



Open Data Kit: What We Learned

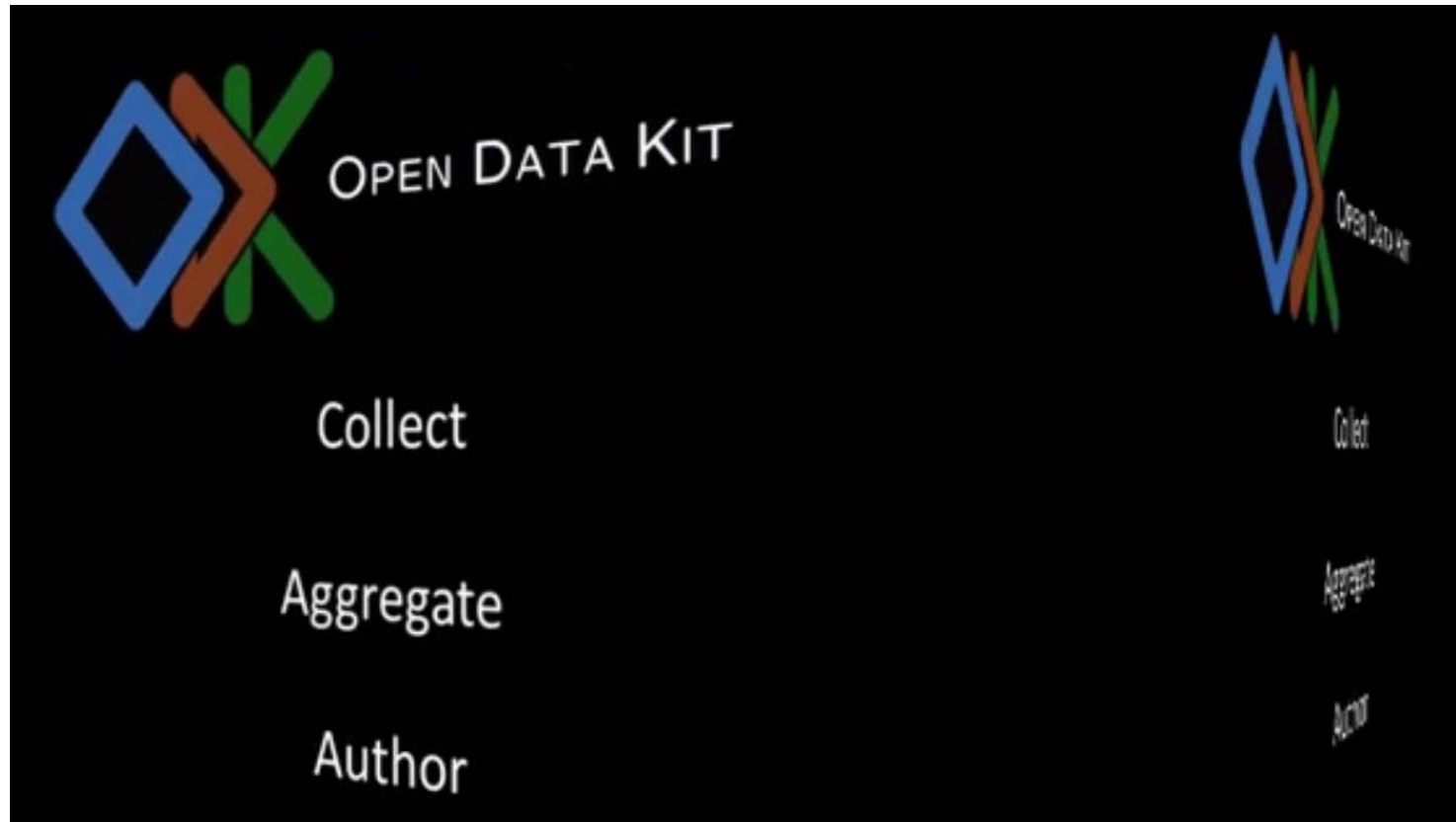
Francis Smart

MICHIGAN STATE
UNIVERSITY



Video Introduction

View 5 Minute Video



Found at <https://opendatakit.org/>

Benefits of Using Mobile Data Collection

- Data can be uploaded as it is collected/needed.
- Error checking can be embedded into the collections device.
- Data is both collected and digitized simultaneously, reducing labor costs and potential transmission errors.
- Media data can be easily collected and integrated such photos, videos, audio recordings, or GPS.
- Complex data can have embedded instructions.
- Verification checks can be embedded into the device (photo of field/family/GPS etc.)

