CAPI COLLECTIONS IN THE PACIFIC – USING SURVEY SOLUTIONS

27/02/2020

Using Survey Solutions

Using Survey Solutions for CAPI Censuses and Surveys including technical specifications, planning and execution of activities.

Toga Raikoti & Phil Bright
Pacific Community

Contents

CAPI Collections in the Pacific – USING SURVEY SOLUTIONS

SURVEY SOLUTIONS

This document details the steps required to:

- Prepare for a CAPI collection
- Planning and budget
- Conducting training
- Conducting field work
- Data processing – editing and tabulations

This document was written based on the current version of Survey solutions as of the date of production of this document and it is suggested that users refer to https://support.mysurvey.solutions/ for newer information and releases of upgrades to the software.

INTRODUCTION TO CAPI

1. Introduction

CAPI (Computer-Assisted Personal Interview) technology has been widely used in the world by National Statistics Offices (NSOs) for their data collection to improve data quality and timeliness in their statistical outputs. The NSOs in the Pacific Region have been introduced to the technology since 2016 to address the common challenge in statistical collection in the region – collection of good quality data.

Through the collaboration of the Statistical for Development Division (SDD) with the World Bank and the NSOs in the region, CAPI was introduced to a few NSOs for their census and survey collections and this has created more interests and demand in the use of the technology.

Looking back at history, the region had been using paper-based questionnaires for most of their data collection until the 2010 Round of Census when scanning technology was introduced to a few countries. There were some success stories for a few countries while some had a few issues and took longer to process and this leads to the introduction of the CAPI technology, with the intention to adopt it for the 2020 Round of Census.

Some countries had been using CAPI in some of their small surveys and collections through different organizations, but Vanuatu, Tonga and Tokelau were the first countries to use the Survey Solutions CAPI technology for their Population Census. A few other countries had follow suit for other surveys like HIES, Labor Force and Market Surveys.

It is anticipated that CAPI will be widely used in the region and this has been adopted by some countries for their HIES, Agriculture, Disability, Labor Force and a few other smaller scale surveys.

OVERVIEW OF SURVEY SOLUTIONS

1. Why SDD chose Survey Solutions

There are a lot of data collection software and tools that are readily available in the market and some common ones include:

a. World Bank Survey Solutions (SS)
b. Survey123
c. CSPRO Android
d. SurveyCTO

Based on evaluations of these software and meetings/discussions with their respective support teams, the World Bank Survey Solutions software was identified as the best 'Solution' for efficient data collection for the NSOs in the region.

The World Bank named their software 'Survey Solutions' with the impression that it is the best solution for Survey data collection in terms of the data quality checks and monitoring processes involved through the interactions between the Interviewer with their respective Supervisors and the Headquarters in the office.

This interaction is seen as the best solution to the problems faced in the fieldwork of censuses and surveys during a paper-based interview, where supervisors and NSO Census staff have no control on what the interviewers are capturing in the field and errors are only identified when the interviewers have returned the questionnaires to the office.

With the integrated GPS and mapping facility in Survey Solution, Supervisors are able to monitor their respective Interviewers to ensure they are enumerating in their correct Enumeration Area and fully capturing every information that is required.

World Bank providing their cloud Server free was another reason why Survey Solution was selected due to its security features where data is secured and encrypted. Many countries (some with millions population) are using this facility and so far have no problems with the security of their data.

As for data quality, Survey Solutions allows for all necessary checks to be included in the questionnaire design – enabling conditions, skips, logic and consistency checks, and this helps the interviewers during the interview as it ensures that correct data are collected and verified at the same time.

The Support Services provided by the Survey Solution Team through their online Help and Resource website is very efficient and reliable in fixing problems and issues encountered during the collection.

PLANNING FOR A CAPI COLLECTION USING SURVEY SOLUTIONS

1. Equipment/Infrastructure

1) Tablets

1) Survey Solution is designed for tablets with minimum Android 5 and 7-inch screens. Lower size tablets could be used as well but users may find difficulty reading the fonts and maneuvering around the screen. For more information on the tablet requirements could be found here
2) Ideally, tablets should have 7" or 8" screens, which makes them easy to store and carry in the field. A mobile phone screen is too small, and a 10" screen is a little cumbersome (larger screens also drain batteries faster). 10" tablets are relevant for Supervisor roles as the screen is wide and shows the full application, in comparison for smaller screens which provides compact screens of the application.
3) Tablets should allow for at least 3G Sim Cards for internet data
4) Minimum 4000 mah battery. Tablet MUST last at least one full day in the field before charging is required.
5) High quality GPS – both GPS and GLONASS are required, not only A-GPS (this uses the telephone network which is not very good quality and will not work in areas with no 3g connection).
6) Must integrate with Mobile Device Management (MDM) software – which allows for tablets to be remotely setup and monitored, as compared to setting up 2500 tablets manually (roughly 40 person days of work).
7) Good quality aftermarket ruggedized case with integrated screen protector.

