

Collecting data with smart forms on mobile phones

Yaw Anokwa

<http://nafundi.com>

Hi, I'm Yaw.

Last June, I finished my PhD in Computer Science at UW with Gaetano Borreillo.

As part of that work, I helped to create Open Data Kit (or ODK), a free and open source platform for collecting data with smart forms on mobile phones and tablets.

In this talk, I'll show what some of the tools are and give you a few examples of their use at scale. I'll also show some of the new ODK tools that are being developed at UW.

I'll end with a quick look at some of the work I'm doing at Nafundi, which is a software company I started after I graduated last June.

I've got two goals today.

I want to convince you that ODK is fantastic for anyone who wants to collect data accurately, quickly, offline and at scale.

I also want to convince you that technology can help solve important problems -- even in the most challenging environments.

I want to make this informal, so interrupt to ask questions at any point, and I'll leave some room at the end for even more questions.

Collecting data in low-income regions is difficult.



This picture is of a village where I did some work. It's at the edge of Lake Victoria in Eastern Uganda.

I think it's a great example of the places where organizations like IHME need to collect data.

You can see in a place like this there is almost no infrastructure to enable communication with the rest of the world.

There's no electricity, there are no land lines, and when it rains, there are no roads.

We didn't even see any Coca Cola for sale, which is the one thing you can usually find even in the most remote of places. That's how you know when you've really left the map.

So how do you collect data in a place like this?

Paper is common, but can be very limiting.



You probably use paper. Here is an example of a nicely organized records room at a hospital in Kenya.

Even in this best case scenario, there are problems because paper is very limiting.

It's hard to search paper.

It'll take you a few hours just to tell me how many of these patients have malaria.

With paper, it's hard to have data validation.

If someone enters an age of 40 instead of 4 or enters the data in day-month-year instead of month-day-year, now you have lots of data cleaning to do.

With paper, it's hard to capture multimedia that would make your data richer.

If you are surveying the efficiency of hospitals in an area, it'd be nice to have the GPS locations and images as part of that record.

With paper, it's to get the data analyzed in real time.

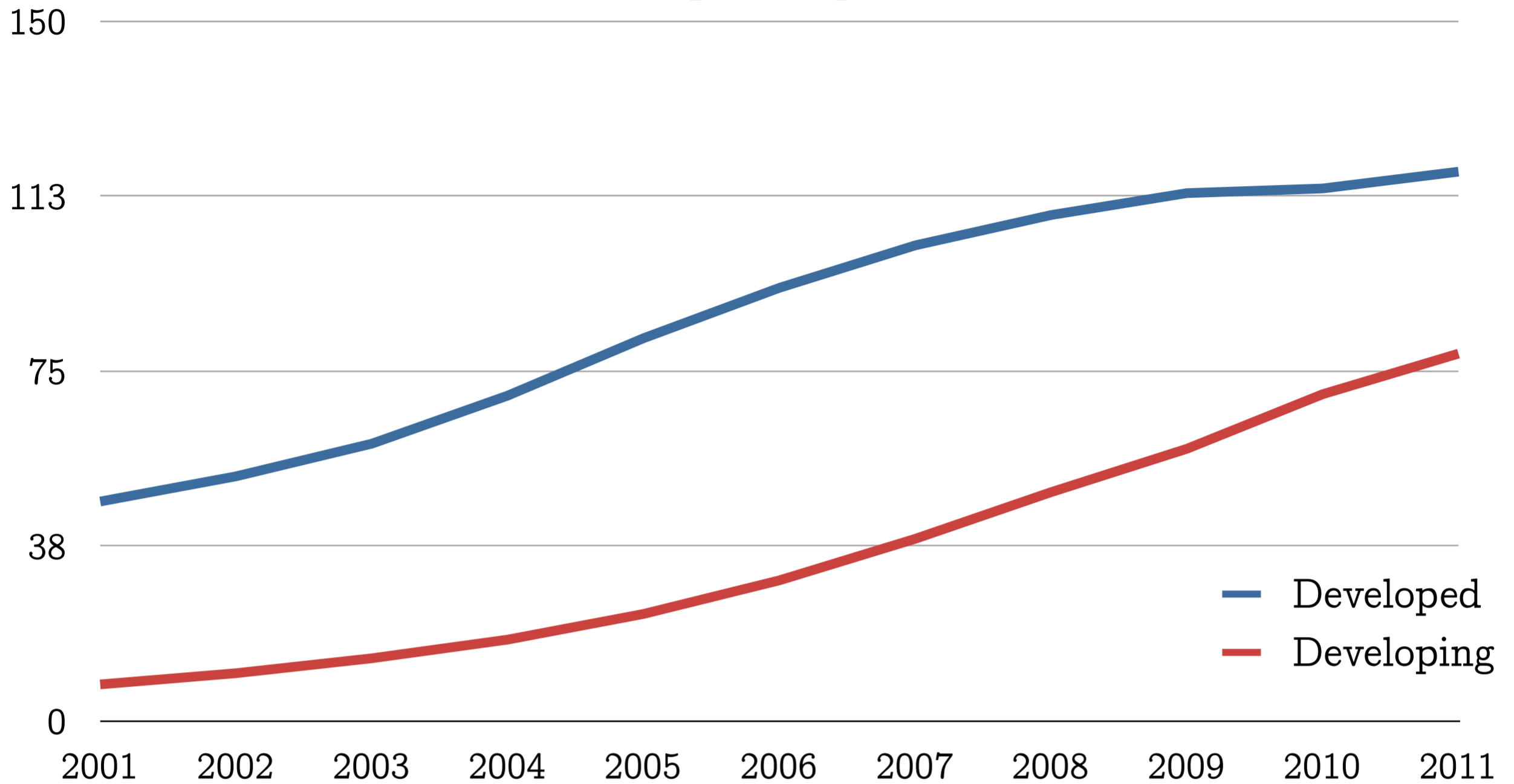
You don't want to wait a few months after the study is done, digitize all the data before you know something went wrong.

Our research group saw these problems a lot when we did our fieldwork, and so that was the driving motivation behind our work -- addressing these limitations of paper.

We chose to do this using smartphones and internet or "cloud" servers to try to replace paper entirely.

Growth of cellular technology enables innovation

Mobile subscriptions per 100 inhabitants



It turns out there is incredible growth of cell technology all around the world that enables this kind of innovation.

So in this graph, the blue curve is developed countries and the red curve is developing countries. And this is over time from 2001 to 2011.

To me the growth in the developing countries is amazing! And what's more with the increase in mobile subscriptions also comes an increase in Internet access.

Let's take Kenya as an example.

Currently 26% of Kenyans have Internet access and 99% of those people get onto the Internet through the cell networks -- so either on their phone or using a USB dongle.

And these aren't always feature phones. In Kenya, the price of the basic smartphone, something like an original iPhone, is about \$50-\$60. And it sells well at that price.

