Statistics South Africa
Preferred supplier of quality statistics

The use of Geographical Information Systems, Global Positioning Systems and automated demarcation technologies in surveys and census mapping at Statistics South Africa.

The Regional Workshop on
Census Cartography and Management, Lusaka.
08 – 12 October 2008.

Mr Carel Basson

Professional: Geographic Frame, Geography Division, Statistics South Africa.

carelb@statssa.gov.za

This paper discus Statistics South Africa's experience in geo-technologies and geo-information usage for surveys and census activities from 1996 until now.

Introduction

The use of digital data collection is revolutionising the way surveys are being done internationally. In keeping up with current methodologies and new technologies, Statistics South Africa has decided to implement the use of Global Positioning Systems (GPS) technologies and hand held computers in conducting, monitoring and reporting in its surveys. The Geography division of Statistics South Africa (Stats SA) is the primary driving force behind this initiative which is quickly gaining momentum and stirring inputs and expectations amongst other user divisions within Stats SA. The aim is to eventually move towards a totally digital system of data collection, thus totally eliminating the paper trail. This is inline with Statistics South Africa's vision of being the preferred supplier of quality statistics, not to mention achieving savings in costs and the environmental impact because the digital system is re-usable.

Because of the higher volume of high accuracy spatial- and attribute data collected for the surveys and the Dwelling Frame project, the Geography Division of Stats SA is now able to do more advanced spatial GIS analysis and automation of laborious tasks, such as Enumeration Area (EA) demarcation is now also a reality.

This paper will make reference to the Labour Force Survey (LFSR) as pilot in implementing this process. An explanation of the LFSR will be given as background in order to understand the context for development. The paper will then move on to briefly discuss the use of GPS and General Packet Radio Service (GPRS) technology in the Dwelling frame project. An overview of the automated demarcation process will also be done.

In addition to this, the paper will focus on how GPS was used to monitor and report on the progress/data quality of the LFSR survey. Finally a summary will be provided with regards to problems encountered and the lessons learnt.

LFSR Survey

The Labour Force Survey is used to measure the health and wellness of the country's work force. It also measures employment and unemployment rates in the country. Currently, the survey is being conducted twice a year using contractual fieldworkers. The survey is under going a restructuring process in terms of methodologies, processes and techniques for data collection. There will also be the appointment of a permanent fieldwork team that will now conduct the survey four times a year as opposed to twice a year.

The frequency of the survey arose out of a demand for more current statistics during the year with regards to the labour force for a variety of reasons. As a result, the data collection and processing time had to be reduced in order for an appropriate turn around time from data collection, -processing and analysis. For this reason it was decided that PDA devices would be used for data collection with information being transferred back to head office via GPRS. This would effectively minimise the amount of processing time needed as the information captured on forms would not need to be manually captured from paper.

The survey consists of three phases: publicity, listing and data collection (also called enumeration). Data collection or enumeration is the actual interview phase where the fieldworkers will visit the sampled dwellings that have been provided by the methodology section.

Currently all survey fieldwork is facilitated by paper forms. Due to the decreased timelines available to the LFSR team, Stats SA, for the first time, implemented the use of handheld computing devices nationally in the beginning of 2007; however, only GPS points were captured. This formed the basis of a pilot study in the use of GPS and handheld computing to aid the survey collection process. Currently paper and digital systems are running in parallel, though the digital is progressing with further tasks being developed and added in phases.

Dwelling frame project

For the first time in South African census taking history, Stats SA is recording and using the position of every dwelling in the country in its preparation towards Census 2011. Such a frame of dwellings is not currently available in the country, and there is also no complete address register, with many rural dwellings never having been assigned an address. Stats SA has prioritized the creation of a geo-referenced dwelling frame in preparation for and monitoring of operations, to ensure that all dwellings in the country are visited during enumeration. The longer-term vision is to establish a continuously maintained geo-referenced register of dwellings, establishments and buildings.

The development and maintenance of a dwelling unit frame is of great interest to various government role players as an enabler of physical service delivery, emergency response services, billing systems, property valuation rolls and financial services to citizens. In areas where addresses are available, such as large urban areas, the dwelling frame will integrate, standardize and continuously update such a frame by partnering with local municipalities as data custodians.

The methodology to geo-reference dwellings makes use of high resolution aerial photography and satellite imagery. This imagery is acquired through cost-sharing arrangements with other government departments. An office exercise takes place to point capture each dwelling from the photos using GIS, followed by a field exercise to verify points and capture new points using GPS technology, as well as capturing of approximately 30 attributes for each built structure. This information is then integrated into the organization’s spatial database and updated on the website.

