MAKING THE MOST OF FORM DESIGN IN REMOTELY MONITORED PROJECTS - *A DATA QUALITY ORIENTED APPROACH*

The unexpected arrival of the Covid-19 virus at the start of this year has forced many sectors to adapt their activities and to plan around restrictions which strictly limit travel and social interaction. **When data collection processes are remotely managed, there is often a greater reliance on technology to replace some of the quality checks which could otherwise be done in person**. The risks to the quality of the data are therefore greater but well-designed forms which make use of the functionality available through ODK-based tools such as Kobo, ONA and SurveyCTO among others can provide reliable and accurate data and allow remotely monitored projects to proceed while minimising as much as possible the risks associated with travelling between different locations.

Building on CartONG guidance documentation on checklists for mobile data collection and designing forms with data quality in mind, **this document provides details of the main approaches which can be taken to ensure that data collected in a remote context is as accurate and reliable as possible**.

METADATA

Metadata is **data which is automatically collected by the mobile device** rather than being recorded manually by the enumerator. It includes information about who collected the data, at what point in time and in what location. This limited information can give a very strong indicator of the quality of the rest of the data collected and, as it is recorded automatically, it is not prone to human error.

Start & End times

Why use it?

Start and end times can be used to **calculate the duration of the survey** (by subtracting the start time from the end time). For many form designs, each record should be expected to take an approximately similar length of time and large variations in duration can be worth checking.

What to look for?

- Where **surveys take unrealistically long times** it may be worth following up with the enumerator to find out why. Long time periods of many hours can imply that data may have been modified after the initial data collection (in some cases this may be part of the process) or simply that the enumerator does not know how to finish the survey in the proper way.
- Of greater concern may be **survey durations which are unrealistically short**. This can imply that the questions are actually not being asked to the intended person or that the data is being recorded on paper and then entered into the mobile device at another time.
- When checking for short times it's important to **exclude survey entries for interviewees who did not consent to taking part** as these are not relevant for this purpose.
- It should always be considered that **the initial rounds of a survey will take longer than later rounds** (as the enumerators are still becoming familiar with the specific process). For surveys where many questions are not asked to all interviewees, there may be more variation.

How to use it?

In ODK based tools such as Kobo, **start and end times can be automatically recorded**. The duration can be calculated either in the form itself or in excel after the data is exported. To know how to include metadata in an ODK form please see this example.

	A	B	C	D
1	**type**	**name**	**label**	**calculation**
2	start	start		
3	end	end		
4	today	today		
5				
6	calculate	survey_duration	Duration of the Survey	${end} - ${start}

Example of an ODK form with a calculation included for the duration of the survey

GPS

Why use it?

GPS coordinates can be used to **ensure that data was really collected in the expected location and that geographical sampling rules were properly observed**.

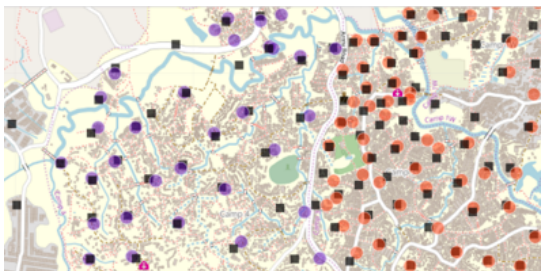
What to look for?

- Check that all of the data collection took place in the **expected area and locations**.
- Check if **sampling methodologies have been followed**. If there is a clustered sampling method being used then distinct groups of similar sizes would be expected to appear in the data. A systematic sampling approach should create evenly spaced individual locations and a random sampling method would generate a more unpredictable pattern. Unexpected coordinates could mean that sampling is not being respected or that GPS points are not accurate. It is usually worth checking if there is a valid reason for the different locations.
- If some **enumerators are failing to record GPS coordinates** at a much higher frequency than others, then it may be worth checking the reason why.

How to use it?

In some locations, normally rural areas or inside buildings, or if there are technical problems, it may not be possible to gather GPS coordinates, therefore the best design practice is to **make the GPS question non-mandatory on all forms**. Before collecting GPS points, **consent must be given** and recorded, in addition to any consent already given for the overall data collection, as there may be specific objections. This example demonstrates how it can be included in an ODK form.