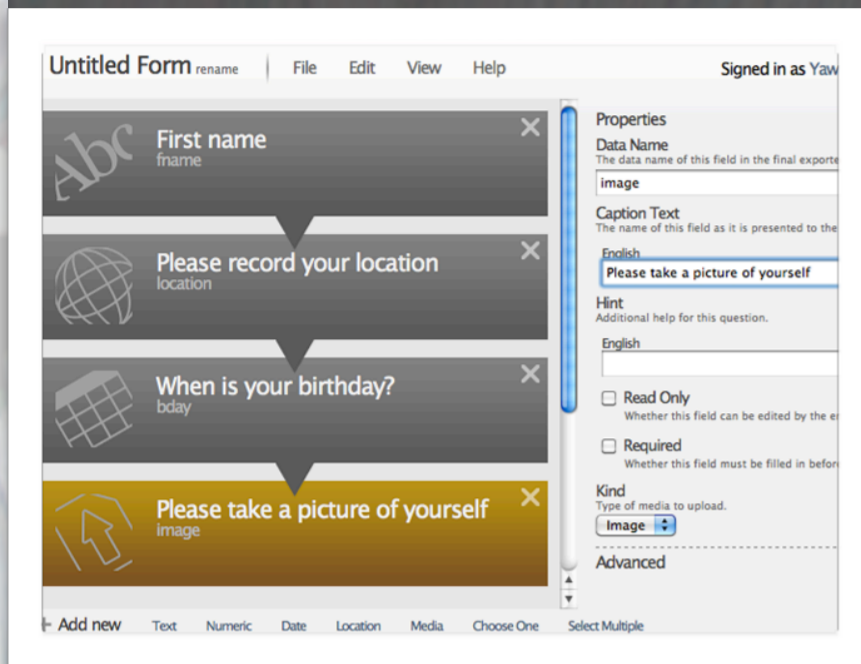
The punchline here is that when we started ODK five years ago, there was this same growing Internet access and smartphone usage in these developing areas. And on that infrastructure we thought we could build systems that address the limitations of paper I talked about in the last slide.

And as researchers, we felt if we could do data collection with these phones and server, we could help folks solve some big problems even in these challenging environments. So that's what drove much of what has become ODK.

So what is ODK?

ODK uses phones and servers to digitize data collection.

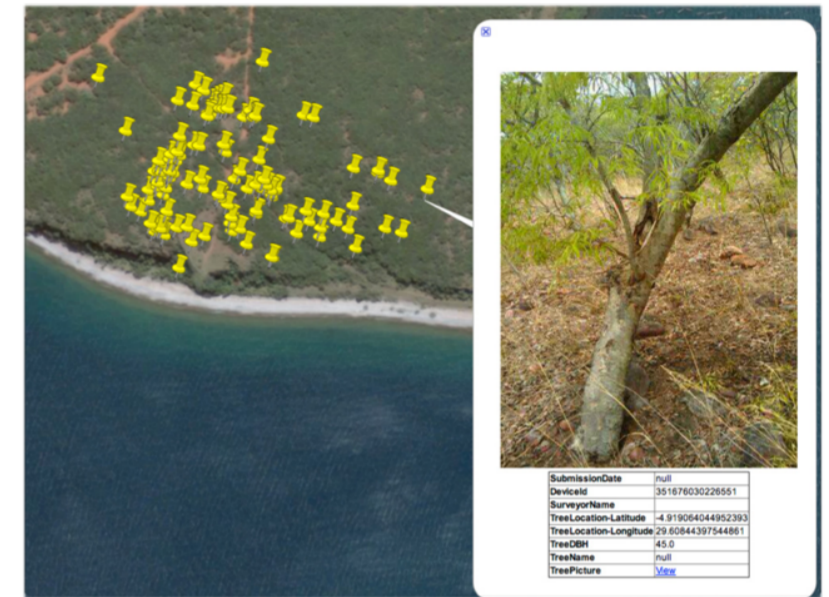
1. Build form



2. Collect data



3. Aggregate results



1371 → 1540 ↓

1541 → 1730 ↓

ODK is a free and open-source set of tools which use smartphones and cloud servers to digitize data collection.

It's great for field staff who need to collect data accurately, quickly, offline and at scale.

ODK can be used to implement large socio-economic surveys, collect geo-tagged vaccine data, or even triage patients with complex decision trees and embedded videos.

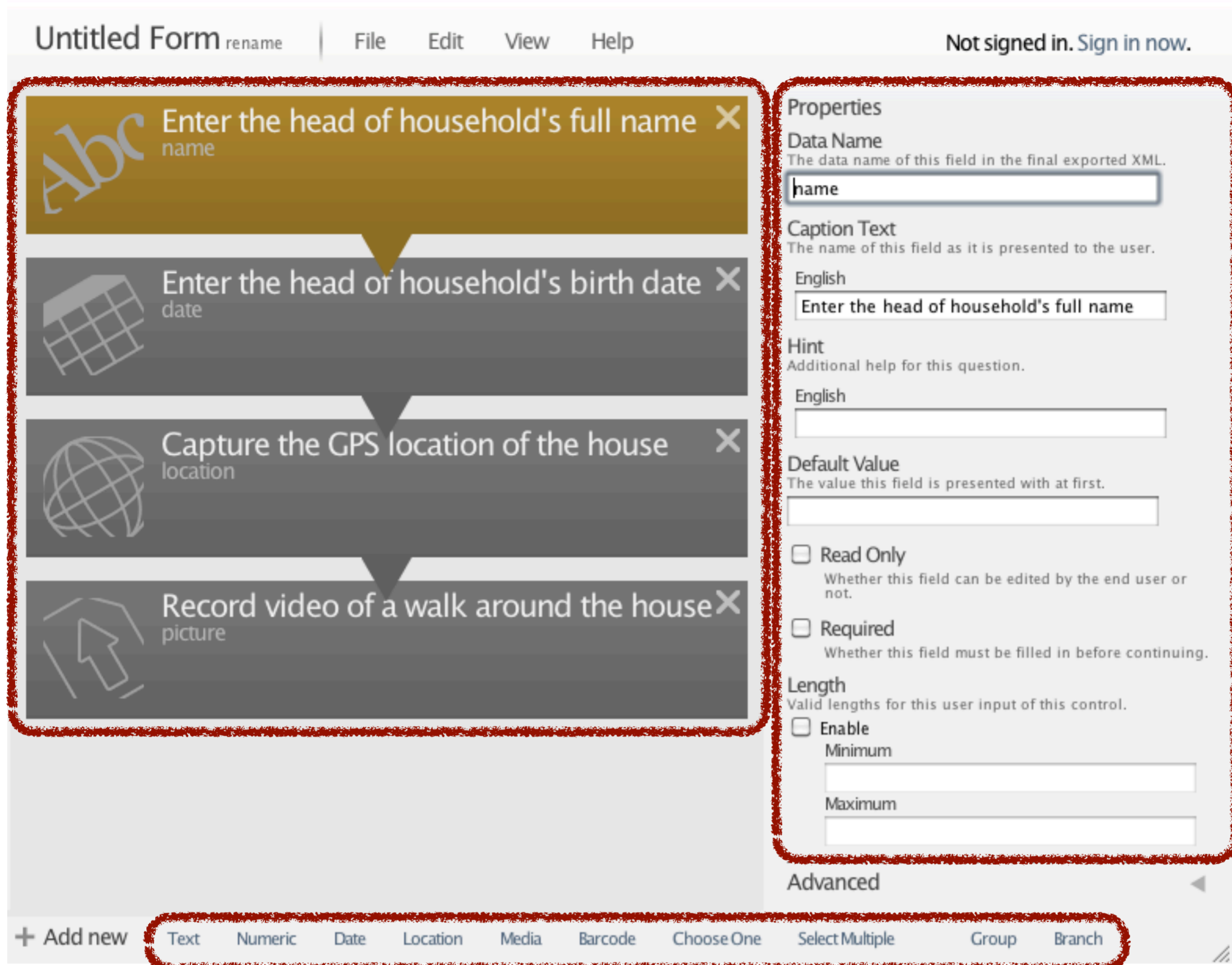
ODK provides an out-of-the-box solution for users to:

1. Build a data collection form or survey;
2. Collect the data on a mobile device and send it to a server; and
3. Aggregate the collected data on a server and extract it in useful formats.

