

## Operation “I want GIS now”

I have spent the past few weeks doing a practical exercise of taking field data from a live project and attempting to spatially represent the data outputs and summaries it in a GIS program called Mapinfo. The objective was to see the practical problems that I would have in doing so and also see why DW has so far failed to successfully exploit the tool. I won't attempt to make any long term objectives in this document but will rather concentrate on a realistic 3 month shock action plan (called 'I want GIS now') to kick start the process. At the end of that three month, if the objectives are realised, then I suggest the hiring of an expert consultant to make a further 1 year action plan.

I would advise you to first of all read the report, “Assessment of GIS projects for development workshop Angola”, written by a consultant called Joss Swennenhuis in 1999 to get some background information and he made some excellent recommendations most of which were not realised. You can contact me for an electronic copy of the report ([moloney@angonet.org](mailto:moloney@angonet.org)). His main recommendation was that we improve our data management structure as without proper data there can be no mapping to GIS. Once the data is good and the maps exist then the spatial data mapping process is relatively simple. Therefore in this document I will concentrate more on data management than GIS although I will make strong recommendations on what approach to take to ensure efficient and reliable mapping of data.

### ***The raw data to GIS output exercise***

At the moment DW has excellent maps of Luanda sitting on Mapinfo system on the one hand and lots of raw data from various projects collected over the years on the other hand. However there is a problem in the middle with successfully analysing and presenting this data in a format that can be useful to the individual project and to the organisation as a whole.

The best current information that I could find in the organisation was the data being collected by the SCSP team in Luanda on latrines. There were over 2000 records in an excel spreadsheet of families in a number of sectors which detail some basic information on the family, as well more details on the access to latrine facilities and the recording of any intervention by DW to build or improve the family latrine. At the end of each page was a summary of the page information which in turn is summarised by quartiereiro and sector. The excel summary process was labour intensive, awkward and probably unreliable.

I chose 3 simple indicators, from many, for the sake of the exercise to represent in GIS, that is (i) population density as described in the number of members in the family, (ii) percentage representation of no families with latrines and the (iii) indicator that represents the ownership of the property.

I broke the task into 3 distinct parts:

#### **1. Excel to Access Conversion Process**

Conversion of the excel summary sheets into a flat Access database structure with strong key fields to allow easy summarisation and modification of main indicator fields to restrict option to permitted values only (ie For the property ownership field the only values allowed are (i) Owners, (ii) Rented (iii) Others). I assumed that summary requirements would be based on location (ie we need to see the number of families per quartiereiro, or the percentage of families that rent property in a particular sector and in order to do this we required a much more reliable way to represent the location of the property. As this is a requirement for all of DW's data collection activities I have created a separate database that holds a structured coded representation of the areas of Luanda. (See below).

I created an input form to allow more reliable of data as well as output report with summaries of the main indicators that can summarise from province down to quartiereiro level and also some simple graphs of my chosen indicators. I also made use of crosstab queries to highlight inaccuracies and inconsistencies in operator data entry.

#### **2. Creation of Area Coding System**

Creation of a separate Access database to give a strong structured coding system to areas of Luanda. We adopted the current INE coding system down to comuna level and extended it down three more levels through to quartiereiro level. Thus a province is represented by a 2 letter code (Luanda is 04) down to quartiereiro level which is represented by a 12 letter code. As an exemple the quartiereiro 040904031601 represents

Province:	Luanda	04
Municipality:	Cazenga	0409
Comuna	Hoji-ya-Henda	040904
Bairro	Ilha Madeira	04090403
Sector	Sector 16	0409040316
Quarteirao	Quarteirao 1	040904031601

I would see the elaboration and maintenance of this database as the most important link in summarising data and further transferring it to Mapinfo. As well as the data there are a number of routines that allow one to access and update the data. The tables that I setup in this database were linked into the latrine database and used to represent the location of the family/property.

