Experience – Training & Pilot of Questionnaire

The 2017 After Access Survey was conducted in 10 African countries: Ghana, Kenya, Lesotho, Nigeria, Mozambique, Rwanda, Senegal, South Africa, Tanzania and Uganda. This report is a brief summary of fieldwork activities. The report focus on few countries where significant challenges were experienced.

South Africa

The South African training and piloting was kickstarted on the 3rd of May 2017. The survey interviewed 1815 valid household and 400 valid business across 79 Enumeration Areas (EAs). The fieldwork was done by initially six teams of fieldworkers, each team consisting of four enumerators and one supervisor. The teams were based in Cape Town, Johannesburg, Pretoria, Polokwane, Durban and Mthatha. Research ICT Africa in conjunction with Ikapadata, the consultant, provided training in Cape Town and ikapadata in the other locations. Each training workshop typically lasted four days: two days for going through the instruments and role-play, one day for fieldwork logistics and protocols and one day for piloting and a final feedback session.

Electronic devices

Backend development: This involves the scripting of an automated import, airtime distribution, quality control, data cleaning and reporting (Slack and online dashboard) solution.

Questionnaire development: The survey used SurveyCTO which, like Ona, is based on ODK. The same XLSForms (almost) was used interchangeably across these platforms.

Each enumerator was equipped with an Android tablet, an external GPS device, print collateral to show gatekeepers, a T-shirt and hat with the RIA logo, print versions of the instruments and show cards.

Listing

The listing was done using an electronic listing form. A new listing form for every structure was generated and capture every household within a structure. Using the generated information a script that creates a randomly ordered list of all households and one of all businesses is run. In the case of households the first 24 households on the list are the ones that must be visited; replacements start at row 25 and must be visited in consecutive order.

These lists are preloaded onto the server, meaning that all the household info will be available in the household and business forms. The forms containing household and business information which includes EA, Structure ID and Unit ID. These information is complemented by an uploaded list of Google Maps. The Google Maps is sent to the entire team so they can track down the households using GPS. The entire process, from team leader sending us the notification and us sending back the sample list and link, takes less than 15 minutes.

Fieldwork

Fieldwork was scheduled to last 10 weeks but in the end took 14 weeks between the 15th of May and 21st of August 2017. The main reason for the delay was the longer-than-anticipated average time spent in each EA. The original fieldwork schedule and budget was based on the assumption that a team of four fieldworkers and one supervisor would do the full listing, 24 household interviews and approximately 6 business interviews in an EA over the course of two days. Using the timestamps associated with each listing and interview we now know that the average EA takes three to four days days to complete, which has significant consequences for fieldwork planning and costs.

The other major challenge having an impact on the fieldwork schedule was the lack of businesses in the sampled EAs. Based on a response rate of 67%, the fieldworkers would have needed to list at least 10 businesses per EA, on average, in order to meet the target of 500 valid business interviews. In reality, the enumerators only counted 6.6 businesses per EA, with a few EAs exceeding the required minimum by far and many others falling short of it or not having any businesses at all. As a result, the teams had to revisit some of the EAs with businesses that were unavailable the first time around (e.g. because they were closed at the time of the first round of interviews).

Another factor that had a negative impact on survey progress was the amount of racism encountered by the fieldworkers. This is not a new phenomenon in predominantly white and wealthy neighbourhoods and was taken into account during recruitment and assignment of EAs to teams, but it has now deteriorated to a level where fieldworker safety has become a real concern and response rates in better-off neighbourhoods have fallen to extremely low levels. Future iterations of the survey should probably take this into account as one needs to consider using alternative survey modes to compensate for the lack of white, better-off respondents willing to participate in the survey. In some areas such as Johannesburg, the team suffered from internal communication problems and it was dissolved and replaced with half of the remaining members of team Pretoria. Team Mthatha, as predicted, struggled with the remoteness of their EAs- they spent 3 days looking for a particular EA in the deepest Transkei which turned out to be inaccessible by car.

Mozambique

Fieldwork in Mozambique was conducted by Francisco Mabila's group of (former) students and university colleagues. It was done in two stages: the first lasting from 20th of July to 29th of July and focusing on Maputo, the second after a break necessitated by the national census, lasting from 20th of August to 6th of October and focusing mainly on the rest of the country. The enumerators, who were divided into three teams of four fieldworkers and one supervisor, conducted a total of 1224 household and 521 business interviews across 50 EAs. Training was provided by Christoph Stork with the support of Jan Schenk of ikapadata. Dr Schenk returned just before the second survey round for a refresher, based on the experiences and results from the survey in Maputo

Listing

Similar procedure to the South African survey was followed. Online listing which involves the capturing of every household and business in an EA using a mobile form was done. The fieldworkers create a new form instance for every structure and then identify every household and business in that structure by using names, house numbers, visual cues and the GPS location of the structure. In Mozambique they also numbered structures using chalk, but this proved unnecessary in South Africa.

