the easy guide to

MOBILE DATA
COLLECTION

A guide by **Nirab Pudasaini** and **SCHOOL OF DATA**

TABLE OF CONTENTS

GETTING STARTED

So you have a project in mind, and it requires collecting data using mobile phones. Or maybe you already started collecting data, but your pen and paper method is showing its limits. And on top of that, your experience of mobile data collection is close to zero. Have no fear! You've arrived at the right place.

The spread of cheap smartphones, which started during the 2010s, allowed the potential of mobile data collection to be unlocked. A multitude of projects, from <u>urban mapping</u> to the <u>surveying</u> of remote communities, dropped their previous cumbersome methods (pen and paper, specialised suite of hardware tools…) to switch to collecting data using the now-ubiquitous smartphone.

But a tool can only do so much. Correctly collecting data in general requires a mix of knowledge and skills that no smartphone will replace. And if your project is of any significant size,creating an interdisciplinary team with the right mix of skills will be vital to its smooth and successful execution.

But before detailing the role of each team member, let's review the **when**, **where** and **how** of collecting data.

WHEN, WHERE AND HOW

Obviously, you can only collect data to which you have access. The first part of defining a data collection strategy is understanding when or where will the act of data collection will take place.

- → **WHEN** applies to something that goes through different steps of a chain: for example, a fish is first caught on a fishing boat, then brought to a harbour, sent to a storage room etc.

- → **WHERE** applies to the location where you might have to go to do the data collection: you will not be able to bring the same equipment on a boat or in a office building.

Once you have your answer, you need to define **HOW** you will collect the data. It could be :

- → using one or several existing database(s),

- → using forms, questionnaires and interviews,

- → using direct observations.

Direct observations is making direct measurements of the subject being studied (like the number of buildings in need of repair after an earthquake), rather than using indirect measurements, such as databases or testimonies from people knowledgeable in the field. Most often, direct observations are made by human observers, sometimes with the aid of specific equipment:

- → pollution sensors to measure air pollution,

- → Geiger counters to measure radioactivity,

- → smartphones to record visual observations and, potentially, sound.

! As you may have guessed, this guide will focus on direct observations using smartphones to collect data. Any additional or process-specific content available can be found on the dedicated website, in the +RESOURCES section.

SETTING UP YOUR TEAM

Collecting data using smartphones has many advantages in terms of ease-of-use and cost-reduction, especially over the paper based method:

→ they are available in sparsely populated areas, which means that the act of collecting data can be, if needed, <u>crowdsourced</u> by people living in the field;

→ they often feature at least a basic camera and sound recorder, which allows for multimedia data collection;

→ the results of the data collection can be sent in real time, provided an internet connection is available.

But there are many ways in which smartphone data-collection can go wrong, making the whole process costlier and more difficult than it should be! This is why a dedicated preparation phase is vital.

To give you the best overview of the tasks involved in collecting data, we've divided them into the four key roles that feature in a data collection team. These functions aren't rigid, of course: your team members may be able to cover multiple roles or there may be a sub-team focused on a particular area. Your choice of <u>resource allocation</u> should encompass the level of expertise of each member, the scope of the project, and the budget available.

I am the...

Project Manager

" I am responsible for overall management of the survey. I define the project goal and I'm responsible for all the major decisions that need to be taken during the project.
I'm also responsible for communication amongst the survey team, to make sure that everybody is on the same page. I report the project status, maintain a record of lessons learned and lead retrospectives. "

TIMELINE OF THE TASKS

BEFORE ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨ ⑩ ⑪ ⑫ ⑬ ⑭   DURING ⑮ ⑯ ⑰ ⑱ ⑲ ⑳   AFTER ㉑ ㉒ ㉓
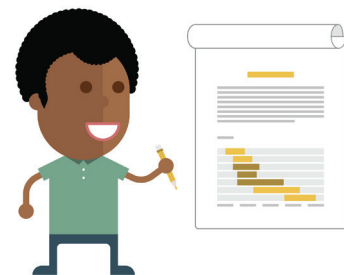
BEFORE

① I design the project along with the other stakeholders: this involves deciding on a clear goal (the objective), identifying a question that the project will seek to answer, and defining the requisite data to be collected. Example: *What is the state of Kathmandu buildings after the earthquake?*

② I write down the work plan, which details the necessary steps of the data collection project, their sequence and their coordination.

