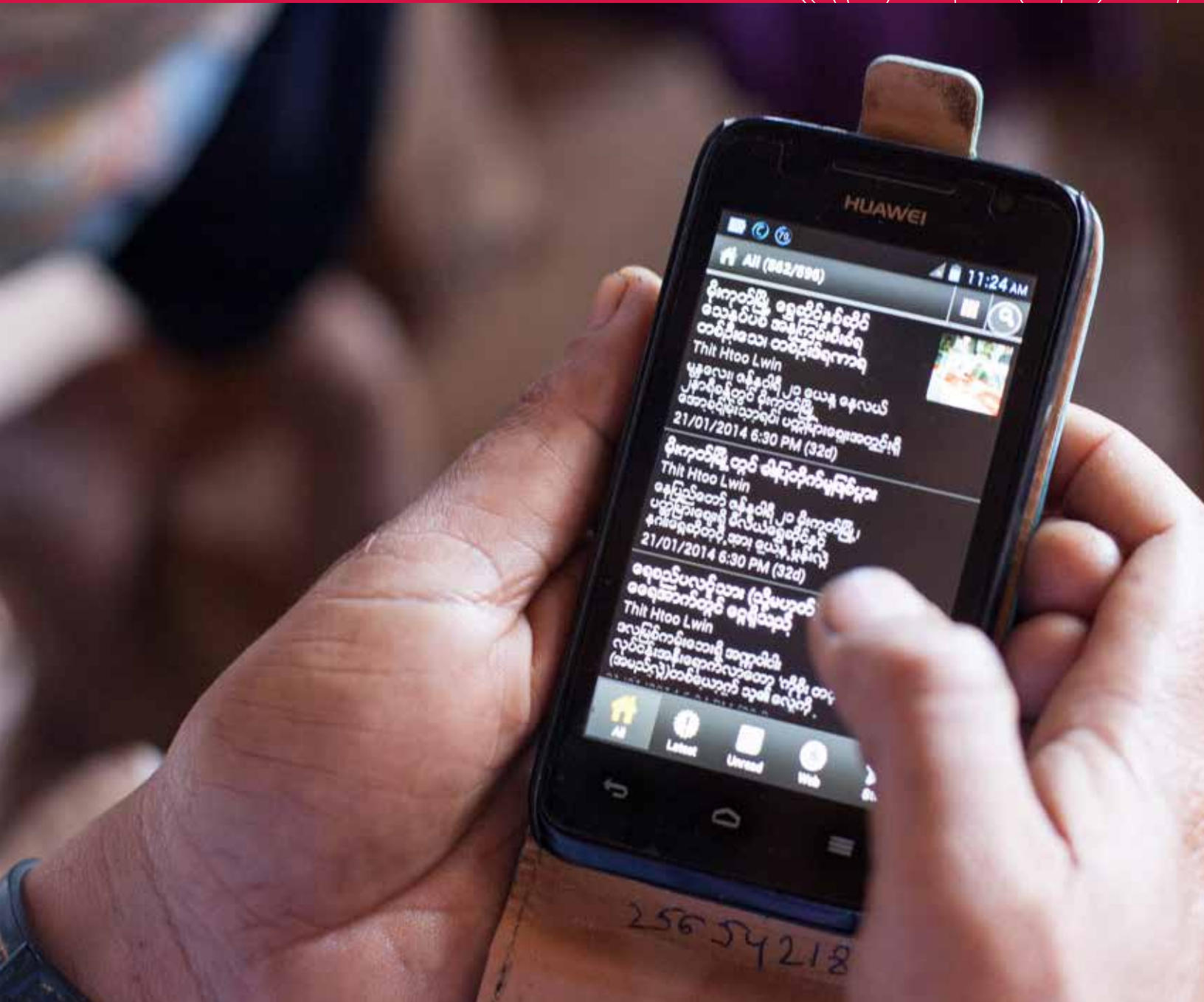# PAPER-TO-MOBILE
DATA COLLECTION:
A MANUAL

USAID
FROM THE AMERICAN PEOPLE

mSTAR

fhi360
THE SCIENCE OF IMPROVING LIVES

U.S. GLOBAL DEVELOPMENT LAB

PAPER-TO-MOBILE
DATA COLLECTION:
A MANUAL

Written by Erin Satterlee, Leela McCullough, Michael Dawson, and Kelly Cheung

U.S. GLOBAL DEVELOPMENT LAB Powered by USAID   mSTAR   fhi360 THE SCIENCE OF IMPROVING LIVES

CONTENTS

CONTENTS & FOREWORD

SECTION I

SECTION 2

SECTION 3

CONCLUSION

ANNEXES

RESOURCES

FOREWORD

Advances in information technology have opened exciting new avenues for how research is conducted and data collected, with huge implications to monitoring and evaluation (M&E) frameworks and program design. One of these advances comes in the form of Mobile Data Collection (MDC), which makes innovative use of mobile technology to replace traditional pen and paper data gathering, and is being increasingly adopted globally.

Used correctly, these technologies equip the information gatherer with a powerful tool that can be used to improve data collection speed, reduce cost, and increase quality. For institutions engaged in longer term projects or programs, the ability to collect data faster, potentially in real time, allows for increased responsiveness to changes on the ground. They also allow survey supervisors and the enumerators to keep better track of the gathered data, which in turn enhances performance, enables the identification of potential problems, and reveals interesting lines of inquiry while they can still be addressed, as well as improves accountability of the various actors.

Today, the most common use for MDC is in the implementation of quantitative surveys. As such, those will be the focus of this manual. However, it is good to keep in mind that MDC is well suited for collecting a number of types of quantifiable data that USAID or program staff might utilize. Examples of these would include surveying through SMS, Interactive Voice Response, or web-based surveys.

This toolkit is written in order to give USAID program and contracting staff, implementing partner M&E and program staff, and researchers a better idea of what mobile data collection is, to evaluate its suitability and added value to a given project, and to give guidance in how it can best be put to use. It is directed at USAID Missions and implementing partners engaged in monitoring and evaluation, and offers the resources necessary for making an informed choice and an effective transition process from paper and pen to mobile. It will lay out the benefits of MDC use for implementing partners, as well as incorporating these methods in M&E related solicitations (RFAs and RFPs), when appropriate.

The toolkit is organized around three major sections:

1 What is mobile data collection and what benefits does it offer?

2 Evaluating the suitability of mobile data collection tools for a given project;

3 How to successfully build the capabilities needed for transitioning to mobile data collection.

Although mobile data collection activities encompass the whole range of data collection activities including program design, implementation, data aggregation and analysis, and dissemination of results, this toolkit concentrates on the aspects of implementation that need to be addressed early in the planning stages of data collection efforts.

CONTENTS & FOREWORD

SECTION I

SECTION 2

SECTION 3

CONCLUSION

ANNEXES

RESOURCES

SECTION

1

MOBILE DATA COLLECTION EXPLAINED

- The Benefits of Mobile Data Collection

- How Exactly Does Mobile Data Collection Work?

- Limitations of Mobile Data Collection

- Frequently Raised Concerns

1.1 Introduction

Mobile data collection (MDC) refers to the utilization of existing information technology products such as phones, smartphones, and tablets (hardware), and a number of different possible programs (software), for data gathering.

Instead of recording information on printed paper using a pen, which is then manually entered into a database for analysis, data is input into a device which is then capable of exporting directly into a centralized database (which can be done using the Internet or a local computer). These methods are also sometimes referred to as Computer Assisted Personal Interviewing (CAPI).

At its core, MDC is best thought of as simply an additional method in a data collector's toolkit. It is a powerful method when applied correctly and to the appropriate types of questions, but it does not necessarily signify any fundamental revolution in traditional data collection. The field of MDC has applied existing technology to existing methods, resulting in many cases in an improved process. Although it requires some new capabilities from the planning and implementation standpoint that may feel overwhelming in the beginning, it should not be considered anything altogether alien. MDC simply allows program staff to conduct their inquiry of survey questions more efficiently. From the data collector's perspective, it is little more than switching one of the many tools that are used during the course of a research project or in a monitoring system.

MDC is not a panacea for correcting the effects of poor M&E design, badly selected indicators, or inappropriate analysis. It will not remove any limitations imposed on data analysis from sources such as sampling frame, sample size, lack of a measured baseline or control group, response/non-response bias, etc. Like any tool, it is only as effective as the staff applying it.

1.2 The Benefits of Mobile Data Collection

Although mobile data collection tools are not necessarily revolutionary, they offer numerous benefits compared to traditional methods. Some of the main benefits are elaborated on in the sections below.

CONTENTS & FOREWORD

SECTION 1

SECTION 2

SECTION 3

CONCLUSION

ANNEXES

RESOURCES

MAIN BENEFITS OF MOBILE DATA COLLECTION

| Increased speed | Reduced cost | Increased accuracy | Increased functionality | Accountability |

1.2.1 Benefit – Speed and efficiency

Using MDC methods, data can be ready for analysis in a significantly reduced amount of time. As data is entered directly into an electronic form that can be exported into a statistical package or information management system, access to data can even become instantaneous. With the use of Internet technology, access may not only be real time but also global.

Depending on the choice in software, data can also be presented automatically in visual forms that track the most important indicators without requiring manual analysis by a statistician.

In addition to faster turnaround of data due to eliminating separate data entry, the speed with which actual interviews are conducted can be reduced, particularly for longer, more complex surveys. As the interviewer is automatically prompted with the correct questions to ask, they do not need to concentrate on skip patterns or spend time checking that each question has been asked and all required fields have been filled.

From a program perspective, the faster turnaround time is significant as it allows better responsiveness to issues arising on the ground. This is especially important in cases where data collection is continuous, as rising or declining trends can be effectively identified and addressed while doing so is still possible.

EVIDENCE: A randomized controlled trial in Peru comparing methods of entering TB laboratory results found that data entry using PDAs required 70% fewer working hours and reduced turnaround time from 25 to 12 days.[1] A World Bank study also found that average interview length was reduced by 3.6%.[2]

1.2.2 Benefit – Cost

Although the adoption of MDC tools has some upfront costs, which will be elaborated on in the following sections, it can also lead to significant savings, particularly through the elimination of the data entry process. Securing an acceptably small error rate with traditional survey methods often requires "double data entry," where two separate workers enter the same data and one supervisor identifies differences between the two. Eliminating this redundancy often adds up to significant cost reduction. The larger the size of the survey, the larger the savings from this activity become, as manual input requires a set amount of time per observation.