If there is one thing to take away from the talk this is it. If you were planning on taking a nap during the talk, you may now do so guilt-free.

Anyway, so what are these tools that you use to Build, Collect and Aggregate?

Build: Drag and drop prompts for form creation.



ODK Build is an web application where you drag and drop prompts to create forms.

It runs in the browser but can also be used offline.

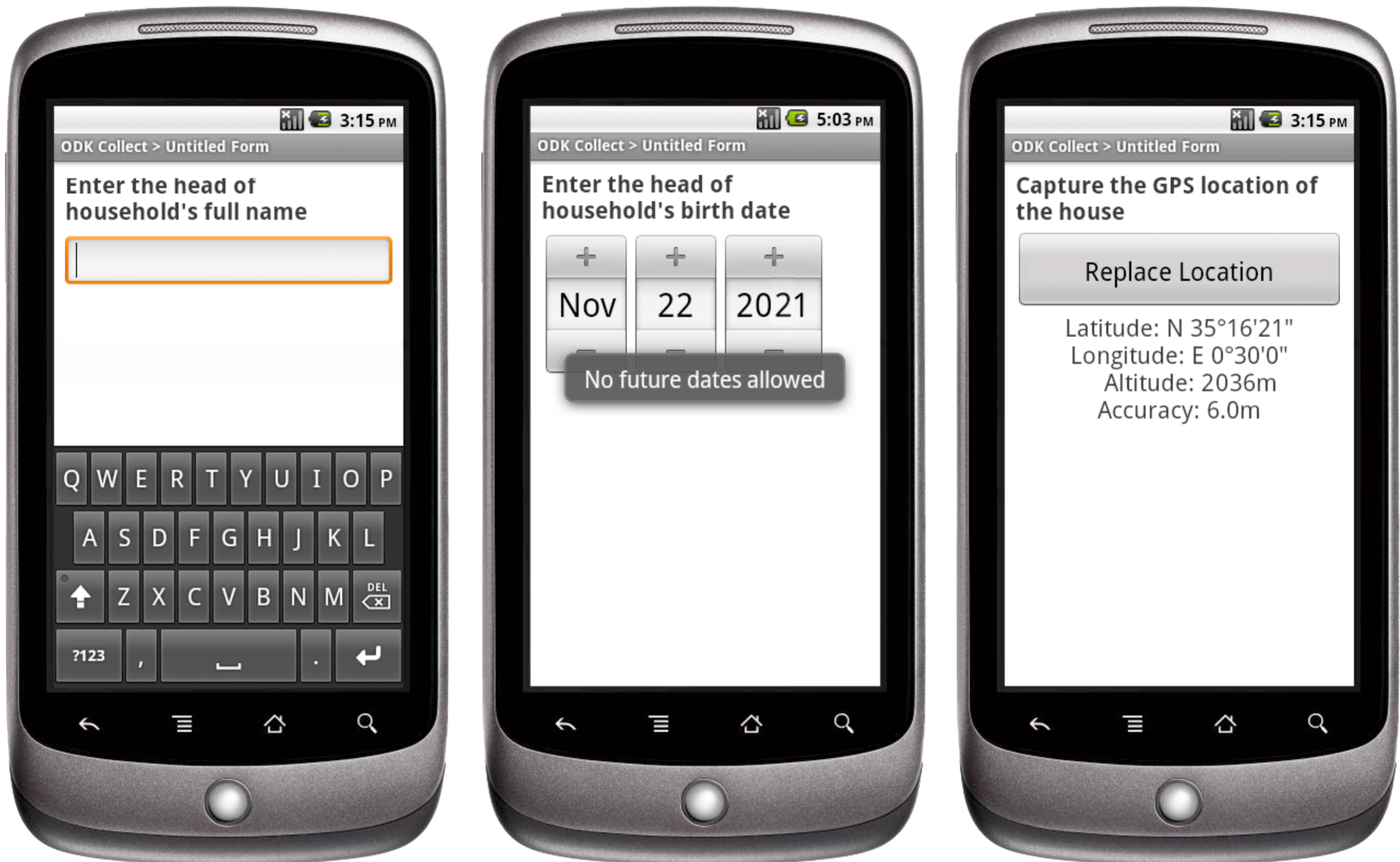
To design a form, you drag and drop each prompt the user will interact with from this button pane to the canvas.

Each prompt has a set of properties which users can edit here.

The prompts are pretty powerful. For example, you can have multiple languages for each prompt, make prompts relevant based on previous answers, or make prompts loop some number of times.

When you are done, that's what goes to the phone.

Collect: Display prompts for data collection.



ODK Collect is an app that display prompts to collect and deliver different types of data.

Collect runs on the Android operating system so you can use it on phones, tablets and netbooks.

