HOW TO USE OPEN DATA KIT (ODK): A BRIEF TUTORIAL

Monalisa Ghosh
monalisaghosh16@gmail.com
and
Shyamasree Dasgupta
shyamasree.dasgupta@gmail.com

GCP/JU/15/02
Dated 23/11/2015

Global Change Programme
Jadavpur University
188 Raja S C Mallick Road, Kolkata 700032
Phone: +91 33 64147760

Website: http://www.juglobalchangeprogram.org

This working paper should be cited as:
## Contents

1. What is Open Data Kit (ODK)? ................................................................. 1
2. Components of ODK .............................................................................. 1
   2.1 ODK Form (Basic format of XLSForm) .............................................. 1
      2.1.1 The survey sheet ....................................................................... 2
      2.1.2 The choices sheet ..................................................................... 2
      2.1.3 The settings sheet .................................................................... 3
   2.2 ODK COLLECT .................................................................................. 3
   2.3 ODK AGGREGATE ............................................................................. 4
3. Question types ..................................................................................... 11
   3.1 Numbering (integer, decimal, string) ................................................ 11
   3.2 Multiple choice questions ............................................................... 12
   3.3 Display of a text (note) ................................................................. 13
   3.4 Grouping questions ....................................................................... 13
      3.4.1 Nesting groups within groups .................................................. 14
      3.4.2 Skipping .................................................................................. 14
      3.4.3 Repeats .................................................................................. 14
   3.5 Ranking questions ........................................................................... 15
4. Functions ............................................................................................ 16
   4.1 Hints .............................................................................................. 16
   4.2 Relevant .......................................................................................... 16
   4.3 Required ......................................................................................... 16
   4.4 Appearance ................................................................................... 17
   4.5 Multiple language support: ............................................................. 17
5. Conversion of XLSform to XML file .................................................... 17
6. How to upload forms on server .............................................................. 19
   6.1 Manually data upload ................................................................... 19
   6.2 Send all of finalized forms ............................................................ 21
7. Troubleshoots ..................................................................................... 21
8. Limitations ......................................................................................... 21
1. What is Open Data Kit (ODK)?

Open Data Kit (ODK) enables users to capture and instantly digitize information of a variety of formats, eliminating the need for paper questionnaire surveys and data entry. It allows preparing a digitally programmed questionnaire, facilitates intense monitoring of the collection process and gathering of data immediately after survey in a format prepared for analysis. It thus eliminates the need for paper surveys and significantly reduces survey time and time required for data entry. The smartphones and similar devices are equipped to support ODK software.

2. Components of ODK

The ODK software suite consists of 3 different programs including ODK Form, ODK Collect and ODK Aggregate.

First we setup our server that is ODK Aggregate in our laptop. Then we search for the "ODK Collect" app from Google Play on the device. We install the latest version to the mobile device. For more information about ODK Collect and its requirements, visit the https://opendatakit.org/use/collect/. The questionnaires are subsequently saved to the phone’s SD memory, where it can be accessed without internet connectivity. Questionnaire in ODK Collect is ODK Form which is an xml file. There are several ways to create forms for ODK. One of them used in our Project is XLSForm. After completion of XLSForm we convert it into xml file.

XLSForm is a standard form created to simplify the authoring of forms in Excel in a readable format even for non-technological persons. They are simple to get started with but also allow for complex XForms by someone familiar with the syntax described below. Here we have inscribed those syntaxes used in the context of our questionnaire. For further requirements, visit http://xlsform.org/.

2.1 ODK Form (Basic format of XLSForm)

Each Excel workbook usually has three worksheets: survey, choices and settings.
2.1.1 The survey sheet

This worksheet gives the form its overall structure and contains most of the content of the form. It contains the full list of questions and information about how they should appear in the form. Each row usually represents one question; however, there are certain other features described below that one can add to the form which enables the user to incorporate complex type of questions (tables, ranking questions).