The attributes of each dwelling that are captured cover elements of what would be included in a housing census, and the possible contribution of the dwelling unit frame to the collection of information usually undertaken in a census of housing is being explored. Since the characteristics and locations of establishments and other buildings (e.g. schools, clinics) are also recorded during the capturing of dwelling units, the frame can be explored as a source for other statistical collections, such as information on informal businesses, as well as in the compilation of building statistics.

GPS

When Stats SA started investigating the use of GPS, it was discovered that there existed a variety of different devices offering different options with regards to battery life span, memory, GPRS connection, etc. The investigation revealed that the organisation’s GPS requirements would ultimately depend on the business needs of the organisation. All organisations would however need to take into account the following criteria;

- ✓ Device specification: for example rugged or standard PDA
- ✓ Available budget
- ✓ The required accuracy for the fieldwork application.

At Stats SA, accuracy was the defining point. Since Stats SA collects household level data, it was decided that we needed an accuracy that would enable us to capture the location of a particular household. This would also result in the creation of a geo-referenced master sample database for households which could also supplement the work done with the dwelling frame project which aims to provide an address based system for every household.

As a result of this, it was decided that Stats SA would require an accuracy of 1m or less, hence we set about acquiring a sub meter accuracy signal. As mentioned, a normal GPS machine can obtain an accuracy of around 3 meters on its own and Differential Global Positioning System (DGPS) needs to be purchased to increase the level of accuracy.

Sub meter accuracy is usually obtained by means of a backpack which is connected to the GPS logger on one side and an antenna on the other. The logger is the actual GPS unit on which the information is captured whilst the antenna is used to communicate with the DGPS transmitters. This is unscrambled in the processor that is housed within the back pack and this then sends the sub meter accurate coordinates back to the logger which stores it on the data card or memory which can then be transmitted via GPRS or 3G using the existing telecommunications network or they can be stored on the device and downloaded onto a Personal Computer (PC) when back at the office. Within the backpack are also batteries that power the whole system.

It is important to note that the average lifespan of a battery is 6 – 10 hours so staff will need to charge the batteries on a daily basis. Stats SA decided that in order to minimise downtime and that since human error is possible in that staff will forget to charge the batteries, all units were supplied with a second battery which was to be used should they not have been able to charge the first battery.

To summarise: for the LFSR Stats SA secured the purchasing of 300 handheld PDA's with built in GPS capability. For the Dwelling Frame, Stats SA plan to acquire a complete digital data collection solution. That should include the handheld devises, the DGPS devices and the system at head office to receive the data captured in the field and to send data to the instruments in the field.

Cheap GPS devices are available from about R2000 but these are merely devices that show you your coordinates. Stats SA feels that in our endeavour to go paperless, this will be of little value to us, so we are investigating the use of handheld PDA or handheld mobile computers that has GPS capabilities. This will allow us to capture coordinates as well as capture attribute information on the device. After investigation, the device chosen for the LFSR met the following criteria:

1. Built in GPS receiver
2. GPRS/EDGE connection to transmit information wirelessly
3. Fast processing speed of 520mb (CPU, other PDA average 200mb)
4. Memory (256MB, upgradeable with SD card)

5. Alpha numeric keyboard (a bonus option to those that don't like the touch screen keyboard)
6. Battery life of 8-10 hours

How is the GPS used?

Currently for the LFSR, the fieldworkers are using both handheld and paper based data collection methods. The devices are currently used just to capture GPS coordinate information. The associated attribute information is captured using a paper based listing form. For the Dwelling Frame project no paper is used. The spatial data and the attribute data are digitally captured and send to the office. The basic methodology of capturing the data on the unit is the same for both projects.

When the fieldworker first starts up the GPS they need to register on the device and the PSU (Primary Sample Unit for the LFSR) number or the EA (Enumeration Area for the Dwelling Frame) number so we have a record of who was working with the device and in which area. This is useful in monitoring which fieldworkers are constantly experiencing problems and in which areas. The supervisor or area manager is then informed who can then investigate the problems.

The tech support team is constantly monitoring these types and other issues when monitoring data coming in via GPRS. As soon as a fieldworker gets to a structure, they need to capture a GPS point for every single structure, with the only exceptions being hostels or flats (apartments) in which case a single GPS point is captured. The remaining units in the block of flats for example, are linked to this single GPS point, i.e. many records linked to one point. Only a single point is required because only one point will be needed to navigate back and collect data from all the people in the flat-building or hostel.