3   I budget the project. Mobile data collection projects have particular costs that need to be considered (<u>hardware</u>, data transmission, <u>data aggregation</u>, software maintenance) along with more general costs (insurance, personnel...).

4   I identify the best device for the survey: smartphone vs tablet? Battery-life needed?

5   I make sure that the training is properly conducted and that <u>enumerators</u> are sufficiently skilled to carry out the data collection.

DURING

15   I ensure that enumerators are actively collecting data and communicate progress to stakeholders, as appropriate.

AFTER

21   I supervise the analysis of the data.

22   I consolidate the results in a report that is presented to the stakeholders.

BEST PRACTICES

BE CLEAR ON THE GOAL
Throughout the project you will need to take major decisions and being clear on your goal throughout will help you with this, as well as saving time in the project's later stages.

DEFINE A <u>PROJECT WORKFLOW</u> AND PROTOCOL
This serves two purposes: it ensures that all your stakeholders are on the same page, that they have similar expectations as regards the expected rate of progress. Secondly, it streamlines the reporting part of your role: everyone knows what to expect and when.

PICK THE RIGHT DEVICE TO FIT YOUR DATA COLLECTION NEEDS
There are lots of mobile devices on the market and choosing the right one can be difficult. You'll need to consider points such as smartphone vs. tablet, battery life, <u>GPS</u> accuracy, camera and <u>app compatibility</u>.

PROVIDE DIFFERENT OPTIONS FOR UPLOADING DATA

If your survey requires media inputs, data uploads might cause concerns around cost (data charges) and logistics (limited coverage of phone carriers). One approach, available in software like KLL Collect, is to upload the data right away, then only later submit the media attachments. Another one is to use swappable memory cards and assigning a runner to collect them for later upload.

HAVE A PLAN B

Things won't always go as planned, so you'll need to have a contingency plan for everything: what if a device stops working? What happens if a server goes down? What if the GPS doesn't work as expected? Is there a process for backing up the data?

MAKE SURE THAT A MOBILE DATA COLLECTION PROCESS IS THE RIGHT CHOICE

Mobile data collection might not always be the best idea. Sometimes paper based survey or web based survey makes more sense than using mobile based data collection. This is the case for small surveys with low data entry cost where real-time updates are not necessary or data collection projects where mobility and internet connection are not an issue.

I am the...

Survey Designer

" My mission is to construct questions that elicit just the right data from the survey's respondents. I have to make sure questions are not ambiguous and are as straightforward as possible, encompassing complexity and subtleties in a way the respondents fully understand their meaning. My role also has a large visual design element: I ensure the survey looks appealing and is engaging to complete. "

TIMELINE OF THE TASKS

BEFORE   **DURING**   **AFTER**

1 2 3 4 5 6 7 8 9 10 11 12 13 14   15 16 17 18 19 20   21 22 23

BEFORE

6   I examine the research design and project plan and identify the key information the survey needs to cover.

7   I begin drafting the different fields that will need to be filled, or questions that will need to be answered.

8   I design the survey for smartphones, making sure it is fully functional for the platform.

9   In collaboration with the rest of the team, I assess the quality of the survey questions during the pilot study.

DURING

16   In collaboration with the rest of the team, I provide support to the enumerators in the field, as necessary.

BEST PRACTICES

QUESTIONS: THE FEWER, THE BETTER
If the survey requires interviewing people, a good number of question is 10 or below. Similarly, the more fields enumerators have to fill, the more errors they will tend to make.

