# Getting Started with ODK

Max Planck Institute for Evolutionary Anthropology
Department of Human Behavior, Ecology, and Culture

Version 0.1
Managed by Anne Pisor

January 11, 2017

## 1 Introduction

This document was designed to help you get oriented with the Open Data Kit (ODK). ODK is a set of open-source software for the design and implementation of simple surveys, including the upload of collected data to the cloud or a server. These surveys can be created on any Windows, Mac, or Linux machine, then implemented with Android tablets[1]or smartphones. While the surveys ODK creates are not particularly fancy, it does allow implementation of things often crucial to field anthropologists, including the implementation of repeated questions (particularly useful for focal follows, or when completing household demographic surveys), question counterbalancing and randomization, and the capture of photos, GPS, voice, and even signatures (can be useful for affirming informed consent in literate populations). ODK enables surveys to be edited and uploaded from computer to Android device without an internet or data connection, handy when researchers are in the field. Further, ODK benefits from a community of users active in forums.

### 1.1 How does ODK work?

ODK Collect, one of the tools of the larger system, reads a survey coded in Java in an XML file and presents the user with a graphical representation of that survey. However, you don't need to know how to code in Java to use ODK: in the course of this introduction, you will learn how to create an XML file from an Excel sheet.

When a researcher collects data from a participant, ODK Collect creates a new XML file based on the survey, with fields populated with the participant's responses. This file can be uploaded directly from the ODK Collect via a wifi or data connection; alternatively, the XML file can be transferred from your Android device to an external hard drive or a computer via a USB connection. When you are ready to analyze these data, they can then be converted to CSV and downloaded using ODK Aggregate.

[1]An Android emulator can also theoretically run ODK, but both the ODK site and forums tell us that the emulator is slow and buggy.

# 2 Downloads

To begin the process of familiarizing yourself with ODK, we suggest you begin by downloading the following software:

## 2.1 Download to your computer

- **Microsoft Excel or Apache OpenOffice Calc.** Both Excel (included on some devices or available via subscription) and Calc (free for download) can create .xls files. The easiest way to create a survey for use in ODK is via an .xls (or .xlsx) file, which can then be converted to XML using the below software. Read on for how this works.
- **XLSForm Offline.** Nafundi (founded by one of the co-creators of ODK) has created a free software for validating the .xls or .xlsx files in which you create your survey (i.e., it will tell you what is wrong, and where, if you have an error) and convert them to XML files. A donation to Nafundi for downloading the product is optional.
- **LINKS Codebook.** The LINKS Codebook is a program that will accept an XML file and create a codebook from it. In other words, your survey, along with all its questions and attendant multiple choice responses, follow-up questions, etc., will be converted into a printable format. This is the only document in the ODK pipeline that is remotely printer friendly, even if the codebook itself is a little long and unwieldy.
- **Optional: ODK Briefcase.** ODK Briefcase has some nice features. Most importantly, itallows you to export repeat questions as CSV files, which is not possible when using ODK Aggregate as hosted on a server. See more in a later section. It also *used* to allow data to be pulled directly from ODK Collect on your Android device to your computer for processing, so you wouldn't even *need* a server to host ODK Aggregate (more details below). However, those capabilities only exist on Android 2.x or earlier – that is, OSes earlier than Honeycomb, released in 2011. Ah well.

## 2.2 Download to your Android tablet or smartphone

- **ODK Collect.** ODK Collect is the app that reads your XML survey and displays it graphically on your Android device. It is the app you will use to collect your data. We will introduce it further below.
- **ES File Explorer.** ES File Explorer allows you to easily view the folders on your Android device, including the ODK folders where your completed surveys will be stored.

# 3 Preparing your survey

Next, we recommend you take a peek at the ODK welcome page. It will give you a sense of the flow of information, from survey creation to data output.

## 3.1 Which XML editor to use?

You'll notice that the first step in the process is to build your survey! (We have some recommendations for how to design these things: see our website.) There are three options for this on ODK: first, if you know how to code in Java (or want to learn), there are details here about how to edit the XML directly using JavaRosa; second, if you like drag-and-drop graphic interfaces and will have a reliable internet connection throughout form design and editing, Build is an option; third, if you are not familiar with Java-based code, will not have a reliable internet connection, but would like to leave a trail of breadcrumbs as you design your survey, consider XLSForm. As we expect XLSForm will make the most sense for the most people, we will discuss XLSForm here.