The use of GPS for the listings was crucial as team leaders were notified on Slack (a company-wide messaging app) by our backend system if a fieldworker did a listing outside an EA. In Mozambique, the fieldworkers were warned directly on the device if they tried to do a listing outside the EA, thanks to a geo-fencing solution developed by ikapadata (see https://medium.com/@jan_22556/geofencing-in-xlsform-odk-e892b94164ac for details).

Figure 1: Listing outside EA

The listings were submitted in real time and visualized almost instantly on an online map accessible to the in-office quality control staff as well as the team leaders on the ground. Once a team leader felt that they were done with the listing of an EA, she would contact the quality control staff via Slack. After checking the map, quality control would either ask the team to go back and complete the listing by pointing out gaps on the map, or confirm that the listing was completed. Once the QC staff was satisfied with the listing, they ran a Stata command developed by ikapadata specifically for RIA's sampling requirements. It basically creates two lists of all households and all businesses in a random order and produces the following outputs: two printable lists of households and businesses with addresses, names and descriptions, two CSV files with the same information and GPS coordinates (for uploading to Google Maps) and another CSV file with additional variables (for uploading to SurveyCTO as a preload). QC staff then sent the printable lists to the team leader via Slack, so she could assign households and businesses to the individual fieldworkers. The fieldworkers had access to the Google Map with the listing on their devices and could tap on a household on the map to view the household/business information and receive directions to the location of the household/business.

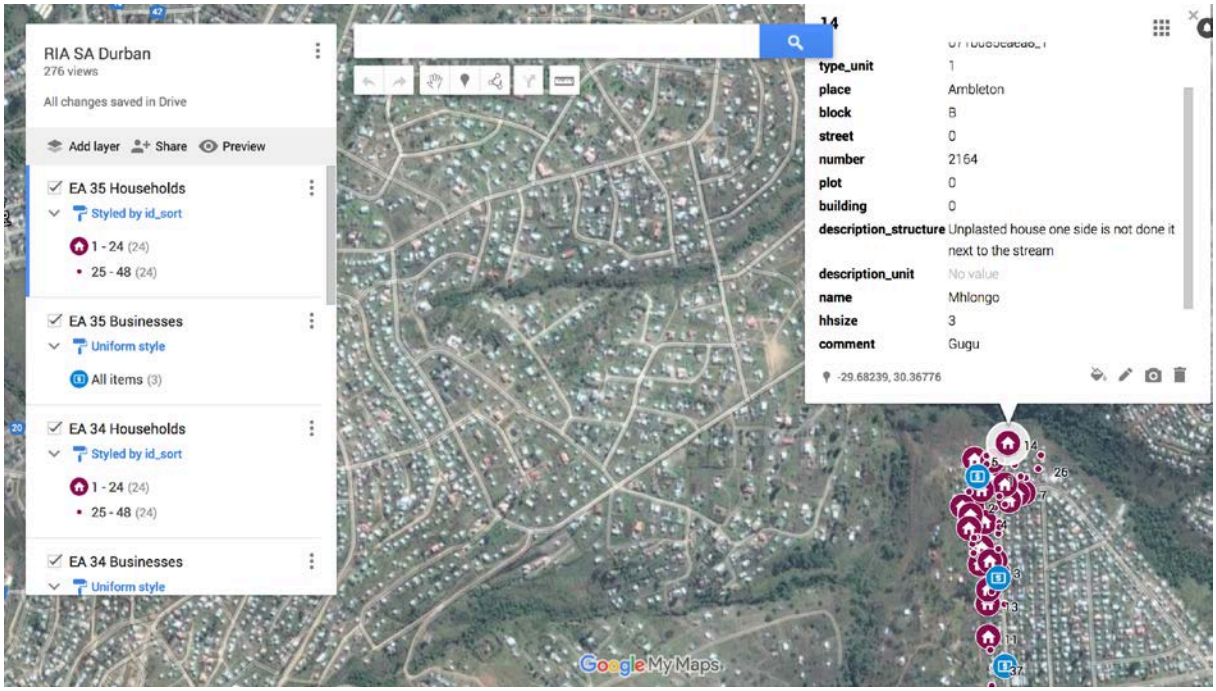


Figure 2: Google Map for Fieldworkers

At the household/business they could compare the information they received from the respondent with the information captured during the listing straight in the household/business mobile form. The form also warned them if their location was more than 20m away from the location where the listing of the household/business took place.

