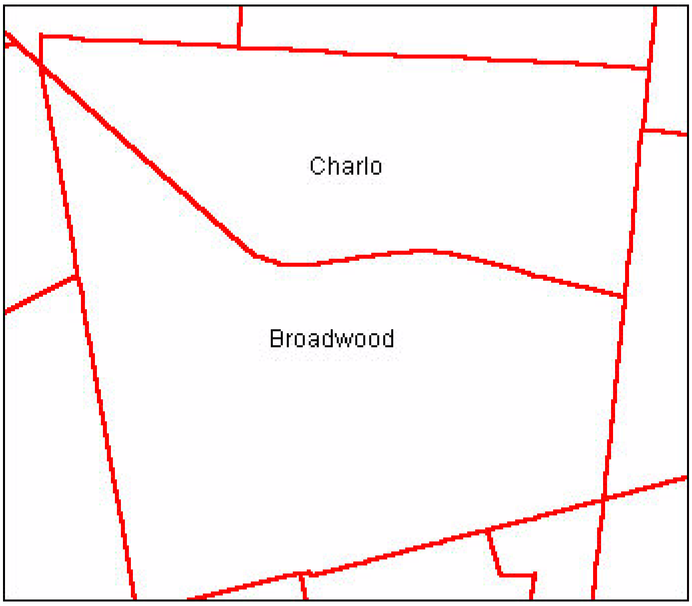
An example of an overlapping suburb is Charlo. According the paper map, Charlo contains of 2 regions, one called Springfield and the other is called Broadwood.



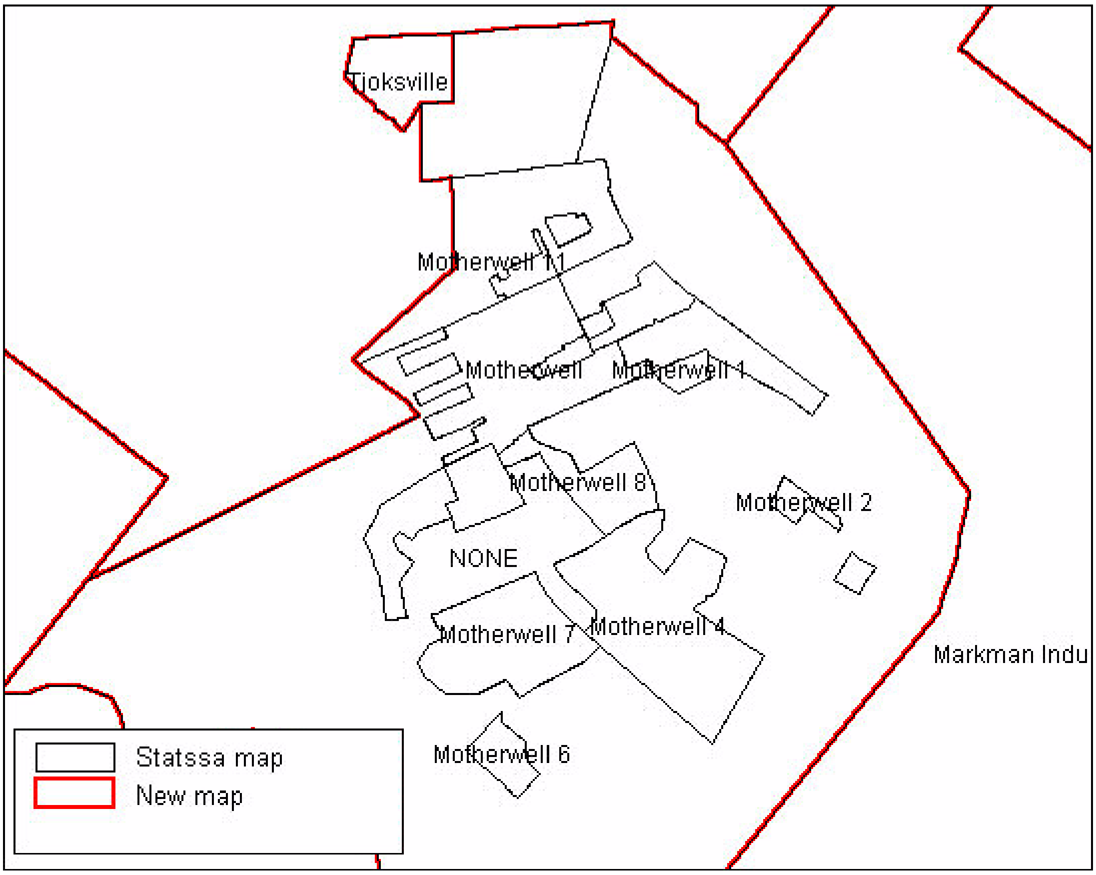
On the Statssa map Charlo does not exist as a sub place.