## 3.2 Familiarizing yourself with XLSForm

XLSForm has a nice introduction to their format. We recommend you read through the introduction with a spreadsheet open, whether Excel (.xls or .xlsx) or Calc (.xls), and try implementing the different options as they are introduced. ...the XLSForm introduction isn't *perfect* however: see our GitHub repository for an annotated version of this introduction, which you may wish to keep open for your reference.

Our Git repo also includes examples of XLSForms we've created for our own surveys. These may serve as templates for your surveys, and may be uesful examples of anthropology survey creation in practice.

### 3.2.1 Naming your survey

There are some restrictions as to what your survey can be named: ODK Aggregate, which we will discuss later, cannot manage files with names that (1) begin with a number, (2) are longer than 10 characters, or (3) have spaces in the name. See here, Form ID Guidelines section, for a full list of constraints and name your file accordingly. We suggest something that will be easy for you to keep track of, such as an abbreviation of your survey name and the date you edited the form (e.g., for a focal follow survey, FF-JAN0716).

### 3.2.2 Validate your XLSForm

When you have a form ready, whether a play example or the real thing, pass your .xls or .xlsx document through an XLSForm validating program. Validating programs will scan the document for errors, then flag the location and nature of the error for you to fix. In general, this is best to do when you have an internet conncetion, as you can Google some of the more vague errors. A web-based one is available here, though it is imperfect: it does not always report the nature or location of the error. Instead, XLSForm offline (available for Windows and Mac; also designed by Nafundi) is much better: it is more accurate with the location and nature of your errors, and, as the name says, it works offline!

Both the online and offline validation software also do something convenient: they convert your .xls or .xlsx form to XML. Bingo! To look at the XML file generated by the validations software, use a text editor with XML language support (like Notedpad ++). This will later come in handy for peeking at your filled-out surveys when you have no internet connection.

### 3.3 You have an XML version of your survey. Now what?

If you are without an internet connection, or want to avoid using ODK Aggregate, read the first section below. If you plan to use ODK Aggregate to deploy your survey, read the second section.

#### 3.3.1 Drag and drop

Mount your Android device on your computer using the appropriate USB cord. Then navigate to the notification bar on your Android (top of the screen): a notification there will say something about your connected USB. Click on the notification and select whichever item talks about "file transfers" (the exact wording differs by device). A folder displaying the contents of your Android should appear automatically on your computer. If it does not, select "This PC" on File Explorer (Windows). For Macs, things can be more complicated: see this page for some help. Browse for the file titled "odk" on your Android. (If you can't find it, open ES File Explorer on your Android device and search for "odk" to see where it is stored on your device, then use that information to locate it via your computer.[2]) Inside the odk folder you'll see a folder titled "forms." If you drag and drop your XML file into that folder, it will be available via ODK Collect on your Android device (see details below)[3].

#### 3.3.2 Uploading your survey to ODK Aggregate

To create an ODK Aggregate primary account and activate other users who will have access to your survey, see our section on ODK Aggregate. Once you have set up ODK Aggregate, log in to your Appspot and click on the Form Management tab. At the top left, you will see a button called Add New Form. Click it. For Form Definition, select Choose File and navigate to your the XML version of your survey[4]. Once you click Upload Form, the form should display on the Form Management tab, with the Downloadable and Accept Submissions buttons selected. These will enable you to do (1) and (2), below.

## 4 Setting up ODK Aggregate or your own server

You can use ODK Aggregate to (1) deploy your survey, (2) accept incoming surveys via 3G/LTE or wifi, and (3) churn out CSV files from filled-out XML surveys.

There are a couple reasons to bypass ODK Aggregate entirely. First, you can avoid paying for a server. Second, you can store sensitive information only on your own devices (of couse, we

[3]We have experienced an additional hiccup at this juncture: while the majority of the files reflected in ES Explorer were displayed on the computer, the odk folder was not. This happened because we had not yet launched the ODK Collect app *ever* on our Android device after downloading. If you experience this problem, disconnect your Android device, launch the ODK Collect app, then try mounting your Android device on your computer again.

