

Field Report:
Second Quality of Life Survey
conducted on behalf of the
Gauteng City Region Observatory
(GCRO)

DOCUMENT NO: 1
VERSION: 0.1

CONTACT: Adhir Nursayhe
EMAIL: adhir@dataworld.co.za

DATE: 09 March 2012

Table of Contents

1.	<i>Introduction</i>	3
2.	<i>Primary data collection tools</i>	3
3.	<i>Piloting of primary data collection tools</i>	4
4.	<i>Data Collection</i>	4
5.	<i>Quality control and quality assurance</i>	10
6.	<i>Conclusion</i>	11

1. Introduction

The Gauteng-City Region Observatory (GCRO) commissioned Data World to conduct its Second Quality of Life Survey, with surveys being conducted in second half of 2011.

The Gauteng City-Region Observatory (GCRO) was established in 2008 as a partnership between the University of Johannesburg (UJ), the University of the Witwatersrand, Johannesburg (Wits) and the Gauteng Provincial Government (GPG), with local government in Gauteng also represented.

The objective of the GCRO is to inform and assist the various spheres of the Gauteng government in building and maintaining the province as an integrated and globally competitive region.

The Second Quality of Life Survey must comprehensively represent the whole of Gauteng, which consists of 10 municipalities, which in turn covers 508 wards. Data World was contracted to undertake 15000 surveys across this sphere.

Among the main aims of the Quality of Life Survey, is to inform the GCRO as well as provincial government and other relevant parties with regards to the perceived states of the municipalities within Gauteng, with focus on the quality of the lives of people who live within these municipalities.

2. Primary data collection tools

The survey instrument (questionnaire) which was used was provided by the GCRO. The instrument was similar to the questionnaire used for the initial Quality of Life survey, with new questions being added only where questions from the previous survey were removed. This was done with the intention of keeping the duration of the survey the same as the initial one.

The survey instrument was a 20 page questionnaire, broken up into 12 sections. The bulk of the possible answers were pre-defined, such that most of the survey could be answered using a combination of tick-boxes or by writing down a number answer from a predefined set. To this end there are not many open – ended questions in the survey.

The survey instrument was reformatted by Data World to ensure optimal flow, as well as to cater for the technology platform which was used to conduct the surveys.

The surveys were conducted using digital pen technology. The technology works in conjunction with a gps enabled cellular phone. One of the primary reasons for the use of this technology is that data is sent directly from the field to in-office databases, with GPS co-ordinates being attached as the location of where the survey was sent from. The technology allows for the storage of actual handwriting strokes, and converts these via HWR (handwriting recognition) modules. This vastly reduces capture time, as in most cases there is no need to fully recapture the hardcopy survey once it is returned from the field, instead the data capturers need to verify that the information has been received and recorded correctly by the digital pen and recognized correctly in the databases. Capture is only required where forms were not sent via the digital pen, or where field collectors have not completed forms correctly, in terms of sending forms once the survey is complete.

3. Piloting of primary data collection tools

Piloting of the survey allows for testing of the actual instrument. Piloting is necessary to ensure:

- Mutual exclusivity of response categories
- Understanding of the questions by respondents, that is minimizing complexity of phrasing while not compromising the intention of researchers when formulating the question.
- Appropriateness of questions
- Certain possible fieldwork errors which are noted during piloting can be addressed during actual fieldworker training.

Initial piloting of the surveys was done in-house by Data World, and factors such as time taken as well as question understanding were analyzed. Further piloting was done over a period of 2 days by the initial field teams, and these surveys analyzed further to ensure correct flow in terms of questions being answered, household and respondent selection.

4. Data Collection

Data collection is reliant on accurate and reliable infrastructure, which includes the following:

- Appropriate methodology
- Selection of suitable resources for fieldwork
- Adequate training
- up to date / real time monitoring fieldwork

4.1 Fieldwork methodology

The sampling methodology used was broken up into 4 stages:

- Generation of random starting points
- Stand Selection
- Household selection
- Respondent selection