In Mozambique we discovered a grave case of erroneous fieldworker behavior during listing. Some fieldworkers marked households or business which were unavailable (i.e. nobody at home or business closed at time of listing) as "empty" which was an option reserved for structures that were truly empty (e.g. ruins or unfinished buildings). This happened although the use of the "empty" option was explicitly discussed during both rounds of training. Also, this sort of mistake did not occur during the pilot and only started to creep in gradually as the survey progressed, making it difficult to detect. We managed to contain the possible selection bias introduced by this behaviour (basically households or businesses categorized as empty never appeared on the sample lists) by identifying the wrongly categorized households and businesses in the data and even revisiting many of them in Maputo. Nonetheless, the question must be asked to what extent digital listing enabled this sort of behaviour, and why we included an empty category in the first place. The answer to the latter is that we simply copied the options from the paper version and it did make sense to have it digitally, as it allowed fieldworkers to GPS mark empty structures on the map so our quality control staff could see that they did not forget about a section of the EA as otherwise empty structures would have appeared as gaps on the map.

But given the experience we recommend dropping the "empty" option from the listing form and let team leaders and quality control staff communicate to explain gaps on the map as they occur.

Uganda

The Uganda After Access survey was first commissioned in April 2018. Fieldwork was conducted by Knowledge Consulting Ltd (KCL) between the 7th of May and 4th of August 2018. Ikapadata was commissioned with the quality control the Household and Individual ICT Access and Use Survey in Uganda.

Fieldwork

The survey interviewed 1419 valid household and 696 business across 64 Enumeration Areas (EAs). The fieldwork was done by four teams of fieldworkers, each team consisting of four enumerators and one supervisor. The teams were trained centrally in Kampala and subsequently assigned to different parts of the country (Central, Eastern, Northern, Western). Onkokame Mothobi from RIA, Ali Ndiwalana from KCL and Jan Schenk from ikapadata provided the training in Kampala from the 10th-14th of April. The training workshop lasted four days: two days for going through the instruments and role-play, one day for fieldwork logistics and protocols and one day for piloting and a final feedback session.

Each enumerator was equipped with an Android tablet, an external GPS device, print collateral to show gatekeepers, print versions of the instruments and show cards.

Fieldwork was scheduled to last 8 weeks but in the end took 14 weeks. The main reason for the delay was the need to revisit each EA again due to irregularities in the original sampling by the teams on the ground (see Limitations below).

Quality Control and Data Management

Listing

A digital listing solution from previous RIA surveys in South Africa and Mozambique was used. The solution involves the capturing of every household and business in an EA using a mobile form. The fieldworkers create a new form instance for every structure and then identify every household and business in that structure by using names, house numbers, visual cues and the GPS location of the structure.

Quality Control

Quality control consisted of the employment of automised data checks by the backend combined with automised messaging, regular checks by dedicated quality control using the online dashboard and live dataset as well as backcheck calls.

The backend flagged “suspicious” submissions and notified team leaders and QC via Slack based on a series of criteria, including duration, GPS accuracy, distance from listed household/business, duplicate ID/contact number as well sampling inaccuracies.

The online dashboard gave all relevant stakeholders access to the main dashboard showing an online map with all interviewed households, graphs depicting survey progress and indicators of data quality. The team leaders also had access to their team-specific dashboards where all indicators were broken down by team members.

Figure 3: Online Dashboard

If necessary, QC would contact the team leader to enquire about irregularities in the data, and the team leader would talk to the fieldworker in question and sometimes accompany them to an interview to ensure that they adhere to the fieldwork protocols.

Data Management

All interview submissions were sent via the cellphone network after each interview (meaning that the data would still be available even if a tablet got lost or stolen) using Secure Sockets Layer (SSL) technology and stored on a server hosted by the provider of the mobile data collection platform, SurveyCTO.

Limitations and Lessons Learnt

After the fieldwork teams had visited all EAs, the quality control team noticed that the spelling of respondent names, which are purposefully captured twice (household roster and at the end of the individual interview) was often identical, even in cases where it included unusual symbols, capitalisation or spaces before or after the word, indicating that the names might have been copied and pasted between the household roster and the end of the form. On further investigation, it became apparent that the response rate throughout the survey had been unusually high - almost all households had been available and almost all randomly selected respondents had also been present and available at the time of the visit.

After raising the issue with the client, who raised it with KCL, it was decided to send an independent team of fieldworkers, organised by research consultancy IB&C under the supervision of Anicet Munyehirwe, to four of the EAs, to confirm that the information collected by the fieldworkers was correct:

- 34 households were found, available at the time of the visit and all information matched,

- 41 households were unavailable at the time of the visit (a few refused, a couple have moved since the original interview, but most were just not at home),
- 7 households could not be found again and nobody at the location recognised the name of the respondent/family,
- 4 households did not remember any interview and did not know anyone with the name of the original respondent,
- 2 respondents were replaced within a household,
- 8 households had a larger household size than indicated in the original interview (one way to avoid having to replace households is to only include the household members that are present at the time of the interview in the household roster).