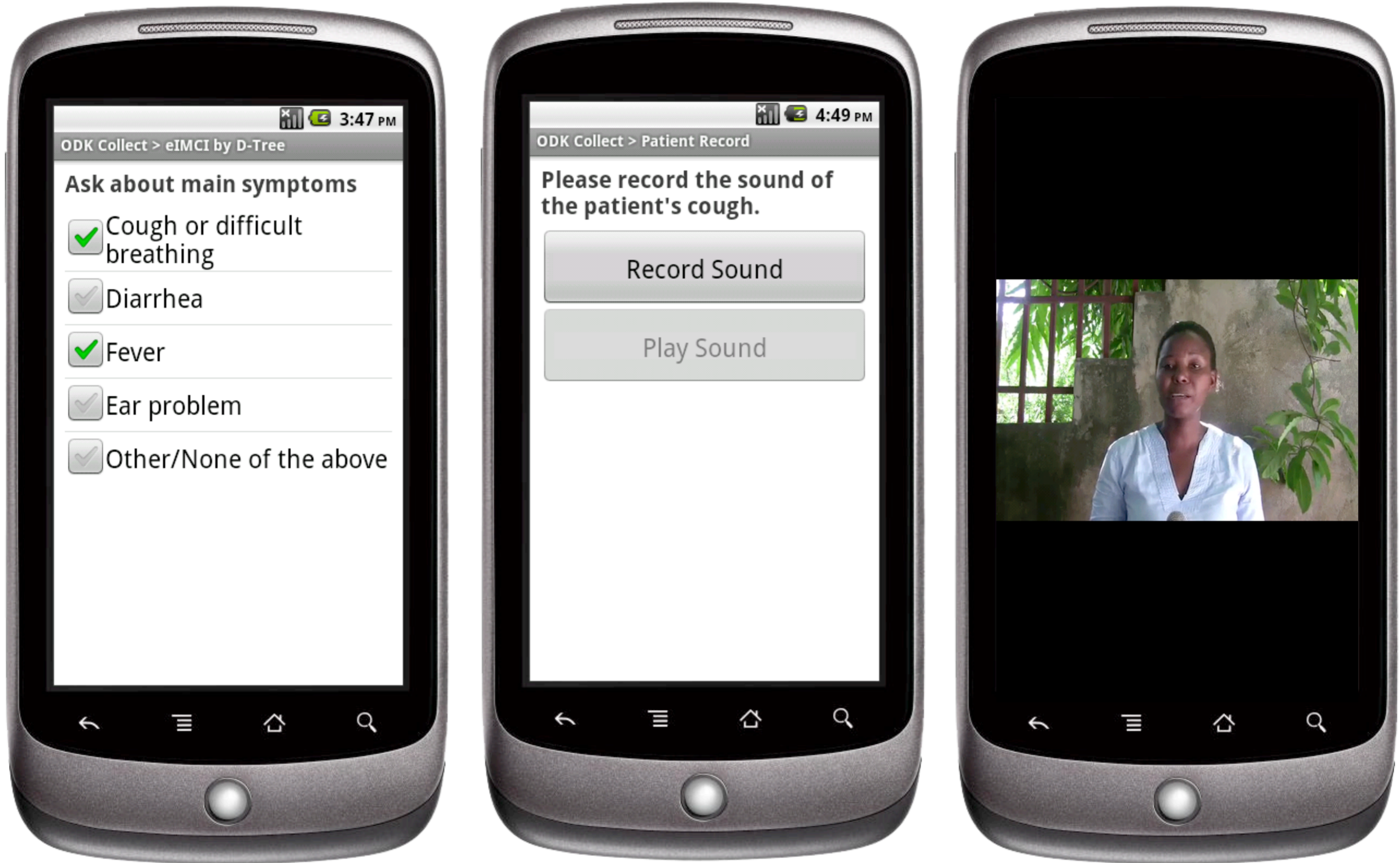
It runs great offline you can collect text, numbers, dates, and you can put constraints on all prompts.

So for example, birth dates can't be in the future.

You can also capture GPS location using one click.

We even use the camera on the phone to scan barcodes.

Collect: Multiple languages and multimedia ability.



The forms themselves are really powerful. So here is an example of IMCI in swahili.

IMCI is the integrated management of childhood illnesses. It's basically a triage protocol for children under five.

My Swahili is weak, so let's switch to English. With ODK Collect you can do this on the fly.

And if I enter that the child is coughing and has a fever, in the next screen,

I can record the sound of the cough and then I can show the mother a video about how to treat fever.

So there is this collection of data, but also a delivery of information using complex logic.

Any data I gather can be stored offline and then be sent off to a server.

So that's ODK Collect. And where does the data go?

Aggregate: Host data and provide extraction interfaces.

The screenshot shows the ODK Aggregate web interface. At the top, there are navigation tabs: "Submissions" (selected), "Form Management", and "Site Admin". In the top right corner, there is a "Log Out" link and the email address "wbrunette@gmail.com". Below the navigation, there are buttons for "Filter Submissions" and "Exported Submissions". The main content area shows the "Form" as "Geo Tagger v2" and the "Filter" as "none". There are buttons for "Visualize" and "Export". On the left side, there are buttons for "Save", "Save As", and "Delete". Below these, there is a checkbox for "Display Metadata" and a "Submissions per page" dropdown set to "100". There is also a "Filters Applied" section with an "Add Filter" button and a "Hide DeviceId" button. The main table displays a list of submissions with the following columns: "Image", "Location Latitude", "Location Longitude", "Location Altitude", "Location Accuracy", and "Description". Each row has a red "X" icon in the "Image" column, indicating that the image is missing or failed to load. The data in the table is as follows:

Image	Location Latitude	Location Longitude	Location Altitude	Location Accuracy	Description
	47.65434975	-122.30498975	21.29999924	6.708204	HUB construction
	47.64834739	-122.29989853	-20.29999924	5.656854	Docks at WAC
	47.65335942	-122.3255423	-14.89999962	6.708204	Foot of Latona
	47.64634424	-122.33644953	-7.0999999	6.3245554	Home
	47.64540379	-122.33636588	-7.19999981	4.472136	Kite hill... Gasworks
	47.62708792	-122.33274967	-8.0	6.708204	Chandler's Cove
	47.65824883	-122.31314593	31.0	5.0	The Ave
	47.66945725	-122.30360415	25.39999962	5.0	Ravenna Park
	47.68088862	-122.3291259	48.59999847	5.0	Greenlake
	47.68999464	-122.3554331	68.59999847	7.2111025	Greenwood

ODK Aggregate hosts the submitted data and provides interfaces such as spreadsheets, maps, and queries.

We don't run one big server, you download an installer, and it configures one just for your organization.

This way you control where your data lives.

Aggregate: Stores or forwards data to external systems.



Deviceld	351676030226627
SurveyorName	Shadrack
TreeLocation-Latitude	-4.9192410707473755
TreeLocation-Longitude	29.60762321949005
TreeDBH	57.0
TreeName	Myombo
TreePicture	View

ODK Aggregate can either store or stream your data to other systems, like Google Earth.

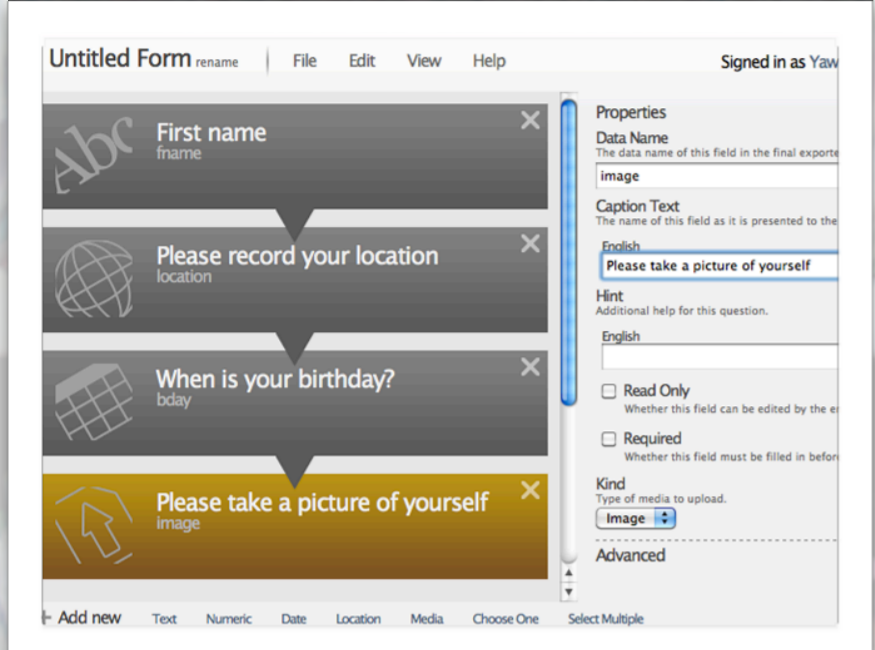
In this example, forestry workers with the Jane Goodall Institute in Tanzania, submitted data from Collect to Aggregate and then exported to Google Earth.

Managers could then click on each yellow point and get the data that was submitted.