Generation of Random starting points

Random starting points were generated at the lowest primary sampling unit level, which in this case was at ward level. An initial required minimum number of surveys required was provided by GCRO, and referred to as the initial sample frame. Data World had worked out that each field collector should have a minimum of one starting point, and based on this it was calculated that each starting point should lead to at least 4 surveys being undertaken. The starting points were generated using distance parameters for urban and peri urban densification, with urban areas having more starting points per unit than peri-urban starting points. As the points were generated randomly, they would not always make sense (points may have fallen in an open field, in someone's back yard, etc). To correct this, a GIS analyst went through the starting points to clean and standardize them as far as possible, that is moving them to the correct sides of properties, moving them to the correct sides of streets. Although the starting points were created at a ward level, Data World created and used a reduced ward boundary layer, which entailed buffering inwards of 30 metres. This in effect created a layer which had voids for the outer 30 metres of every ward. This was done prior to the creation of the random points, with primary reason being to minimize the likelihood of a field worker crossing over ward boundaries in error while conducting interviews.

Once checked and corrected, each starting point was geocoded to obtain a gps co-ordinate as well as street address. These were used in the data collection maps which were allocated to each field team, so as to ease the locating of the starting point. Each starting point was also allocated with a directional arrow, to show the field teams what direction they would need to go in once they had arrived at the starting point. The directional arrows were created individually, on the basis that fieldworkers should as far as possible not cross each other, and very importantly to keep field workers away from the boundaries of the ward (hence minimizing the crossing over into other wards as far as possible).

All the data collection maps were set into batches per area, with overview maps created as well. Each map was created with aerial imagery as a background, and field workers trained how to use the aerial views as landmarks when conducting fieldwork. Each data collection map contained the ward number as well as municipal code for the area being sampled. Cover sheets were created for each map, and these contained starting point numbers, GPS, address, as well as the number of interviews required per starting point.

Stand Selection

Stand selection is the process whereby the actual properties to be sampled are selected. These were thoroughly communicated to fieldworkers during training, and verified by plotting gps co-ordinates of received surveys. The fieldworkers were instructed to always use their initial starting point as the first stand selected for survey, and thereafter count 5 stands in the direction of their provided directional arrow, crossing over the street at the 5th stand. This was done to allow for the surveys to be spread out as evenly as possible, as well as allow for both sides of a street to be surveyed as far as possible. Fieldworkers were further trained on stand substitution, to follow a set of rules and allow for other stands to be used (ie not every 5th stand being surveyed).

The accepted substitution from GCRO was that in the event of an initial starting point stand being unoccupied, a stand to the immediate left or right of it could be used. This would in turn lead to a shift in the 5th house rule. For subsequent 5th house selected stands, if the household is unoccupied or required respondent not available after 3 attempts, substitution to the left or right of the 5th selected stand was allowed.

In the event of the sampling area being a block of flats, the 5th house rule was applied by treating every flat as a separate stand, in essence meaning within a block of flats every 5th house would be targeted for interview until the required number of surveys for the starting point was achieved.

Household selection

Household selection was required in cases where there was more than one dwelling on a stand, or more than one household occupying the dwelling unit selected.

This was done using a dice rule, where the fieldworker would the 1st dwelling / household on the left as number 1, second as number 2, etc. Rolling of the dice would indicate to the collector which dwelling / household needed to be interviewed. In cases of up to 6 stands this is fairly straight forward, as there are 6 numbers on the dice. In cases of more than 6 stands, the dice would be rolled more than once, with both dice scores being noted. The combination sequence of these dices scores would indicate which household should be interviewed.

As with stand selection, substitution was allowed when the required household / respondent was not available after 3 attempts.

.Respondent selection

The target respondent criterion for this survey was respondents within the household who were over 18 years of age, except in the case of child headed households, where the head of the household would be interviewed.

The target respondent was selected using the birthday rule, which would entail listing the names and birthdays of all household members who were 18 years or older. The birthday rule is applied by considering the date of actual fieldwork, and then calculating which household members birthdays is the closest date after the date of fieldwork. Substitution was allowed, by picking the household member whose birthday was the 2nd closest, 3rd closest, etc in cases where the selected respondent was unavailable after 3 attempts, or if outright refusal was an outcome from the selected respondent.

4.2 Recruitment and training of field resources

Recruitment

Field workers were selected based on a set of criteria included :

- Residing within Gauteng
- Familiarity with the areas being surveyed (fieldworkers were broken up into teams based on areas they resided in)
- Communication – field workers had to be fluent in English, and at least one other official language (most of the field workers could speak English and at least 2 other languages).