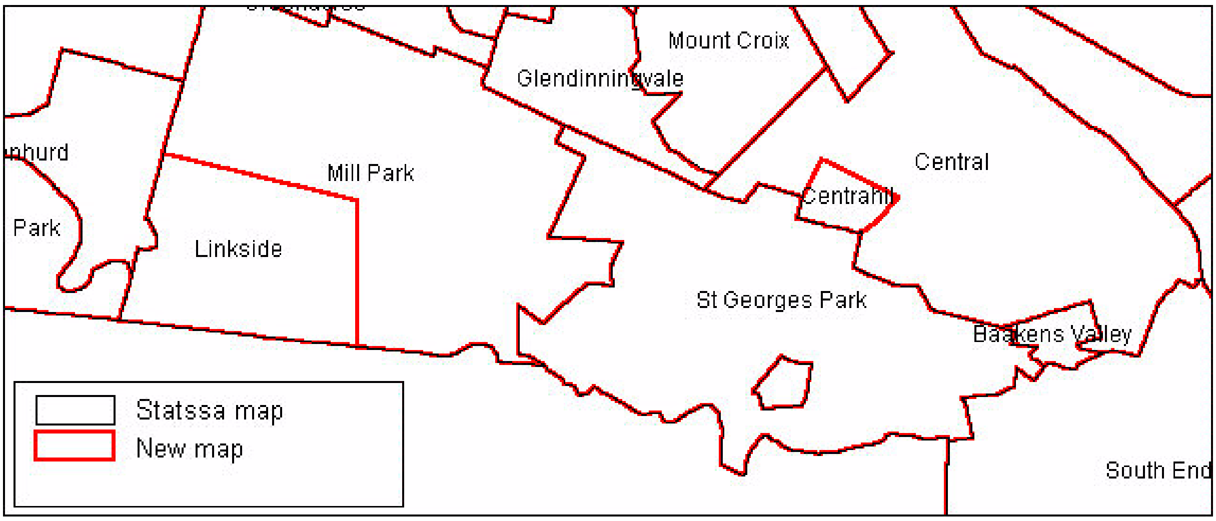
In the ECMAC table there are SMMEs in Charlo and in Broadwood. A subjective decision was made to change the name Springfield in Charlo.



The Statssa table contained a lot of suburbs with different extensions, for example Motherwell 1 to 11, while in the ECMAC table Motherwell is mentioned without an extension. In MapInfo these Motherwell 1 to 11 are combined in a new region called Motherwell.



Sometime sub places were split in order to create 2 or more new suburbs. For example, Linkside did not exist. Mill Park was split in order to create a physical region called Linkside. Same thing was done with Central Hill and Central.



A consequence of only changing names and shapes of suburbs in the Statssa map and table to names in the ECMAC table in order to make the ECMAC table mapable, is that when new suburbs are added to the ECMAC table in the future they might not automatically match to the ones in the Statssa table. Making the complete Sapo table for Port Elizabeth, the Eastern Cape or the whole of South Africa match the whole Statssa table would be far too time consuming for now.

Appendix 3, Bench mark study

During the benchmark study the desktop solutions of four GIS software suppliers are compared: MapInfo, ESRI, Intergraph and Microsoft. ESRI and Intergraph have the biggest worldwide GIS market share according to Daratech's GIS industry report (2002), respectively 35 and 13 %. These numbers represent their whole product, services and data line. MapInfo has only 6% of the total market share, but MapInfo Professional is considered one of the two greybeards, together with ArcView of the desktop mapping world (Marshall. 1999).
MapInfo's success is built on developing, marketing and supporting location-based solutions for the customer relationship management (CRM) market (http://www.gismonitor.com/news/pr/110702_Daratech.php). All three GIS software suppliers provide scalable systems that can extend from the individual desktop to the whole enterprise.