The survey worksheet has 3 mandatory columns: type, name, and label. The type column specifies the type of entry we are adding. Different types of questions are discussed under section Question types (Section 3). The name column specifies the unique variable name for that entry. No two entries can have the same name. The label column contains the actual text one see in the XLSForm.

2.1.2 The choices sheet

This worksheet is used to specify the choices for multiple choice questions. Each row represents an answer choice. Answer choices with the same list name are considered part of a related set of choices and will appear together for a question. This also allows a set of choices to be reused for multiple questions (for example, yes/no questions).

The choices worksheet has 3 mandatory columns as well: list name, name, and label. The list name column lets one group together a set of related answer choices, i.e., answer choices that should appear together under a question. The name column specifies the unique variable name for that answer choice. The label column shows the answer choice exactly as one want it to appear on the form. This choices sheet corresponds to the survey sheet mentioned above.
One should keep in mind when authoring forms in Excel is that the syntax used must be precise. For example, if one writes `Choices` or `choice` instead of `choices`, the form won’t work.

### 2.1.3 The settings sheet

This worksheet allows one to further customize the form. An example of `settings` worksheet is below. The column headings in this example `settings` worksheet do the following:

- **form_title**: The title of the form that is shown to users. The form title is pulled from `form_id` if `form_title` is blank or missing.
- **form_id**: The name used to identify the form submission. The form id is pulled from the XLS file name if `form_id` is blank or missing.
- **submission_url**: For encrypted forms, this url specifies the server where finalized forms are submitted to.
- **default_language**: If the survey is to be undertaken in any local language then one needs to keep the “default_language” blank. If it is kept “English” then the question will always appear in English on the smart phone.

#### 2.2 ODK COLLECT

ODK Collect enables us to display the ODK Form on a smart phone in a predetermined logical system which can be used to carry out the survey. The program also supports multiple languages. Then the
surveys taken on the phone can be sent wirelessly to the server ODK Aggregate via either wifi or a mobile internet connection.

Now we describe steps that we follow.

First we install ODK Collect application from Google Play Store on the smart phone. This creates a folder named “odk” in “File Manager” under the “SD Card”. There are four sub-folders under odk. We save XForm in the folder named “forms”.

Once the app is installed there will be an icon on the smart phone named “ODK Collect”. To begin the survey, click on “ODK Collect”. A window will open with multiple options. One can start the survey by touching Fill Blank Form. The questions will appear in sequence. To move on to the next question, simply swipe the screen. At the end one should be careful about saving form and then exit.

Multiple facilities are available in ODK Collect. We can edit saved forms, delete saved and blank forms.

Finally we can send final forms to our server using internet connectivity.

2.3 ODK AGGREGATE

This is our server where we upload or store our data. To have a server we need to follow following steps.

First download ODK Aggregate file of latest version in laptop or PC from https://opendatakit.org/downloads/ for windows i.e. windows-installer.exe.

Next after double click upon above file; a box will appear that is given below.
Now we run this software. After selecting Next button we accept license agreement.

Next a box will appear which ask for a parent directory where usually ODK Aggregate folder will be saved. Generally we select Desktop for our convenience but one can select a location of her own choice.

After that it’s required to choose a platform for our webserver. We have three options below among which we select Google App Engine.
But before that there is very important step. We need to create an application in Google. First go to https://appengine.google.com/ site and we need a gmail account to create an application. Due to upgradation of technology, appearance of the process of creating application may change from time to time. In our case, previously we just went to the above page and put an Application ID and select create application button. Now what we need to do is that we have to start with same https://appengine.google.com/ and through a gmail account we start creating an application.

When we select Create Application a page will appear like this

Here we select Google Developers Console and arrive at next page with several option will come out like this below
Next we get started with **Deploy a Hello world app**. A box pops-up like this below

Next we write a project name of our choice. We agree with terms of services and select **Create** button.

Close the current tab and open [https://appengine.google.com/](https://appengine.google.com/) again.