Initially 40 field staff were recruited, with this figure being increase to 60 staff within the first 3 weeks of fieldwork. A further 50 staff were recruited during the second month of fieldwork, bringing a total of 110 field staff – which included 17 team leaders / supervisors. These counts exclude resources from Data World who would go out on the field to do further monitoring as well as provide field support (there were 5 additional resources who worked on this).

Training

Training was conducted over a 2 day period for each batch of fieldworkers. Field workers and team leaders were trained on completion of the survey in detail, including how to further explain questions and the application of skip scenarios. During in-office training, field workers would survey themselves as well as each other. They were rigorously trained on stand, household, and respondent selection.

Training manuals were provided to all field workers and team leaders, which also included descriptive examples of the “code by observation” questions, such as house type and material of construction. The training also included items such as etiquette and safety issues. A separate training manual was provided which provided information as well as support for use of the digital pen technology being used.

For training on the digital pen technology, completion of the survey was explained in detail with regards to flow required, which entailed the basic requirements for a survey to be sent successfully. This included that the collector always had to ensure they had recorded a survey number, as well as ensure they ticked the “Send Form” box once complete.

The client (GCRO) attended the initial training session which was held in Edenvale, Johannesburg, and provided feedback as queries were raised by the field workers. This also provided a degree of clarity and further understanding to the field workers, as they better understood the requirements of the survey as well as how to go about explaining the content and questions further if required.

When further training was provided to new teams, Data World provided the training in full. All the training sessions consisted of in-office training as well as field training which was conducted after in-office training, where by the collectors would go out supervised and conduct surveys in sample areas.

4.3 Fieldwork

Fieldwork commenced on 15 August 2011. Initially there were 5 teams, consisting of 9 field workers and 1 team leader / supervisor per team. The teams generally were broken up 50:50 in ratio of male and female. The second fieldwork teams began on field on 01 September 2011, these were an additional 10 field workers, who were broken up into 2 teams of 5 each. The third fieldwork team began fieldwork on 17 October 2011. These were an additional 50 resources, which were broken up into 10 teams of 5 each.

Fieldwork was initially conducted from Mondays through to Saturdays, from 10:00am until 7:00pm. The reason for the late finish time was to allow for more household heads to be included in the selection, as working class people are available more after normal office hours. This was the same reason that Saturdays were used for surveying as well. From The middle of October 2011, the final field team began doing surveys on Sundays as well, however this is not very well received by the public, as many feel that Sunday is their day of rest, or their day off.

A set of "mop-up" teams was assembled by reassigning fieldworkers, as the major work began tapering off, to revisit wards and make up shortfalls in the number of required surveys as best as possible. The first mop-up teams (10 field workers) began in October 2011, with another 20 field workers conducting mop-ups from the November 2011.

The bulk of the fieldwork was complete by the end of November 2011, with a few wards being visited in early December as initial visits, and mopping up going on until 15 December 2011.

Access and other issues

Data World ran a quarter page advertisement in the Saturday Star newspaper which contained details of the survey, as well as contact information for verifying identity of fieldworkers. The advert also contained a picture of the typical field worker, which depicted the uniform they wore consisting of a branded jacket and branded yellow bib / vest worn over their shirt. Even though this awareness was created, access was noted as one of the major issues in conducting surveys.

This was further compounded by the Census 2011 which was being conducted, as there was a series of advertisements run which warned households to not allow other people into their properties. Whilst a lot of effort was spent on distinguishing and differentiating between the Census and Quality of Life surveys, many respondents would simply deny access, or if they allowed access would then question why similar questions were being asked between the surveys.

The issue of safety and respondent perception and willingness to allow strangers into their home was an issue, and attempts to send in teams of different racial composition also did not help much. It was noted that there was more willingness of lower income ward respondents to allow fieldworkers in, with refusal being higher in middle income areas, and the highest in high income areas. Higher income areas also tend to have higher walls and tighter security, which makes access even harder. Wards which contain security complexes and apartments are also hard to gain access to, as there is unwillingness to co-operate on the part of the body corporate and estate managers.

Suggestions were made by various communities, which included that fieldworkers should get police clearance from the police station in the area they were working in. Many communities also do not want to accept a general letter which details the survey, and would force field teams to have letters drafted which pertained to their actual ward numbers and suburb names.

In certain cases, field teams were arrested for being within boomed off or gated communities, and even chased out of suburbs by residents brandishing firearms and other weapons. There were also certain wards where outright refusal was given, or respondents and people in charge would tell the field teams that they can only come back next year to complete the surveys.