1 Blaya, J. A., et al. (2009), Personal digital assistants to collect tuberculosis bacteriology data in Peru reduce delays, errors, and workload, and are acceptable to users: cluster randomized controlled trial.

2 Schuster, C. and Brito, C. P. (2011), Cutting costs, boosting and collecting data real-time – Lessons from a Cell Phone-Based Beneficiary Survey to Strengthen Guatemala's Conditional Cash Transfer Program, World Bank.

Additional savings can be made on printing costs, which can be significant as survey size and interview length increase (imagine a 30 page questionnaire administered to a sample of 1,000 respondents). As any and all mistakes made in the survey can be corrected using a computer and updated in the collecting device used, the need for reprinting materials no longer applies.

Potential cost savings can also arise from the flexibility and customizability of mobile tools. For instance, imagine gathering data in an area where five different languages are spoken. With traditional pen and paper, estimates of language proportions (if such data is even available) must be taken into account while printing the questionnaires. To be on the safe side, some extra copies of each questionnaire will probably be necessary. With mobile tools, all languages can be downloaded on the collecting device and the data is input automatically in the same dataset category regardless of the language version.

From a program perspective, investment in a common platform that can be used in multiple projects can further spread costs and reduce the burden on a single study or round of data collection.

EVIDENCE: Proof for the cost efficiency benefits of mobile data collection is robust. A World Bank study on the question found that average interview cost was reduced by up to 71%.[3]

Filter logic: refers to the survey coding that routes only certain applicable questions to a respondent based off the answers to previous questions. Also referred to as "skip logic."

1.2.3 Benefit – Data quality

The improvements to data quality from MDC tools stem from two distinct sources. Firstly, the survey development process allows for a large degree of customization (the extent depending on software, see section III), and the insertion of a number of quality checking tools. **Filter logic** can be utilized, which eliminates the possible confusion arising from questionnaire instructions. Secondly, as data does not need to be separately transcribed from paper to a computer, there is no possibility for data mis-entry during this phase.

Furthermore, MDC tools allow for validation of the data being entered during an interview. For example, questions requiring a numeric answer can be forced to only accept a numeric answer, which can also be forced within a range (such as between 1-99 for age) to reduce the chance of input errors. Answers to questions, such as date of birth and age, can be cross-referenced to validate each other and prompt the enumerator to clarify any inconsistencies. Forms that are not filled can be coded to not allow submission.

In addition to the questions programmed into the survey, some MDC tools automatically collect metadata such as timestamps. These allow for the survey supervisor to monitor interview start and end times to ensure that they match with expected interviewing behavior, and thus helps keep enumerators more accountable.

EVIDENCE: A study by Caeyers et al. showed that in a 152-item household survey conducted in Tanzania, routing errors were reduced from an average of 10 per survey to 0 using mobile data collection. Furthermore, the number of mandatory questions left unanswered was reduced from 4 to 0.5 per survey.[4] The randomized controlled trial from Peru also showed that data entry errors were more than halved by the use of PDAs, from 6.1% to 2.6%.[5]

3. Ibid.

4. Bet Caeyers, Neil Chalmers, and Joachim de Weerdt (2012), Improving consumption measurement and other survey data through CAPI: Evidence from a randomized experiment.

5. Blaya et al., op. cit.

1.2.4 Benefit – Customizability and added tools

Though the exact extent will depend on the software choice (see section III), MDC tools generally offer a wide range of customization possibilities. These include but are not limited to different types of questions (all that can be asked using pen and paper), programming in mathematical equations that can transform data according to pre-set functions, the utilization of GPS technology, and the addition of multimedia features such as graphics, audio, and video.

Examples of these added features include the ability to use audio, graphics or video to improve comprehension of survey topics, and using a camera function to capture a photo of a school while also capturing its location via GPS to create a map that allows users to see images as well as location of survey sites.

EVIDENCE: An initial analysis of Mozambique household mapping using PDAs with GPS, which gathered the location of 4855 houses dispersed in 32 villages in 8 districts, was completed in 10 days using 18 field observers. The data was aggregated, analyzed and presented in less than a day. The average mapping of individual houses in an enumeration area, spanning several kilometers, was completed between 3.5 to 4.5 hours.[6]

MDC IN ACTION

In an evaluation of the impact of the social drama *Daremo* on the behaviors of Somaliland mothers, BBC Media Action found that playing the theme song of its programming from the mobile device made a significant difference in the recognition of the program for listeners who did not necessarily know the name of the radio program although they were often regular listeners.

1.2.5 Benefit – Control and flexibility of fieldwork

Faster access to data improves not only analysis and program responsiveness, but improves staff ability to react to unexpected changes. For example, since surveys in the MDC devices can be updated (given even limited Internet connection), it is possible to fix small issues in questionnaire design such as typos, badly phrased questions, or filter logic. For longer studies, further questions can even be added to investigate interesting findings. The costs for such responses are negligible, as lead staff or field supervisors can make any alterations on their computer.

Effective supervision of enumerators is also a common problem in fieldwork, and can be substantially mitigated by the use of MDC tools. In addition to the automatic metadata and timestamps, the additional feature of GPS (varies by hardware selection) allows tracking the enumerator's route to further validate that the supervisor's directions were followed. This enables staff to give informed feedback to the enumerator teams and improve their performance.

1.2.6 Benefit – User friendly

Using the MDC technologies also empowers enumerators and builds their capacity for utilizing information technology. The possession of MDC devices can even enhance credibility of the data collection or research and staff in the eyes of the local community. As the programming guides the enumerator through the interview, they are able to

6. Krishnamurthy, R., Frolov, A., Wolkon, A., Eng, J.V., and Hightower, A. (1991), Application of pre-programmed PDA devices equipped with GPS to conduct paper-less household surveys in rural Mozambique.

CONTENTS & FOREWORD

SECTION I

SECTION 2

SECTION 3

CONCLUSION

ANNEXES

RESOURCES

spend less effort in making sure that all the correct questions are asked, and are thus able to better concentrate on other relevant aspects of interviewing and maintaining good rapport. The more complicated the conditions for asking a question (for example, only ask if three preceding questions have a certain answer), the more evident this benefit will be.

An enumerator collecting data with a mobile device does not need to carry around large numbers of survey material; all that is required is the device itself and perhaps a charger or backup battery. This reduces the hassle and burden of traditional pen and paper methods.

On the other hand, it must be ensured that the possession of the data collection device does not endanger the enumerator when operating in highly insecure areas. MDC has been conducted successfully in a number of potentially risky

> **MDC IN ACTION**
>
> In its study of impact of its *Reinsertion Reintegration* programming with current South Sudan SPLA soldiers, UNESCO researchers were surprised by the excitement of soldiers to participate in surveys using smartphone technology. Rather than becoming resistant to questions as the researchers feared, soldiers lined up to have their responses recorded on smartphones. The ease of data collection resulted in fieldwork ending two days sooner than was originally anticipated.

locations due to hardware choices appropriate for the local context and clear safety and equipment management protocols.

EVIDENCE: A World Bank study found that 88.9% of enumerators preferred mobile data collection tools. [7]

1.3 How Exactly Does Mobile Data Collection Work?

To reiterate one of the key points of this document, mobile data collection tools are not fundamentally different than any other data collection tool, and follow the same logic of standard data collection and research. As always, designing a survey will begin with the questions that require answering, followed by determining what types of data are best suited for answering the questions, and then deciding what methods are best suited for collecting the required data. No matter what tool is used, data must be firmly underpinned to well-defined and operationalized research questions and purposeful data use design in order to provide quality results.

However, insofar as mobile data collection tools are considered applicable to the project (more in Section II), the process for implementing the data collection has some differences in comparison to paper and pen methods. As you may notice from the representation of the processes below, however, the two data

collection methods are fundamentally similar from the research perspective.

The most significant difference is in the process of preparing tools for the field, training enumerators, and getting the data ready for analysis. Overall, the time from project start to finish should be reduced, but some days will likely be shifted to the front end of the research as programming and additional rounds of testing are required to finalize the collection instruments.

The testing of the survey instrument on a mobile device can take place only after it has been programmed and downloaded, but it is important to note that the content of the survey can be tested as any pen and paper questionnaire. The testing of the survey, thus, differs by the necessity of testing the technical implementation, in addition to pre-testing questions for content.

7. Schuster and Brito, op. cit.

Phases of a data collection project:

CONTENTS & FOREWORD

SECTION 1

SECTION 2

SECTION 3

CONCLUSION

ANNEXES

RESOURCES

MDC
(also see ANNEX A)

Pen & Paper

#	MDC		Pen & Paper
1	Define research questions	⊘	Define research questions
2	Select suitable data	⊘	Select suitable data
3	Develop methodology	⊘	Develop methodology
4	Create data collection material (survey instrument)	⊘	Create data collection material (survey instrument)
5	Test content of survey instrument	⊘	Test content of survey instrument
6	Convert survey instrument into electronic form	• • •	N/A
7	Load electronic survey instrument onto a mobile device and test	• • •	N/A
8	N/A	• • •	Print hard copies of survey instrument
9	Train enumerators on the survey *and mobile data collection*	←→	Train enumerators on the survey
10	Collect data *with mobile devices*	←→	Collect data *with pen and paper*
11	*Export data* over the Internet or *cable* into a database	←→	Enter data *manually* into a database
12	Clean data for analysis	⊘	Clean data for analysis
13	Analyze data	⊘	Analyze data
14	Write up, present, and share results	⊘	Write up, present, and share results

Activity	Time added	Time saved	Cost added	Cost saved
Program survey into electronic form	Per x number of questions		Per x number of questions	
Load surveys onto tablets and test				
Print survey material		Per page and # of observations		Per page and # of observations
Train enumerators	See table below		See table below	
Gather data				
Enter data to database		Per x number of observations		Per x number of observations
Total				

Enumerator type	Added time
Highly tech-savvy	Will learn themselves while going through the survey, no significant added time
Somewhat tech-savvy	Over the course of a few hours or a 1 day training course the enumerator will be functional in MDC.
Not at all tech-savvy	After a 1-2 day training course and subsequent field practice the enumerator will be functional in MDC.