Microsoft Mappoint is not taken into account, as there is no South African version available yet. Microsoft Maps for Excel is added in the benchmark study because ECMAC has licenses for Microsoft Office 98 and therefore free access to Microsoft Maps for Excel.

MapInfo's desktop solution: MapInfo Professional 7.5 and MapInfo's freely available viewer: MapInfo Proviewer are tested in this study. Of ESRI the old desktop solution: ArcView 3.3, which will only be supported for an other year, the new desktop solution: ArcView 8.3, part of ArcGIS, and it's freely available viewer: ArcReader 8.3 will be tested. Intergraph's Geomedia and it's more sophisticated brother Geomedia Professional will be part of this study.

Selection of the software is based on several aspects. Software has to be selected to use during the project, the three-month period from December 2003 until February 2004. During the project the data has to be prepared.
In the future ECMAC and in later instance possible other MACs have to be able to use the GIS software. Possible extension to an Internet GIS and a migration to an ERP system have to be taken into account. Following is a list of considerations.
Software is compared on 5 aspects:

- Functionality
- Ease of use
- Licenses available
- Prices

- System requirements

Functionality
Since attribute information is stored and managed in a Microsoft Access database, it is necessary that the GIS software is able to establish a dynamic connection with tables stored in an Access database: the map has to be automatically updated after changes take place in the information in the client database (Access). Write permissions are not necessary, as data is only read from the database and will not be changed by means of the GIS. A nice-to-have would be the possibility to open Excel or Dbase files.
Dynamic connectivity to Microsoft Access can be defined as number one priority.

In the first place the GIS serves as a means to show statistical information that is derived from the existing NAMAC database in a spatial way. This does not involve complex spatial analyses. Minimum required functionality for the GIS software that will be implemented at ECMAC

- Thematically maps
- Measuring distances
- Information tool
- Zoom and pan
- Switch map layers on and off
- Labelling

In case the system is going to be extended in the future, more functionality might be desired. Future functionality is not yet known, but a possible increase in the amount of desired functionality should be kept in mind.

In order to prepare and edit the spatial data, complex functionality is required during the time of the project. Apart form the above stated functionality, required functionality during the project consists of the following:

- Overlays
- Edit (split and combine features) and create geographical data
- Index geographical data
- Determine centroides

However, due to limited time available an excessive benchmark for the software to use for preparing the data has not taken place. MapInfo Professional 6.0 was selected based on the following criteria:

ArcReader, Microsoft Maps, MapInfo Proviewer do not have the required functionality. Editing in ArcView 3.2 is time consuming due to the fact that all temporary tables and selections have to be saved separately. ArcView 8.2 is extremely slow on the available desktop computer that runs on Windows 2000, due to a rather small 122 MB RAM memory. The available laptop has Windows 98 as operating system and is therefore not suitable.

My personal preference and familiarity with the product has lead to the choice of MapInfo Professional instead of Geomedia. An added advantage of MapInfo is the fact that the MapInfo tab files are the only supported file format in Microsoft Maps for Excel. The functionality required during the preparation phase is not taken into account in this benchmark study.

According to a study undertaken in 1999, MapInfo and ArcView are the front-runners of the desktop mapping solutions. It is stressed however, that, whether you are looking for a general-purpose mapping program with extended consulting support or a data-intensive programmable GIS engine, you will want to look closely at the specifics of each program before deciding which is right for your department's purposes. (Marshall, 1999).

Dr Robert Barr (2003), director of the Regional Research Laboratory, University of Manchester, and senior vice-chairman of the AGI, (http://www.ginews.co.uk/1100_44.html) concludes in a review about Geomedia 4.0 that GeoMedia is a powerful product and one cannot help but be impressed by the design and execution of GeoMedia. It provides a scalable system that can extend from the individual desktop to the enterprise.

Checking the minimum required functionality for the each software package included in this benchmark study provides the following table (**x** means available):

	Dynamic link to Access	Thematically maps	Measuring distances	Information tool	Zoom and pan	Switch map layers on and off	Labelling
MapInfo Professional 7.5	**x**	**x**	**x**	**x**	**x**	**x**	**x**

MapInfo Proviewer	**x**	**x**	**x**	**x**	**x**		**x**
ArcView 8.3	**x**	**x**	**x**	**x**	**x**	**x**	**x**
ArcView 3.3	**x**	**x**	**x**	**x**	**x**	**x**	**x**
ArcReader		**x**	**x**	**x**	**x**	**x**	**x**
Geomedia professional	**x**	**x**	**x**	**x**	**x**	**x**	**x**
Geomedia	**x**	**x**	**x**	**x**	**x**	**x**	**x**
Microsoft Maps		**x**	**x**		**x**	**x**	**x**

Ease of use
At ECMAC there is nobody with any GIS experience. For one person limited time can be made available to be trained in the use of GIS software and to maintain and expand the system in the future. Therefore it is necessary to select software that is easy to use and that can be taught quickly and remembered easily.