BE PRACTICAL
The survey should be designed with the phone screen(s) it will be read on in mind. Because the screen estate is often very small, superfluous elements such as graphics, having the title on every page, the progress bar etc should be avoided. One question per page is a good idea.

OPTIMISE THE USER EXPERIENCE
To avoid errors and allow enumerators to focus on one question at a time, a sequential, single question display should be used. The previous/next buttons should be obvious and visible. Lastly, answer confirmations help avoid mistyping.

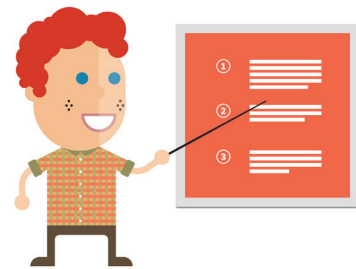### MAKE SURE THAT THE LOGIC OF YOUR SURVEY IS SOUND
Testing your survey beforehand allows you to make sure that mandatory fields can not be skipped, that the survey reacts correctly when incorrect data is entered and that the pre-filled fields display the correct information.

INCLUDE A POST-COMPLETION REVIEW
Wherever possible, the errors in data entry should be corrected in the field, when the context of the error is still there. Good practice is then to allow a review of answers before sending data to the server.

I am the...
Trainer
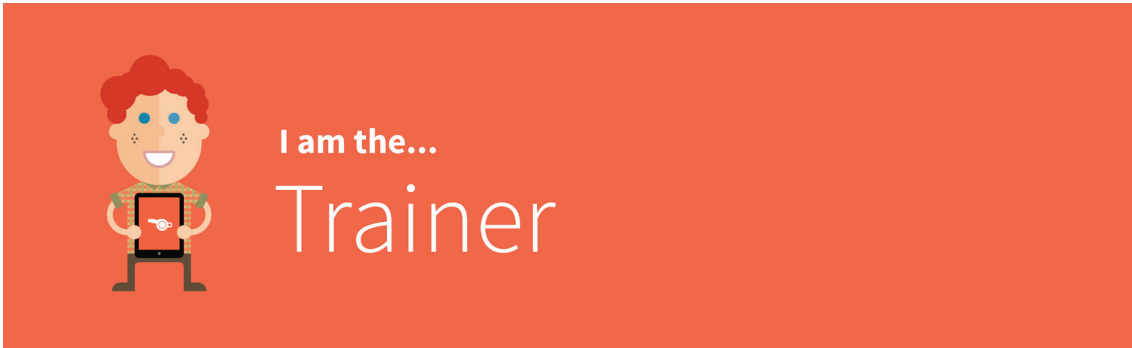
" I run the training sessions that will help interviewers, or enumerators, learn the skills they need to perform a successful data collection with participants. Even if the survey is well designed the quality of the data collected through the interviews depends ultimately on the data collection skills of the enumerators. I make sure they get acquainted with the mobile devices and the mobile app used to conduct the survey. "

TIMELINE OF THE TASKS

BEFORE

1 2 3 4 5 6 7 8 9 **10 11 12** 13 14

DURING

15 16 **17** 18 19 20

AFTER

21 22 23

BEFORE

10 I get to know the survey, the mobile devices and the potential difficulties that enumerators might face on the field.

11 I plan and organise one or more training sessions for enumerators, pre-empting all potential difficulties.

12 I work with the survey designer and data manager on the pilot study to help flag mistakes done by enumerators.

DURING

17 In collaboration with the rest of the team, I provide support to the enumerators in the field, as necessary.

BEST PRACTICES

VARY THE TEACHING FORMATS
The training should include demonstrations, hands-on and Q&A sessions, ideally spread over a few days. Additionally, there should be sessions on core survey content, damage assessment training, proper utilization of fields guides, etc.

SUPERVISE THE PILOT STUDY
At the end of the training, having a trainer supervising the pilot study is very useful to help clear out any doubt enumerators might have and identify the points which might need further training.