Another difference between the two data collection methods is that the enumerators must be familiarized with using the MDC device. However, the widespread coverage of information technology, combined with the intuitiveness of MDC devices and tools, makes training relatively easy. The general survey training can also include MDC familiarization components that are integrated with questionnaire review and fieldwork preparation sessions.

The user friendliness of different software and hardware does vary, and this may be something to consider when making decisions on which to use (see Section III).

Altogether, experiences suggest that the time savings extending from the automation of survey components is more than sufficient to cover any added time in learning how to use the MDC tools. Additionally, these front-end costs are further eclipsed by the effort saved in the data entry phase, where mobile data can be simply exported rather than requiring reentry. This difference becomes more pronounced the larger the datasets become.

It is critically important for the successful use of MDC that a correct assessment of the project's needs (see Section III) is made, the exact tools are chosen accordingly, and integrated in the survey design at the earliest possible stage.

1.4 Limitations of Mobile Data Collection

As with any method, MDC has its strengths and limitations. As illustrated in this section, the benefits are numerous and the tools flexible. They can be used to do many different things. Although things such as bad electricity grids or Internet connections can pose challenges, most can be overcome with appropriate planning. However, it is good to note that as MDC is not a panacea for poor survey design, it is also not a tool that will enhance all data collection efforts.

One of the main limitations of the mobile devices is the difficulty with which text can be typed in, as keyboards commonly lack 10-finger typing functionality and systems are not amenable to making side notes with ease as open ended or qualitative interviewing may require.

For these reasons, MDC tools do not currently extend well to qualitative methods, where large amounts of open text need to be gathered. In such projects, the ability to make quick notes with pen and paper still tends to be more efficient and effective. However, mobile forms may be used to supplement qualitative data by collecting information such as GPS points and photos.

1.5 Frequently Raised Concerns

This section will provide examples of concerns that organizations often raise about using MDC, or as reasons not to, as well as effective answers to address them:

- *Beneficiaries won't like it*
  It is quite common for organizations to argue that their beneficiaries would not like to be interviewed using the electronic devices. Although it may be true that some individuals might have some initial suspicion regarding the device, no systematic evidence to support this concern has been found. MDC tools have been successfully applied in some of the toughest security conditions in communities that are traditionally skeptical of outside interventions.

- *Enumerators cannot learn it*
  Another common misconception is that using the mobile devices is extremely difficult and enumerators without extensive training will not be able to use them for interviews. Although some general limitations, such as literacy, apply (as they do with paper and pen surveys), enumerators have been successfully trained from amongst the least educated groups of people. As evidenced in a World Bank study, 88.9% of enumerators preferred this method.[8]

- *Electronic devices are unreliable*
  Devices operate on batteries that run out. They drop on the ground and break. Software freezes. These are among the common concerns people have regarding MDC devices; while legitimate, these concerns can be addressed relatively easily in the preparation phase.

  If electricity is an issue, choose long-lasting battery hardware, take extra batteries, and plan places where recharging can be done. Devices do break from time to time, so take an extra device with you just in case. If you are operating a phone, buy shock absorbing cases. Software may cause concerns from time to time, but as long as the programming phase of the project is handled correctly, these issues are generally minimal.

- *Paper and pen has worked in the past*
  An organization may be set in its ways and comfortable with the status quo. And, of course, paper and pen methods do work. However, MDC often works better. The evidence from numerous studies shows significant decreases in time and effort, combined with higher quality.

- *It is too expensive*
  Numerous studies have provided evidence that the investment required for utilizing MDC pays itself back and much more. Of course, this requires a sound assessment of the research needs and the choice of appropriate tools, on which you will find guidance in the coming sections. However, as with any technology, the more you use it, the cheaper it becomes, as the capacity that is built up can be applied to more studies in the project or program. An initial investment in equipment may seem daunting, but the equipment may be reused for multiple rounds of data collection across projects. If the equipment is purchased with USAID funding, an implementer may propose in their disposition plan that it be granted to their organization or to another organization or project with similar data collection needs.

- *It will be too difficult*
  Utilizing MDC does require some new capacities, both hardware, skills, and competencies. However, these are not necessarily extensive, and can be secured by a number of ways. Section III of this manual, however, will guide you through the process and help you make the right decisions through the transition.

- *Our project does not need it*
  This is the concern that has the most potential legitimacy. It is true that not all research projects or M&E frameworks benefit from MDC. The amount of benefit depends purely on the research question and whether the type of data required is that which MDC tools are best suited to collect. To evaluate the suitability of MDC tools for your research study, project, or program, consult Section II.

8. Schuster and Brito, op. cit.

CONTENTS & FOREWORD

SECTION 1

SECTION 2

SECTION 3

CONCLUSION

ANNEXES

RESOURCES

PAPER TO MOBILE DATA COLLECTION

CONTENTS & FOREWORD

SECTION 1

SECTION 2

SECTION 3

CONCLUSION

ANNEXES

RESOURCES

SECTION

2

IS MOBILE DATA COLLECTION USEFUL FOR YOU?

- Evaluating the Benefits from Utilizing MDC Tools, Single Study

- Evaluating the Benefits of Transition to MDC Tools, Project & Program Level

- Evaluating the Costs of Transition

2.1 Introduction

As with any new tool, there are transitional costs involved before any of the benefits of Mobile Data Collection can be realized. The costs are related to securing the proper hardware and software for the project, as well as developing the skills required to manage such components. The benefits, as elaborated on in the previous section, relate particularly to speed, quality, as well as cost. Done correctly, MDC should improve both quality and speed of data collection, but may increase the cost in some projects, especially upfront.

While it is possible to utilize MDC tools on a single project, it can also be associated with a larger programmatic shift in the tools used for M&E. In development, each research study generally happens within a project's M&E framework, which again

should be developed at the outset of a program. In some cases, transitioning to MDC tools (or building such capacity) may make sense only for a certain evaluation, sometimes for an entire project, and sometimes the capabilities would benefit an entire program. Although, again, a comprehensive list of when higher-level organizational transition becomes beneficial is impossible to lay out, some general rules for evaluating this will be developed.

It is important to note that the benefits and costs assumed in this section apply only when the transition is made correctly, and guidance towards this will be provided in Section III. There are also some aspects in specific projects that may require special attention in the design and implementation phase, and these will be highlighted in this section.
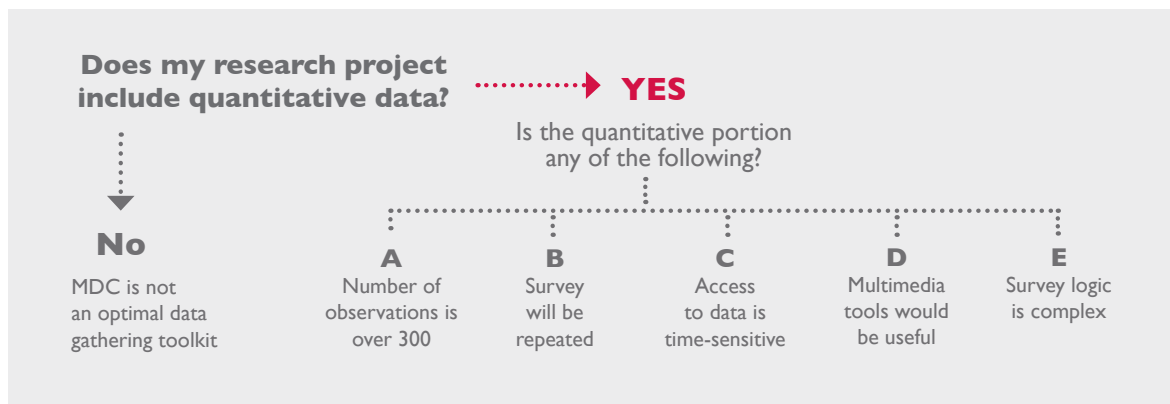
2.2 Evaluating the Benefits from Utilizing MDC Tools, Single Study

Before assessing possible benefits for organizational transitions, it is essential to understand when it is that MDC tools are most useful, and in which kinds of research studies their use is limited, at best. This offers the core framework for assessing any higher level benefits.

As noted in the previous section, MDC offers very powerful tools for the gathering of certain types of data more effectively and efficiently, primarily data of a quantitative nature. Evaluation of the possible benefits from transitioning can only be estimated after the first steps of designing the M&E framework for the project has taken place. Program staff must first and foremost define the relevant research questions, how they will be evaluated, and what kind of data is required for addressing these. Only after this can the benefits from MDC be explored.

The graph on page 21 summarizes key aspects of the data and how they relate to the benefits of transitioning to MDC:

The first question to consider is whether the research questions call for quantitative data. For example, when there is insufficient information regarding the question to form testable hypotheses, qualitative methods for exploratory research may be most suitable. Quantitative methods would be unsuitable as too many questions would require an "other" answer, which would make data analysis laborious and lose out on the depth behind those answers. Some questions that wish to go deep into the motivations, desires, and understanding of the respondents, require qualitative tools.

CONTENTS & FOREWORD

SECTION I

SECTION 2

SECTION 3

CONCLUSION

ANNEXES

RESOURCES

Does my research project include quantitative data? ‧‧‧‧‧‧➤ **YES**

Is the quantitative portion any of the following?

No

MDC is not an optimal data gathering toolkit

A
Number of observations is over 300

B
Survey will be repeated

C
Access to data is time-sensitive

D
Multimedia tools would be useful

E
Survey logic is complex

Using MDC tools for qualitative questions would likely be inappropriate (see also subsection 1.4), though this relates to the limits of quantitative research rather than MDC itself. For such cases, traditional pen and paper, or typing directly into a text file on a computer would typically be more useful. However, as noted in the previous section, although MDC tools are not typically as useful for gathering qualitative data, they can offer additional metadata (such as time, length, verification of location, etc.) that can improve the quality of the other methods being used.