2) Bulk SMS software (Frontline SMS)

1) This is another technology which could be organized with the service providers to allow for easy and quick communications between everyone involved in the Census operations – Office, HQs, Supervisors and Interviewers.
2) It allows for the sending of text messages to everyone at the same time if any information is to be conveyed to those in the field. Information on changes in processes, logistic and administrative arrangements, application updates, tablet issues and so forth could easily reach everyone and ensure everyone is aware of what is happening.

3) **CUG** – Closed User Group allows for everyone involved in the Census to make free calls among each other. This enables sharing of information and addressing issues faced in the field. If this is used then Frontline SMS can be essentially used for free if one of the sim cards in the CUG is used.

4) 3G Sims for data

1) There are various options of what 3G Sim Data Cards could be used and this can be determined by the budget available for the cost, and most importantly, the questionnaire size – the longer the questionnaire, the more data usage.
2) A few options that have been used by countries include having fixed size data cards – and the cost is charged on the fixed data, irrespective if the data is fully utilized or not. This is then topped up if the interviewers used up the data. Another option is using pooled data where they are charged on data usage or they are charged a standard rate with unlimited data.

3) Estimating questionnaire size –this could be checked using android data use settings or installing android app such as Data Usage Monitor. This will show and monitor the data usage during interviews – a good apps to see how data is being used.

2. Budget considerations – cost of tablets

1) Considering what has been explained above, the cost of tablets will require necessary budget planning and decisions. The budget plan should include all these additional costs for these supportive applications and equipment.
2) A big chunk of the budget will be on the cost of the tablets, so a few things will need to be decided first to fully determine how many tablets to be purchased. These will include:
    a. How many Enumeration Areas (EAs) to be covered. The ideal scenario is for one interviewer per EA, but in some cases, countries have assigned more than one interviewer per EA, which we do not recommend. For these cases, then the number of Interviewers determines the number of tablets.
    b. Ratio of interviewers per supervisor - This will determine the number of supervisors that are to be appointed.
    c. Number of Headquarters – if they are going to use the tablets for checking questionnaires
3) Purchasing of tablets takes time, based on the quantity to be ordered and it is highly recommended that the budget for the tablets have to be made available quickly so procurement orders could be processed. Experience from countries that have used tablets shows that it takes about 5-7 weeks to order tablets (depending on the quantity)..

3. Interviewers/Supervisors/Headquarters Numbers

1) The number of Interviewers, Supervisors and Headquarters has to be determined based on the number of EAs in the data collection.
2) The ideal ration which is based on the workload of Supervisor checking the completed questionnaires from the Interviewers, together with the expected number of rejections and length of questionnaire, it is suggested that one Supervisor should be assigned at most 3-4 interviewers. Having five or more interviewers per supervisor will be too much workload for the supervisors and will result in delays in approving questionnaires and additional more work to the interviewers to verify rejections with the respective households. Depending on the length and complexity of the questionnaire, the number of interviewers could then be increased, for example, in a Listing collection with only a few questions, then the number of interviewers in the team could be increase to 4-6.
3) In the Survey Solutions HQ application, HQs does the final checking of the questionnaire, hence, the number of HQs will depend on the number Teams to monitor. To ease the workload of the HQs, it could be organized on how many Teams to be checked by a specific HQ.
4) Depending on what headquarters will do, they can spend as much time as a supervisor checking the questions. So 3 Teams for 1 headquarters will be impossible for the HQ to handle so it is always recommended to limit the number of Teams for each HQ.