The results strongly indicated that there had been serious incidents of non-compliance with sampling protocols across the board, and a decision was taken to send the fieldwork teams back into the EAs to verify all interviews and conduct replacement interviews (using the replacement lists for each EA) for households that could not be verified.

The verification process brought the following to light:

- 1084 of the 1529 original household interviews could be positively verified (stayed in sample).
- 418 households could not be positively verified (dropped from sample).
- 27 households were not revisited (dropped from sample).
- 335 households were replaced (added to sample).

Obviously, the revisiting of all households meant a great expense in terms of time and resources, which is regrettable. To avoid this situation going forward, we suggest the following measures:

- Assign team leaders to fieldworkers they do not know and make remuneration of team leaders dependent of data quality.
- Dispatch regional coordinators to the field, to conduct unannounced spot checks and backchecks with interviewed households.
- Include the overall response rate as a key indicator for quality control.

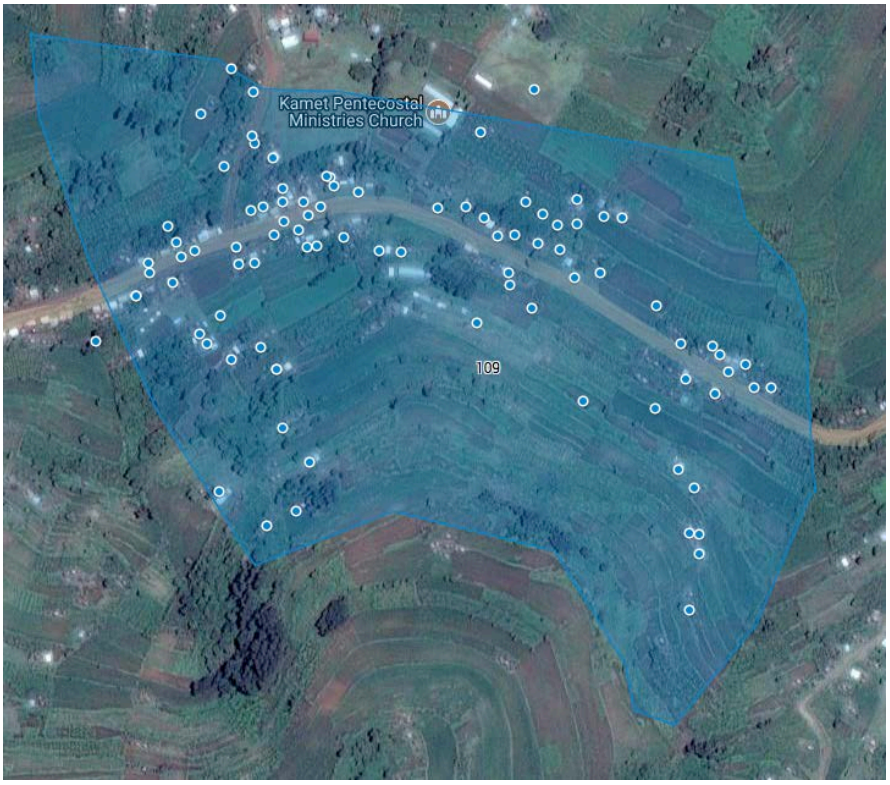
Discrepancies between Survey and Census Data

During data cleaning, significant discrepancies between the number of households per EA counted during the listing and the census figures became apparent - the total number of households in the surveyed EAs is 8908, but only 5853 were counted during the listings. There is no definite answer to how these discrepancies come about, but based on the fact that listings were closely supervised by ikapadata's quality control team, and after revisiting the satellite imagery for the most severe cases, it seems unlikely that they are a result of fieldworker negligence. The EAs with the greatest differences between survey and census figures are listed here, with additional context, so the analysts will be able to make their own judgements on how to proceed with the calculation of population estimates:

EA 109

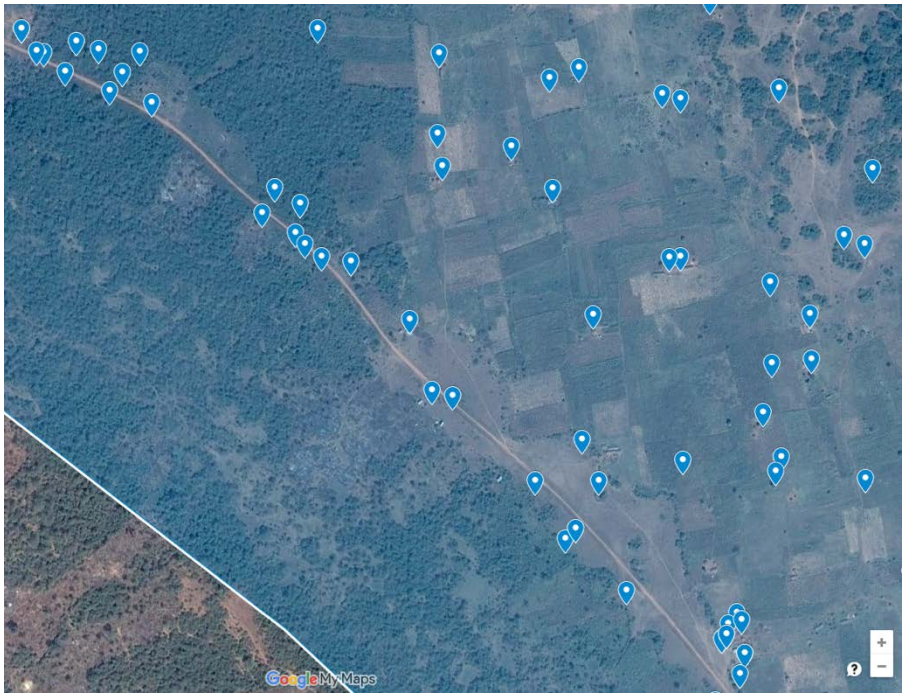
This EA is unusual because the number of counted households (92) significantly exceeds the number of households according to the census (23). There was no reason to question the listing at the time (there was only a request to identify one of the buildings - it was a primary

school), and looking at the map, it is clear that no households outside the EA boundaries were included in the listing.



EA 130

According to census figures there are 609 households in EA 130 - the fieldwork teams only counted 163. Looking at the map, it is difficult to imagine that there are over 400 households that were not counted by the fieldworkers. In fact, many of the dwellings that were recorded by the fieldworkers are very difficult, or impossible, to make out even on the satellite imagery - meaning that the fieldworkers did not rely on satellite imagery alone but counted what they could find on the ground.



While the map shows only a part of the entire EA (the EA is too large and the resolution too low to show the entire EA without losing too much detail), the rest of the EA looks very similar to what is shown here. There are a few huts that seemed to have gone unlisted, but QC enquired with the team leader at the time, and she reported that these were empty as many people had moved away.

LISTING EA 130
Enumerator: Rose
Date/Time: 07may2018 10:22:00
Structure ID: B55
Description: There is a small tree opposite neighbors to kinda Fortune
WARNING: Submission outside EA boundaries. Please contact fieldworker ASAP.

Kobusingye Eve 6:34 PM
@jan we were able to finish listing and are hoping to do the interviews tomorrow morning

There was no network coverage in the EA so I think you are continuing to receive the submissions