If the quantitative portion of your data collection meets none of the five criteria from the previous graphic, the benefits from MDC tools will be minimal at best. The start-up costs for setting up MDC capabilities may indeed be prohibitive, and pen and paper methods more appropriate. However, keep in mind in that if the project or program builds the capacity for the purpose of supporting multiple studies, the start-up costs may become dispersed so as to make it beneficial to adopt MDC tools for research studies that would not justify the investment on their own.

The three primary criteria for assessing MDC usefulness are the number of observations, repeated or continuous survey need, and time-sensitivity of data. Some additional benefits may be reached using MDC tools if the survey requires multimedia or other added features, or very complex filter logic.

2.2.1 Number of observations

Limiting the scope to a single project, there are multiple aspects that increase the benefits of transitioning to MDC. Firstly, if the number of respondents is over 300 (2a), the savings from data entry alone are significant. It is worth remembering that with using paper and pen methods data entry does not benefit from economies of scale, as each and every observation needs to be manually entered. As such, disregarding the setting up of the data entry process, the time per survey is not significantly different between a survey of 100 and 1,000 respondents.

With MDC tools, increasing the size of the survey always reduces the time it takes per survey to transfer from interview to database, leading to significant economies of scale. After a certain point, the savings from this alone will cover the costs of all the investments that were required to building the capacity.

2.2.2 Repeated or continuous surveying

Secondly, although the number of the survey respondents may be small, continuous or repeated surveys (2b) may increase the number of observations to a degree that makes MDC desirable. For instance, baselines, mid-term evaluations, and end-of-project studies should use the same methodology to ensure comparability of the data time points.

For other projects, data gathering may be more continuous and observation numbers rise regardless of respondent numbers.

2.2.3 Quick access to data

Thirdly, and particularly for continuous surveying projects, the quick access to data (2c) may be an essential need. For example, when surveying institutions such as hospitals and their attendance rates, or crime statistics, using MDC tools can provide quick response capabilities as data can be ready for analysis in real time. By the time that data becomes accessible using paper and pen methods, the situation may already be at a stage where responding to it is quite costly or even impossible. Depending on the number of observations, MDC may not in this case bring savings, and may even increase costs. However, the benefit from improved speed may make it an acceptable trade-off.

There are additional program implications of rapid data collection: analysis and programmatic course correction may improve programs as a result of accessing data in a timely manner over the program life cycle. Integration of real time data collection into programming is beyond the scope of this toolkit, but it is important to note that this technology can represent a shift in the way projects are monitored and ultimately allow timely action to adjust activities.

2.2.4 Survey logic complexity

Fourthly, some surveys require (or would greatly benefit from) complex logic on which respondents answer particular questions. Paper and pen methods often have a limit on this complexity imposed by what enumerators can be expected to effectively handle. Generally, human-controlled filter logic cannot be expected to be reliable much beyond skipping the next question in chronological order. Having conditions in later questions that require multiple particular answers to questions in previous sections would greatly increase risks of the wrong people being asked, and at the very least, the interview flow being disrupted by checking previous responses.

MDC allows for great complexity to be introduced with high reliability and no interruption to the flow of conversation. Although it may increase costs, improvements in the quality of the data may lead to a trade-off that is desirable.

To know whether your survey would benefit from using complex survey logic, you must assess whether you would like to ask multiple follow-up questions from certain groups that cannot be isolated in sampling (for example, due to information on the criteria not being readily available.)

2.2.5 Added features

Finally, if the survey's quantitative portion requires any of the numerous possible added multimedia or GPS features enabled by MDC, utilizing the technology may be advantageous, or even necessary. For example, if your project would benefit from verification of enumeration locations or people via pictures or GPS, backing up paper registries, the use of multimedia tools such as graphics or music (particularly useful with illiterate populations), or capturing voice or video, you should strongly consider MDC tools. As discussed in Section I, these added tools can be used both to gather new types of data and improve the quality of data by increasing enumerator accountability or interviewee comprehension.

Program staff planning to make use of these features, however, should make sure that they make the right choices in terms of software and hardware when acquiring the capabilities for transition.

CONTENTS & FOREWORD

SECTION I

SECTION 2

SECTION 3

CONCLUSION

ANNEXES

RESOURCES

2.3 Evaluating the Benefits of Transition to MDC Tools, Project & Program Level

As a project or program officer, you may have multiple ongoing research studies. In such cases, capacity building and MDC costs may be dispersed over multiple projects but may also require higher levels of buy-in across both your organization as well as the implementing partners.

A quick assessment of your research study, using the tools laid out above, should give you an idea of how much potential benefit there may be in including MDC into your program. It should also allow you to get an idea of the requirements, staffing, and institutional capacity that will be required to utilize MDC.

Another consideration is whether MDC should be coordinated at a national level with the USAID country office, or if leaving MDC integration to the project-based implementing partners would be preferable. For missions with many programs that could benefit from MDC, a centralized platform could enhance uptake, as well as ease the transition process for partner organizations. Alternatively, MDC uptake at an individual program level can allow for more customized considerations. Overall, the capacities, goals, and project requirements will dictate the best implementation models.

2.4 Evaluating the Costs of Transition

Although the different choices made in building MDC capacity will have huge effects on the costs incurred in the transition, it is helpful to understand some of the basic sources of those costs. MDC capabilities require investment in the following three areas: skills, hardware, and software. In addition, these investments occur in two distinct phases: the costs related to the creation of MDC capability and those related to maintaining them. Note that different choices explored further in Section III have significantly differing amounts and distributions of costs between the three main sources as well as the two types.

Understanding the sources of the cost is particularly important for programs that need to make decisions on which parts require building MDC capacity.
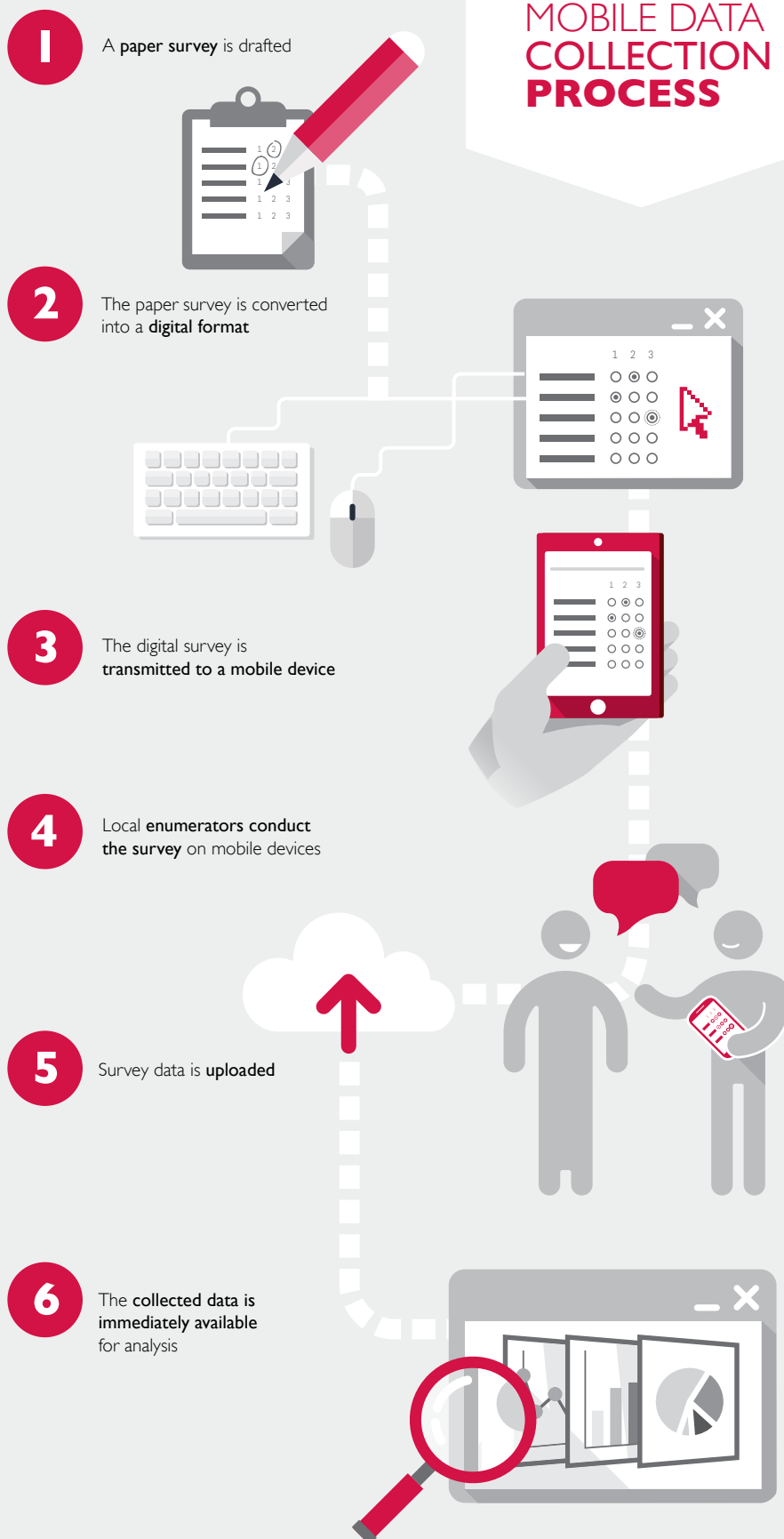
2.4.1 Start-up costs

In order to successfully conduct research projects using MDC tools, the staff or researcher(s) require certain skills. As preparations for the fieldwork are done with a computer, basic computer knowledge

is mandatory, and advanced proficiency is desirable. Surveys need to be converted into a digital form and tested. The more complexity that is added to the survey, the more skills the person conducting this conversion needs. However, some software is easier to use than others. If this aspect of the MDC capacity is foreseen as a problem, the choice of software should reflect this. Many open source tools are available and should be evaluated to see if they may meet the project's MDC needs; some of the expertise needed for programming may also be outsourced (more on that in Section III.)