4. Recruitment of Interviewers / Supervisors / Headquarters

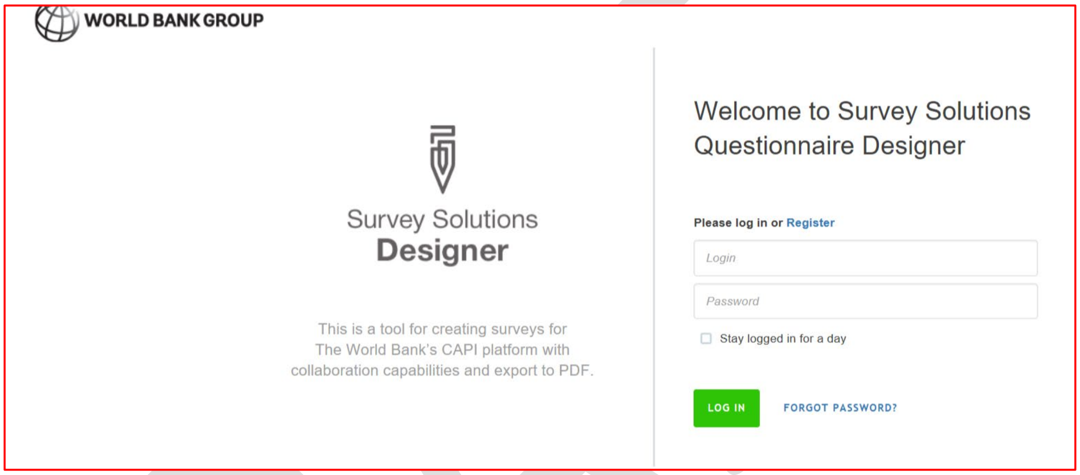
1. The use of technology requires knowledge and expertise in this field. It is highly recommended that people recruited should have at least some knowledge on how to use a tablet – the different functions and buttons available, configurations and simply maintenance and safekeeping.

2. They should also have good communication skills, good understanding of the questionnaire and concepts and are familiar with field enumeration operations.
3. With their knowledge of the technology, they are able to share and assist other colleagues in the field in addressing issues faced.

DEVELOPING QUESTIONNAIRE WITH DESIGNER

1) Questionnaire Designer Application

1) You will need to register into the Designer to use it. Go to the link https://solutions.worldbank.org then click on 'Register'.



More information on registration could be accessed here

2) The Designer is very user-friendly where you have the freedom to design your questionnaire to suit your need and requirements.
3) The most important thing is to ensure that there is consistent flow of information in your questionnaire with clearly defined enabling conditions and filtering questions.
4) There are public shared questionnaires in the designer, which you can access and see how other users design their questionnaires.
5) more information on questionnaire design could be accessed here or full documentation could be accessed here

2) Include detail on validation rules

1) There are different ways of writing validation rules to check a particular error and the designer has the freedom to use whichever code is easy and understandable. Some of the validation syntax provided below could be simplified more and will serve the same purpose of what needs to be validated

2) Range checks – these ensures values of a response to a question are within a certain range – anything outside the range will give an error. Categorical questions will have these range checks as the question will only allow for values in the categories, but Numeric fields like Age, ranges have to be included like 0-110 for the age range.

3) Enabling conditions – these are usually referred to as 'skip' questions where a question is only enabled based on a condition or a response to a previous question or a filtering question. For example, for fertility questions, these will only be enabled to all females aged 15 and above. The syntax is very easy to understand: sex==2 && age>=15

4) Complex validations – the validation checks are written in C# language and may require some programming skills for the complex validations. Most of the functions and commands could be readily accessed via internet. As for a Population Census, apart from the normal enabling and simple validation checks, some common complex validations used include:

   a. Relationship
      i. Checking the Head of the Household that he/she is always the first person and should only be one Head
      ii. Age difference between Head and child should be 15+
      iii. Head and Spouse age should be 15+
      iv. Cannot not have more than one Spouse
      v. Head and Spouse sex should not be the same
      vi. Age difference between Head and Parents
      vii. Age difference between Head and grandchild
   b. Education
      i. Checking education level (current and completed) against age
      ii. Completed level should always be lower than current level
   c. Economic Activity
      i. Check on Economically Active and Inactive population
      ii. Occupation and Industry coding
      iii. Unemployed conditions on available and looking for work
   d. Fertility
      i. Check on children ever born against children alive, living elsewhere and died
      ii. Female head who has children in house should at least have children in the fertility questions
      iii. Mother of Head should at least have a child (given birth)
      iv. Mother of Spouse should at least have a child (given birth)

5) More on these validations could be accesses through the PDF version of the Questionnaire generated from the Designer.