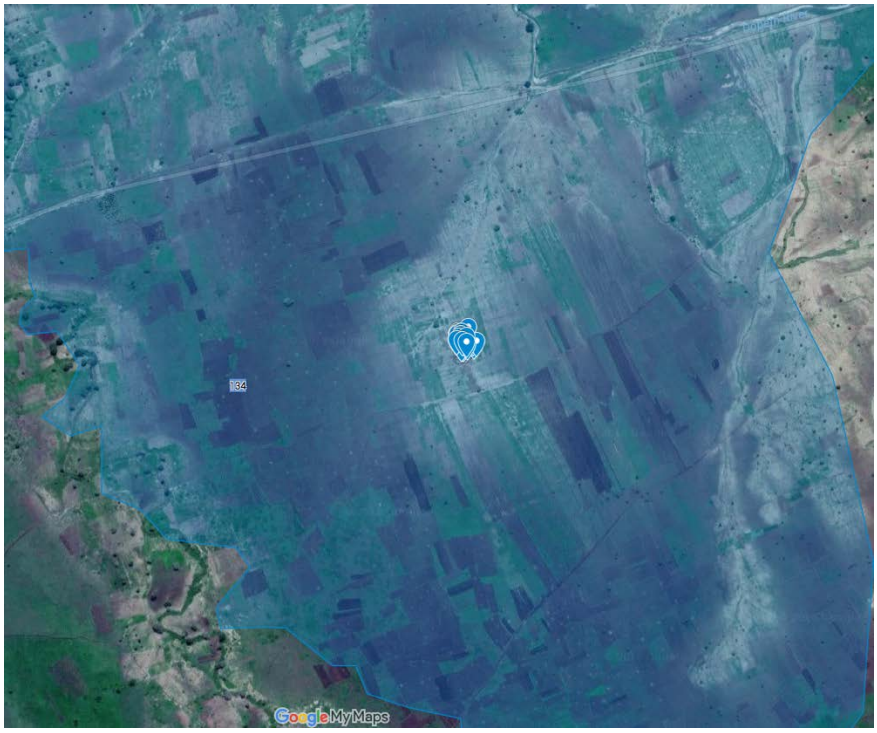
The GPS accuracy was also not good

Accuracy was 10m

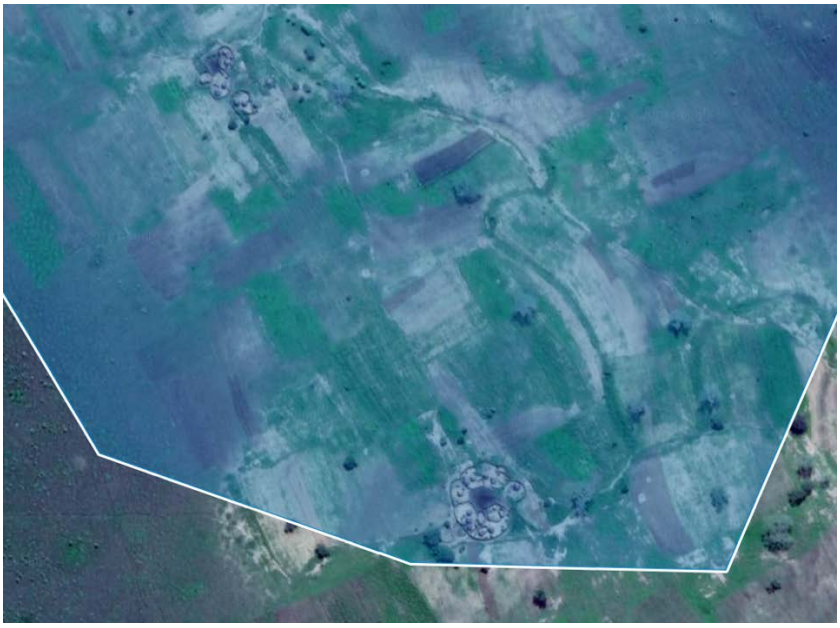
Kobusingye Eve 6:50 PM
Plus many houses were unoccupied because people shifted

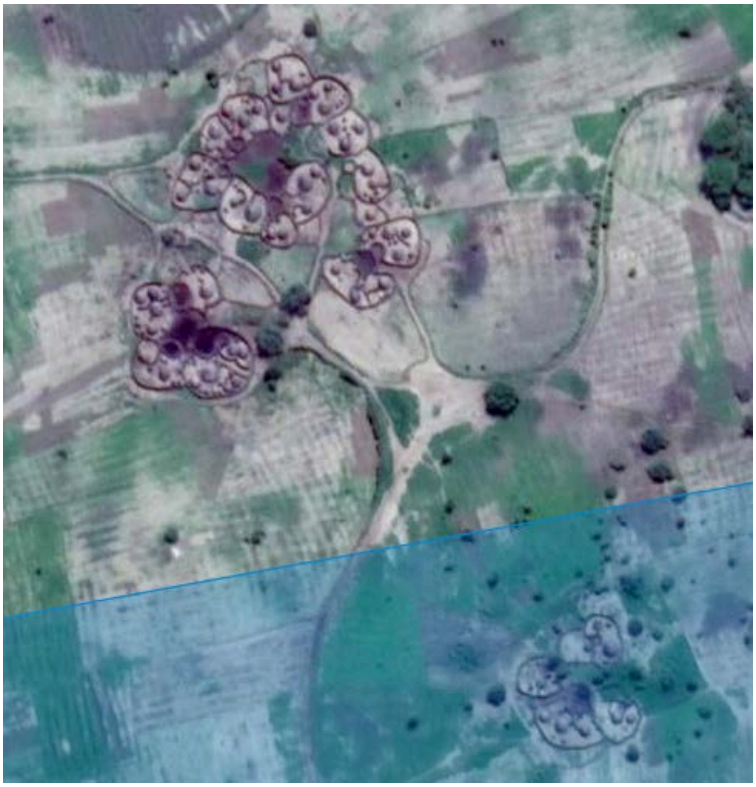
EA 134

EA 134 has 156 households according to the census, but only 22 were counted. It is a very large and remote EA, but all listed households are clustered within a small village consisting of huts.



At the time of the listing, it did seem like the EA was empty otherwise, but a closer look now reveals that there are three more clusters of huts that were not listed.