Major Benefits of Using ODK

- ODK is free. The cost of mobile data collection software is generally exorbitant.
- ODK support free. Most companies provide support for their products only at significant additional costs.
- ODK is open source. ODK is created by public research institutions for public research.

ODK Open Data Kit

Created by University of Washington- Computer Science and Engineering

<https://opendatakit.org/>

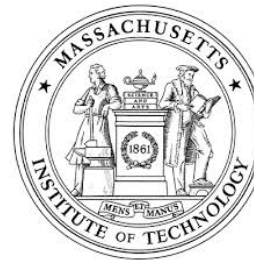


Selected ODK Users, Projects, or Funders



USAID
FROM THE AMERICAN PEOPLE

BILL & MELINDA
GATES foundation



wellcometrust



Implementation List

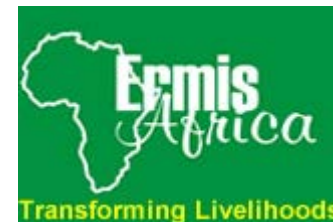
- Has projects in 37 countries and 1 at the International Space Station
- ODK is deployed on tens of thousands of devices submitting hundreds of thousands of forms a day.
- 52 Companies list ODK form building and deployment services on the ODK website.

Selected ODK Implementation Companies



EarlySail

Contract Product Development: Mobile Apps and Messaging, Portals, Unified Communication



Clearly ODK has been successful but...

Despite the seeming immense success of ODK, we experienced a number of challenges implementing the software.

The large number of companies offering to assist with the deployment of ODK, may suggest that we are not alone in this. Consulting with a prominent one of these companies here in Maputo resulted in an astronomical price quote (25,000 USD) for *building* an application much simpler than our needs.

In this presentation we will discuss the ODK tools we used, the problems we faced, and the solutions we came up with.

Outline of Presentation

1. Discuss the various components:
 - a. ODK collect
 - b. XLS Form Builder
 - c. ODK Aggregate
 - d. Briefcase

2. General challenges
 - a. Documentation – scattered and incoherent
 - b. Software options diverse
 - c. A-typical project

3. Problems and solutions
 - a. ODK collect
 - b. XLS Form Builder
 - c. ODK Aggregate
 - d. Briefcase

4. Alternative software options
 - a. CSPro Android
 - b. SurveyBe

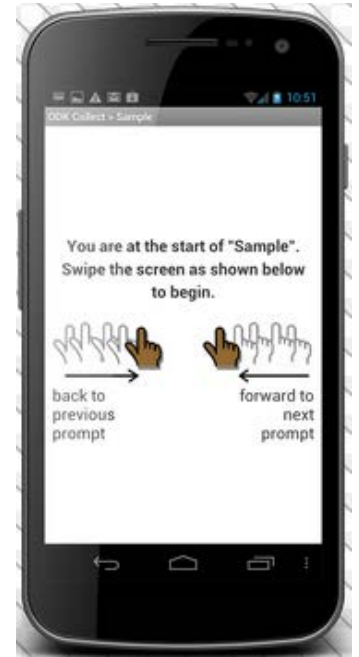
1a. ODK Collect

Easy swipe right to left navigation.

Easy installation (android apps store).

Can deliver text, images, video, and sound.

Can collect text, video, sound, GPS.



1b. XLSForm Builder

Build forms in excel.