TESTING THE QUESTIONNAIRE

1. Testing your Questionnaire with the Tester

1) The Tester Application is a tool which allows you to test your questionnaire while designing. It allows you to see the flow of the questions and the information being collected, the enabling conditions (skip) are working and also the filtering and validation checks are working properly.
2) The tester is only a tester, which basically means that data is not stored and saved but only for testing of the questionnaire.
3) The tester could be downloaded from the Google Play Store in your tablet and installed for free.

2. WB Demo Server

1) The World Bank Survey Solution has its own Demo Server which could be used to test your questionnaire, interviewer, Supervisor and Headquarter processes.
2) The server link is: https://demo.mysurvey.solutions and uses the accounts:
   i. Username – Headquarters1
   ii. Password – Headquarters1
3) This will allow you to create interviewer and Supervisor accounts for testing purposes. This allows the opportunity to see how the full process works – the interaction between Interviewer, Supervisor and HQ on the completed questionnaires as it is synchronised by the interviewers, checking by the Supervisor and rejecting the questionnaire if there is a need for things to be fixed and addressed and approval of questionnaires.
4) Please note that since these are Demo Servers, questionnaires of other users will be also available for access and caution has to be made that your interviewers and Supervisors are fully aware of which questionnaire they are working on.
5) Also an important note to remember that the Demo Server is cleared by World Bank at the end of each month so questionnaires and accounts will have to be uploaded and created in a new month.

PREPARING FOR PRODUCTION SERVER

1. **Make sure all testing is done, and double checked with demo servers before going to production server. All content stays on server even if deleted**
    1) Once you are certain that your questionnaire is ready and fully tested, then you will need to request for the Production Server which will be used for the actual survey or census. This will be the real (live) version of the server and this is normally setup to be used for the Pilot and final training so they familiarize themselves with the Server Name.
    2) There is a standard naming convention used by Survey Solutions for the Server Name which will always be in the form: PREFIX.mysurvey.solutions; where you specify PREFIX. Prefix should be no longer than 8 characters long, should start from a letter and can contain alpha-numerical characters. Examples of prefixes: zimbabwe1, blfs2016, ugandahbs
    3) It is highly recommended that you name your server as simple and short as possible for easy typing and configuration purposes. The longer the server name, the longer it will take to type when the Interviewers log in for the first time, or when there is a need to reconfigure a tablet. A good example of a server name could be ***fjcens17.mysurvey.solutions***, which refers to the Fiji census 2017.

2. **Contacting SPC to make server request (SPC is the Pacific focal point)**
    1) For coordination and monitoring purposes, World Bank, in collaboration with SDD, allows for SDD to be the focal point for Server Request, hence, any request for a Server will have to be requested through SDD to togar@spc.int. But a few countries have requested directly to WB for their server, which is fine but this limits SDD's involvement in providing assistance to the countries if there is any issues faced.
    2) In requesting for a Server, a few information are required for the Survey Solutions Support Team to let them determine the Server for you. Relevant information will include:
        i. Name of Survey / Census
        ii. Survey mode – sample or census
        iii. Sample size if it is a survey as this will determine the size of the server allocated
        iv. Survey start date
        v. Survey end date
        vi. Name of questionnaire in designer
        vii. Specify prefix for the sever – e.g fjcensus17
        viii. Implementing and Funding agency – their websites links is required
    3) Usually, the Server request takes a least a couple of days for the Survey Solutions Support Team to action on and they normally agree to the schedule dates as per each request.
    4) They normally allow an extra 30-days grace period after the End-Date to close the Server which gives ample time for the countries to finalize their Census/Survey to finish off their project.
    5) Survey Solutions also sends out a warning notice when the Server is not being used (idle) for a couple of weeks – this happens in cases where Countries have acquired the Server but then they had delays in their schedules which causes delays in training and testing. This is a good reason to always requests for the Server when everything is finalized and confirmed and also using the Demo Server for trainings and practical purposes.