[3]Note that you can use the same method to get the survey onto multiple devices if, for example, you are without internet but there will be several researchers collecting data for the survey. Just mount each device to your computer and drag and drop the XML survey file per the instructions above.

[4]Want to include pictures, audio, or even a census from which you can select participants? All these data can be included as media. See the Adding Media to a XLS Form, Where to Put the Media Files, and Additional Notes sections here. See our census section for more details.

recommend you back it up *somewhere* besides your computer every single day, like on flash and external hard drives). To do everything from here on out without ODK Aggregate, be sure to read the Drag and Drop and Virtual Machine sections.

If you want to proceed with ODK Aggregate, you have to decide where you want to host it. This is an important decision, as whichever service hosts ODK Aggregate will also be hosting all your data – in the social sciences, this information is often sensitive and should be well-protected – and will have a price tag, regardless of the option you choose.

The first option is the default option described on the ODK site, the Google Cloud Platform. (You may recall that Google and the University of Washington paired up to create ODK. That's why GCP is the default.) Note that the GCP provides free usage, but only below a limited amount of daily usage, and requires fees above that level. A second option, which we use at MPI-EVA, is to host ODK Aggregate on a local server with Tomcat. (If you are at MPI-EVA and have decided to use our server, see this section for instructions.) The local server option definitely requires some IT support if you're not familiar with server management – and, of course, the server itself costs money, as does electricity and internet. We're happy to talk to you about how to make the local server option happen. Other alternatives, which we have not vetted, include hosting via Amazon Web Services EC2 and Ona.io. Yet *another* option is to not use a server at all, but keep everything locally on your computer by using a virtual machine to process your data, again provided by Nafundi. We will detail this option more below in a later section.

## 4.1 Google Cloud Platform

Setting up an account for using ODK Aggregate is relatively straightforward using the instructions provided on the ODK site. There are only minor pieces of information that are out of date or are insufficiently clear, which we have flagged on the annotated how-to on our Git Repo. We suggest at least having our annotated document open while you get Aggregate up and running.

## 4.2 Local server with Tomcat

Instructions on how to install Tomcat and get ODK Aggregate running on your server are provided here.

If you are part of MPI-EVA and have an MPI-owned computer (i.e., have access to the intranet), the first step is to install virtual machine software on your computer via the Intranet while on site at MPI-EVA. Instructions are in the IT section of the Intranet. Once this software is installed, when you are in the field and have a good data or wifi connection, you can log on a virtual machine at the Institute. Then, using the instructions detailed on this Google Earth tutorial, pull your completed surveys, one by one (unfortunately!), into ODK Aggregate. Remember that you need to transfer the files onto the virtual machine to do this: dragging and dropping the files onto the virtual desktop is a quick way to do this.

# 5 Using ODK Collect

Now everything is ready: time to get your survey on your Android device and start putting it to good use!

## 5.1 Activate your ODK account on your Android device

Open ODK Collect on your Android device. On the top of the screen, right-hand side, click the drop-down menu (three dots); select General Settings. Set the values on this menu to the following:

- **Platform.** The platform selected should be ODK Aggregate. If it isn't, click on Platform and choose ODK Aggregate.
- **Configure platform settings.** On this section, you'll need to enter the URL for your Appspot.
  - **Google Cloud Platform.** If you're using GCP, from our annotated ODK Aggregate how-to, recall that the URL for your project is https://...../appspot.com, where the empty spot is your project name. See the appropriate area of the Google Cloud Platform window, per the top of page 4 in the annotated Aggregate how-to.
  - **Tomcat.** The URL is the DNS name that you chose, or that was given to you (if at MPI-EVA), as consistent with the Tomcat install instructions.

  Next, enter the username you created for yourself. Remember, that's not your superuser user name – the superuser cannot collect data – but rather the *second* user name you created for yourself, complete with data collection permissions. Then enter the password you assigned that account.
- **Username.** Enter your user name here too...
- **Password.** ...and your password here too.