Complexity of the software and the amount of functionality available in the software have a strong reverse relationship to ease of use. In some programs it is possible to hide part of the functionality, through which it appears easier to use. The availability of good documentation and a clear help file also influence the feeling of user friendliness. When a maintenance license is obtained users might have access to a helpdesk, which can help when you are stuck with one problem.
All software is Windows-based, so general computer skills form the basis for working with the software. Ease of use is also dependent on personal preferences and therefore it is difficult to make objective statements on the ease of use of the tested products.

An effort to rank and group them based on personal opinion gives the following list:

Very difficult to use (training 5 days, knowledge of GIS concepts necessary, regular use needed in order to keep up):
ArcView 8.3 and Geomedia Professional

Difficult to use (training 3 days, knowledge of GIS concepts necessary, regular use needed in order to keep up):
MapInfo Professional, ArcView 3.3 and Geomedia

Easy to use (training 1 hour, no GIS knowledge necessary):
ArcReader, MapInfo Proviewer and Microsoft Maps

GeoMedia and MapInfo's work is organised into workspaces. A workspace allows work, including connections to warehouses, queries, defined coordinate systems, map and data windows and queries, to be grouped together into a single unit that keeps track of everything that the user has done. If users discipline themselves to use proper network-wide file naming conventions the workspace becomes a powerful tool for collaborative working. The workspace definition file is quite small and can be sent to another user; he or she in turn can re-open the workspace and see it exactly as the first user had left it.

Perhaps because of its power, Geomedia is somewhat intimidating compared to MapInfo Professional and ArcView 3.x. (Barr R, 2003).

Licenses available
The license and the software should not only be available on the PE Technikon to prepare the data and to design an environment to work with for ECMAC, but it should also be available for future use by ECMAC. ECMAC has a limited budget; therefore the prices of software licenses have to be compared. At PE Technikon the following GIS software is available for the preparation of the data:

- ArcView 3.2a
- ArcGIS 8.2 (1 month evaluation license)
- ArcGIS 8.3, students license available from the 22 January 2004.
- MapInfo 6.0
- Geomedia Professional 5.1 (1 year evaluation license)
- Microsoft Maps for Excel

Free GIS software available on the Internet:
- ArcExplorer
- MapInfo Proviewer

Prices
All prices are quoted in Rands. Prices are derived from South African GIS software resellers. Prices of the web based solutions of ESRI and MapInfo are added in order to give an impression of possible future costs.

<u>MapInfo Professional 7.5</u>

1 license	17.581
10 licenses	16.346 per license
Maintenance + upgrades	4.084 per annum per license

<u>MapInfo Proviewer</u>

1 license	free

<u>MapInfo's MapExtreme</u>

Intranet license	80.000 + developing costs

<u>ArcView 8.3</u>

1 license	14.572
10 licenses	10-15% discount per license
Maintenance + upgrades	4.600 per annum for the first license
	2.800 per annum for the following licenses

<u>ArcView 3.3</u>

1 license	11.036
10 licenses	10-15% discount per license
upgrade to ArcView 8.3	6.000

<u>ESRI's ArcIMS</u>

Intranet license	60.000 to 68.000 + developing costs

<u>ArcReader</u>

1 license	free

<u>Microsoft Maps</u>

1 license	free

<u>Geomedia</u>

1 license	12.737
Maintenance + upgrades	2.400 per annum

<u>Geomedia Professional</u>

1 license	63.642
Maintenance + upgrades	12.000 per annum

System requirements

At ECMAC most computers have Windows XP as operating system, Office 98 and 240 MB RAM.

For preparation of data PE Technikon has a desktop computer with Windows 2000 as operating system, Office 98 and an internal memory of 122 MB RAM.
My laptop has Windows 98 as operating system, Office 97 and 256 MB RAM Internal memory.