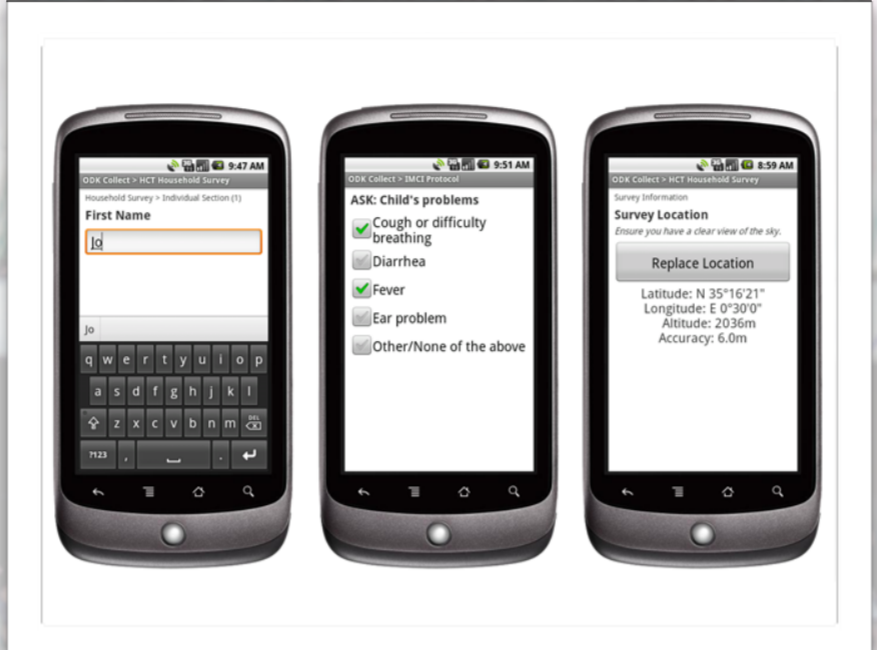
400 → 540 ↓

541 → 700 ↓

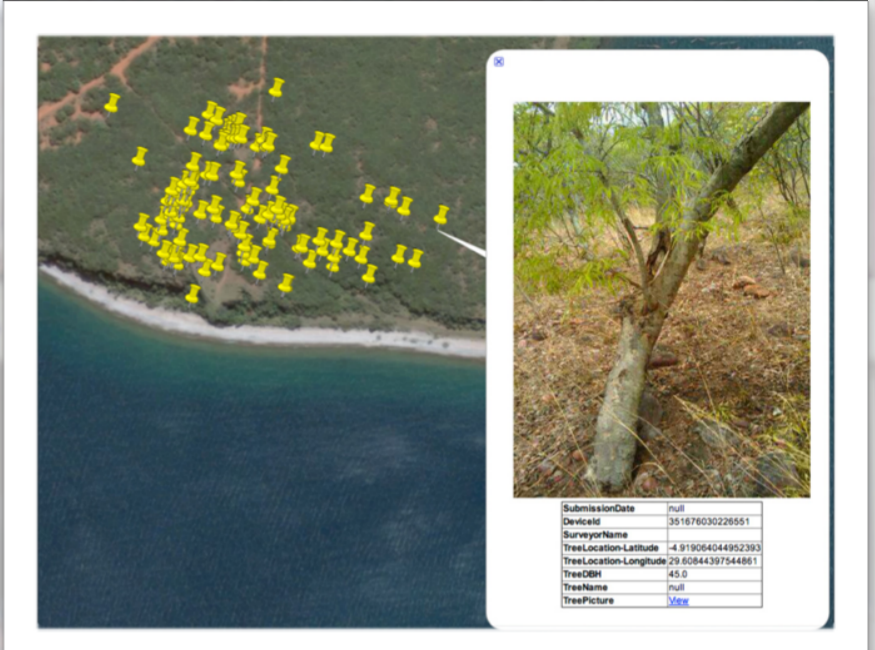
1. Build form



2. Collect data



3. Aggregate results



1371 → 1540 ↓

1541 → 1730 ↓

So that's what ODK is. A platform that will help you build a form, collect data, and aggregate some results.

While we provide an end to end solution, everything is designed to be modular. You can pick and choose which components you use. That's why it is called a kit.

For example, Modi Labs at the University of Columbia has some really great ODK-compatible tools that you can use instead of the ones I just showed you.

We'll start with building forms.

XLSForm: Build and share forms with Excel

type	name	label	hint	constraint
		7. What is your occupation?	[Classify respondent on occupation. If other, list occupation stated] Researchers: please make sure not all selected respondents belong to the same occupational category (if at all possible).	
select_one occupation	a7			
text	a7_working_class	Working class? Please specify.		selected(\${a7},'8')
text	a7_other	Other? Please specify.		selected(\${a7},'66')
end group	a7_group			
begin group	a8_group			
		8. Please could you tell me your own education level?	Researchers: please make sure not all selected respondents belong to the same educational level (if at all possible).	
select_one educational_level	a8			
integer	a8_primary	Last year of primary school?		selected(\${a8},'2') and . >= 1899 and
integer	a8_secondary	Last year of secondary school?		selected(\${a8},'4') and . >= 1899 and
text	a8_details	Further studies? Please specify.		selected(\${a8},'6')
end group	a8_group			
select_one yes_no	a9	9. Do you have access to a radio within your home?		
begin group	a10_group			
select_one listen_radio	a10	10. How often do you listen to the radio?		
text	a10_other	Other? Please specify.		selected(\${a10},'66')
end group	a10_group			

So for form design. There is XLSForm, which you can use to build forms with Excel.

So it's like ODK Build, but instead of drag and drop, you just enter them into a spreadsheet.

Each row is a prompt the user sees.

So this prompt's data type is a number or an integer.

And the wording the user sees is asking for the last year of primary school.

And you can see some of the validation constraints here. It is making sure the year is greater than or equal to 1899.

It's not as easy to use as ODK Build for beginners, but for long forms and for users familiar with Excel, it's a great alternative.

So that's XLSForm.

enketo: Collect data offline in the browser.

Makubaliano iliyo toka kwenyi utafiti inayo husu Idadi Ofisini S.E.L.F. * ⓘ

Ok. Tafadhali kuendelea.

A. Utafiti

1. Namba ya Mulengwa *

2. Siku ya kujibu ku maswali ama maulizo

Utafiti kuhusu wakaaji

5. Una umri gani? * ⓘ

Mwisho wa utafiti

Namba ya mutafiti *

There is also enketo, which you can use to collect data in a web browser.

So it's like ODK Collect, but you can use it on your iPhone with Safari or your laptop with Internet Explorer.