SETTING UP PRODUCTION SERVER

1. Copying across questionnaire template

1) Once the approval for a server has been granted, a username and password for the Server administrator will be provided to be responsible for the Server.
2) The administrator will have admin rights and privileges for the uploading of new questionnaires as well as deletion of questionnaire, creation of HQ accounts as well as Supervisor and interviewers if required and other admin settings of the Server.
3) Once you are sure that your questionnaire is ready, then this can be uploaded to the Server.
4) The server creates different versions of the questionnaire depending on the number of times the questionnaire is uploaded. The questionnaire will be updated now and then during the training and testing phase, hence, any new upload of the updated questionnaire will create newer versions of the questionnaire. Uploading the questionnaire for the first time will create Version1 of the questionnaire.
5) Please be aware that different versions of the questionnaire could be loaded in the Server but the risk is that Interviewers and Supervisors sometimes get confused on which version they are working on. There could be potential loss of data due to different versions as there will be different versions of the data as well and more work will have to be done to ensure that data are consistent between the different versions.
6) Also note that deleting questionnaire is also risky as Interviewers may still have not synched their Interviews so before any deletion of a version of the questionnaire, ensure that the Interviewers have synchronized their interviews and no questionnaire is open or in 'Started' mode in their tablets.
7) It is highly recommended that only one version of the questionnaire is to be used in the Server and other versions to be deleted.
8) Important to note as well that deleting a questionnaire means all data collected from that questionnaire is also deleted. You need to download and backup the data first before deleting a questionnaire.

2. Setting up usernames + passwords (bulk upload vs one by one) Include links to SS knowledgebase

1) Setting up of usernames and passwords is tedious work especially for a big survey or census as it takes up a lot of time to create accounts individually.
2) Creation of accounts for 20-30 interviewers could be easily created one by one but more than 50, then it is suggested to use the bulk upload tool in Survey Solutions. Information on bulk upload could be accessed here.
3) It is always recommended that the list of Interviewers, Supervisors and Headquarters are finalized immediately after the final training has been completed. This will allow more to time to confirm the upload and check that every Interviewer are correctly assigned to their respective Supervisors.
4) Once the list is finalized, their respective accounts will never be changed. In cases where a member of a team leaves due to resignation, quitting, absenteeism and so forth, it is always suggested that any replacement person will have to use the same username of the person he or she is replacing but will need to provide a new password so the person who has left will not be able to log in again in another tablet.
5) Usernames accounts could be deleted but this will be archived in the server and could be retrieved again if need be. No account can be created again if it has been deleted.

TRAINING FIELD STAFF

1. **How to run training with tablets – include info of mirroring android screen – airdroid, bluestack etc**
    1) Preparing for training of field staff involves a lot of work and organization to ensure the field staff are well prepared and having a good understanding of what is involved.
    2) Training using tablets sometimes becomes tedious and stressful if you do not have the technological support and equipment to supplement your training.
    3) There are Android applications available (freed downloads) which could be loaded into laptops or PCs to allow for projection to the big screen for the trainees to see. Some common apps include:
        i. Bluestack
        ii. Airdroid
        iii. Nox
        iv. Vysor
    4) Projectors are essential during tablet training for visual presentations as this greatly helps the trainees to see what is happening.
    5) The most important aspect of training with tablets is practice, practice and more practice. It is always recommended that training to be done for 2 weeks and a big chunk of this will be on testing and interviewing practices of the questionnaires.
    6) Most of times, trainees have difficulty understanding the concepts and definitions of the questions in the questionnaire but this becomes clearer and understandable when they do the practical work.
    7) A note to remember that Survey Solutions Application and software are updated every month and this results in training materials and presentations becoming out-of-date due to the changes. As the Survey Solutions Applications are upgraded frequently, the use of the Survey Solution Support site is to be referred to during the trainings for these new updates. Training materials becomes obsolete and irrelevant due to frequent changes and upgrades to the Software and its applications, hence, the need to refer to the Survey Solutions Support site for the new updates and materials.

2. **Include important points for:**
    1) Interviewer
        i. Each Interviewer will be provided a tablet that is registered under his/her name, together with the SIM Card number and the tablet unique IME and Serial numbers.
        ii. During the training, the Interviewers will be involved in a lot of practical work and this is to ensure that the Interviewers fully understand the questions and the information required for each question. It is recommended that PRACTICE, PRACTICE and more PRACTICE is undertaken for the Interviewer to fully grasp the use of the technology and understand the importance of collecting good quality data.
        iii. Ensure the Interviewers fully understood their Enumeration Area boundaries (EA) with the help of their Maps. This will ensure they do not overlap to another EA boundary and also allows for them to be within their boundary while they are capturing their GPS points.
        iv. Together with the practices and testing of tablets, Interviewers will need to be trained on Interview techniques, communication and presentation and most importantly, on the concepts, definitions and classifications used in the questionnaire.