You will see a page like this
If you click on burnished-core-99606, we arrive at the page below. Here click on Application Settings under Administration and finally arrive at a page where at the end there is a Create button.
Within few seconds application will be created after click an on **Create** button with a success message.

So, this is the process how we create an application through Google as we need this for our web server.

In ODK aggregate we are at a stage where we select **Google App Engine** as our platform. Next we enter a name of **instance**.
In the next page we have to put **ODK Aggregate Username** of our choice.

![ODK Aggregate Username](image)

Next we required to enter the **application id** which we created before

![Application ID](image)

We already have this **application id** when we created application in appengine.google.com. In our above example it’s **burnished-core-99606**. This is very crucial as it is going to be our server’s url i.e. [http://burnished-core-99606.appspot.com](http://burnished-core-99606.appspot.com). We don’t required to select the box above against which it is written, “Yes, a Master/slave...........application id”.

Click on **Next** ODK Aggregate will start configuring. Then a page will ask for Email and Password where we insert our Gmail id and password which we created for Google app-engine and select **upload**.

Wait 3-10 minutes for the script to run. Once it's finished, we will get a "Success" message: **END-SCRIPT SUCCESS**.
If there is any mistake in any step, at the end it will show END-SCRIPT-FAILUER. That means process is hindered due to some wrong information. But the good thing is that the problem occurred is written in output message in this stage. After identifying the problem we have to re-run the Setup Wizard by going back to installed ODK Aggregate and double-click the installer to run it again.

Before re-running the wizard again we should delete the folder that has been created on our desktop as ODK Aggregate. Otherwise, we will be restricted when we insert destination of ODK Aggregate folder which we are supposed to do in our second step.

This is the theory through which we have tried to depict entire process of using ODK in details and hope, now you can apply ODK by yourself. But it may happen that you may found some difficulties when we practically try to do this.

3. Question types

This is the most important part in forming a questionnaire. XLSForm supports a number of simple question types. These are just some of the simple question type we use under type column in the survey worksheet:

<table>
<thead>
<tr>
<th>Question type</th>
<th>Answer input</th>
</tr>
</thead>
<tbody>
<tr>
<td>integer</td>
<td>Integer (i.e., whole number) input.</td>
</tr>
<tr>
<td>decimal</td>
<td>Decimal input.</td>
</tr>
<tr>
<td>text</td>
<td>Free text response.</td>
</tr>
<tr>
<td>string</td>
<td>Under this input words, numbers, decimals all are allowed.</td>
</tr>
<tr>
<td>select_one [options]</td>
<td>Multiple choice question; only one answer can be selected.</td>
</tr>
<tr>
<td>select_multiple</td>
<td>Multiple choice question; multiple answers can be selected.</td>
</tr>
<tr>
<td>note</td>
<td>Display a note on the screen, takes no input.</td>
</tr>
<tr>
<td>date</td>
<td>Date input.</td>
</tr>
<tr>
<td>time</td>
<td>Time input.</td>
</tr>
<tr>
<td>today</td>
<td>Automatically captures date and time of the day of survey from android</td>
</tr>
</tbody>
</table>

For further types of question, visit http://xlsform.org/.

3.1 Numbering (integer, decimal, string)

If the response is in number like 1, 2, 3, 4 etc, ‘integer’ question type is normally used. Remember this type won’t function in ODK Collect if one tries to give an input of any decimal like 1.3 etc. In that case
**decimal** type is used. So be careful, when inserting question type. If both integer and decimal appear in response, in this case one should use ‘**string**’ question type. Under **string** type one can write in both words and numbers.

<table>
<thead>
<tr>
<th>type</th>
<th>name</th>
<th>label: English</th>
</tr>
</thead>
<tbody>
<tr>
<td>integer</td>
<td>members_no.</td>
<td>103. How many members are there in this household?</td>
</tr>
<tr>
<td>string</td>
<td>dist_XXXXX</td>
<td>113. How far XXXXX High School is located from your home?</td>
</tr>
</tbody>
</table>

### 3.2 Multiple choice questions

**select_one** (select only one answer) and **select_multiple** (select multiple answers) questions both can be used in XLSForm. With this type of question we need to incorporate available options for that particular multiple choice question in our **choices** worksheet.