Many respondents also say that they are not interested, or say that they do not have the time to complete a 40 minute survey.

All of the above issues have an impact on final surveys completed per ward, and even with an oversampling factor of 33% being applied, and a total of almost 25000 surveys being attempted; certain wards have few surveys and result in high weightings being applied.

Specific cases and wards with issues:

74202001 – This ward is situated around the Vaal Dam, and is made up primarily of holiday homes. As a result, where dwellings are occupied, they were occupied by either domestic workers who do not reside there, or holidaymakers who are from other areas altogether. This ward was excluded from the sampling and subsequent weighting due to this.

79700080 – Field worker was held up and attacked, hardware (digital pen and cellphone) were stolen.

79700100 – field workers were forcibly removed from the area, area is made up predominantly of complexes and boomed suburbs.

79800020 – field worker was caught in the middle of a shooting where 2 people were killed. This was in the middle of an informal settlement in Soweto.

79900093 – Ward contains prison, access was denied to the prison area. Surveys were conducted in other parts of the ward.

79800096 – Ward made up predominantly of estates and complexes, access denied even after negotiation with estate management.

79900101 – ward is made up mostly of vacant and agricultural plots. High refusal rate as well as limited households to select from.

79800112 – ward is made up mostly of complexes and estates, access denied by security.

74803001 – mining complex made up mostly of vacant land and mine dumps. Access was also refused by mine management.

79700099 – mining complex made up mostly of vacant land and mine dumps.

79900098 – high income area with security patrols forcing fieldworkers out, high refusal rate

79800097 – area is made up of complexes and smallholdings, collectors were mostly denied access

79900104 – low cost and informal area, collectors were chased out by residents

74202004 – area made up mostly of vacant land and smallholdings

79800104 – area has mostly high income and boomed off streets, access denied

79800102 – area has mostly high income and boomed off streets, access denied

79800012 – area is in Soweto, field workers were chased out by residents

79800090 – middle to higher income area with boomed streets, collectors were denied access, high rate of refusal

79800076 – this area is in Alexandra township. Field workers were chased out by residents.

79900059 – this area is made up of about 50% vacant land. Where there are households these appear to be middle to high income with refusal. Field workers were also asked to leave by security.

74804024 – this area consists mostly of vacant land and farm plots.

5. Quality control and quality assurance

Data World made use a variety of quality assurance and control procedures at the various stages during the cycle of the surveys.

In Field checks:

Field workers would check that their surveys were complete and skips applied correctly. Team leaders would check field worker surveys at the end of each day on the field, and mark possible issues before they are sent back to supervisors.

Supervisors conduct a check on the surveys to ensure completeness.

Random field checks on teams and field workers were conducted. These were done either by contacting the team leader prior to the visit, or by tracking the field workers movements (using the gps co-ordinates being sent by the digital pen and cellphone when forms were sent). These surveys would then be checked on the field by asking the respondent a few of the questions from the survey they answered, to confirm accuracy.

The field checks would also take into account stand (5th household selection) to ensure it was being applied as best as possible, as well as household and respondent selection.

In addition to the checks conducted by Data World, a QA team headed by Ross Jenkins (on behalf of GCRO) also conducted field checks, with mostly positive results. Where issues were raised, retraining was provided, and if necessary, problematic surveys were removed from the sample.

In Office checks:

Once forms are returned to the office, they are checked against electronic database to confirm completeness. If forms have been sent successfully via digital pen then they only go through a partial capture process.

Forms which were not sent successfully via digital pen go through a full capture process.

Both the partial and full capture process contained geocoding map screens, whereby the capturer could search for the address to verify the co-ordinate which was sent, and to fully geocode where a co-ordinate did not come through correctly.

Quality assurance was also done via telephone, following a similar process to field QA, in that a set of questions would be checked against what was collected by the field worker.

6. Conclusion

The overall data collection was successful, bearing in mind the large sample size and area under consideration. The key issues which were noted and affected both sampling as well as timeframes were access, as well as the high rate of respondent refusals. The total sample size of successful surveys was 16729, which exceeded the requirement of 15000 surveys. The successful surveys have all been through rigorous QA and back – checking processes, which gives a high level of confidence in the quality and accuracy of the data. We are confident that the collected data will result in informative and meaningful reporting and analysis for the GCRO.