In addition to the conversion of surveys into an electronic form, people with data cleaning and integrity checking skills are needed to sort through the database created when data is exported. Some software outputs messier data than others, increasing the importance of this capacity. The organization will also need either enumerators able to use the MDC devices or a person who is able to train enumerators to use them for different research projects. However, this capacity is generally not difficult to build up as the tools are very intuitive in practice. This capacity is only

MOBILE DATA COLLECTION PROCESS

1 A **paper survey** is drafted

2 The paper survey is converted into a **digital format**

3 The digital survey is **transmitted to a mobile device**

4 Local **enumerators conduct the survey** on mobile devices

5 Survey data is **uploaded**

6 The **collected data is immediately available** for analysis

maintained as long as the personnel with the skills are retained, and can be "lost."

Hardware purchases consist of the second large aspect of building the capacity for MDC research. Enumerators require a device for the collection of data (either a phone, a tablet, or other mobile device) and the number of enumerators will determine the number of required devices. Beyond this, the person converting the survey and analyzing the data needs a computer. The amount of this cost varies greatly and is determined by the needs of the project, both in the required number and functionality of the devices (see Annex C.) Outside of losing the devices through breaking or their standard life-cycle, hardware is not "lost", but may be transferred from actor to actor (though shipping can be costly and purchasing new phones can even be cheaper than shipping in some cases.)

Software costs, much like hardware, depend greatly on the choice of the provider. Some software is free, some requires an up-front investment, and some is pay-as-you-go. As you consider the needs of the project, particularly using the questions provided in the figure in the beginning of this section, the software choice will become easier. The same software can, in most cases, be used by multiple projects.

2.4.2 Running costs

Maintaining the capacity for MDC tool utilization has some limited running costs. Firstly, the skills required for running the technological side of the research must be retained. If the person with these skills leaves the organization, the hiring process for a replacement may incur additional costs. It may also be desirable to use enumerators with existing MDC device skills to reduce training requirements in the future, and this retention may impose some extra costs.

CONTENTS & FOREWORD

SECTION I

SECTION 2

SECTION 3

CONCLUSION

ANNEXES

RESOURCES

**Does your organization have
any of the following:**

Staff with
computer/IT skills? — **YES** →   **Potential for
reduced cost**   ← **YES** —   Existing
hardware?

Hardware can, and does, break. It is wise to budget some extra devices for when ones break unexpectedly (for example, due to impact) and plan for general wear and tear. As with the initial procurement costs, the choice in devices will have implications to both likelihood of unexpected breaking and general length of usability.

If software with a pay-as-you-go payment model is chosen, there will be some running costs related to the software as well. However, many software options do not incur such costs.

2.4.3 Existing capabilities that may reduce costs

Although some of these costs may feel substantial and new, it is quite possible that your organization already possesses at least some of the building blocks for MDC capability. For example, it is quite likely that your organization will already be using computers, and possibly even company phones. Computers only require the installation of specific software, and thus a part of the hardware procurement costs are avoided. Phones, provided that they meet the criteria in terms of functionality, can also be brought on as MDC devices (see Annex C.)

Your organization is also likely to employ staff with varying levels of IT skills. Computer proficiency and a keen willingness to learn new things goes a long way in creating the capability needed for a successful MDC transition.

In some cases, even personal phones can be used for MDC, though it is important to set out clear arrangements for what happens when one breaks. Procedures for removing the data from the phone's memory must also be established to secure confidentiality.

CONTENTS & FOREWORD

SECTION 1

SECTION 2

SECTION 3

CONCLUSION

ANNEXES

RESOURCES

SUCCESSFULLY IMPLEMENTING MDC

SECTION 3

- Assessing Research Project Complexity

- Choosing the Right Tools

- Organizational Needs

3.1 Introduction

Having reached this part of the manual, you will have hopefully determined that mobile data collection tools offer some benefits to your research study, project, or program. The next vital step is choosing the right tools from within the huge variety of options available to you within the MDC world.

As in the previous section, making good decisions requires that you understand your needs. The successful implementation of MDC tools begins, as any research methodology, with sound and precise monitoring or research design. In development work, this is generally set out in the M&E framework. This, in turn, is determined by the needs of the project; when making decisions regarding your use and implementation of MDC tools, always relate the questions back to those specific needs.

Knowing what is needed from the data will allow you to correctly identify not only the benefit of using MDC, but also exactly what type of MDC tools best fit your needs. MDC comes in a large variety of shapes and sizes with different levels of complexity, costs, benefits, limitations, and so on. Trying to transition to MDC by concentrating on the tools would inevitably lead to an information overload, which can only be navigated by firmly subjugating the tools to the researcher's needs.

This section will guide you in evaluating the complexity of your research project, the implications this has on the MDC tools you should use, and in conjunction with assessing your wider program or organizational needs, makes some recommendations on how you should build the capacity for using the tools. The choice of MDC tools will be largely determined by the complexity of the research project, whereas program/organizational needs are more relevant when deciding on how the capacity for using those skills should be created.

3.2 Assessing Research Project Complexity

In order to be successful in choosing MDC tools for a project, it is vitally important to assess and thoroughly understand the goals of acquired research data. Although the MDC tools only become part of the research in the implementation phase, it is crucial that their use is already accounted for and incorporated early in the design phase. A research project that begins considering MDC tools only at the point when they are actually needed is very likely to be unsuccessful in utilizing them.

Project complexity, for the purposes of an MDC tool requirement assessment, relates broadly to two distinct entities: the complexity of the actual data being gathered (the survey) and what the project research/M&E framework requires from the collected data in the analysis phase. They are separated here for easier understanding, but they are not independent from each other and overall complexity should never be assessed without reference to both. In addition to these, location difficulty is also an important consideration, referring to the relevant physical conditions within which your data collection takes place.

3.2.1 Survey complexity

When assessing survey complexity, there are five main components to be consulted. These include the number of respondents, the number of questions, filter logic, language, and need for additional features. Note that even though the overall survey complexity may be low, even a single "high" or "medium" will have implications to the choice of tools.

The size of the survey, both in regard to the number of respondents and the number of questions being posed to them, is a question of key importance to MDC tool choices, particularly for the software. However, estimating them is a process that is not done differently for MDC, but follows exactly the same procedures as it would for an M&E framework using paper & pen data gathering. Similarly, assessing the need for multiple languages should not differ from traditional methods.

As noted before, filter logic refers to the program prompting questions to the interviewee depending on previous answers. In its simplest form, filter logic would include a function like "if person responds no, skip next question." In increasing levels of complexity, it may include skipping whole sections, and filtering questions depending on multiple previous answers.

At its most complex, it may also include algorithmic functions that filter questions calculated from the ranges of previous numeric answers.

Filter logic is an issue that may appear more difficult to estimate in the beginning. However, some key questions can help in making an informed choice with filter complexity. In particular, if the research question includes multiple groups of interest, it is likely that certain sections may relate to only some of these groups. This would require at least medium complexity filter logic. If the project hopes to pinpoint very particular subsets within its target population for some specific lines of inquiry but does not limit the whole data collection effort to them, complex filter logic is likely to be necessary.

The need for some of the additional features offered by MDC tools should only be used if they are required by the M&E framework and project data needs. For example, although the use of GPS and taking photos (for example, of the respondent's house) can be used to improve enumerator monitoring and the quality of the data, it can also raise some confidentiality issues as response data can be combined with response location, thus enabling the identification of the respondent. In such cases, the level of data encryption offered by the software should be considered.

Table 1: Survey Complexity

	Number of Respondents	Number of Questions	Filter Logic	Language	Added Features
Low	<300	<50	Max 1 question skip	English	None
Medium	300–1,000	50–120	Section skips	Multi-language, Latin script	GPS
High	>1,000	>120	Filters based on multiple variables	Multi-language, Non-Latin scripts	GPS and graphic design tools

CONTENTS & FOREWORD

SECTION 1

SECTION 2

SECTION 3

CONCLUSION

ANNEXES

RESOURCES

Questions that incorporate graphic tools or music, on the other hand, take a longer time to create than traditional word-based questions, and need further testing to make sure that they are indeed widely understood. However, they may of course be necessary, particularly with illiterate populations.

3.2.2 Complexity of Data Analysis

The analysis phase can require very different things from the data. These needs, in turn, impose requirements on the tools for data collection. As one of the main benefits of MDC tools comes from the speed with which they can bring data to the analysis phase, making the right assessment of what will be needed is of the utmost importance.

The first two considerations, the type of data and the database platforms, refer to what the program staff hopes to get from the data. The project may only include one data collection effort and an output of a few graphs and frequencies, in which case the low complexity of Excel may be sufficient. However, when the output requires any statistical methods, exportability to only Excel may be problematic

as transferring the data to a statistics package and preparing it for analysis is time-consuming. In addition, if any databases are anticipated to require merging, as would happen in, for instance, longitudinal data, this is something that must be properly prepared for during the MDC design. Full SSS compliance (public sharing of data) refers to the possibility of easily combining existing datasets and moving data between different platforms and analytical programs.

The latter two, field analytics and data access, relate to when access to data is required. Particularly for larger surveys taking place over a number of weeks, having access to the gathered data is extremely important in order to properly respond to possible problems. For some projects, manually checking data will be sufficient and possible, but some programs also offer built-in summary statistics that can be viewed nearly in real time. Of course, this requires that data can be uploaded from the field in regular intervals. Some M&E frameworks may also benefit from continuous uploading of data in order to keep the project responsive to time-sensitive needs arising in the field. The more complex the data becomes, the more important it is also to consider the ease of data cleaning and ensuring effective quality control.

Table 2: Complexity of Data Analysis

	Type of Data	Database Platforms	Field Analytics	Data Access & Logistics
Low	Single, one-off study	Excel	None	Data brought to office
Medium	Repeated (e.g., baseline/impact)	Single statistics package (SPSS/Stata)	Data manually accessible	Manual upload from field
High	Longitudinal or multiple waves	Full SSS compliance, ability to merge datasets	Built-in real time summary statistics	Automated continuous upload and access

CONTENTS & FOREWORD

SECTION I

SECTION 2

SECTION 3

CONCLUSION

ANNEXES

RESOURCES

3.2.3 Complexity of Local Context

Most data gathering takes place in the field, and the local context and conditions have implications on the choices you should make when designing the fieldwork.