One line per entry. (Documentation <http://xlsform.org/>)

note	A_note	A. IDENTIFICAÇÃO DA EXPLORAÇÃO					
begin_group	A_group1	CONFIDENCIALIDADE E LIBERDADE DE PARTICIPAÇÃO					field-list
note	A_note1	Este inquérito forma parte de um estudo do Instituto de Investigação Agrária de Moçambique (IIAM) em colaboração com a EMBRAPA do Brazil e a Unive					
note	A_note2	A. IDENTIFICAÇÃO DA EXPLORAÇÃO					
select_one zona	A01	A01. A exploração localiza-se em					
end_group	A_group1						
begin_group	A_group1.5	Associação					field-list
select_one A02	A02	A02. Associação	cf1=\${A01}			minimal	
integer	A02a	A02a. Por favor entrar o código da associação se não consta na lista		.<1000 and > 99	não pode ser um numero menor de 99 ou maior de 1		
end_group	A_group1.5						
begin_group	A_group2	A. IDENTIFICAÇÃO DA EXPLORAÇÃO					field-list
select_one A03	A03	A03. Número do agregado familiar/Nome do Entrevistado	cf1=\${A02} or cf1=0		yes	minimal	
text	A03a	Se for outro, especificar:					
calculate	HH_key	HH-key					
calculate	A04a	A04a. Numero de Contacto					
calculate	A06	Localização/ Descrição da zona da machamba:					
calculate	A07	Coordenadas de GPS da machamba onde a entrevista foi realizada (UTM):					
end_group	A_group2						
begin_group	A_group3	A. IDENTIFICAÇÃO DA EXPLORAÇÃO (CONT.)					field-list
integer	A04	A04. Número da visita					
note	A04a_note	Numero de Contacto: \${A04a}					
note	A06_note	Localização/ Descrição da zona da machamba: \${A06}					
note	A07_note	Coordenadas de GPS da machamba onde a entrevista foi realizada (UTM): \${A07}					
integer	A04a_int	A04b. Se o número de contacto não aparece aqui em cima, indique o numero de contacto aqui					
select_one A05	A05	A05. Tipo de exploração baseada na listagem					
text	A06a	A06. Se a localização/ descrição da zona da machamba não apareceu em cima ou precisa de ser modificada, por favor escreve a localização aqui					
text	A07a	A07a. Se as coordenadas de GPS da machamba onde a entrevista foi realizada (UTM) não foram registadas, ou precisam de ser colhidas de novo, faz aqui					

1b. XLSForm Builder

- Upload forms at opendatakit.org/xiframe

XLSForm

XLSForm (formerly XLS2Xform) is a tool to simplify the creation of forms. Forms can be designed with Excel and XLSForm will convert them to XForms that can be used with ODK tools.

See [XLSForm](#) for more information.

File: P1_B1.xlsx



Download

Preview in enketo

1c. ODK Aggregate

ODK Web Server handling data submissions, form distribution, and data downloads.

(installation <https://opendatakit.org/use/aggregate/>)

The screenshot displays the ODK Aggregate web interface. At the top, there is a navigation bar with 'Submissions' and 'Form Management' tabs. The 'Submissions' tab is active, showing 'Filter Submissions' and 'Exported Submissions' options. The main content area shows a table for 'Part 2_V11' with columns for various data groups and their counts. The table has a header row with columns: A_group1 A_note1, A_group1 A_note2, A_group1 A01, A_group1 A02, A_group1 A03, A_group1 A03a, A_group3 HH_key, A_group3 A04, A_group3 A04a, A_group3 A04a_note, A_group3 A06, and A_group3 A06_note. The first row of data shows counts: 4, 413, 22, and 9499, with a red 'X' icon in the first column. The bottom right corner of the table area contains the text 'Proxima do Vale Mupoa pd ap lady'.

Part 2_V11											
A_group1 A_note1	A_group1 A_note2	A_group1 A01	A_group1 A02	A_group1 A03	A_group1 A03a	A_group3 HH_key	A_group3 A04	A_group3 A04a	A_group3 A04a_note	A_group3 A06	A_group3 A06_note
✖		4	413	22		9499	3				