The GPS point is usually captured as close to the front door as possible. The device is set up in such a way that the fieldworkers will not even know about sub meter accuracy. All they will see is either a red spot or a green dot when they get to the front door and look at their devices. The red dot means there is no sub meter accuracy available at the spot and it will not allow them to continue. They then need to move around in small circle around the vicinity to see if they can obtain a green dot which implies that sub meter accuracy is available and they are now prompted to either capture a single point or capture a point and link multiple records to it. If the green dot is not found within a minute, the device times out and captures a ‘zero point’ and allows the fieldworker to proceed to the next structure for capture. This zero point is recorded as a value for when sub meter accuracy was not found. In traditional areas, the GPS point is captured as close to the main entrance of the homestead that is used by most of the members of the household. The whole idea of capturing a GPS point is that the fieldworkers can come back later to the dwelling during the data collection phase and collect information.

All this information is sent back wirelessly via GPRS to our head office in Pretoria in real time where the technical support guys are constantly monitoring this and other problems as well as calling fieldworkers who have queries.

Another major benefit of using GPS is that for the first time Stats SA was able to actually physically monitor and see where fieldworkers were capturing information. Previously there was no method to monitor if fieldworkers actually were in the field when working. By using the GPS points we were able to quickly identify areas in which the fieldworkers were capturing outside the boundaries or in the wrong area or even identify areas where they did not capture all the information. In such cases, their supervisors or themselves were called and this was quickly rectified, thus saving time and increasing the validity and confidence in our data.

The number of GPS points captured is also used as a measure of the amount of work that a fieldworker is doing. All this information is being sent back wirelessly in real time so monitoring is up to date and always current. If the fieldworker is out of GPRS range then the information is stored on the device and as soon as the device moves into an area where signal is present, all the information is sent wirelessly. Basically any area with cell phone coverage capable of talk has GPRS coverage. Both MTN and Vodacom were approached with regards to the coverage areas and surprisingly most of the country is covered with a minimum of GPRS coverage and coverage from both network providers is similar. MTN and Vodacom have only implemented the higher speeds of EDGE, 3G and HSDPA in the main urban cities and these footprints have limitations. Stats SA is also currently investigating the viability of satellite telephone communications as solution for areas where no cell phone coverage exists such as large areas in the Northern Cape Province.

GPS also aids in reporting progress. As it uses real time GPRS, managers are constantly aware of how many structures were listed, which areas have completed data capture, which areas are lagging behind, etc.

Problems experienced

As promising as this technology is, there will be problems encountered when implementing new technologies.

The first was the reluctance of users to accept that this technology would be useful, given the capital investment needed. This was a problem when introducing this new technologies to staff members, but they soon realised that this will be an invaluable tool in creating increased confidence in the data.

Stats SA is fortunate to have the knowledge and skills of consultants from Stats Canada who helped and guided us in this process. However, as we all know, the datasets that we have in South Africa and the datasets in Canada is totally different in terms of accuracy and quality. Hence, what they want is not always possible at all times. For example: they suggested the use of road centre line maps which fieldworkers can use to orientate themselves. In the rural areas this was not possible - we often had blank maps generated for certain areas.

For interest sake, every fieldworker was also given a map pack for the area that they were going to work in. Due to poor quality and inaccuracy in third party datasets, maps contained data that was either shown as within boundary on the map but in reality was outside the PSU boundary and the other way round. We found during the LFSR project that fieldworkers used these as landmarks and then orientated themselves around them. If it was a school for example, it is a prominent feature and is easy to locate. This problem can't really be solved unless the quality and accuracy of the data is improved. This is part of the motivation behind

the Dwelling Frame project, because one of the outputs of the project is a geo-referenced structure dataset.
During training fieldworkers are made aware that this could be the case and they need to orientate themselves using other features on the map as well and not just one feature. This could also be the reason why some of the GPS points were captured out of boundaries as a fieldworker could have identified a feature which was out of their area of capture and captured around this point thinking it was within boundary.

When they went out to do publicity, the community was not informed that the fieldworkers would have the GPS devices with them when they are coming around for listing. Thus the first couple of days, fieldworkers worked quite slowly whilst getting to grips with the new technology as they only had about 2 days for training and also explaining to all the nervous community members what the devices were and what they were doing.