For practical tips and considerations for collecting GPS points, there is guidance material available here. In regions where there is instability and the risk that data could be intercepted or used in a way that puts the interviewee at increased risk, data including GPS coordinates may need to use additional security measures such as encryption in order to safeguard the data and be able to provide accurate reassurance to the interviewee on the safety of the process. Alternatively the GPS accuracy can be reduced to avoid the identification of individual households (see link for further details).



In this image, you can see the planned survey locations (in black) and the actual survey locations (in other colours, according to the team doing the survey). On the left hand side of the map, you can see the locations which were missed by the purple team and in the centre of the map you can see that some locations were hit twice by the red team.

Enumerator name

Why use it?

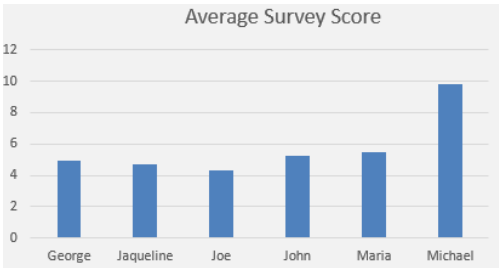
Enumerator names are a useful way of being able to disaggregate the data.

What to look for?

- **Unusual data patterns** linked to just a small number of enumerators who may need some additional training or who may be facing issues when collection data.
- **If incorrect data is discovered** after data collection it may be possible to **identify a small number of enumerators responsible and exclude their data** from some analysis and maintain the reliability of the rest of the data set. For multi day surveys this is something which could be checked after the first day or two with follow-up training where relevant.

How to use it?

While not strictly "metadata" as this is just a single choice question (see ODK select_one example) the **importance of using the correct enumerator name** (or number) and whether or not they are allowed to use more than one phone or tablet during the period of data collection would need to be explained to the enumerators. In cases where a full list of enumerators is not known in advance, it may be necessary to use enumerator numbers instead and then distribute these numbers to staff as they join the team.



© CartONG

In this example, one enumerator has gathered data with much higher average scores than the others. These cases should be investigated to see if they are due to logical reasons or if the survey is being implemented in an inconsistent way.

Device ID

Why use it?

Device ID can be used as a final safety net when disaggregating data.

What to look for?

- If there is a suspicion that **data is being falsely entered** under the identity of another enumerator, this can sometimes be spotted by looking for inconsistencies in the device ID.

	A	B	C	D
1	type	name	label	hint
2	deviceid	deviceid		
3				

© CartONG

- Where phones or tablets distributed for data collection purposes are at risk of being sold or used for other purposes, the device IDs will give an **indication if enumerators have replaced them with personal devices** for the purposes of data collection. This is particularly important if the original devices had additional security measures and makes it difficult to ensure that all devices can be cleaned of sensitive data once the survey is complete.

How to use it?

See the **ODK form screenshot above** *(What to look for)* as well as this example.

Audit

Why use it?

The "audit" functionality in ODK forms can be used to get an idea of how the user interacts with the form. **Metadata is recorded based on individual questions** rather than just once at the start or the end of the form. For SurveyCTO users it should be noted that additional audit functionality is possible as well as the standard ODK audit functionality. These features allow for the programming of checks to indicate subsets of the data with unlikely patterns in the responses or the setting of minimum times for asking complicated questions which should not be rushed, as well as other features such as audio checks.

What to look for?

- It can be useful for identifying **questions which may be causing confusion** as they will have longer than expected duration times. It can also be used to check that consent questions are being asked to the beneficiaries in a proper way. If the consent text is not read out then this question will have a shorter than expected time.
- Forms where **some but not all sections were answered in the expected location**
- Questions where the **response has been modified after the initial answer** (see line 9 in the example audit data export in the image below).