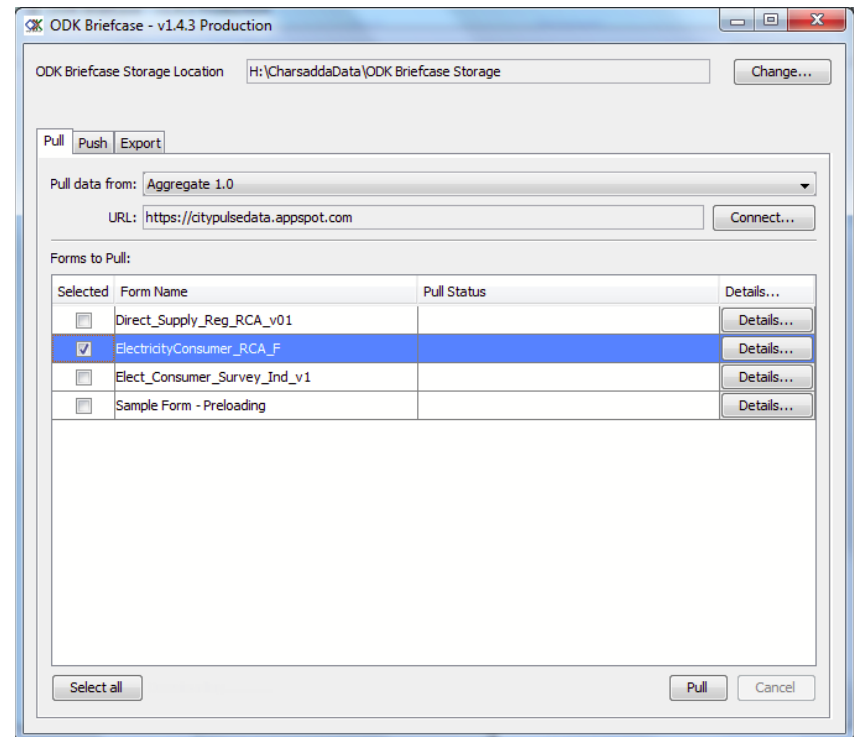
1c. ODK Aggregate (continued)

- Track incoming data directly in real time.
- Map incoming data to google maps.
- Graph data to google fusion tables as it comes in.
- Track incoming data.



1d. Briefcase

- Briefcase allows the direct downloading to desktop of submitted forms from ODK Aggregate.
- Forms are collections of csv files which need to be managed directly at this point with your choice of data software.



2. General Challenges

- Documentation is scattered (and only in English)
- There are too many software products available under 'ODK': ODK Build, ODK Collect, ODK Aggregate, ODK Form uploader, ODK Briefcase, ODK Validate, ODK XLX2XForm, ODK Tables as well as companies producing ODK knockoffs such as CommCare ODK, Acquee ODK.

2. General Challenges (continued)

Too many alternative methods to accomplishing same task. To build a survey form for example:

1. XLSForm

name	A_serie	A_IDENTIFICAO DA EXPLORACAO
begin_group	A_group0	CONVENIENCIA DO LIBERDADE DE PARTICIPACAO
text	A_serie0	Este questionário serve para o seu estudo de Identificação de Atividades
text	A_serie1	A IDENTIFICAO DA EXPLORACAO
select_one area	A17	A17. A atividade localizada em:
end_group	A_group1	
begin_group	A_group2	Associação
select_one ASS	A12	A12 Associação
text	A12a	A12a Por favor entre o código de associação se não consta na lista
end_group	A_group3	
begin_group	A_group4	A IDENTIFICAO DA EXPLORACAO
select_one ASS	A15	A15 Sistema de agrupamento familiar/prime do Estabelecimento
text	A15a	A15a Se for caso, especificar
checkbox	ISS key	ISS key
checkbox	A16a	A16a Titulo de Comercio
checkbox	A16	A16a Descrição de zona de mercado
checkbox	A17	A17 Coordenada de GPS do mercado onde a entrevista foi realizada (UTM)
end_group	A_group5	
begin_group	A_group6	A IDENTIFICAO DA EXPLORACAO (CONT)
text	A18	A18 Nome do meio
text	A18a	A18a Tipo de Comercio (UNAS)
text	A18a1	A18a1 Localização: Descrição de zona de mercado (UNAS)
text	A17a	A17a Coordenada de GPS do mercado onde a entrevista foi realizada (UTM)
text	A18a2	A18a2 Se o sistema de contato não aparece aqui em caso, indique o nome
select_one ASS	A15	A15 Tipo de reprodução baseada na origem
text	A16	A16 Se a localização descrita de zona de mercado não aparece em...
text	A17	A17