Linking work done on paper to GPS points captured on devices was challenging. A system was designed on the principle that a pre-printed record number for each unit of information on paper would be the link to the captured GPS point by a record number generated when ever a point was captured. The device generates the record number automatically; thus the fieldworker was tasked with checking every time that that number displayed on the device was similar to the number of the record completed on paper. The system was not tested sufficiently and during listing it was found that fieldworkers were not diligently keeping track of paper and device. Fieldworkers found themselves ‘out of synch’, the paper number and device number not being similar. This had immense impact because there was no way to determine when the work went out of synch, and thus all links of captured GPS points to paper records, were in these cases irrelevant. The paper is still sound, though. This is a great problem if device and paper are out of synch at records over 100 –some listing books go over 1000!
For this reason the Dwelling Frame project is capturing and sending all the attributes digitally. It saves a considerable amount of time in the downstream processes.

Fieldworkers were required to re-capture GPS points, following the work done on paper if out of synch occurred. This was solved by forcing a software update from head office via

GPRS to all devices that allowed them to continuously check the record number on the device matched the record number of paper. This was updated to all devices in the field wirelessly without the devices having to be returned to head office or even to the provincial office. The final result was about 15% of all data sets captured (PSU's) were out of synch.
The next listing will be an implementation of the same system with improvements in all areas. An improvement to the paper-device link is a method that forces the fieldworker to check the number displayed on device and printed on paper is similar, every single time a point is captured. This is done by a choosing the correct number from a drop down list of four numbers. Only one is correct, and the fieldworker cannot continue without choosing the correct number.

The biggest problem was that not all places had a sub meter signal. We obtained about 80% of the points when compared to the paper records. So for the majority of the survey, we were able to obtain a green spot and capture a point. High buildings, trees and valleys weaken the signal so it will not be possible to be sub meter accurate. At that stage there was nothing we could do about the red dot or zero points being captured. However, for the next round of development we have decided to implement a sliding system that would enable fieldworkers to first capture a sub meter point then if no sub meter accuracy is found in that place, the system will look for the next best accuracy with it being recorded in the attribute table what level of accuracy was recorded.

All problems were relayed to a technical support section and a unique reference number was created for tracking calls in case a fault needed to be reported to a supervisor if they were not satisfied with the help received from the technical support staff.

GIS

One of the most important phases in a Census lifecycle is the preparation phase which includes the division of the country in to manageable workload units. These units, called Enumeration Areas (EAs) are demarcated according to certain specifications to ensure an optimal environment for planning and the execution of downstream processes, especially the enumeration phase.

According to Richta and later Bloomfield [1] [2], technology evolves in three stages: tools, machine, and automation. This evolution, he says, follows two trends: the replacement of physical labour with more efficient mental labour, and the resulting greater degree of control over one's natural environment, including an ability to transform raw materials into ever more complex and pliable products.

This principle also applies to the Stats SA Enumeration Area (EA) demarcation process. Since 1991, through 1996 and 2001 Censuses right up to the current demarcation phase for census 2011, the process of EA demarcation has evolved from a mainly human interpretation of the demarcation rules using paper copy maps and hand drawn EA boundaries, to a more technology driven application of the demarcation rules from a well stocked digital geo-database.

Bloomfield, Masse. *Mankind in Transition; A View of the Distant Past, the Present and the Far Future*, Masefield Books, 1993.
Bloomfield, Masse. *The Automated Society*, Masefield Books, 1995.