Here is an example of a **select_one** question:

<table>
<thead>
<tr>
<th>type</th>
<th>name</th>
<th>label: English</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>select_one</strong> religion</td>
<td>religion</td>
<td>107. Would you like to choose your religion from following options?</td>
</tr>
<tr>
<td>list_name</td>
<td>name</td>
<td>label: English</td>
</tr>
<tr>
<td>religion</td>
<td>hin</td>
<td>Hindu</td>
</tr>
<tr>
<td>religion</td>
<td>mus</td>
<td>Muslim</td>
</tr>
<tr>
<td>religion</td>
<td>chr</td>
<td>Christian</td>
</tr>
<tr>
<td>religion</td>
<td>oth2</td>
<td>Other</td>
</tr>
<tr>
<td>religion</td>
<td>no_rel</td>
<td>No religion</td>
</tr>
</tbody>
</table>

**Note** that **religion** in the **survey** worksheet must match the **religion** in the **list name** column in the **choices** worksheet i.e. options written after **select_one** under survey sheet and list name under **choices** sheet both must be same. This ensures that the form displays the correct list of answer choices for a particular question.

Example of **select_multiple** question:

We can also add multiple choice questions that allow multiple answers to be selected, like so:

<table>
<thead>
<tr>
<th>type</th>
<th>name</th>
<th>label: English</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>select_multiple</strong> enrolsc</td>
<td>enrolsc</td>
<td>112. Which are the schools where the children of the household are enrolled?</td>
</tr>
<tr>
<td>list_name</td>
<td>name</td>
<td>label: English</td>
</tr>
</tbody>
</table>
3.3 Display of a text (note)

Sometime it is required to display note and we don’t need any response against it in this case this question type is used which only display a note.

Example is given below.

3.4 Grouping questions

If multiple questions are related to one particular issue, then often they are put in a group. Example is given below.
Notice how **end group** doesn’t require a name or label, because it is hidden in the form.

### 3.4.1 Nesting groups within groups

Groups of questions can be nested within one another:

<table>
<thead>
<tr>
<th>type</th>
<th>name</th>
<th>label:English</th>
</tr>
</thead>
<tbody>
<tr>
<td>begin group</td>
<td>worries</td>
<td>Worries</td>
</tr>
<tr>
<td>text</td>
<td>worries, txt</td>
<td>109(a). What are your 3 most important long run worries in your household?</td>
</tr>
<tr>
<td>begin group</td>
<td>top_3worries</td>
<td>109(b). Top 3 worries</td>
</tr>
<tr>
<td>note</td>
<td>note1</td>
<td>Please give a ranking of the following worries.</td>
</tr>
<tr>
<td>select_one_numbers</td>
<td>poverty</td>
<td>Poverty/Low income</td>
</tr>
<tr>
<td>select_one_numbers</td>
<td>education</td>
<td>Education</td>
</tr>
<tr>
<td>select_one_numbers</td>
<td>water</td>
<td>Water</td>
</tr>
<tr>
<td>select_one_numbers</td>
<td>sanitation</td>
<td>Sanitation</td>
</tr>
<tr>
<td>select_one_numbers</td>
<td>health</td>
<td>Health and hygiene</td>
</tr>
<tr>
<td>select_one_numbers</td>
<td>rd_transport</td>
<td>Road condition and transport</td>
</tr>
<tr>
<td>select_one_numbers</td>
<td>unemployment</td>
<td>Unemployment</td>
</tr>
<tr>
<td>select_one_numbers</td>
<td>oth1</td>
<td>Other</td>
</tr>
<tr>
<td>text</td>
<td>other1</td>
<td>Please specify, if ‘other’.</td>
</tr>
<tr>
<td>end group</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The first **end group** one see closes the **Top 3 worries** group, and the second one closes the beginning **Worries** group. When working with groups and one keep getting error messages when trying to converting the form in xml file, double-check that for each **begin group** one have one **end group**.