There are some other customizable items on this list which may be of interest. All are available under General Settings. Among these are the option to send finalized forms automatically, useful if you have a consistent data or wifi connection and are using a server (whether GCP, Tomcat, Ona.io, or Amazon) to host ODK Aggregate. Another is "default to finalized version," which affects whether or not a survey is automatically sent to the Edit Saved Form folder (default) or the Send Finalized Form folder (essentially waiving your ability to easily edit it).

## 5.2 Get and complete a form

Return to the main page for ODK Collect. Click Get Blank Form. The app will spend a few seconds connecting to the Appspot server, and will then display the survey you uploaded to your Appspot project[5]. Select the survey you wish to download to your phone and click Get Selected on the bottom right.

[5]Take a quick peek at the Google Cloud Platform tab, if you're using GCP. See that plot on your dashboard? See that little spike in activity? That was generated by you requesting to see the surveys hosted on Appspot. If you scroll down, you'll also see the number of hours of your daily allotment of instance hours you have used. Seem like you have burned through a surprisingly high amount already? Yup. More reason to decide whether

On the main menu, select Fill Blank Form. Select your survey from the menu. The first page that appears instructs you on how to swipe left and right to nagivate the survey. Proceed through the survey: see if all your questions look as you expect. If there are any surprises, head back to your .xls sheet and repeat the steps above.

To navigate to a specific section of the survey, click the arrow and dot on the menu bar, at the top. Different sections of your survey will be displayed; navigate to a given section by clicking. To complete the survey – for example, if someone wishes to stop participating, or if a focal follow is complete – click the back button and select Save Changes. You may also save changes as you complete the survey – always a good idea – by clicking the disk icon on the menu bar as you proceed.

When you complete the survey, you will reach a page saying "You are at the end of .....," where the blank space is your survey name. ODK Collect will give you the option to mark your form as finalized. To decide whether or not to mark the form as finalized when you save and exit, read on.

### 5.3 Submitting completed forms, or saving them for later

By default, ODK Collect saves completed forms on your phone, rather than attempting to send them to your designated server (as specified here). When you complete a survey, it will appear as a saved form on the main menu. Click Edit Saved Form to finalize it, or, if new data became available, edit the survey further. (This is the perfect option if a participant needs to take a break, or if you realized you missed a single question and need to ask again.) Notice that the survey is named using the current date, and indicates the last time it was edited[6]. Convenient. When you are done editing, you can click the "Nagivate to end" button at the bottom right, which will take you to the end of your survey and allow you to save and exit.

When a form is finalized, either because you marked it as finalized when you exited the survey, or because you completed the editing process and marked it as finalized, it will appear in the Send Finalized Form section of the ODK Collect home screen. You have two options here. You can either download your data directly to your device or send the survey to your server over a wifi or data connection.

## 6 Downloading your data

Once surveys are in the Send Finalized Form folder, you may either (1) enter the folder and send the surveys to the server hosting ODK Aggregate, or (2) download the completed surveys to your computer. Or both![7]

you'll use a local server, the Google Cloud Platform, or a third option sooner than later.

[6] This is how the file itself is stored on your phone, as you'll see if you open ES File Explorer: navigate to the ODK folder, open instances, and peek inside. You'll see an XML file with the name of your survey, plus the appended date and time.

[7] Just make sure that you haven't checked Delete After Send (on the General Settings menu) if you plan to transfer the XML survey files to your computer via USB after uploading to ODK Aggregate!

### 6.1 Downloading your data from ODK Aggregate as housed on GCP, Tomcat, Amazon, or Ona.io

Wherever your ODK Aggregate project is hosted, navigate there using a computer and log in with your username and password. On the Submissions tab, you'll see your completed survey, including the time stamp of completion. If you click Export at the top right, ODK Aggregate will prompt you to indicate the file type and filter (i.e., participants to exclude) for export. For most statistics programs, CSV is your best bet[8]. Even for programs like SPSS, you can open the CSV file in Excel or Calc, save it as XLS, and import it to SPSS.

### 6.2 Downloading your data via USB to ODK Aggregate VM (Virtual Machine)

Revisit the previous section on USB mounting for how to succesfully connect your Android device to your computer. Use File Explorer (Windows), Finder (Mac), etc. to browse the contents of your Android device. In odk/instances, you will find your completed survey. Drag and drop this onto your computer.