The Past

The mission for the Geography section for Census 2001 was to improve what was started for Census 1996: to make it more accurate. Many lessons were learnt from Census 1996 and all suggestions from Users were taken into consideration. There was much technological advancement since the 1996 census. The task to accurately demarcate EAs is a huge one and therefore it was divided into 14 sub-processes. 20 Meter resolution satellite imagery was purchased for certain traditional areas and the urban fringes. The 1996 EAs were overlaid on the satellite imagery and through a process of inspection, areas of change were determined. These changed areas triggered the acquisition of more detail aerial photography and guided our field teams. This was known as the ‘Change Detection’ process and was explored for the first time this census. The ‘Data Collection and Integration’ process was the first in a string of processes. Through this process not only did we acquire the data but Stats SA has built close working relationships with various data suppliers and data producers. Through the data integration process we as Stats SA have excellent knowledge of various Government and private sector data and the integration of various data sets, received in a variety of formats and standards. It had to be integrated into one standardise system that posed challenges of its own, but the demarcation methodology required this and we very soon found ourselves having a very huge integrated spatial database. The 90 odd page demarcation manual, containing several demarcation rules, had to be converted into an automated process that facilitated the accurate and efficient updating of EA meta-data. Utilities were built to ensure that the integrity (no duplication, no overlaps or under shoots, no area in the country left out) of data captured was maintained at all times. Some of the EA attribute-data captured was EA-Type, EA-sub-Type, Geography Types, dwelling unit estimates, entities on the EAs like institutions, farms, etc., contact details of persons, etc. The EA history tracker kept the history of changes from the 1996 EA to the end-point of arriving at the 2001 EA. This could be used to compare the two census data sets. The 'Data extraction' process was put in place to divide the work into manageable units. Due to the large volumes of data in the database, data had to be chunked and passed on from process to process. This enabled the completion of chunks per process. The 'EA Validation' process was the process that assessed the 1996 EA against the latest digital spatial data backdrop from the spatial database. Here we made use of the integrated data received from many sources to take a decision about the EA, to demarcate the EA. If the spatial backdrop data answered the rules posed for an EA, the EA was demarcated on screen using GIS tools. If the digital data could not answer the requirements posed, this initiated the next process 'Fieldwork'. Fieldwork involved GPS capture of EA boundaries and supportive backdrop data like major roads and dwelling points that will enable the enumerator to find the EA. Fieldwork also involved getting more recent photography. A new technique of flying over an area with a video camera in the plane, called videography, was used. This enabled the rapid capture of up to date pictures of the area. A lot of the videography is as up to date as July 2001. As we assess the 2001 EA set now, 83% of the EAs were demarcated using the available spatial data whilst 17% were done by fieldwork GPS or videography. Every EA was thereafter 'quality controlled'. The QA process checked that the demarcation followed the rules, that all areas were covered, that all capture errors like slivers were removed. The creation of Supervisor Units was the next phase. A GIS utility was developed to group together more or less 5 adjacent EAs based on their EA type. 15 658 Supervisor Units covering the entire country was created. Once the Supervisor Units were created, the Enumerators' Summary Book maps were printed. Each book contained 2 maps, one on the cover used for orientation to find the EA and the other the detail map of the EA. A total of 180 000 A3 maps were printed for the books itself. Another 15 658 maps were printed for the Supervisors. 600 A0 maps called the Index Maps were printed showing the EAs, Supervisor Units, Place names: these maps were used in the Regional Offices for census fieldwork planning and census recruitment. Since the summary books are a historical record, the maps were laminated. A total of 90 kilometers of laminating film was used to laminate the maps for the 81 000 books made. Laminated maps were bound to the books and the books were delivered to Stats SA's provincial offices. The task was completed by 28 August 2001. All maps were saved in Acrobat format, which enabled quick response to re-print in the event of lost books.

The process was not hassle free. The major problem was having enough time. We were always fighting against impossible deadlines. The first project plan we did by ourselves show us that we would finish after the census night! Several interventions were put in place to streamline processes, which enabled us to finish on time to allow the Census Listing process to proceed. The database is definitely the foundation and pillar of Census 2001. Recruitment was done against a Supervisor Unit. The Census Administrative System was validated against this database: most importantly we knew how many EAs there were and where they were. During census enumeration, a web-based mapping tool was developed to track the enumeration progress. Maps were plotted twice a day showing regional progress. The database was updated with feedback from Census Listing, Enumeration and the Post Enumeration Survey. The census attribute data was integrated to produce spatial data products for Users.

Table 1 shows the GIS sub-processes, the main short-comings and the main improvements.

TABLE 1

GIS Sub Processes	Main Short-Comings	Main Improvements
Spatial Database	Insufficient disk space. Timeframes and level of staff skill resulted in data errors.	Re-assess requirements. Then get more server space. Continuous maintenance, proper permanent staff structure in place, proper training for staff.
GIS IT Infrastructure	Insufficient disk space.	Re-assess requirements, then get the correct IT system in place, system must be scalable and reliable.

GIS Sub Processes	Main Short-Comings	Main Improvements
Data Acquisition, Loading and Verification	Poor quality data received from data suppliers. Data from suppliers not timeous. Re-sampling of images done more than once. Large amount of storage was needed for images.	Adequate server space. Continuous maintenance. Proper data management.
Data Extraction and Supply	Insufficient disk space. Poor process of data distribution to Provincial Offices. Lack of IT support at Provincial Offices.	Re-assess requirements. Then get more server space. Re-think the process of data accessibility for the Provincial Offices. Investigate web technology. Support services need to be available.
EA Validation	Lack of detail on maps. Recommendations on control sheets confused Provincial Demarcation staff. Logistics of passing maps and validation files between parties was problematic.	Continuous maintenance of the EAs. Direct on-screen demarcation has proved to be faster.
EA Office Capture	Provincial Demarcation poor. Missing information from Provincial demarcation. Conflicting instructions or demarcation thinking. Skill level of staff demarcating. Logistics mis management. EA size too large.	Proper permanent well trained staff. Make use of direct on screen capturing. Methodology should be based on digital approach.
Quality Assurance	Standards across all parties had to be implemented. The QA process fixed errors instead of rejecting due to deadlines. Difficulties with the New Municipal boundaries. Incorrect Place name data. Vacant EAs not demarcated.	Continuous maintenance. Trained personnel.
EA Integration and Re-	Insufficient time.	Continuous maintenance.