	A	B	C	D	E	F	G	H	I
1	event	node	start	end	latitude	longitude	accuracy	old-value	new-value
2	form start		1550615022663						
3	location tracking enabled		1550615022671						
4	question	/data/name	1550615022682	1550615097082	37.423	-122.084	14.087		John
5	location permissions granted		1550615068610						
6	location providers enabled		1550615068665						
7	location tracking disabled		1550615095914		37.423	-122.084	14.087		
8	question	/data/age	1550615097082	1550615097655	37.423	-122.084	14.087		20
9	question	/data/name	1550615097656	1550615102351	37.423	-122.084	14.087	John	John Smith
10	location tracking enabled		1550615099271		37.423	-122.084	14.087		
11	question	/data/age	1550615102351	1550615107630	37.423	-122.084	14.087		
12	end screen		1550615107631	1550615109199	37.423	-122.084	14.087		
13	form save		1550615109199		37.423	-122.084	14.087		
14	form exit		1550615109199		37.423	-122.084	14.087		
15	form finalize		1550615109199		37.423	-122.084	14.087		

How to use it?

For details of **how to include the audit function in an ODK form and how to access the audit data afterwards please see this example.** Please note that this feature is currently only available via the mobile app and does not yet work via Enketo webforms. For ethical reasons all enumerators must be informed beforehand that the form they are using includes audit functionality.

MEDIA

Multimedia are a way of capturing field reality for the enumerator under a different angle and to give other insights on the situation. Keep in mind that additional consent is required if people appear in an image, video or audio capture. In most cases a **strong system of managing media files is needed** in order to be able to inform beneficiaries if multimedia concerning them are used beyond the initial objective or to allow for the deletion of this data should it be requested at any point. **Refer to the guidelines your organisation** should have on such a topic to know more.

Images

Why use it?

Collecting images and then reviewing a sample of the images can be a very **effective way of verifying the data collected** and identifying areas where data may not be entered accurately.

What to look for?

- This can be particularly effective in WASH or shelter activities where **photographic evidence** of the quality of new construction work can be verified.
- Where enumerators are **unsure how to categorise** a type of latrine or other structure, photos can be used for **later verification of the data collected**.
- Photos can be used to **verify the use of proper equipment**, for example for the purposes of infection prevention.

© CartONG

In this example, a photograph or video can be used to verify if the recorded information matches the technical specifications of the installed system.
Photos may be checked for a subset of the data in order to verify the overall data quality of data collection or where necessary they may be checked for each data entry.

How to use it?

When images are collected via ODK based tools, each image is **connected to the data collected which is relevant to it**. An example of a form collecting images can be found here.

Videos

Why use it?

Videos can be used to **record evidence of activities and levels of participation**. Like pictures they are occasionally used for **communications purposes or shared with donors for accountability purposes**.

What to look for?

- Videos can be a useful method of **collecting information that would be difficult to demonstrate with individual images**.
- Videos of distributions and other activities can also be used to **verify if procedures are being followed** as expected.

How to use it?

It should be noted that collecting media and especially videos can only be used where there is **sufficient bandwidth for syncing data** and only if the mobile devices have sufficient **storage capacity** for the expected number of video files. An example of how to include video functionality in an ODK form can be found here.

Audio

Why use it?

For **qualitative questions** or **situations where it is not appropriate to enter data into a phone or tablet** in the traditional way, audio recordings can be used.

What to look for?

- **Qualitative data or information which is not in response to a form question** but which should be treated as a priority.
- Audio recordings, where permitted, can also allow for **later verification of the data** entered into the mobile device.

How to use it?

An example of how to include this feature in an ODK can be found here.

The icons used in this document were created by Creative Stall, darkassbola, Heztasia, Kelsey Armstrong, Mello, phi, Prettycons, Sergey Novosyolov, shwepes, Stefano Vetere, Trendy and Wahyu Unggul Sejati of the Noun Project.

AFD Agence Française de Développement
H2H Network Humanitarian Action Support
UKaid from the British people

This project was co-funded by the French Development Agency (AFD) and the H2H Network's H2H Fund, the latter supported by UK aid from the UK government. Nevertheless, the ideas and opinions presented in this document do not necessarily represent those of the H2H Network, UK aid or AFD.