### 3. **Spatial Representation of Data in Mapinfo**

The spatial representation of the areas that we work in are held in a series of tables in Mapinfo that represent municipalities, communes, bairros, sectors and quarteiroes. I experimented with a number of options of mapping the summary data held in the databases including modifying the source Access database to allow Mapinfo to geocode the source tables and thus maintain live data. However this is not an advisable method as two different database platforms are sharing and updating data and thus greatly increases the chance that corruption could happen especially on a peer to peer network not to mention that database become unwieldy and grow large quickly. The best trade off between simplicity and speed was to manually encode the spatial tables representing the respective area in an extra column with its respective code as per the location database. (This would be a once off operation) It was then a simple matter of opening the Latrine database or summary and running a simple query to allow Mapinfo to pick up the data. Once the data is in Mapinfo and was spatially represented then thematic maps were easily created. This process is easily mechanised, and thus easily learned by a relatively inexperienced user, and also avoids clutter in the Mapinfo workspace as predefined queries can be created from Access on the fly, printed and then deleted.

## **Current Situation**

So, we basically remain with almost the same problems as we had a few years ago during the Joss consultancy although I'm sure that we now have much more detailed maps. We remain with poorly structured data management within the organisation and with virtually no basic database design expertise. There is currently no functional database within the organisation. Coordinators and their staff have virtually no experience in entering and summarising data. There seems to be no methodology for designing and building data models on the part of projects based on required project outputs. Without all of this there can be no output to GIS.

So, it appears, the area that we need to attack is how we, as an Organisation, approach the problem of analysing, designing and building our databases and what are the resources required to reach that objective.

## **Steps to good database design**

### **Analysis**

The first step in the data design process will be a written specification of the required database by the program manager or coordinator of the nature of the required data and how they wish that data to be presented/output. Depending on the complexity of the requirement this can be further elaborated with the use of diagrams and sample graphs. The complexity of the required system will also determine what profile of system analyst/designer is required. Most simple databases can be designed in house with a bit of practice. Available funding will also of course determine what can be achieved.

For example if there are going to be less than 50 registers in the database then it may be better to set the information up in a spreadsheet.

In order to move onto the next stage, a realistic picture of the required system should become clear and one can start to think about how the database will be structured. One should not move onto the next step if the objectives are not clear. Up to this point we have not even opened a database but already we should have defined what our ultimate goal is. The design and building steps is merely the vehicle that takes us to our objective.

## Design

At this stage the analyst/programmer will turn the requirements specification into a detailed design by creating a series of tables that represents the structure of the required data as closely as possible. Well-designed tables at this stage will greatly simplify the rest of our task. There are a number of techniques that database designers use to ensure correct database design among them normalisation (simplifying data down to its lowest common denominator) and creating logical unique key fields. Also we can build certain rules into our design to ensure data integrity. (Boring stuff)

At this stage we will run some tests with sample data to ensure that we are on the right track. We will also write our base queries that summarise and extract data according to the analysis phase requirements.

## Building/Coding

Once the tables are designed and base queries completed we are ready to design our data entry forms as well as the output reports. We use forms to allow input and viewing of on screen data. We can build in rules in the data entry forms to ensure data integrity. (i.e. that a data can only be inserted in a data field and a number in a numeric field) We also normally try and hide the tables from the end user as an inexperienced or malicious user can seriously damage our existing data. Thus, a person who enters data in an access database doesn't need to know anything about how the database works other than to type and follow a few basic rules.

The reports that we design will normally be capable of listing the data in the output required by the user or manager. This can include a complete listing of every record or can be a summary according to a predefined requirement. Again little skill is required to print a predefined report.

We also normally try and have a basic menu system that allows the user to navigate through our forms the system.

## Testing Implementation

Once the programmer has completed the building phase he will attempt to install and deploy the database in a real live work situation. Users will input data and run reports to ensure the correct functionality of the system. If any problems are found then they will be investigated and corrected.

## Required

Improvement of key personnel skills in key areas

- **Project coordinators** and **program managers** should be encouraged to walk through a start to finish process of managing the conception, design, implementation and supervision of a database system. Anybody who is at the Program manager or coordinator level already should have most of the skills necessary to do this as the database definition is quite similar to the project definition phase. I see no need for such personnel to become expert in database design but there is definitely a need for a more results orientated attitude and to practice setting up data models.
- Each project should have a person (**project database coordinator**) who has an intimate knowledge of the project and its reporting requirements and who provides reports to coordinators and is responsible for collating data should have a the following minimum skills:
  - In Windows
  - Knowledge of backup procedures