GIS Sub Processes	Main Short-Comings	Main Improvements
numbering	Difficulties with the New Municipal boundaries.	
Supervisor Units	Grouping of EAs only based on EA Type, more attributes should be used.	Continuous maintenance.
A3 Map Production for enumeration	Timeframes extremely tight. Map Specs done while maps were been printed. Delivered data was not always correct. Software not able to handle high volume printing. Due to poor performance from previous processes, staff idle.	Time allocation for this process must be more. Continuous maintenance.
A3 Map QA	Not planned for, thus no personnel planned for this processes, personnel shared with other processes. Poor demarcation surfaced regularly, Municipalities of EAs sent back to previous processes.	Proper planning.
Lamination, Binding and delivery of 09-books	Not planned for from the beginning, no personnel planned for this, made use of GIS personnel for a logistical function. Binder not properly fitted on books, thus 09-books got hooked on each other and pulled apart.	Improve binding method. Proper planning.
A0 Index Maps	Resources shared with Lamination, Binding Process. Maps arrived too late at the provinces. No training provided on how to use the maps.	Make use of better technology (web) to get information to Regions.
Map Reading Training	Teaching time too short. Videotape training inefficient. Assessment of whether the training was understood not done efficiently. Problems with data on the maps, misinterpreted by Enumerators.	More time allocated.
Progress Monitoring – Web Tools	Process not part of the planning, adhoc process.	Maps on the web.

GIS Sub Processes	Main Short-Comings	Main Improvements
	Maps produced only for data available and when available. PC used as web server.	Proper planning to include as part of CAS.
GIS Provincial Support	Poor delivery of maps, data, etc. to provinces. No time to too proper QA. Too little time to properly assess EAs. Some provincial offices did not have GIS persons or Demarcation Co-ordinators. Data quality problems. Supervisor Units too large or far apart. Methodology problems.	Proper planning. More time. Better plan for data assess for provinces.

Timelines were extremely tight in this project. It was the first time that Stats SA made use of GIS technology in the pre-census processes. There was no time to really test the processes nor was there time to test any sub-processes. All processes should have been tested for the pilot census; however this was not done.

Since Census Mapping and GIS processes were so integrated, it stands as proof that the two projects cannot be split. However, the merger was not problem-free. The main problem came from not having a manager overseeing both and looking after the interests of both projects.

The project struggled due to ‘poor project management discipline’. Progress reports were not updated; outputs from processes were not timeous, etc. In April 2001 when a proper timeline was setup the project began to show progress.

We worked in a de-centralised manner, with the Provincial element, contractors not on site, etc. This in itself caused many bottlenecks, for data and other material delivery, problem solving, etc.

This approach to EA demarcation was new. There was a strong resistance towards the new approach; a lot of staff wanted to go back to 1996 methodology.

Although every attempt was made by each and every one to make the EA base as accurate as possible in the timeframes given, there are abnormalities that exist. We have to strive to create a credible data collection base and in doing so we will improve our reporting to our stakeholders.

Short-comings/ Problems

Data

- Some EAs having been flagged as complete by Census Mapping and included in the Provincial Census Mapping Monitors' daily total reports as being complete, were in fact incomplete.
- Missing maps.
- Problems with Provincial demarcation resulted in disregarding Province recommendation in many instances and re-demarcating on-screen at head office.
- Conflicting instructions: on one map a boundary following a different path as indicated on a map of the adjacent EA.
- Metros were too large to work with as a unit, they had to be split into smaller entities
- Large EAs i.e. merged to produce huge EAs due to misleading supporting data and lack of aerial photo interpretation e.g. features clearly visible on the ortho-photo were ignored such as vacant EAs with population.
- EA size was often too big.
- Snapping to unclean vector backdrop was often difficult.
- Large datasets slowed down the correction process – some Municipalities had to be split.
- Missing or incomplete or inaccurate attribute data on the status reports.
- Sectional title data not accurate or missing and not updated by the Provincial Census Mapping Teams.