3. ODK Build



2. XML scripting

```
<translation default="true()" lang="default">
  <text id="static_instance-B13-2">
    <value>Regularmente</value>
  </text>
  <text id="static_instance-B13-1">
    <value>Ocasionalmente</value>
  </text>
  <text id="static_instance-B13-0">
    <value>Nunca</value>
  </text>
  <text id="static_instance-L17-4">
    <value>Membro da Familia</value>
  </text>
</translation>
```

4. Commcare ODK etc...

2. General Challenges (continued)

- An agricultural production survey seems to be an atypical implementation of ODK.
- The typical input tables (multi-row, multi-column) were not an option for ODK.
- Having one question per screen display is the default.
- The excel document became excessively long (with over 1600 rows).

3. Outline of Software Specific Issues

A. ODK collect

- ii. Auto-fill
- iii. Multiple items per screen groups
- iv. Skip items
- v. Soft errors

B. XLS Form Builder

- i. Slow upload
- ii. Form conversion failure
- iii. Xml data inefficiency

C. ODK Aggregate

- i. Difficult to use interface

D. Briefcase

- i. Error Downloading forms

3a. ODK collect

Problem: Auto-fill was hard to make functional for repeated panel data.

Solution: None yet achieved.

Problem: Single items per screen.

Solution: Grouping nearly all of our questions.

Problem: Cannot automatically skip questions when on same screen with logic.

Solution: Created error reports.

3a. ODK collect (continued)

Problem: There is no soft required option.

Solution: Wrote soft code that conditionally displayed manually.

Problem: The more data checks we add the slower the software runs. Inputting data into the software directly is more time consuming.

Solution: Enumerators have had to adapt.

3a. ODK collect (continued)

Problem: If a repeat prior to the current repeat is deleted, the program can crash.

Solution: Save frequently while data collecting/entering, and then we've had to often transfer information from last repeat to the one you are trying to delete to prevent the program from crashing.

Problem: Can no longer edit questionnaires after they are sent to the server. Sometimes questionnaires are sent too early.

Solution: Add a required check a box if the questionnaire is finalized. Prevents sending of the questionnaire until enumerator is ready.

3a. ODK collect (continued)

Problem: Cannot edit questionnaires after server submission. If we wait to edit before submission to the server, we cannot see it on the server for hours or days after.

Solution: Rapid editing of surveys as they are filled to increase submission time.

Problem: The more data checks we add the slower the software runs. Inputting data into the software directly is more time consuming.

Solution: Enumerators have had to adapt.

3b. XLS Form Builder

Problem: Slow upload. The larger the excel form became the slower the upload.

Solution: Download windows version much faster conversion

(<https://github.com/UW-ICTD/xlsform.exe/>).

Problem: Form fails to convert without reported error.

Solution: Download windows version much faster conversion ([xlsform.exe](#)).

3b. XLS Form Builder (continued)

Problem: xml data format is often inefficient. Often the uploaded forms (200K) are much smaller than the downloaded ones (+1 MB). This can be a problem with slow connection speeds.

Solution: Not much can be done on the users side.

Problem: Some unexpected errors are significant enough to crash ODK collect.

Solution: We pretested forms with real data and corrected in form builder updating our form collection version as needed (we are on 28).

3c. ODK Aggregate

- Interface is a little clunky but really not too many issues with ODK Aggregate.

3d. Briefcase

Problem: The issue that we faced with Briefcase is that it can take a very long time to download data and that the download can freeze for unknown reasons and fail to work even if started fresh.

Solution: Delete or rename folder with downloaded data. Start over. Better internet connection may reduce the chance of download interruptions.

4. Alternative software options

a. CSPro Android – Currently in Beta testing
(<http://www.csprousers.org/category/android/>)

b. SurveyBe

Is a professional mobile survey collection system. For our purposes (complex agricultural production survey) it would have likely been easier, but costs more.

Conclusion

- ODK has achieved a wide degree of success by many measures.
- However, ODK needs unification of software and documentation.
- In its current state it takes a highly sophisticated user in order to program complex surveys.
- More detail on how to build basic and complex surveys to follow in an upcoming workshop (slide presentation).