- Ability to navigate in windows explorer and to find and store documents in a network environment
  - In Excel
      - Practical knowledge of the Data facilities of Excel including:
        - Summary of columns and rows
        - Sorting of tables
        - Use of pivot table reports
        - Ability to create simple graphs
    - In Access
      - Ability to create a simple table
      - Ability to summarise data using an access query
      - Ability to graph a query
      - Ability to create a report
    - In MapInfo
      - Run a simple query to display output summarised information from excel and access in predefined maps
    - Ability to organise/coordinate a data collection activity including:
      - Supervise and train data collection staff
      - Supervise and train data entry staff
      - Create forms suitable for efficient data collection
      - Supervise accuracy of data collected and input
      - Feedback of collected data to data collection staff to improve quality
- All **senior project staff** should have a minimum set of computer skills especially in excel and word. (I can define a minimum requirement if required) They should also be kept up to date on ongoing project data collection activities and reporting, even if it is not in their direct area of responsibility, as well as be aware of the techniques and tools used to achieve this.
- Recruitment of **Institutional level database manager** with database design and programming skills. In addition to *above skills* of project database coordinator the database manager needs to:
  - Manage the location code database that is essential for our summary and mapping requirements.
  - Liase with program managers and coordinators to ensure efficient progress in their respective programs
  - Assume role of project database coordinator in small projects that have modest data collection and reporting requirements.
  - Provide ongoing training to project staff especially to project database coordinator
  - Comprehensive knowledge of MapInfo database queries, table import skills and spatial mapping techniques.
  - Liase with GIS map expert to ensure reliable codification of vectorised layers
  - Some web database knowledge.
- Strategic use of expatriate consultant and trainers to solve more complex analysis and design issues. Before consultants are hired the program manager and coordinator should provide a comprehensive TOR to maximise benefits. Program managers and coordinators should try to accompany these missions in order to improve skills in this area.

Organisation of server to hold live database files with proper set-up of permissions and backup procedures.

Realistic budgeting for project data collection, input and reporting including support staff and hardware.

Better definition of the role that the current GIS department plays and a more results orientated appraisal system. I would suggest the following:

- Integration of current GIS department and newly created institutional data manager role into a new department called say “DW Data and GIS services”.
- Clear demarcation of tasks and responsibilities according to individual skills and areas. I would consider the skills required to produce maps and to map data as two distinct areas of expertise.
- Better supervision by Director and program managers. Program managers and coordinators have to plan with Data and GIS department a more results orientated approach that ensures that projects receive what they are paying for.
- A more realistic and client service orientated approach that provides solutions tailored to requirements of projects. This could be linked to billing.
- Yearly consultancy mission to infuse new knowledge and approaches and highlight internal weaknesses.

### ***Plan Of Action for Operation “I want GIS now”***

<b>Required</b>	<b>Action</b>	<b>Responsible</b>
Standard location database that defines codes for provinces, municipalities, comunas, bairros,sectors and quarteiroes.	Start by updating locations database for Luanda so that we have a comprehensively coded database of Municipalities, Comunas, Bairros, Sectors and quarteiroes	Data and GIS department
Facility to easily spatially represent data from many sources in Mapinfo	Coding of Mapinfo Municipality, Comuna, bairro, sector and quarteirao tables in line with above location database	Katu
Attempt to summarise and map existing project data with dual objective of reporting and training.	Each project to identify data that they would like to summarise. Tony to attempt to put something together and make recommendations in each case.	Project Coordinators and tony
Recruitment of a database technician	TOR for technician based on above outline. Grade ?	?
Staff training: Program managers	To discuss but clearly related to experience and on the job training.	?
Staff Training: Project database coordinator	To discuss but related to project funding availability	?
Staff Training: Senior project staff	To discuss	?
A more service orientated approach to work.	Redefinition and reorganisation of existing GIS department.	DW management
Setup and adopt a directory structure system and organise backup facilities.		Tony?

### ***Conclusion***

DW has poor data management facilities both at project and institutional level. As is clear from the above document the problems lies in data management and the key word is management. Without proper management planning, focus and follow through in this area then there can be no proper supervision and

reporting of project activities and achievements. However we can very quickly put in place the structure to correct most of these failings and with a concerted effort quickly start getting some results.