Staff

- Skilled Staff: The nature of the methodology and the fact that everything was linked to a GIS meant the need for skilled, well-trained staff. Both Census Mapping and Census Geography lacked a well-trained highly skilled kernel of staff for this job.
- Logistical miss management regarding the movement of files: Files were shipped from H/O to the Provinces and then back to H/O and on to the contractors. This led to files getting lost or misplaced.

- Provinces fell behind schedule and the digital update had to be performed on-screen without any check or recommendations form the Provincial staff.
- Missing institution data had to be collected if possible – initially the relevant provincial staff was contacted at the Province, but the turn round time was too long (up to 3 or 4 days). Provincial Staff was relocated at the capture points; this was also not efficient so it was more efficient to contact the relevant municipality directly by telephone.
- Missing institution information not derived from the map, other sources of information or from field verification, or simply not attached to the relevant map.
- Flat occupancy data not provided by Census Mapping.
- Little holiday and hotel accommodation information researched.

Skills

- Place names were not always inserted in the right fields: due to misunderstanding and hierarchical confusion by Census Mapping and contractor staff.
- Mixing of EA types e.g. not separating informal/formal; industrial/urban formal etc
- Attribute Data file for large Municipalities took too long to open – slowed process.
- Attribute Data tabs was awkward and did not follow sequential capture logic.

Improvements to implement after 1996:

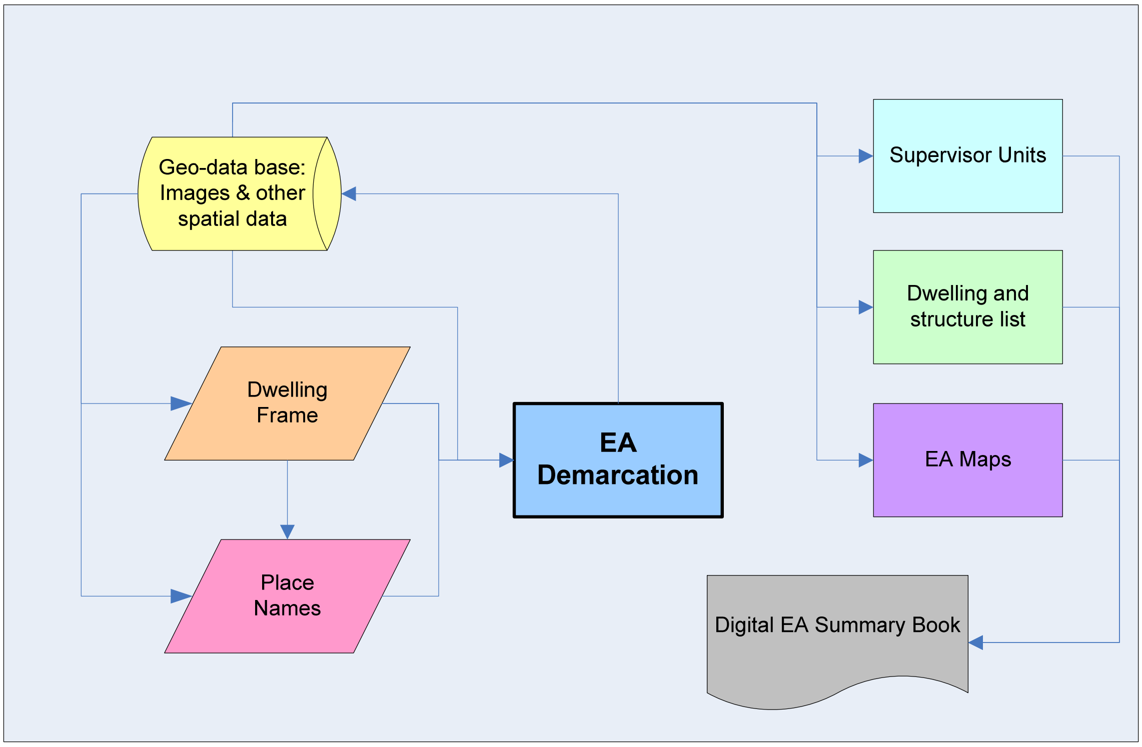
- The set up of a permanent employment structure
- The methodology should be based on a digital approach
- Allocate more time for training
- To reinforce training
- To set up a decentralised census mapping process
- On-screen digitising is faster because of the following reasons:
    1. No EA maps had to be printed by Contractors, which relieved pressure on their side.
    2. Data supply and general logistics from Head Office to the provinces and Consultants and back again was eliminated, which means no "turn around times", data loading or courier mishaps were not slowing down the process.