2) Supervisor
    i. A Supervisors is the 'Team Leader' for a group of Interviewers and they are responsible for the field operations and logistics of these Interviewers.
    ii. Supervisors should ensure their Interviewers have all their equipment needed for the enumeration and these include Manual, Maps, ID Cards and other equipment like bags, umbrellas and so forth.
    iii. Supervisors should ensure that the interviewers are fully aware of the Enumeration Areas (EA) and their respective boundaries, the number of households expected to be enumerated.
    iv. The welfare of the Interviewers is also a priority to ensure safety while they are visiting their EAs. Organizing transportations, addressing issues faced by the Interviewers, arranging community meetings and awareness if needed and other functions as and when required by the interviewers are some of the other responsibilities undertaken by the Supervisors.
    v. Supervisors organize debriefing meetings with their respective Interviewers for sharing information, addressing issues with the regards to the tablet and fieldwork and gauge progress on what is actually happening on the field.
3) Headquarter
    i. Similar to the Supervisor role, Headquarters expects approved questionnaires from the Supervisors to be thoroughly checked but this does mean that HQ's will just have to approve. They will still need to browse and make final checks on the questionnaires to ensure full coverage of the whole questionnaire.
    ii. Headquarters will assist Supervisors if issues are not addressed at the Supervisor level.
    iii. Headquarters provides information to Supervisors if any information is needed to be conveyed with regards to fieldwork operations, questionnaire updates (if any), issues to be fixed and other relevant information that needs to be transferred.
    iv. Headquarters to organize weekly briefings with Supervisors on progress and addressing issues faced in the field.

3. General information about field

1) Each Interviewer will have his own username and password. This will need to be used only by the interviewer and they (Interviewer) are not supposed to share their account with another Interviewer.
2) Each Interviewer should only be assigned one tablet and he/she cannot use another tablet or log in into another tablet.
3) The ideal situation is to have one Interviewer assigned to one EA for easy monitoring of work.
4) There had been cases to minimize workload on Interviewers that more than one Interviewer could be assigned to an EA but this need good collaboration between the Interviewers, Supervisors, Headquarters and Mapping staff so they are aware of this. The two Interviewers will need to collaborate strategically on how to cover all the households in the EA so they do not overlap in their work or miss a household. This has to be made known to the Supervisor, HQ and Mapping so they are aware while monitoring that a particular EA is being enumerated by more than one Interviewer.
5) Survey Solution recommends that Supervisors use a laptop or computer to check on the questionnaires completed by the Interviewers. The reason for this is that of the size of the screen where they have easy maneuvering and scrolling of the questionnaire and easy access to the

different sections and questions in the questionnaire. Survey Solutions technical staff are working on this application to upgrade it so it if user-friendly when using the tablets.

6) The Supervisor application can be accessed from anywhere as it is web-based and there is internet connection. This gives them the freedom to monitor and check on the Interviewers in their own time and place.
7) A number of Interviewers will be assigned to a supervisor, which means that a Supervisor can only check the questionnaires from his or her respective Interviewers.

4. Create fun evaluations during training using Survey Solutions to trainees. Can do the same at the end of training to get feedback from trainees. Allows real-time display and discussion of results.

1) Some practical interviews could be done for the trainees to fully understand the process involved in interviewing a household. In Tonga, someone acts as a Head of the household and someone to be an Interviewer and they put up a drama or play on how an Interview is conducted – from the time the Interviewer knocks on the door and introduces himself right to the end of the interviewer when he thanked the household for their cooperation. This develops their communication and interview skills in capturing reliable and correct information from the respondents.
2) During training, each Team make training presentations on one of the Sections of the questionnaire to gauge how they perform and their understanding of the topic. This helps those who will be training Interviewers in understanding the topics and questions being asked in the questionnaire
3) Having Evaluation and Exams (tests) midway and end of the training gives the opportunity to gauge the knowledge of the trainees and their understanding of the processes. Feedback from the trainees serves as a measurement of the level of knowledge and understand and provides the opportunity to revise and modify training schedules and materials to address on the feedback from the trainees.