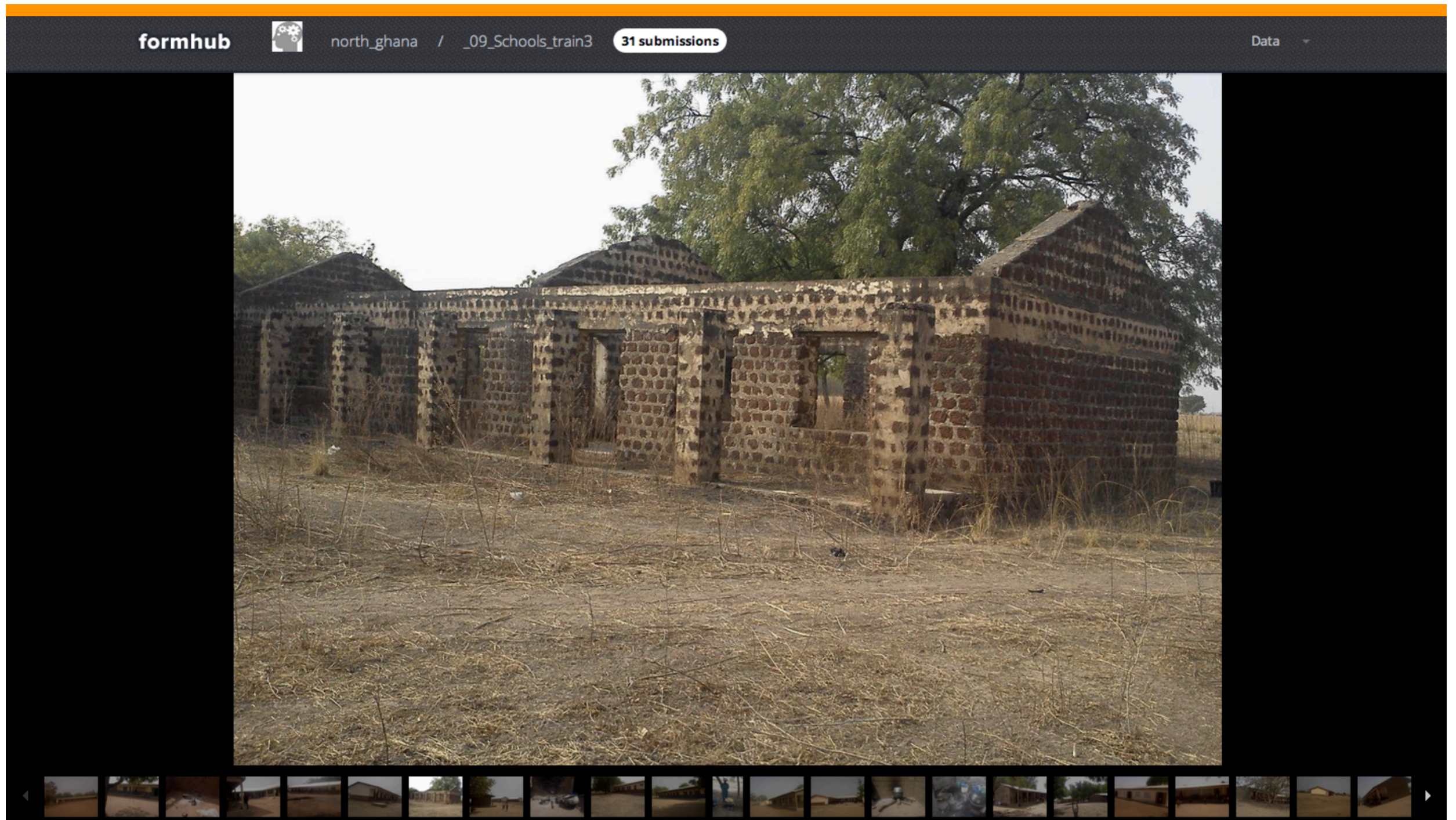
It works great offline, it has all the branching and validation that ODK Collect has, and you can even change languages on the fly.

For example from English to Swahili.

It doesn't have all the multimedia features that ODK Collect has, but if you, for example, have data entry clerks who want to enter a backlog of paper forms while you use phones to collect new data, then it's perfect.

With enketo, all the data will go to the same place!

Formhub: Aggregate and visualize results.



A picture can say a thousand words. View the photos taken in a photo gallery for each survey.

The last tool from Columbia I want to show is called Formhub. It's a really nice alternative to our ODK Aggregate server.

It has a totally different look and feel. For example, it has these great visualizations like automatically showing you where your data was submitted from.

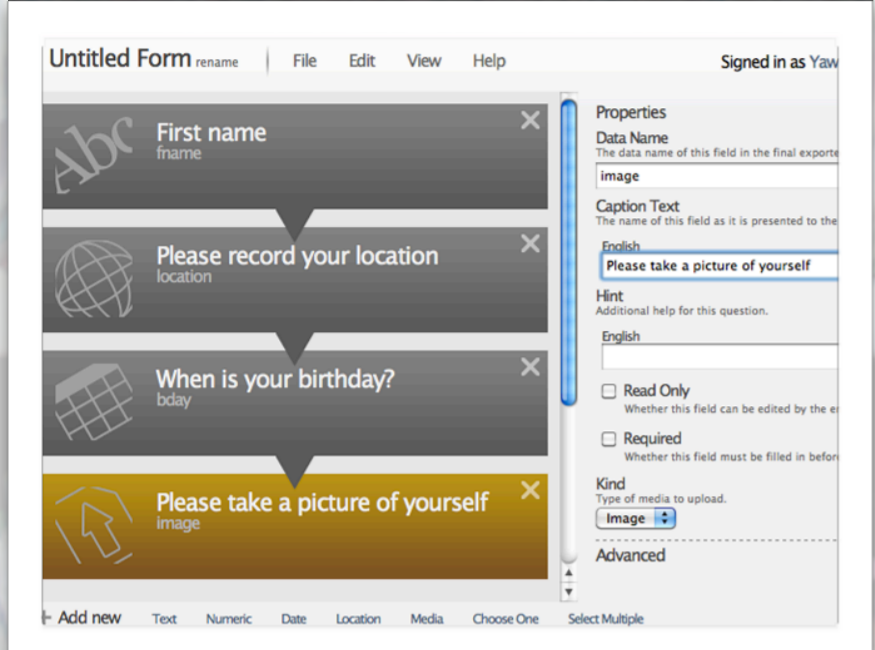
And they also have really amazing photo browser.

The key difference here is that there is only one FormHub website. So if for example the MOH wants you to run the server in country, or you want to control where the data goes, you might be better off using something like ODK Aggregate.