The meanings of these aspects are quite self-explanatory, but their implications for tool choices will be expanded upon within the coming subsections. IT infrastructure will mainly relate to choices in software, whereas electricity and security are most important when it comes to hardware choice.

Table 3: Complexity of Local Context

	IT Infrastructure	Electricity Infrastructure	Security Situation
Low	Stable high-speed internet	Electricity sockets commonplace	Low risk of crime
Medium	Stable low-speed internet	Electricity available, but not common	Medium risk of crime
High	Unstable low-speed internet	No electric grids	High risk of crime

3.3 Choosing the Right Tools

The choice of the right MDC tools for the project is crucial for receiving the benefits that they have the potential to create. The reduction of start-up and running costs are likely to be reversed if complex projects are attempted using basic tools. On the other hand, more complex tools can be brought to simpler projects, though the amount of investment cost may make their use sub-optimal.

3.3.1 Software

This manual's purpose is not to recommend any particular software over another (any such recommendations would soon become obsolete as different software options develop.) Your choice of software should be mainly guided by the needs arising from the survey complexity. When making

the decision, it is worth going through all these aspects point by point to make sure that the software can meet those needs.

In terms of location difficulty, and insofar as uploads from the field are required, the availability and quality of the Internet is also an important aspect to consider. The more difficult the conditions, the more important it is that the chosen software can compress data files to a size that can be transmitted even through unstable and slow Internet connections. Different software also has different protocol for dealing with short breaks in connection. Some do not register these breaks and, thus, some data does not get transmitted properly, whereas other software checks the integrity of the file and retransmits when necessary.

Beyond these needs that the software must meet, there is a general decision that warrants additional explanation. This is whether to use open-source freeware or a market-based solution. Freeware can, in most cases, reach similar levels of complexity as paid software when it comes to questionnaire design, filter logic, multimedia tools, analytics, etc. There can be a major temptation to try to reduce costs by not investing in software and going for the free options.

Although many complex projects have been successfully carried out with free open-source software, a person utilizing them should be prepared for additional capacity building. While user-friendly freeware survey builders do exist, some level of knowledge in writing syntax is required from the person transforming more complex surveys into digital form. Program staff must also be able to run multiple different programs in the course of the study, from a survey creator to data collector to aggregator and analyzer. Finally, open-source software may have certain online resources for peer-to-peer support, but there is no guarantee of support if any problems should arise.

None of these considerations are necessarily reasons not to use freeware tools. However, they are aspects that should be taken into account when assessing the skills needed for successfully implementing MDC tools in M&E and research.

3.3.2 Hardware

There are many types of mobile devices with widely varying capabilities and costs available in the market today, which can make selecting a device feel like an overwhelming task. Models range from simple and cheap to high-end and expensive, and the technology develops and changes rapidly (see Annex C). However, for most data collection needs, even a standard low-end smartphone will be sufficient.

The hardware in general should not be a driving factor in guiding your MDC design. As MDC in itself is subservient to the project and research design, hardware should be mainly determined by the needs of running the software most suited to the data gathering. As you ask the critical questions regarding software and make your decision, ensure that whatever hardware you would like to use is compatible with it.

However, there are some additional considerations regarding hardware that go beyond the ability to run the software. One of the most important of these relates to the data gathering environment. When working in areas with low electricity infrastructure, having a long battery life and the availability of adapters to utilize different electricity sources becomes critically important (car chargers, solar chargers, etc.). Secondly, screen sizes vary and some software may require a certain size in order to run smoothly. Issues may also arise with durability and resistance to shocks and the elements. Finally, in areas with high crime rates, high-end phones and tablets are not advisable, and in some situations where noticeability should be minimized, phones are preferable to tablets.

Some of the needs arising from the complexity assessment of the project may also have implications for hardware choices. For instance, if GPS or graphical elements are required, the phone or tablet should have built-in GPS and a screen large enough to clearly present the graphics. Also, if uploading from the field is required, a stable 3G or Wi-Fi function becomes necessary (though this more often depends more on the IT infrastructure in a given area).

The best route to take, especially if the hardware will serve in more than one single project, is to pilot a few different phones or tablets to find the one that

is best suited for the particular project environment. Pay special attention to the features needed for your project, as the quality of certain features such as GPS may vary dramatically from one model to another.

3.3.3 Skills

The third major component of the MDC capacity is the skillset required to make the correct decisions on software and hardware, and effectively implement the MDC tools in research. The complexity of the project, and thus the software, will naturally impose demands on program staff designing and implementing the data collection. The choice and skills required by the software also have implications for these choices.

Three main options can be identified for this:

* Building internal capacity through DIY training (not generally recommended due to probable high degree of trial and error)
* Hiring an external consultant to conduct trainings and build the capacity of the organization for managing MDC tools. These options both

create institutional knowledge and expertise that is sustainable and can be used on multiple instances. However, the required investment can impose high costs. As such, it is most useful in cases where the project has a long lifespan and the organization deals regularly in data collection.

* Outsourcing some or all of the MDC elements of the data collection to an external consultant or firm that specialize in MDC. Although outsourcing can also be costly, these costs are likely to be less than building the capacity internally, and the consultant/firm can be held accountable for any mistakes. As the organization will accrue only limited institutional knowledge, this option is better suited for situations where capacity is needed only for distinct studies. However, partnering with a research firm specializing in MDC may also reduce costs and ensure a steady supply of expertise.

In order to make an informed decision on this question, the element of organizational needs must be considered.

3.4 Organizational Needs

Whether and how to build internal capacity, or whether to seek outside support for designing and conducting MDC research depends largely on the wider needs of the project, program or organization. As a general rule, high investment in organizational capacity that becomes redundant after a given project, as well as dependence on external actors for capacities that require low investment and have constant demand, are undesirable. Both can become relevant in MDC. The table below helps illustrate the key questions and recommendations for capacity building:

In addition to these aspects, remember to consider the level of computer proficiency of your staff, as well as the time frame before the research fieldwork must be carried out.

As noted in the previous subsection, more complex tools can be successfully applied to simpler projects, but not vice versa. Therefore, when deciding on how to secure the capacity for conducting MDC, keep in mind all the possible uses for MDC in your organization. The capacity you build should be determined by the most complex project your organization will undertake, rather than the simplest.

CONTENTS & FOREWORD

SECTION 1

SECTION 2

SECTION 3

CONCLUSION

ANNEXES

RESOURCES

The more often your organization conducts research that could apply MDC tools, the more incentives there are to create the requisite capacity in-house. On the other hand, the higher the complexity of the survey tools and methodology, the larger the incentives are for also involving an entity with further specialization in research using MDC tools. The additional benefit of outsourcing or hiring an external consultant to guide the organization through the transition is that they would either provide or advise on the choices regarding software and hardware, relieving some of the stress from the organization itself.

In Annex E, there are some critical questions that you can use to evaluate a consultant or firm and their MDC tool design and implementation offering.

Table 4: Building Organizational MDC Capacity

Level of research project complexity	How often the organization undertakes research	
	One-off/rarely	Frequently
	How to secure capacity	
Low	Train staff member with basic knowledge of MDC tools	Have all research staff trained in basic MDC by external consultant
Medium	Hire specialist consultant to advise M&E team on MDC	As above, but also hire an MDC speciality in the organization
High	Hire specialist consultant to advise M&E team on MDC	Partner with a research firm specializing in MDC or build strong internal capacity

Concluding Remarks

....

Mobile data collection tools offer many exciting and innovative possibilities for improving the quality of M&E and research projects. From the project manager to the researcher, it enables cheaper and quicker access to a better quality of analyzable data than traditional pen and paper methods, and allows increased responsiveness to needs arising in the field. In this sense, it is a toolkit with implications beyond research, as when they are effectively built in to an M&E framework, the tools can greatly support different programmatic needs and aspirations. It is a toolkit that is becoming increasingly utilized by different actors globally with positive results.

Despite its innovativeness and technologically advanced tools, it is crucial to identify the fact that all the underlying principles of MDC are exactly the same as data collection with any other tools. Successful use of the MDC toolkit will still begin with a good formulation and operationalization of the research questions, choosing the right type of data to address them, and analyzing the results with scientific rigor. MDC should be seen as something that empowers researchers rather than something that replaces them.

As tools, they are only as effective as the researcher applying them, and only truly beneficial insofar as they are used in the right kinds of projects. A good understanding of what types of data MDC is best suited for, and making the correct decisions on which tools within MDC to utilize, are of key importance to reaping the full benefits the tools have to offer. As any toolkit, MDC has both strengths and limitations, and a dynamic of costs and benefits for any given situation.

This manual has sought to give the reader the requisite knowledge to start assessing these benefits and costs that MDC tools may hold for their project, and the organization or program, when applicable. It also gives guidance on how to go about designing and building the capacity to use them, so that the implementation of these tools becomes successful. The decisions that need to be made are numerous and the wide variety of options may feel overwhelming at first, but it is our hope that this manual has been able to demystify some of the aspects of MDC and encourage the use of mobile data collection across projects and organizations moving forward.

CONTENTS & FOREWORD

SECTION 1

SECTION 2

SECTION 3

CONCLUSION

ANNEXES

RESOURCES

CONTENTS & FOREWORD

SECTION 1

SECTION 2

SECTION 3

CONCLUSION

ANNEXES

RESOURCES

ANNEXES

Annex A: Decision Matrix for Platform Selection

As the field of Mobile Data Collection is in a state of constant development, any list that could be provided would risk becoming obsolete quickly as new software options arrive on the market or existing ones develop new capabilities.

In addition, the possible combinations of local contexts and data needs are so large that no comprehensive decision tree can be conceived. Instead, we have provided a list of steps and critical questions that can be used to guide you through the process of choosing the right solution for you.

In building the capacity for MDC, it is generally best to go from project needs to software choice and finally to hardware (Annex B). However, some existing capabilities may make considering hardware first more efficient.

Step 1 (taking stock of existing capacities)

Having decided to make use of mobile data collection tools, the best place to start building the capability for using them is to take stock of your existing capacities. Making good use of both skills and hardware already present in your organization can substantially reduce the costs for transitioning to MDC.