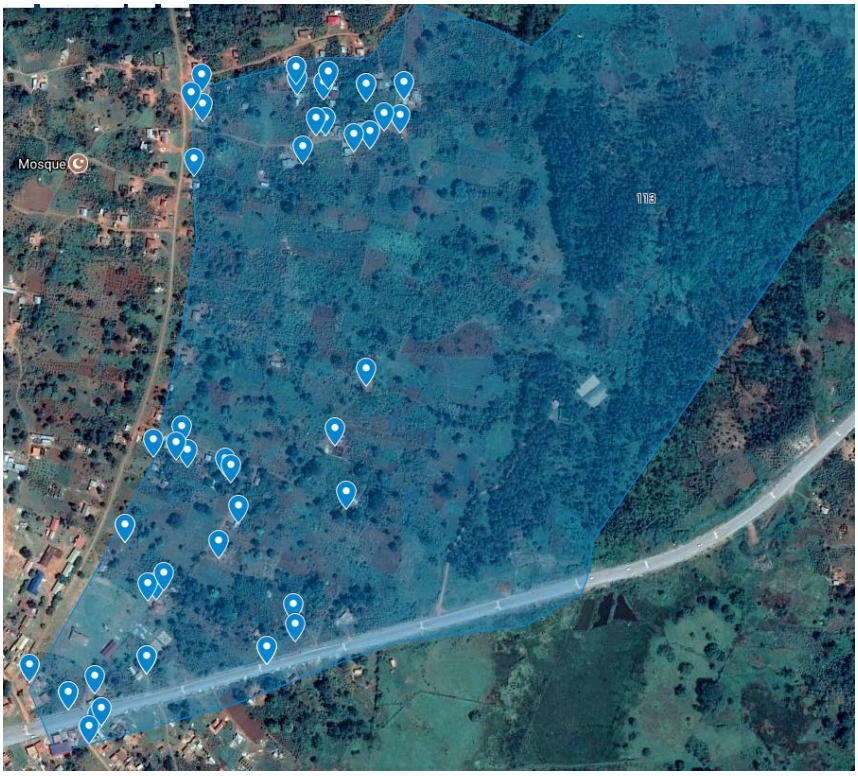




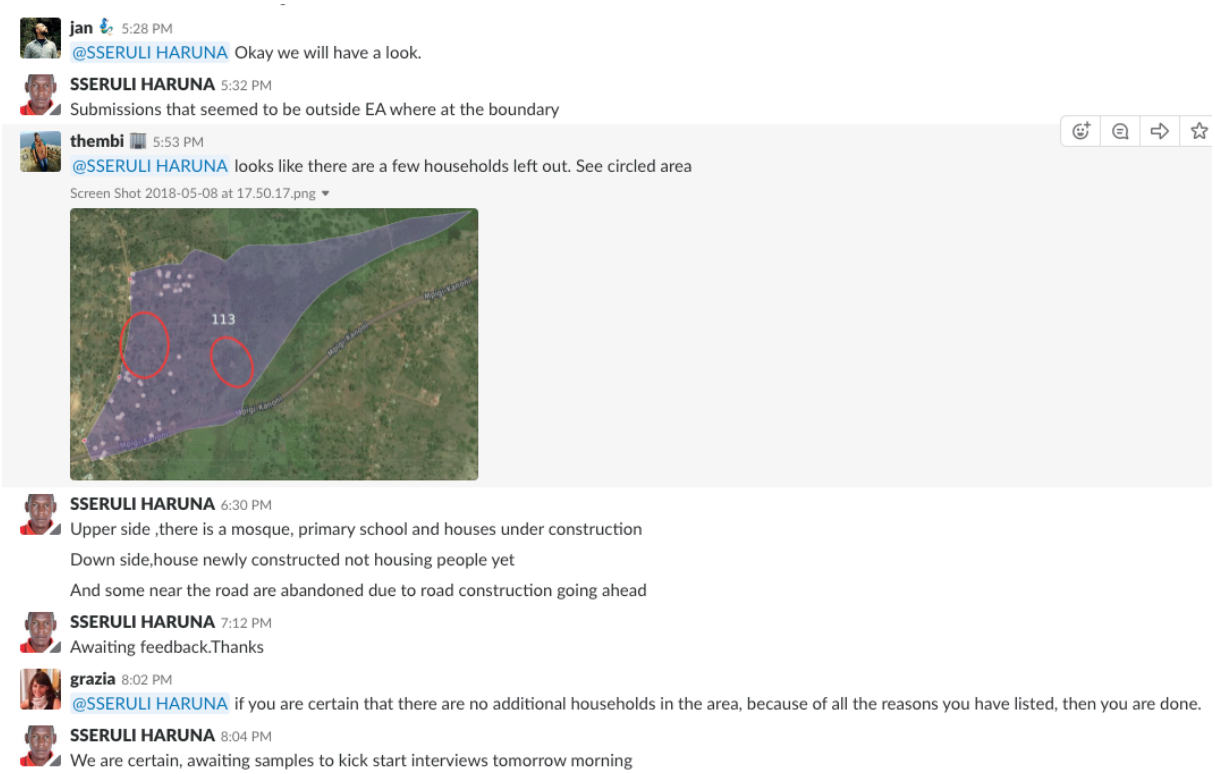
The larger settlement in the last image falls outside the EA. The smaller ones were not included in the sample. However, it does not seem that they are large enough to make up for all of the "missing" 134 households (difference between census and listed households). Short of revisiting the EA, a possible solution would be to assume that the three non-listed villages are of similar size as the one that was listed and base the population estimates on that (i.e. number of households in EA = 4 * 22 = 88).