### 3.4.2 Skipping

One feature of XLSForm is the ability to skip a group of questions by combining the group feature with relevant syntax. If one wants to skip a group of questions all at once, put the relevant attribute at the beginning of a group. We didn’t use this feature in our present survey even though we mentioned it as it comes under usages of **relevant function** (discussed under Functions).

### 3.4.3 Repeats

A user can repeat a group of questions by using the **begin repeat** and **end repeat** construct:

<table>
<thead>
<tr>
<th>type</th>
<th>name</th>
<th>label:English</th>
</tr>
</thead>
<tbody>
<tr>
<td>begin group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>end group</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In this example, the Name, Age, Gender, Marital Status, Highest Education and Occupation are grouped together in a repeat group, and the user can repeat this group as many times as required by selecting the option in the form to start another repeat.

The label column is optional for begin repeat. Assigning a label to a repeat group will add the label as a title to the block of repeat questions in the form.

Instead of allowing an infinite number of repeats, one can specify an exact number of repeats. Details can be found in [http://xlsform.org/](http://xlsform.org/).

### 3.5 Ranking questions

Sometimes we need to record respondent’s preference ordering over few choices. This type of example is given below.

In above example respondent has to rank the 3 options according to her preference. We form it as a group question with appearance of list of numbers: 1, 2, 3 which are included in choices sheet against list_name **numbers**.
4. Functions

It has been written before that type, name and label, these three are mandatory in survey worksheet. Now we consider few more functions which are used to improve pattern of questionnaire such as hints, constraint, relevant, require etc.

4.1 Hints

Sometimes we want to add a small hint to a question on the form, instructing the user how to answer the question, but one doesn’t want the hint to be part of the question itself. It’s easy to add hints to questions in XLSForms. Simply add a hint column and add the hint message. In appears in italics. See below for an example.

![Example of hint column](image)


4.2 Relevant

One great feature of XLSForm is the ability to skip a question or make an additional question appear based on the response to a previous question. Below is an example of how to do this by adding a relevant column for a select_one question, using example from our questionnaire before:

![Example of relevant column](image)

In this example respondent is asked, “Do you know about Arsenic?” If the answer is yes, then the next question appears whether he is aware of the diseases caused by the arsenic contamination in drinking water. $\{ \} $ around the variable name “knw_as” from previous question equals (=) ‘yes’ i.e the alternative which the variable takes from choice list.

4.3 Required

Another column ‘required’ is added to restrict the surveyor from moving to next question if she tries to skip any question, in other words marking them as required means the user will unable to move on to the next question or submit the form without entering an answer for that question.

To make questions required, see the example below:
4.4 Appearance

The **appearance** column allows one to change the appearance of questions in the form. The following table lists the possible appearance attributes and how the question appears in the form which we have used in our questionnaire.

<table>
<thead>
<tr>
<th>Appearance attribute</th>
<th>Question type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimal</td>
<td>select_one, select_multiple</td>
<td>Answer choices appear in a pull-down menu.</td>
</tr>
<tr>
<td>field-list</td>
<td>Groups</td>
<td>Entire group of questions appear on one screen (for mobile clients only).</td>
</tr>
</tbody>
</table>

In our questionnaire we use this function as below:

![Image of XLSForm]

One can find many options of appearance attributes which are listed in [http://xlsform.org/](http://xlsform.org/).

Functions that one can use in XLSForm according to requirements of questionnaire are **Formulas, Calculation**. It was not required for this present context, so we didn’t use any of these two. Details of these two functions can be found in [http://xlsform.org/](http://xlsform.org/) link.