DON'T ASSUME ANYTHING
The training should cover everything, from an explanation of the context of the survey to the way to take useful photos, to details of the mobile device's features: GPS, Auto-Rotate, System Date and Time…you generally can't predict with what your enumerators will be familiar.

LEAVE ROOM FOR QUESTIONS AND FEEDBACK
Paper copies of the survey may be useful in the first instance, so that enumerators can annotate and make comments on the fields or questions of the survey. Encourage everybody to ask questions, even if they fear that the question could be stupid!

PREPARE ENUMERATORS FOR THE WORST
It's almost inevitable that something will malfunction during the survey. Enumerators should be able to reset or reinstall problematic software.

I am the...

Data Manager

" I work with the Survey Designer to ensure the questionnaire used in the interviews will generate relevant data for the analysis phase. I also create <u>variable mappings</u> and define the format in which output is generated. It is my responsibility to track the progress of data collection and to communicate it to the rest of the team. "

TIMELINE OF THE TASKS

BEFORE	DURING	AFTER

1 2 3 4 5 6 7 8 9 10 11 12 **13 14**   **15** 16 17 **18 19 20**   21 22 **23**

BEFORE

13 I set up the technical infrastructure which allows the reception, storage, security and analysis of the collected data.

14 I monitor the quality of the data collected and the responsiveness of the technical infrastructure during the pilot survey.

DURING

18 I keep track of the data being sent by the enumerators, making sure that the data received is not flawed.

19 In projects where the data should be analysed and visualised in real-time, I manage this process continuously.

20 In collaboration with the rest of the team, I provide support to the enumerators in the field, as necessary.

AFTER

23 When the data collection process is complete, I run the final analysis, export and create reports, in collaboration with the Project Manager.

BEST PRACTICES

DO NOT TRY TO MAKE THE SURVEY WORKFLOW TOO STRINGENT

It is very tempting to add lots of validation rules into the data collection app itself, to reduce data validation later. But keep in mind that this may make data collection problematic, in the event of outlier data being encountered in the field.

THINK ABOUT THE END USER

Make sure you are well familiar with the end user of the data: how will they be using the data and the tools they use. This will allow you to tailor the data output.

TEST, TEST, TEST

Fill in test data and generate mock outputs. This will reveal issues in the data and allow you to take corrective action, as necessary.

AUTOMATE THE VISUALISATION OF THE PROGRESS

It is not only useful to track the evolution of the survey, but it will save you a lot of time. Some open source data collection tools allow data visualization and progress monitoring without any work.

THINK ABOUT DATA ACCESS PERMISSIONS

Sometime you might need to give different level of access to different people, to make sure that only the right people are able to view and change the data. You should make sure that the data collection tool you are using is capable of doing this.

COMMON MISTAKES IN THE FIELD

Obviously, you can only collect data to which you have access. anything is constrained by the available access to the subject being studied. The first part of defining a data collection strategy is understanding when or where will the act of data collection will take place.

- Enumerators accidentally delete photos from gallery.

- Enumerators upload trial data in an actual dataset.

- Enumerators take blurred images and some pictures are taken in low light.

- Enumerators type data with spelling mistakes. Example: 0 (zero) is spelled as O (alphabet O).

- Enumerators change the default storage location to internal memory in the devices provided to them. This causes confusion as half the data is stored in removable memory and half the data is stored in device memory, causing panic among enumerators who believe the data is lost.

- Enumerators use an additional device but forget to configure it properly.

- Typos and mistakes appear, due to enumerators hurrying.

- Enumerators try to record the GPS location from inside a building, which results in poor accuracy and delays.

- Enumerators try to submit high volume data (with photos) from a place with low bandwidth internet service, which can result in frequent upload unsuccessful errors. This can raise frustration levels and give the impression that there is a problem with the data upload.

- Enumerators accidentally delete files, using the file manager, which are essential for survey software to function.