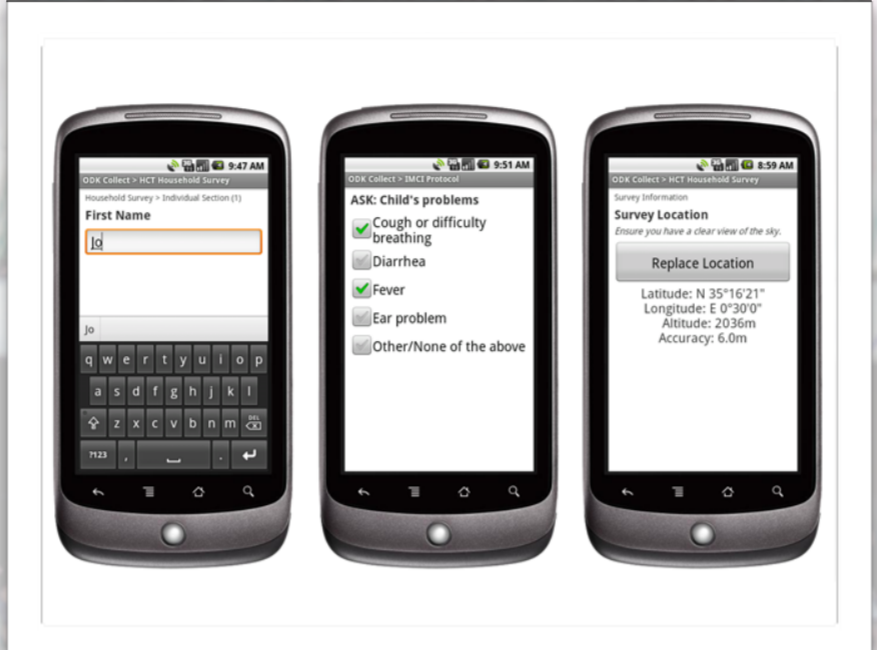
400 → 540 ↓

541 → 700 ↓

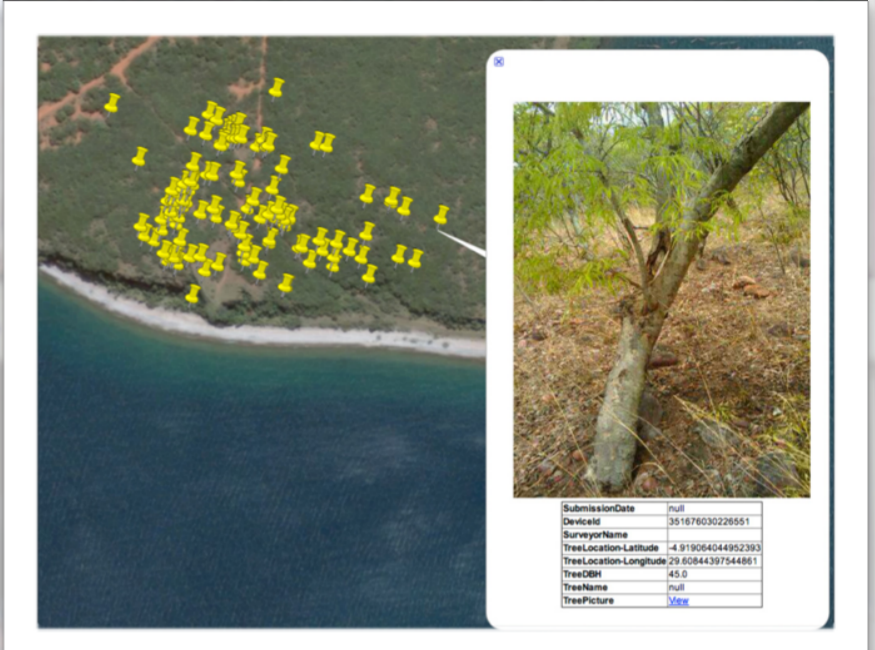
1. Build form



2. Collect data



3. Aggregate results



1371 → 1540 ↓

1541 → 1730 ↓

So hopefully that gives you a sense of what we mean by a kit.

- If you are an Excel expert, use XLSForm.
- If you need to collect data in a browser, use enketo.
- If you don't like ODK Aggregate, try FormHub.

They are all free. They all work together seamlessly together. You can pick and choose individual tools and you are never locked in.

Now that's nice and all, but when people find out ODK came out of a research project, and that everything is free and open source, they start asking if the software actually works.

Well, yeah! ODK is used by tens of thousands of people who have collected millions of forms.

Let me give you some examples of a few projects.

Jane Goodall Institute uses ODK for forest monitoring.



The Jane Goodall Institute uses Open Data Kit on Android phones for their REDD+ forest monitoring projects in Tanzania and Uganda.

This project is in partnership with the Google Earth team.

In these communities, trees are chopped down to be sold or used as firewood. This is a problem because if there are no trees, then there are no chimps.

With ODK, JGI can empower communities to transparently document the health of their forests.

And because the process is transparent and efficient and near-real-time, the community can then show the data on Google Earth as the proof they need to sell carbon credits and get income.

They no longer need to sell the wood.

The ODK team helped JGI with their first few weeks of deployment.

LOAN OVERVIEW

REPAYMENT SCHEDULE

LENDER COMMENTS



PAYING BACK

This loan has been fully funded!

A loan of \$375 helped Ruth buy irrigation pipes.

16% repaid

[Find a Loan](#)

<u>Repayment Term:</u>	18 months (more)
<u>Repayment Schedule:</u>	Irregularly
<u>Pre-Disbursed:</u>	Aug 30, 2012
<u>Listed</u>	Sep 7, 2012
<u>Currency Exchange Loss:</u>	Possible
<u>Default Protection</u>	Not Covered

Kiva allows you and I to give a small loan to an entrepreneur anywhere in the world. They do this through a network of micro finance groups that administer these loans.

The key to convincing people to loan money is providing lots of information about the entrepreneur and updates about whatever it is their doing.

Some of Kiva's partners use ODK to make the process of gathering all this data easier.

With the phone, they can take pictures, track loan status and get updates to the Kiva website quickly and accurately.

The ODK team was not involved at all with this project.

Carter Center uses ODK for monitoring elections.



The Carter Center uses ODK for their election monitoring. They recently used it in Egypt.

With ODK, observations from polling stations across a country to be transmitted to headquarters immediately, allowing a richer picture of an election to emerge in real time.

And that real time is key to being able to determine quicker if an election is credible.

The ODK team was not involved at all with this project.

Carbon For Water collected over 1,000,000 forms with 4,000 ODK-powered phones in 6 weeks.



Carbon For Water wanted to distribute LifeStraw Family water treatment units to 4 million people.

And as part of that distribution they want to take images and location of each barcoded unit, along with some socio-economic data.

To do this, they used 4,000 ODK-enabled phones to collect over a million forms in six weeks. Which is incredible.

They've found that ODK enables real-time collection of image, GPS and survey data at scale.

Interesting side note, one of this project's members is an astronaut. And apparently they have the internet in space now, so he was using ODK on the space station. Pretty impressive.

Even more impressive, the ODK team was not involved at all with this project. These guys found ODK and deployed it without our help. We found out about the project after they finished.

AMPATH's health workers have used ODK to counsel and test over 775,000 people for HIV.



AMPATH is one of the largest HIV treatment programs in Africa. They have about 2 million people in their catchment area and are treating 130k HIV positive people.

As part of their care, they have 300 community health workers who go house to house and do home-based counseling and testing for HIV.

These guys are equipped with ODK-powered smartphones. The phone's logic helps them collect socio-economic data and also decide when to do tests for HIV or TB. All the data gets sent to their OpenMRS medical record system.

This image shows one of those workers using the camera on the phone to identify a patient by scanning a barcode.

We helped these guys pilot the system for two weeks, then they took over. The project started in early 2010 and as of a few months ago they had visited 775k individuals.

Users found the system easy to use and thought it allowed them to collect higher quality data but at much lower cost (compared to PDAs and paper).

Electronic collection also facilitated earlier reporting, which means more people get on treatment much faster.

The ODK team helped AMPATH with their first few weeks of deployment, but the project has been running now for 3 years without our help.

Reproductive Health Vouchers

**Helping to save the lives of
thousands of mothers and of babies through
access to quality health care.**

I want to use RHVouchers as a last example.

A reproductive health voucher is a card that entitles a poor, pregnant woman to subsidized maternity care.

These vouchers are used around the world, but do they actually help improve the quality of care?