A1 Existing capacities

Existing capacity	Y/N	Implications
Does your organization already have a stock of devices (phones/tablets/etc.) that can be used for data collection?		If yes, consult table B1 on hardware needs (skipping question 1). If your devices meet all the requirements for the project, you may want to limit your software choices to those compatible with the hardware operating system (provided that software meeting your needs exists for that operating system). If no, proceed to the next question.
Does your organization have people with skills for using particular programs for data management/analysis/visualization/etc.?		If yes, proceed to table A2, question 1. If no, proceed to table A2, question 2.

CONTENTS & FOREWORD

SECTION I

SECTION 2

SECTION 3

CONCLUSION

ANNEXES

RESOURCES

Step 2 (choosing the appropriate software)

Having mapped out possibly relevant existing capabilities, please list all the software options that you are considering to use. After this, follow the questions in the table below to eliminate any software that does not meet the needs of your project.

A2 Critical questions and implications for software choice

#	Project needs and context	Y/N	Implications for software choice
1.	Does the project require the data to be used by existing data management/ analysis/visualization tools?		If yes, give strong preference to software that are easily compatible with the other software that you have expertise in using. If no, proceed.
2.	Does the survey involve displaying graphics or video?		If yes, ensure that the software has the capability of either displaying such multimedia within the survey or that the application can be resumed without problems if the home screen and multimedia area accessed in the middle of the survey. If no, proceed.
3.	Does the survey require enumerators to take photos that are recorded within the dataset?		If yes, limit your software options to those that support pictures within the dataset. Note, however, that if pictures do not need to be inside the actual dataset (if you simply want pictures from the field), no software support is necessary. If no, proceed.
4.	Does the survey seek to record GPS data?		If yes, limit your software options to those that support GPS data being recorded in the dataset. However, you may also use other means for GPS positions and input the coordinates as a separate question. If no, proceed.
5.	Does data need to be uploaded directly from the field?		If yes, inquire how large the files that the software creates are. You may experience significant problems trying to upload large files from the field. Also find out how the software deals with mini-second losses of connection, as some software does not identify these and uploads lead to small intermittent gaps in the data. These problems are especially pertinent in areas that have low quality Internet. If no, proceed.
6.	Does the survey include a large amounts of filter logic, including section skips or filtering through multiple variables?		If yes, make sure that the software option supports the type of filters that you want to build for your survey. Complicated skips are also easier to code in some software as opposed to others, so try to search for an example of how what you want to do looks like in the software. Also check how the complex skips affect loading times in the platform, as in some software it can lead to significant delays. If no, proceed.
7.	Does your survey require the use of multiple languages or non-Latin script?		If yes, make sure the software supports the use of any script you may need. If you need to use more than one language simultaneously, make sure the software supports this so you do not need to create two separate surveys for the same project. If no, proceed.
8.	Do you estimate that you will need external support in case where there are problems during the data collection process?		If yes, you may want to give preference to software that has a dedicated support service (generally, customer service). However, freeware also often has forums where users assist each other and support may be found. If no, proceed.

A2 Critical questions and implications for software choice *(Continued)*

9.	Will you need to merge the data coming in with other datasets?		If yes, you will want to make sure the data collection forms are compatible with your existing datasets in advance of selecting a software. If no, proceed.
10.	Will you need to build quick illustrations of the data as it comes in?		If yes, make sure you either have a staff member that can visualize the data or limit your software to those that have automated graphical outputs. If no, proceed.
11.	Do the devices need to be operated in areas with a low degree of services and infrastructure?		If yes, refer to implications on question 5. Furthermore, check power consumption estimates, as battery life may be an issue. If no, proceed.
12.	Will the devices be operated in areas with high levels of insecurity and crime?		If yes, we strongly recommend choosing a software that encrypts data as it is entered. If questions can be accessed by anyone who physically holds the phone, confidentiality is at risk. In some areas of high insecurity, respondents may be put at risk by their answers, in which case this consideration is particularly important. If no, proceed.
13.	Proceed by listing the software options that fulfill your criteria and making a choice at which one best suits your need. If there are multiple options remaining, you may use the matrix A3 to assist you in ranking the remaining options and making your final decision. You may want to weight the different aspects of the software differently, depending on the context you are operating in. After choosing your preferred software, continue to Annex B, table B1 to make your hardware choice.		

A3 Software decision matrix

Software	Price by # of surveys*	Ease of coding**	Ease of export into database**	Support by provider**	Compatibility with existing capacities***	Potential for development****	Total
Weight							
Option 1							
Option 2							
Option 3							
Option 4							
Option 5							

*There are different funding models (and freeware) for software. To avoid confusion, it is best to estimate the total price for using the software and divide by number of observations (if not a single project, count by number of observations in a common project)

** Considering these three factors can help to estimate the amount of staff time it will take to learn and use the software

*** If your organization already has significant skills and investment in certain programs used for coding or analysis, it is worth estimating how the MDC software fits with those

**** Although your organization's current project may not need some of the extra functions of a given piece of software, such functionality may be useful in future projects. It is good to have software that allows you to grow.

CONTENTS & FOREWORD

SECTION I

SECTION 2

SECTION 3

CONCLUSION

ANNEXES

RESOURCES

Annex B: Key Features of MDC Devices

This Annex expands on some of the main features of the mobile devices and helps you to choose the right type of hardware. Before choosing your hardware, it is generally advisable to make a choice of software that meets the needs of the project.

To make your choice of hardware, list the different options that you are considering, and follow the table below to ensure that the device you will select has the requisite features and functionality.

B1 Choosing the right hardware

#	Project needs and context	Y/N	Implications for software choice
1.	(skip if you are assessing your current hardware) Is the device compatible with the software you have chosen?		If yes, proceed. If no, choose a different device.
2.	Does the survey involve displaying graphics or video?		If yes, limit devices to those that have the capability and sufficiently large screen to display graphics and video. If your context requires doing this outdoors, give preferences to devices with screens that remain clear in sunlight. If no, proceed to the next question.
3.	Does the survey require enumerators to take photos?		If yes, limit devices to those that have built-in cameras. The quality of the camera depends on the level of analysis you require from the pictures, but note that higher quality pictures also take more space. If you are operating in areas with low-quality Internet, having large pictures may make uploading from the field extremely difficult. If no, proceed to the next question.
4.	Does the survey seek to record GPS data?		If yes, limit devices to those that have GPS capability. Note that some devices take location coordinates by triangulating from telecommunications towers (requires SIM) or through AGPS, both of these forms have limitations that could limit your ability to acquire coordinates. In contexts where towers are not present, make sure the phone can connect to actual satellite positioning using GPS (different from AGPS). Alternatively, buy a separate GPS locator and input coordinates as a separate variable. If no, proceed to the next question.
5.	Does data need to be uploaded directly from the field?		If yes, limit devices to such that can connect to either a cellular network or WiFi (depending on availability and reliability in your context). If no, proceed to the next question.
6.	Does the survey include a large amounts of filter logic, including section skips or filtering through multiple variables?		If yes, give preference to devices that have a decent amount of processing power, as low-speed processors may cause increased loading times during the survey (however, most new smartphones have more than enough processing power to meet these needs). If no, proceed to the next question.

CONTENTS & FOREWORD

SECTION 1

SECTION 2

SECTION 3

CONCLUSION

ANNEXES

RESOURCES

7.	Do the devices need to be operated in areas with a low degree of services and infrastructure?		If yes, give strong preference to devices that have high durability and that are easy to repair (low level of integrated parts, high availability of replacements). Accessories (including chargers) may be difficult to find in some areas so devices with high levels of compatibility with common accessories (micro-USB chargers, for instance) is preferable. If no, proceed to the next question.
8.	Will the devices be operated in areas with high levels of insecurity and crime?		If yes, give strong preference to cheaper devices, as expensive equipment may put your enumerators at risk of being robbed. If no, proceed to the next question.
9.	Proceed by listing the remaining options that fulfill your criteria and making a choice at which device best suits your need. You may use the matrix B2 to assist you in ranking the remaining options and making your final decision.		

B2 Hardware decision matrix

Hardware	Price of device	Functionality*	Durability	Battery life	Compatibility with common plugs/devices**	Usefulness for other purposes***	Total
Weight							
Option 1							
Option 2							
Option 3							
Option 4							
Option 5							

* Although the current project may not need certain functions, future projects may require some further functionality (GPS, multimedia, etc.)

** It is not always possible to find manufacturer-specific accessories (including chargers) in the field. The more compatible the device, the better

*** Your hardware may only be used for data collection a few times a month. When possible, procure devices that can be used for other purposes as well

Annex C: Data Quality Control and Processing

Switching to mobile data collection can provide your organization with multiple tools for improving the quality of your data. These broadly fit into 3 categories: 1. Filter logic and verification; 2. metadata; and 3. additional features.

Firstly, as with any data collection exercise, front-end quality control mechanisms are important in ensuing the quality and reliability of your data. When utilizing MDC, well-designed filter logic (that ensures enumerators ask the right questions), answer range restriction, and response verification (by checking data consistency) all can be used to ensure you receive the best data possible.

Filter logic (also see Section 1 on data quality) has been shown to significantly reduce the likelihood of enumerators either asking unnecessary questions or failing to ask necessary ones. The more complex skips your survey has, the more you will potentially benefit from MDC and the associated reduction of data miss entry. However, if you do apply complex skips, make sure you take into consideration the implications listed in Annex A table A2 point 6 in choosing the software.

For certain types of questions, you may also want to consider limiting and also verifying response options. Some data, particular open numeric entry (such as for age, phone numbers, etc.) is subject to a risk of mistyping. Common ways to minimize these problems are limiting range (for example, age 15-99), provide categories for verification in the next question (such as age groups), or double entry (subject to software capacity, you may code the second question to only accept an answer that matches the first, which is particularly good for long non-ordinal sequences such as phone numbers).

Secondly, in addition to enumerator entered data, you can set the MDC device to auto-collect some key variables, referred to as metadata. These include timestamps such as the length of and interview (and/or any parts of the interview), and start and end times, as well as interviewer identification.