- Whilst playing with and exploring the app, enumerators change default settings. This may result in the app behaving differently from the way in which the enumerators were taught to expect during the training and result in confusion.

- A device is damaged. The most common type of damage in the field is breakage of the device screen. Enumerators do not complete many necessary fields which may not be compulsory in the survey form.

- Enumerators do not fill out some data points, as they think might have already captured it previous questions. For example, in one case enumerators did not capture a photo of the main building of a school, as it had already been captured in a previous photo of the school sign.

- Device auto-corrects, causing errors in the entries.

- Enumerators try to keep a backup of the data but do not copy the whole folder which results in loss of data.

- Enumerators accidentally delete data while transferring data from SD card to laptop.

FAQS FROM THE FIELD

Here is a list of questions that Kathmandu Living Labs gets most frequently from enumerators. These questions are extracted from help desk logs and debrief meetings from various data collection projects.

- I've carried out surveys using the tablet, but I don't see them. What shall I do?
- Can we upload records with partial data?
- Can I edit data which I have already sent?
- What should I do if my form submission/upload fails?
- What should I do if I accidently delete some files and am getting errors uploading?
- What should I do if I can't download the template questionnaire?
- Where in the disk can I find the data I collected?
- The tablet screen is too dark out in the sunlight. What shall I do?
- How can I save battery power?
- What can I do if I can't charge my tablet's battery properly?
- How do I take a screenshot?
- What can I do if my tablet's touchscreen doesn't work properly after I used it for a long time?
- How do I know whether my tablet is using its mobile data or Wi-Fi connection to access the Internet?
- How do I check the mobile data connection status?
- Why can't my tablet connect to a Wi-Fi network?
- What can I do if I can't find any available Wi-Fi networks after Wi-Fi is turned on?
- Does the data connection disconnect after my tablet enters sleep mode?
- How do I transfer files between my tablet and a PC?
- What can I do if no new disk drive is displayed for my tablet after I connect it to a computer?
- How do I close background applications?
- How do I change the default storage location on my tablet?

GLOSSARY

App compatibility: be aware that some applications may not be compatible with your smartphone, or it may happen that the installation of an application will influence the behaviour of another.

Budget: the total amount of money allocated to a specific purpose.

Crowdsourcing: when you obtain needed services, ideas, or content by soliciting contributions from a large group of people, and especially from an online community.

Data aggregation: any process in which information is gathered and expressed in summary form, for purposes such as statistical analysis. A common aggregation purpose is to get more information about particular groups based on specific variables such as age, profession, or income.

Database: an organized collection of data. The collection of schemas, tables, queries, reports, views and other objects.

Enumerators: the person hired to conduct the survey, be it the observation of the damage on buildings or interviews with the participants.

GPS: Global Positioning System. A satellite-based navigation system made up of a network of 24 satellites. Smartphones equipped with a GPS chip are able to track their locations around the globe with a fair amount of precision.

Hardware: machines, wiring, and other physical components of a computer or other electronic system.

KLL Collect: a mobile application created by the NGO Kathmandu Living Labs that you can install on your smartphone for data collection.

Pilot study: a small-scale, preliminary study conducted in order to evaluate feasibility, time, cost, adverse events, and effect size (statistical variability), in an attempt to predict an appropriate sample size and improve upon the study design, prior to performance of a full-scale research project.

Project workflow: the serialisation of tasks to produce a desired outcome, usually involving multiple participants and several stages.

Q&A session: session in which the most commonly asked questions are answered.

Questionnaire: a set of questions with a choice of answers, devised for the purposes of a survey or statistical study

Runner: a person who can quickly pick up goods from a location and take them to another.

Server: computer program or machine that waits for requests from other machines or software (clients) and responds to them.

Stakeholders: person, group or organization that has an interest or concern in a certain project, activity or organisation.

Variable mappings: process in which Data Managers populate research templates with the values of the answered questions.

ACKNOWLEDGEMENTS