All software compared in this benchmark study requires an industry-standard personal computer with at least a Pentium or higher Intel-based microprocessor and a hard disk. System requirements for the web based solutions are not taken into account.

MapInfo Professional 7.5

Operating System	Windows 98 SE, Windows 2000 Professional SP3 , Windows NT 4.0 SP 6a, Windows XP Professional SP1, Windows XP Home
Memory	32 MB RAM, (64 MB RAM for Windows XP) Minimum of a Pentium PC recommended
Disk space	103 MB

MapInfo Proviewer

Operating System	Windows 95, Windows 98, Windows NT with Service Pack 6 Windows 2000 Professional
Memory	MapInfo ProViewer will operate with Microsoft's standard Operating Systems Memory Requirements
Disk space	35 MB

ArcView 8.3

Operating System	Windows NT 4.0 with Service Pack 6a (or) Windows 2000 (or) Windows XP (Home Edition and Professional)
Memory	128 MB RAM minimum, 256 MB RAM (or higher) recommended
Processor	450 MHz minimum, 650 Mhz recommended

ArcView 3.3

Operating System	Windows 98/98SE, Windows Me, Windows NT 4.0, Windows 2000, and Windows XP--Home Edition and Professional (supported with limitations).
Memory	24 MB RAM (32 MB recommended)

ArcReader 8.3

Operating System	Windows NT 4.0 with Service Pack 6a (or) Windows 2000 (or) Windows XP (Home Edition and Professional)
Memory	128 MB RAM
Processor	450 MHz
Disk Space	200 MB

Geomedia

Operating System	Microsoft Windows NT 4.0 SP6a, Microsoft Windows ME, Microsoft Windows 2000 SP3, Microsoft XP SP1
Memory	512 MB RAM recommended
Processor	Pentium III or equivalent microprocessor - minimum
Disk Space	245 MB (according to Intergraph, in practice it works fine with 122 MB)

Geomedia Professional

Operating System	Microsoft Windows NT 4.0 SP6a, Microsoft Windows ME, Microsoft Windows 2000 SP3, Microsoft XP SP1
Memory	512 MB RAM recommended
Processor	Pentium III or equivalent microprocessor - minimum
Disk Space	265 MB 245 MB (according to Intergraph, in practice it works fine with 122 MB)

Microsoft Maps

Software	All versions of Microsoft Office suite except Office XP

Conclusions

MapInfo Professional, ArcView 3.3, ArcView 8.3, Geomedia Professional and Geomedia all have the minimum required functionality plus more intelligent functionality that can possibly be used in the future. ArcView 8.3 and Geomedia Professional are very complex and not very easy to use as a consequence, ArcView 8.3 use a lot of memory. Therefore ArcView 8.3 and Geomedia Professional are not suitable for ECMAC.

ArcView 3.3 is not going to be supported in the future, because of which it is a less favourable option. Geomedia is less expensive than MapInfo Professional. Geomedia is the preferred GIS software solution in this benchmark study.

However, due to the fact that free 1 year evaluation license is available, it is decided to use Geomedia Professional for the first year and switch to Geomedia after that.