FIELDWORK

5. First few days – special checks

1) Interviewer
   i. Interviewers are always reminded to synchronize completed questionnaires as and when they are completed, except for cases where there is poor or no internet connections. This is to ensure security of the data they do not lose the information.
   ii. Interviewers should minimize having a lot of 'Started' questionnaires (incomplete) in their tablets as this sometimes becomes tedious and confusing when there is a lot of rejections involved. Recommended that having at most 3-5 started is manageable. Again, those areas with poor or no internet will have more completed and started questionnaire opened due to inability to synchronize.
   iii. Interviewers will need internet access to log in to the tablets and receive assignments and during the synchronization of completed questionnaires. Once they are logged in, they can continue conducting interviews even without internet access.
2) Supervisor
   i. Ensure all Interviewers have logged in – these will be shown in the Supervisor screen on which Interviewer has logged in and those that have not.
   ii. Ensure the Interviewers are in their correct EA by visiting their respective areas.
   iii. For the Supervisor to begin checking on a completed household, first thing to check first are the tabs on top of the screen when they opened a questionnaire. The three important tabs to check first are the 'invalid', 'unanswered' and 'comment'. Invalids are those questions which have errors; Unanswered are those questions which are supposed to be answered but have no answers (these exclude skipped questions) and the Comments are those questions which the Interviewer has commented on. These three needs to be fixed first before they can continue checking on other inconsistencies.
   iv. Supervisors will have to do a thorough check on the completed questionnaires for inconsistencies, even though most of the validation checks are inbuilt in the questionnaires. After a few number of questionnaires have been checked from each Interviewer, the Supervisor will be able to gauge which Interviewers are very efficient which they may do a random check on these efficient ones and concentrate more checking on the inefficient ones.
   v. To save time, cost and energy, Supervisors will need to be up-to-date in the monitoring and checking so that the Interviewers could do validation work during a rejection, while they are still close to the respective household. Any delays by the Supervisor will cause more work for the Interviewers as they will have to return to a household to verify information if there is a rejection. Imagine the Interviewer is 200 meters away from a completed household and had to walk back to that particular household to verify some information because of the delay in the Supervisor checking on the questionnaire of that household.
   vi. The Supervisor screen has a Report Menu which the Supervisors could use to monitor the progress of each Interviewer in terms of number completed, number rejected and number approved, the quantity completed for a particular day and the speed of completion of a questionnaire. Please note these reports sometimes give a wrong estimate of summary as the system counts a completed questionnaire as one. Any questionnaire that is rejected and resent again by the Interviewer will be treated as another completed questionnaire, hence, the number of rejections to a questionnaire will

be the number of households completed, so a few households will be counted more than one as completed depending on the number of rejections.

3) Headquarters
    i. Similar to Supervisors, HQ's have access to the Report Menu but with additional privileges like Mapping and Charts.
    ii. Through the Map Report menu, HQ's will be able to see the GPS points of each household being collected by the Interviewers. This allows them to monitor the Interviewers to ensure they are within the EA boundaries, overlapping points, wrong capturing of IDs and other locations issues. Even though the Map is not linked to the EAs, this gives a rough indicator of the location of the interviewers on the ground.
    iii. HQ's can download data but for security reasons, the System Administrator assigns passwords for the download where only a number of people can download data and make summary tables on the data being collected at a given time.
4) Field control staff i.e those downloading data to do some mapping
    i. These field control staff are mostly GIS staff who will used the downloaded data to plot on the EA maps.
    ii. With the EA boundary lines, they will be able to determine which Interviewer had gone over his/her EA and this is then given to the HQs to inform the respective teams.
    iii. Sometimes, data collected are correct but the GPS points are wrong due to inability to capture the points and the Interviewers moving away from the area to capture the waypoints – this is then rectified by the Supervisors by rejecting the questionnaire and for the Interviewer to go back to the house to capture the correct GPS points.

ONCE FIELDWORK IS COMPLETE

1. Download final datasets/documentation

1) For the final download of the data, it is highly recommended that all questionnaires are 'Approved by Headquarters' status to ensure full completion of the census/survey.
2) There are four datafiles that could be downloaded from the system:
    i. Main Survey Data
        1. STATA
        2. SPSS
        3. Tabular format (Excel)
    ii. Binary Data – which include the photos taken by the Interviewers
    iii. DDI – for questionnaire documentation
    iv. Para Data - which is the metadata on the interview process and has records of all event and timing
3) Data download could be done by the Headquarters and the Administrator but for security reasons, the Administrator has the privilege to assign passwords so only a few people can be able to do this function. Headquarters could still download the data but will need the password to open the file.
4) A major advantage is that data could be downloaded anytime which allows for the HQ's to generate preliminary results and updates of what has been collected so far. This serves also as a monitoring tool to see the progress and performance of fieldwork in terms of completion rates by EAs, Provinces and Islands. Results from these reports provides good information on problematic areas that will need further assistance and monitoring. In extreme cases, then extra resources could be reallocated to these areas to improve on the collection.