3. Problem solving was much faster because less people were involved and everyone was at the same place.
4. The Provincial Offices did not have to deal with the receiving, filing, cataloguing, dispatching and demarcation of EA maps.

- Working at H/O was more efficient because
    1. Quality control was made easier because demarcation and quality assurance took place in the same venue
    2. Management of resources was not difficult and less time consuming

The future

Figure 1: Up and down stream processes in relation to Census 2011 EA demarcation



Inputs to the EA Demarcation Process

The demarcation process relies on the following inputs from other components within Geography;

- Geo-database: An up to date geo-data base with as recent as possible imagery, topo-graphic maps and other relevant spatial data (for example roads and rivers). A stable frame to run the automated processes on and no changes to the system for at least 3 years.
- Dwelling Frame: Up to date point and attribute data (including number of units & place names) for all dwelling units and other structures.
- Census Mapping Field operations: Up to date point and attribute data (including number of units & place names) for all dwelling units and other structures.
- Place Names: Up to date place name layers per municipality.
- Project coordination: integrative support for example scheduling, meetings management, documentation management, communication and logistical support.

The demarcation process consists of the following sub-processes:

- Planning
- Application Development
- Creation of demarcation line layer
- Office dot-dot
- Automated demarcation
- Manual demarcation
- Verification
- Field data capture
- Quality assurance
- EA numbering
- EA final update

Planning

The suggested approaches are linked to the development of detailed Operational and Financial Plans.

Application Development and Geo-Database

The Geo-database containing national coverage with a variation of data types and acquired dates.

Applications developed according to specifications for dot-dot process, census boundary layer capturing, automated and manual demarcation, as well as automated EA numbering.

Creation of the census boundary layer

In areas where there are not enough usable line features from current available data sources to use as EA boundaries, line features will be digitized off the most recent imagery and classified per line feature type.

Spatial Point Creation

Limited missing points will be captured digitally in the office, using available backdrop.

Automated Demarcation

Using the dwelling frame, place name layer and census boundary layer, EAs will be automatically demarcated using the developed tool as per EA specifications.

Manual Demarcation

EAs that could not be formed according to set specifications for what ever reason will be flagged for human intervention. If the problem is associated with lack of information, the EA will be logged for field data capture.

Office Verification

EAs that fail to adhere to the specifications from the automated demarcation process and QA will be evaluated by a staff member to indicate the problem and the possible solution.

Quality Assurance

All processes will have quality controls and will be monitored.

Detail quality plans will be compiled and documented.

EA numbering

The same intelligence as in 2001 will be followed in the automated EA numbering system: i.e. a code indicating the province, local municipality and a unique number within the municipality. The numbering process will use the *nearest neighbour* method starting from the most south-western EA in the municipality.

EA updates

Change detection will be done using an automated and visual process comparing coverage of between 2006-7 and coverage in 2009-10, using satellite imagery and other more recent photography.

The EAs that changed since demarcation will be flagged and attribute data regarding the change will be generated for use by downstream Census processes. The boundaries will not be changed at this stage.

Conclusion

A greater number of processes are increasingly becoming digital, and with technology rapidly advancing, the evolution of a digital data collection process for surveys is the next logical step in keeping up with this advancement. The LFSR has in many ways provided (and continues to provide) the testing ground for the use of new technologies that can be adapted for larger projects like a population census.

Statistics South have seen the value of using the basic concept of capturing points in order to monitor progress. However the technology is not limited to merely this and Stats SA is committed to ensuring a move towards a totally digital survey. This is however a long road that needs to be covered and we are just at the beginning of our journey.

A lot needs to be done in terms of research and development and setting up the infrastructure that enables us to get data back wirelessly via GPRS. The security of staff and information also need to be considered. Data will have to be encrypted by the device when it is captured and decrypted on arrival. Thought needs to be given to what will happen if a device is stolen before information can be sent back via GPRS, etc. To date only one device was stolen but data from the stolen device was not lost due it being sent via GPRS.

The use of technologies in a statistics organisation such as Stats SA can provide enormous benefits such as saving time and reducing human error and thus building confidence in the content of the statistical releases made by the organisation. The important factors that will determine success or failure is to get a 100% buy in from all that are involved; that includes the decision makers and the people responsible for the implementation of the technologies, and proper and adequate training for the professionals using the technology.

ACKNOWLEDGEMENTS

I would like to acknowledge all the contributions of all present and previous members of the Geography Division, as well as the former GIS and Demarcation section of Stats SA.