EA 113

The EA with the largest ratio of census (454) vs listed households (42) is EA 113. Looking at the map there is no obvious explanation for this massive discrepancy.



There are a few dwellings that seemed to have gone unlisted, but the team leader maintained that these were schools, mosques and houses under construction.



Other EAs with large discrepancies follow the same patterns. Sometimes there are a few non-listed dwellings, which according to team leaders are schools, hotel, or empty houses, but they would not explain the difference in numbers even if they were all fully occupied.

Tanzania

The training and pilot of the questionnaire in Tanzania took place in April 2017 in Dar es Salaam. The exercise was successful amidst some challenges which will be duly discussed, and suggestions made on how to prevent them from happening in the future.

Partnership with Stakeholders

During the in-country visit, the team met with the National Bureau of Statistics (NBS), the Tanzania Communications and Regulatory Authority (TCRA) and the Commission for Science and Technology (CosTech). Whilst these stakeholders gave their commitment to support the study, there seemed to be a lack of coordination amongst them resulting in an overlap in the way they would like to support the survey. It was also discovered at the time when the training had started that there was a cost involved in generating the maps and that the NBS is required by the New Act on Data Collection in Tanzania, to provide logistical support and personnel for any survey to ensure quality control. These services would all come at a cost and there was the need for a Memorandum of Understanding (MoU) to work in partnership. This resulted in a lot of back and forth and many interruptions during the training.

To avoid or limit such delays and interruptions to the smooth flow of the training and piloting of the questionnaire, it is necessary for negotiations between the Research ICT Africa team and in-country partner (s) and the stakeholders in the country to start as early as possible. This will allow some of these issues to come up earlier and be resolved ahead of the in-country visit. It is also important to have a round table discussion with all stakeholders present so that there is clarity and overlaps in terms ways to support the survey can be prevented.

Electronic Devices

The devices to be used for the survey are usually set up during the training and this results in delays in training enumerators on the use of the questionnaire and on conducting mock interviews. This was not only unique to Tanzania but was the case in most of the other countries. In future, it is advisable that the devices be procured and set up ahead of the in-country visit.

Senegal

The training and piloting of the questionnaire in Senegal took place in July 2018. A lot of delays were encountered which led to the training starting a couple of days later than scheduled.

Working with National Statistics Offices

The team met with the Director of the Economy and Telecom Markets from the Regulator's office and the Statisticians from ANSD. Whilst the regulator showed appreciation of the work RIA wanted to embark on, they stated that they will not be able to give their hands-on support as the survey was not being conducted by their statistics office, ANSD. The regulator and the statistics bureau were under the impression that the statistics office will conduct the national survey on behalf of RIA. The meeting with the Statistics office at the time of the training revealed that they were not clear on what RIA needed from them. As a result, the sampling and the maps were not ready during the training.

RIA's in-country partner has a vital role to play in liaising with national stakeholders and representing the interest of RIA and providing accurate feedback. The Country Partner therefore needs to understand and be abreast of all happenings related to the survey. It was clear in the case of Senegal that the Statistical Agency wanted to conduct the survey on RIA's behalf, whilst RIA's approach in mainly to engage the Statistics offices in the sampling of Enumerator Areas and the generation of maps. This resulted in delays to the start of the fieldwork.

Pilot

Whilst negotiations for getting the list of EAs was ongoing, the Statistical Agency supplied the team with the maps of 2 Enumerator Areas (EAs), one rural and one urban to conduct the pilot. In reviewing the maps, it was clear that the map of the rural area lacked details of specific landmarks that would help the field team in conducting the listing. It took a while to locate the rural EA as there was a mistake in the name indicated on the map. The map lacked clear boundaries which made it difficult for us to do the pilot.

It is important to have the list of EAs and the maps a few weeks ahead of the training, so that any challenges or issues are raised and sorted out so that the training will not be disrupted.

Liaising with Local Authorities

An incident occurred in one of the EAs in Senegal during fieldwork. One of the residents reported to the nearby police that one of the supervisors was taking photos of the area, which is a requirement and one of the deliverables for verification at the end of the survey. The police confiscated all the EA maps and ordered that the fieldwork be stopped across Senegal until the claims of the field team was verified by the Statistics office. This caused further delays to the project and of course was alarming for the research team in the field.

This indicates a lack of adequate introduction of the survey nationwide. It is important in future that all the local authorities in the sampled EAs are informed and properly briefed about the project and the role of the field staff. The supervisors and enumerators should also make it an absolute must and a priority to report to the police and any other local authority within the EA upon arrival.