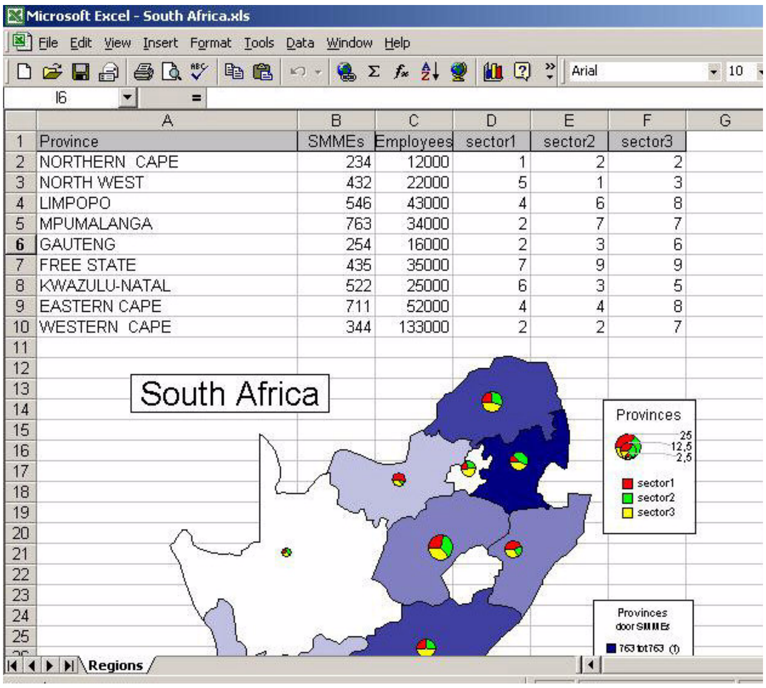
Appendix 4, Microsoft Map for Excel

Microsoft Map is a component that was created by MapInfo and can be added to Microsoft Excel. Microsoft Map is a tool that enables users to quickly create maps from spreadsheet data; because Microsoft Map is an 'add-on' component, it is not typically included with Excel and must be installed before use. It comes with a couple of standard maps. A map showing the provinces of South Africa is not included in the standard set. It is however possible to add your own maps to the set, providing they are in MapInfo .tab format. In order to add a maps a couple of steps have to be taken. I wrote a small manual to do this. Using Microsoft maps is quite simple after the set-up of the maps has taken place. The functionality is rather limited. It is possible too to use the Microsoft Map functionality in Microsoft Access. But in Access, opposite to Excel, but updates do not take place automatically.

A problem I encountered using Microsoft Maps in Excel is that a map is not redrawn properly when the map is too detailed, in other words, when the size of the map is too big. The Province map that came with the Statssa dataset is 1.4 MB. This turned out too big. A workaround is to digitise and recreate the map again in MapInfo with far less detail. The new map is only 9 KB.

Microsoft Map also does not except new maps to be added when the name of the map or its reference field is not unique or when the referenced field is not indexed.

The following picture shows the province map in Excel with fictional statistics showing some of the thematic possibilities.



Province	SMMEs	Employees	sector1	sector2	sector3
NORTHERN CAPE	234	12000	1	2	2
NORTH WEST	432	22000	5	1	3
LIMPOPO	546	43000	4	6	8
MPUMALANGA	763	34000	2	7	7
GAUTENG	254	16000	2	3	6
FREE STATE	435	35000	7	9	9
KWAZULU-NATAL	522	25000	6	3	5
EASTERN CAPE	711	52000	4	4	8
WESTERN CAPE	344	133000	2	2	7

6 References

Barr, R (2003) In: **Intergraph GeoMedia 4.0**, Software review (On-line http://www.ginews.co.uk/1100_44.html)

Bell, S. and Wood-Harper, T. (1992) **Rapid Information System Development: A Non Specialist's Guide to Analysis and Design of an Imperfect World**. McGraw-Hill, London.

Burrough, P.A. (1983) **Principles of Geographic Information Systems for Land Resources Assessment.** Claredon press, Oxford.

Grimshaw, D.J. (1988) **A Review of land and Property Information Systems Practice**, International Journal of Information Management, 8 (June) pp. 81-92.

Grimshaw, D.J. (1999) **Bringing Geographical Information Systems into Business**, Second Edition, John Wiley & Sons, Inc, USA

Heywood, I. Cornelius, S. and Carver, S. (2001) **An Introduction to Geographical Information Systems**, Second edition, pp. 233-253.

McDaniel, K. E. and. Sherman, B. E., (1992), **Introduction to AM/FM/GIS Project life Cycle**. AM/FM International Conference, Montreux, Switzerland.

Peuquet, D. J. and Bacastow, T. (1991) **Organizational issues in the development of Geographic Information Systems: a case study of U.S. Army Topographic Information System**. International Journal of Geographic Information Systems, **5**(3), pp. 303-319.

Somers, R., (1994) **Alternative development strategies**. In Proceedings of GIS/LIS'94, Phoenix, Arizona: 706-715.

Tai On Chan (1998) **The Dynamics of Diffusion of Corporate GIS**, Submitted in total fulfillment of the requirements of the degree of Doctor of Philosophy, Department of Geomatics, The University of Melbourne

Ward, J. Taylor, P. and Bond, P. (1996) **Evaluation and Realization of IS/IT Benefits: An Emperical Study of Current practice.** European Journal of Information Systems 4: 214-255

Woodcock, M. (1993) **GIS Project Management**, UNIGIS, Module, Manchester Metropolitan University, Manchester