4.5 Multiple language support:

It is easy to add multiple languages to a form so that one can represent her own questionnaire in a language of her choice. This is one of the salient features of ODK. We have to simply name the label::language1, label::language2.

5. Conversion of XLSForm to XML file

Finally, when questionnaire is completed convert this final **XLSForm** into *.xml file which can be done by dint of [http://opendatakit.org/xiframe/](http://opendatakit.org/xiframe/). Here one can convert by choosing questionnaire in excel and submit it. This website converts that xls file to xml file which is downloadable.
After submitting a page will appear like this below:
By selecting download we get our final XForm i.e., ODK Form which through mail will be sent to ODK Collect.

This xml file or XForm, is a popular open form standard that allows one to skip from complex functionality of logic in a consistent way.

6. How to upload forms on server

We can upload data by two methods. Either we can upload manually or just send all of finalized forms by using internet connection of our device.

6.1 Manually data upload

Final data is located in instances folder under main odk folder of SD Card. We have to send each xml file via gmail. Then download those files in your desktop or laptop. Then select Submission Admin under Form Management, a page will appear like below.

Click on Manually Upload Submission Data and choose a file from the location where dataxml files are saved. Finally select Upload Submission and data will be uploaded on server.
The other way is to upload data by sending from device. In ODK Collect we have to select all finalized forms and touch Send Selected button. While sending final forms we have to keep our server open. There is another way to upload that is done manually.

Note that we can send final forms if and only if our questionnaire is uploaded on server. How we upload our questionnaires given below:

and server will appear with 2 main headings Submissions and Form Management. Next select on Form Management and then click on Add New Form which will appear under Forms List. Now we can upload our questionnaire XForm by choosing downloaded XForm.

When data will be uploaded, server will look like this below.
6.2 Send all of finalized forms

In ODK Collect we have an option **Send Finalized Form**. Under this all final forms are available. Select **Toggle all** and then **Send selected** buttons.

Now we have to **Export** data in *.csv format which save all data in Excel Sheet.

This is the **final output** of dataset which we need for our analysis.

7. Troubleshoots

First, after reading manuals and theories when we had started with ODK Aggregate we repeatedly ended up with **END-SCRIPT-FAILURE**. We couldn’t understand our mistake from **output message**. After consulting with JU experts we get to know our mistake rather silly mistake. The process was hindered due to obstacle in internet connectivity. While setting up ODK Aggregate we used wifi connection of our Lab which restricted transferring information across web. So this problem was sorted out by using a dongle.

We have a question where respondent is asked to rank over three alternatives and another question where respondent has to select one level of satisfaction regarding criteria of drinking water from each source. Former has three options:1, 2, 3 for each alternative and the later has four levels of satisfaction for each criteria. Respondent has to select one among them. So we want them to appear in a drop down box after each alternative and each criteria. But we didn’t perceive how to do this after trial and error of few methods finally we had to post our problem in ODK Community which is a open forum where people discuss about problems regarding ODK.

https://groups.google.com/forum/#!forum/opendatakit.

8. Limitations

We have already talked about advantages. But there are limitations too.

1. With full battery backup our device (Micromax) allows to conduct survey of maximum 6 households with disabled auto-rotation and lower brightness.
2. Server has restriction on daily usage. So it was not allowed to use after a certain time.

Server shows a message like below:
Problem is that we have 3 repeat questions for which we need to copy it from server and paste on excel sheet. But we are bounded by the available time span.

Note:

While copying select upto last digit does not drag cursor to the end of the box. Like this below:

But do not select like this

Notice that in above example we copied by selecting Copy option which will appear on Left click not by ctrl+c. As ctrl+c will not function in this context.

We figured out two solutions:

Try to keep submissions below 50, even though capacity per page is 100. Export all data and delete data from server when all informations are transported to excel sheet.

We create a second sever with different mail id and all necessary requirements. In our study we make second server and upload manually. We could not send it from device as submission url is different for new server and our questionnaire XForm includes previous submission url. All detail informations are given above with 1st server*.