Timestamps are extremely useful in overseeing the activities of your enumerators. As the number of interviews increases, it will be easy to begin tracking average interview times and identify suspiciously activity quickly. The software can also verify the times of day the interviews are made, which allows for further investigation into possible systematic biases, ie. who is responding to your survey. Even the mention of these capabilities during enumerator training can serve as an effective deterrent against enumerator misconduct.

Generally your enumerators will log in to the software with a unique login, which is then recorded as metadata in the dataset. Although you may want to also have enumerators entering a code in the survey, this metadata is often more reliable and helps interrogate the data by enumerator with more ease by allowing direct comparison between data entered by the enumerator and that collected by the device.

Finally, you may want to use some of the additional features that MDC enables, such as GPS or picture verification for quality control purposes. This data can show you the movement patterns of your enumerators. Similarly, taking pictures of locations where interviews are conducted can be used to verify that correct places are visited. Mentioning GPS or making pictures a requirement for enumerators is also an effective deterrent against cheating in location choice and household selection patterns. However,

CONTENTS & FOREWORD

SECTION 1

SECTION 2

SECTION 3

CONCLUSION

ANNEXES

RESOURCES

it should be noted that GPS and picture-taking may cause risks for enumerators in some low security contexts, and shouldn't be used unless necessary.

In summary, significant improvements with data quality from the data collection phase can be reached by utilizing MDC. Coding in filter logic, applying simple ranges or verifications, and keeping an eye on metadata, are relatively easy and low cost to implement, provided that capacity for coding exist. More explicit methods such as GPS or pictures are also enabled, though using them should follow an assessment on whether they are needed and what implications they may have for the enumerator.

Annex D: Sample Specifications Sheet for Mobile Data Collection

The goal of the survey is to track public opinion across different gender, age, location, and socio-economic strata. The firm will have the familiarity, experience, and capability to conduct the survey using mobile data collection as per the below specifications.

Experience:

- The firm will posses demonstrated experience using mobile data collection in a similar context with associated client references.

Software:

- The firm will have access to a survey software license.

- The firm will have the proven ability to script a survey with complex filter logic in the host language(s).

Hardware:

- The firm will have access to a minimum of 30 mobile devices that can be used for data collection

Training and Fieldwork:

- The firm will have the proven ability to recruit, train, and monitor a team of local enumerators using mobile data collection.

- The firm can demonstrate a system of best practices based on experience operating with mobile data collection technology in a similar context.

Quality Control:

- The firm will have a robust set of fieldwork protocols to ensure data quality, as well as real-time backend data check and verification capabilities.

Remote Monitoring:

- The firm's chosen survey software/hardware will have the ability to record GPS coordinates during data collection and the agency will have a system for remote verification of field data.

Data Management:

- The firm will have the ability to upload, store, and clean all collected field data wirelessly, without the need for data entry.

CONTENTS & FOREWORD

SECTION I

SECTION 2

SECTION 3

CONCLUSION

ANNEXES

RESOURCES

Annex E: Key Questions to Ask Specialized Consultants or Firms

If you have decided to make use of an external consultant or firm to assist in the MDC component of your project, it is important to establish first and foremost exactly what service you want them to deliver. You may be interested in anything from training and support for your staff to fully conducting a research project using MDC. Think through what type of data you will require and how you hope to use it so that you can clearly define your need in your call for applications or upon approaching a consultant/firm directly (example in Annex D). In this Annex, we provide a list of some of the main services you may contract (of which you may want any or all) and some of the key issues to consider when making the hiring decision. Note that different firms use different pricing models, charging per project, per month, etc. Therefore, when considering multiple bids, convert the quotes if necessary so you are comparing like services.

Advisory and support

If you are looking to conduct MDC projects internally in the future, you may want to hire an external consultant or firm to assist you with support in building your organization's capacity while avoiding common pitfalls when choosing the software and hardware, training staff, and/or offering a troubleshooting service during implementation.

Besides the general questions on the firm's or individual's experience in offering similar services in comparable contexts, it is important that you make sure that the platform they are familiar with meets the needs of your project. You may want to make the choice of software within your organization, or this may be a service provided by the consultant:

➔ *If you are choosing the software internally,* refer to Annex A's table A2 for guidance on the key features the software should have depending on your data needs. In this case, it is relatively easy to assess consultant/firm competency by the number of times they have provided similar services in comparable contexts. However, if your data collection includes any advanced tools from the software, ensure that the consultant/firm has experience utilizing them in the field;

➔ *If assisting you with the software choice is among the services you want from the consultant/firm,* you should first map your data needs using table A2 in Annex A to see if you have any specific needs from the software. These should be included in your scope of work or mentioned as you contact a potential service provider. Once a suggestion has been made, you may use the "implications for software choice" column to probe the software and assess its suitability for your project. Be wary that some software may amount to repackaged freeware, so take care in assessing the suggestion;

➔ *For providing troubleshooting services during the project,* extensive expertise in managing fieldwork projects using the chosen software and hardware in similar conditions is of key importance. As a further consideration, however, make sure that you negotiate a flexible contract that allows for support at most times of the day. If the service is provided from a distance, make sure to consider the effects of the time difference on availability. If you need instant data access, ensure they are also available during evenings, when your enumerators are likely to be uploading the data.

Planning and preparation for fieldwork

There are several services that can be procured for MDC prior to the fieldwork. They are mainly related to designing the survey and transferring that survey onto the data collection devices.

Although your organization likely has a good deal of experience in creating surveys, certain aspects of MDC affect some types of questions:

- For example, open text entry eliminates part of the data entry benefits from MDC, as the data will need to be recoded for quantitative analysis. These types of questions should be eliminated whenever possible for standard response options;

- Open numeric entries (commonly age, income, phone number, etc.) are often subject to mistyping, and require quality control mechanisms to be built into the survey for such questions (see Annex C for more);

- Finally, large matrices and tables do not necessarily transfer well to MDC devices, particularly phones. Breaking them up into the constituent questions for the survey is preferable (note that this does not change the number of questions, simply how they are laid out);

- ➔ *If you are contracting for survey design services,* enquire about the consultant/firm's knowledge of these common issues that require consideration in transferring surveys to MDC, and what solutions their software/hardware offer.

After a survey has been appropriately designed, it needs to be **coded into the software** before data collection can begin. Multiple problems can arise during this stage if the appropriate care is not given to the process. For example, mistakes in skip logic may mean that key questions are missed or wrong follow-up questions are asked from individuals. Badly coded questionnaires can also be unstable in some software.

➔ *If you are hiring for a survey that is extremely complex,* requiring advanced tools and highly complicated skip logic, probe the consultant/firm's previous experience in coding surveys with similar requirements. It would also be advisable to enquire what quality control process they have in place for checking the coded survey's functionality.

Fieldwork

MDC projects are by and large similar to implement in the field as normal surveys, and knowledge of local context, logistics, connections with trained enumerators, in addition to applying MDC in similar contexts, are prime concerns for hiring a consultant/firm to handle fieldwork portions. However, issues can arise over provision of hardware, access to the Internet and to electricity, depending on the context.

➔ *If you require continuous access to data,* rather than a dataset at the end of the fieldwork, ask applicants how they will upload data from the field. Also enquire on how the software they are using uploads from poor Internet connections.

➔ *If you do not possess the hardware that can be used for data collection,* ensure that the consultant/firm has suitable devices.

Back-end

This final section refers to those services that relate to what happens to data as it is uploaded. It generally takes some level of data cleaning to prepare a dataset for analysis. For some more continuous data collection projects, you may also want to have access to some real-time analytics and graphics as the data is coming in. You may also wish to have the ability to have the incoming data merge with other datasets.

→ *If you are seeking to outsource data cleaning,* ask what processes they have for detecting abnormalities in the data. Note that data becomes noisier if the survey is badly designed or coded, so consult the above section if the same actor is providing both services.

→ *If you require real-time access to data and/or visualizations,* inquire whether the software and server facilitate such services.

→ *If you need to merge the dataset with other datasets,* ensure that the data produced by the software is SSS compliant.

Summary

As with any research, there are many distinct services in conducting MDC-utilizing projects that you may want to seek external expertise in. Some services can be delivered from a distance where others require personal presence. Mapping your organization's needs in relation to the data, as well as existing and desired capacity will greatly clarify what solution is best for you. Following the questions in this Annex and those in Annexes A and B will help you map out the needs and capacities that are most relevant in successfully utilizing the MDC toolkit, and enable your organization to reap the maximum benefit and value from the technology.

CONTENTS & FOREWORD

SECTION 1

SECTION 2

SECTION 3

CONCLUSION

ANNEXES

RESOURCES

CONTENTS & FOREWORD

SECTION 1

SECTION 2

SECTION 3

CONCLUSION

ANNEXES

RESOURCES

RESOURCES

NOMAD Online Selection Assistant

http://humanitarian-nomad.org/online-selection-tool/

The NOMAD Online Selection Assistant uses a short needs assessment survey to help users narrow down potential mobile data collection platforms to only the options that meet their program specifications.

Kopernik Impact Tracker Technology Catalogue

http://impacttrackertech.kopernik.info/

Profiling dozens of digital data collection apps, geospatial mapping tools, SMS communications platforms, and remote sensors, this catalogue gives readers the information they need to choose the best fit for their unique project needs.

Ways to Practice Responsible Development Data

https://responsibledata.io/ways-to-practise-responsible-development-data/

This e-book explores the meaning of responsible data in the context of international development. It is intended to support thoughtful and responsible thinking as the development community grapples with relatively new social and ethical challenges stemming from data use.

Using Mobile Data for Development

https://docs.gatesfoundation.org/Documents/Using%20Mobile%20Data%20for%20Development.pdf

This paper aims to increase awareness amongst researchers and development leaders of the power of mobile data for development goals, to illustrate in more detail how data is gathered, what type of data exists in mobile networks, and to explore ways to overcome barriers to its use.

Integrating Mobiles into Development Projects

http://www.usaid.gov/sites/default/files/documents/1861/M4DHandbook_August_2014.pdf

This handbook provides a broad overview of mobiles for development, or M4D. The first section, Understand, provides an in-depth guide to everything you might need to know about mobiles. The second section, Implement, gives practical tips and information about integrating mobiles into a project in practice.