DATA PROCESSING

2. Data Structure

1) As mentioned above, data is downloaded in three format – STATA, SPSS and tab-delimiter (ASCII-textfile), which means that the data could be converted to any other format depending on which software the NSO wants to use for their processing.
2) Before processing, it is important to understand how the data is structured which means having a good understanding of how the questionnaire is designed, the variable types of each question, the number of rosters and sections and so forth.
3) For a Population Census data, based on how well the questionnaire is designed, there will always be a Person and a Household Record. This means that for every Household in a Private Dwelling, they will have a Household record with a number of persons.
    i. Vacant households will have no Person record
    ii. Institutions will only have a Person record and no Household record
4) Apart from the single records like Household data, each roster in the questionnaire will have separate data files created for them – which is why we always suggest not to create a lot of rosters/netted rosters when designing questionnaires as this make data linking and merging a bit complex.
5) Survey Solutions automatically creates a unique Id variable in all the downloaded data and this is used as the Primary Key for linking all the data files downloaded.

3. Linking Data

1) For Population Census data, the Household record which is a single record (one for each private household) will contain all the individual questions data (not rostered data) in the questionnaire. These include variables, which we normally use as unique IDs like Province, Island, District, Ward and Village together with the unique generated ID.
2) The rostered data will only have the variables asked in the roster with the generated ID.
3) For linking purposes, you will need to merge these files (using the unique generated ID) so each data file will have the main IDs mentioned in point 1. You may have to rename variables to link the data files using the same variable name.

4. Data verification / editing in STATA

1) Since data is downloaded in STATA format, the following information will be based on doing data editing in STATA. For those who prefer to use other software, the following information will still be relevant and the concepts are similar when editing in another software.
2) If the questionnaire is designed with all the relevant validation checks and enabling and filtering conditions, data verification and editing will be minimized significantly.
3) If all questionnaire are fully completed and they have gone through the Supervisor and HQ approvals, then all the skips and consistency checks inbuilt in the questionnaire will not have to be rechecked again in the final edits.
4) With reference to 2.3 above, this will need to be checked first that every Private Dwellings will have a respective Household record; vacant houses have no Person record and Institutions has no Household record. This should not happen if the questionnaire is well designed but past experiences shows that some interviews were incomplete where only the Persons are filled but not the Household.

5) One biggest advantage of using STATA is its efficient capability to link with Excel where data could be easily exported to Excel for editing and then imported back into STATA.
6) STATA Help functions are so detailed and concise which provides relevant information on whatever command you may want to use for processing.
7) Even though you may need to have a bit of programming skills to write command scripts, STATA has the Menu commands which could be used instead to generate these command scripts.
8) One of the suggested way to start data verification and editing is to run a quick frequency output for each variable. This will allow you to identify the inconsistencies between variables in terms of missing responses, unanswered questions and so forth.
9) These frequencies will allow you to identify if there is any inconsistencies for those filtering and enabling conditions questions where a response to a question should be consistent with the response to the next question based on the enabling or filter condition.
   Example:

```
. tab c1_ever_attend_sch, missing
```

Ever attended school	Freq.	Percent	Cum.
Yes	85,554	85.00	85.00
No	7,246	7.20	92.20
.	7,851	7.80	100.00
Total	100,651	100.00	

```
. tab c2a_highest_level_attended, missing
```

Highest level of school attended	Freq.	Percent	Cum.
Pre-school	3,224	3.20	3.20
Primary	17,974	17.86	21.06
Lower secondary	14,516	14.42	35.48
Upper secondary	38,024	37.78	73.26
Technical and Vocational	4,852	4.82	78.08
Tertiary	6,660	6.62	84.70
Other	301	0.30	85.00
.	15,100	15.00	100.00
Total	100,651	100.00	

   The above frequency example shows some inconsistencies where those who answer 'Yes' in c1_ever_attended_sch should match the numbers in the next question. This will then have to be checked in the edits to fix the inconsistency.
10) STATA is so powerful in generating new variables, recoding of items, extracting variables in small datasets, merging of datasets and easy tabulations and frequencies, which greatly helps in editing the data.