Use the ODK Aggregate VM, a virtual machine[9]to run ODK Aggregate on your computer, which will transform your completed XML surveys into CSV files as described above. These outputted CSV files can, and should, be backed up.

## 7 Accessing responses from repeat data

As ODK Collect spits out data such that each row represents a single survey, it rolls repeated questions – such as anthropometrics for different members of the same household, or different samples during the same focal follow, or the details on the saliva samples collected for the same person on the same day – into a URL, which it sticks in one cell. If you paste that URL in a browser, you'll then see the repeated data.

To avoid this, download ODK Briefcase. Again, the directions on how to do this leave something to be desired: open up our annotated version on the Git repo. Once you have downloaded ODK Briefcase, run the file. Pick a place on your computer when you'd like to store incoming ODK data, in a folder that the software automatically names ODK Briefcase Storage. (Best if you don't stick it in a folder you've named "ODK" anything, because that causes trouble for the software.)

Once the window opens, make sure that "Pull data from" is set to Aggregate 1.0. Then enter (1) the URL of your Google Cloud Platform or (2) the DNS address of your server, depending on which is hosting your ODK Aggregate project. Click Connect. Enter the username and password that you use on ODK Aggregate – that is, the one you created that can do everything, from data collection to site management. Click Connect.

[8]If you have a section of repeated questions, like for taking anthropometrics or reproductive histories for all members of a household, the repeated questions are not exported as part of the CSV per se. Instead, a cell in the column specifying the repeat (e.g., repeat_follow, in our sample focal follow interview) has a long URL in it. At that URL, you can find the multiple rows, one for each iteration of the repeat section. ODK does this so that each row exported represents a single interview... but no, it's not terribly convenient. What we recommend is to use ODK Briefcase and to read our intro to Briefcase below.

[9]This is NOT the same as the virtual machine used by MPI-EVA, described elsewhere in this file.

In the window below, the survey(s) you have synced to ODK Aggregate will appear by name. Click the radio button for the one whose data you wish to download, then click Pull. If all worked, under Pull Status, the word SUCCESS! will appear. Next, select the Export tab. Ensure that the form whose data you wish to access is selected next to Form, and next to Export Directory, click Browse and choose a place where you would like to deposit your files. (We keep ours in the ODK Briefcase Storage folder.) Click Export; if it works, the word SUCCEEDED will appear in the bottom left corner. Now, if you navigate to the location where you exported these files, you'll find both (1) the CSV with one row per completed survey, just as ODK Aggregate produces, and (2) additional CSVs, one for each question with repeats.

You can merge primary and repeat CSV files in program like R by merging by the KEY column from the primary file and the column that begins with SET-OF in the repeat file(s).

# 8 Advanced: Looking up values from a previous interview or census

If you have existing data from a field site – for example, a past interview, or a community census – you can use these data to populate fields in your new form. For example, if you are visiting a household and have already collected data on a given person, you can enter someone's ID number; ODK can search your census for that ID, then populate the name and community fields for that person in your new survey. This can definitely be a time-saver! One particularly handy use is demonstrated in the sample focal follow survey in our our Git repo. You can design your survey such that, when you enter a participant's ID number, ODK Collect throws up their name from your census; if the name doesn't match, you entered the ID number incorrectly, and can swipe back to correct your errors in entry.

Ona.io provides a clear example of how to set up a look-up[10]. Where it says "go to your Ona account," just go to your Appspot site instead; everything else is the same.[11]

[11] Note that the the XLSForm how-to is much less clear about this, as highlighted in our annotated version of their how-to.

[11] This is something we have implemented, if you would like to see an instantiation. On ODK Aggregate, upload a survey – the FF XML file on our Git repo, plus the IDs.csv file as a media file – and then download this survey to ODK Collect on your Android device. If you enter one of the made-up IDs from the IDs.csv file when completing your survey on ODK Collect, you will see that the made-up person's first name appears after the word "PID" on the page that says "CHECK." If the ID you enter does not match, nothing will appear on that page after the word PID.