The Population Council and Marie Stopes are evaluating voucher programs in Kenya, Uganda, Tanzania, Bangladesh and Cambodia to find out.

As one could expect, these evaluations are an intensive process of collecting, organizing, cleaning, and analyzing data.

The process happens over largely rural areas with vast distances between households and health service providers and these guys put together a video about how they use ODK.

<http://vimeo.com/38123850>

The ODK team was not involved at all with this project or this video.

So hopefully that gives you a sense of what ODK is like today.

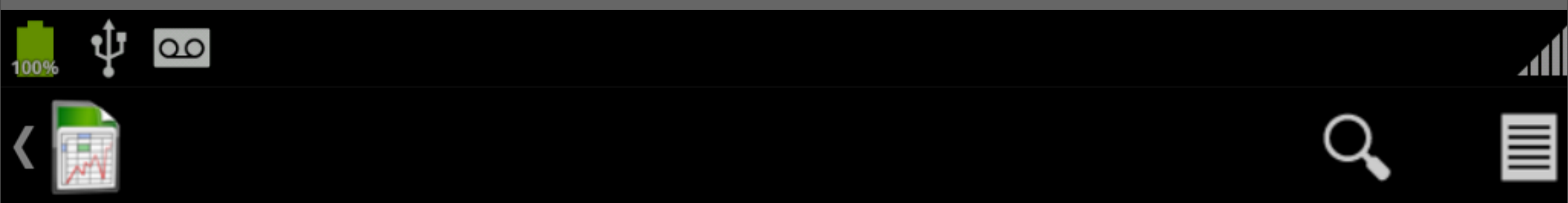
We think the platform is great for anyone who wants to collect data accurately, quickly, offline and at scale. And five years in, we have lots of evidence from tens of thousands of people and millions of forms that it works.

But what does the future of ODK look like?

I wanted to give you all a sneak peak at two of our newest tools in active development by the UW team.

The first is called ODK Tables, and the second, ODK Scan.

ODK Tables: Find, update, share data across devices.



Equipment_ID	Facility_Code	Type	Library_ID	Model	Manufacturer	Serial_Number	Power_Source	R
R-000063	02-30	CREK	E391M	RCW 50 EK	DOMETIC		EK	N
R-000071	02-33	CREK	E391M	RCW 50 EK	DOMETIC		EK	N
R-000077	02-35	CREK	E391M	RCW 50 EK	DOMETIC		EK	N
R-000085	02-39	CREK	E391M	RCW 50 EK	DOMETIC		EK	N
R-000087	02-39	CREK	E391M	RCW 50 EK	DOMETIC		EK	N
R-000088	02-39	CREK	E391M	RCW 50 EK	DOMETIC		EK	N
R-000089	02-40	CREK	E391M	RCW 50 EK	DOMETIC		EK	N
R-000091	02-41	CREK	E391M	RCW 50 EK	DOMETIC		EK	N
R-000096	02-44	CREK	E391M	RCW 50 EK	DOMETIC		EK	N
R-000098	02-45	CREK	E391M	RCW 50 EK	DOMETIC		EK	N
R-000196	03-1	CFAC	E397M	MF 214	Vestfrost		E	R
R-000202	03-5	CFAC	E397M	MF 214	Vestfrost		E	R
R-000229	03-19	CFAC	E397M	MF 214	Vestfrost		E	R
R-000293	03-64	CFAC	E397M	MF 214	Vestfrost		E	R

ODK Tables is a great way to find, update, share data across devices.

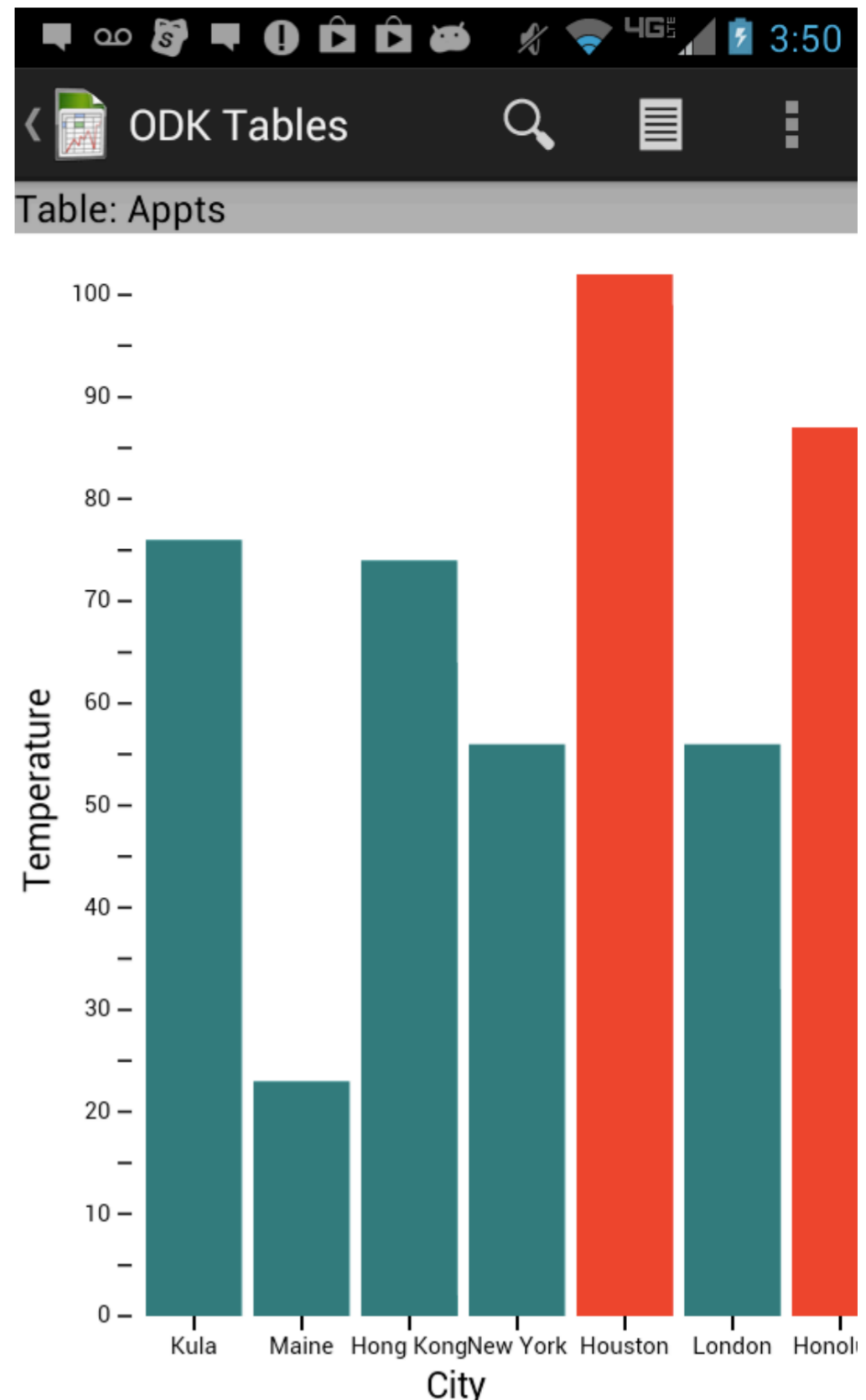
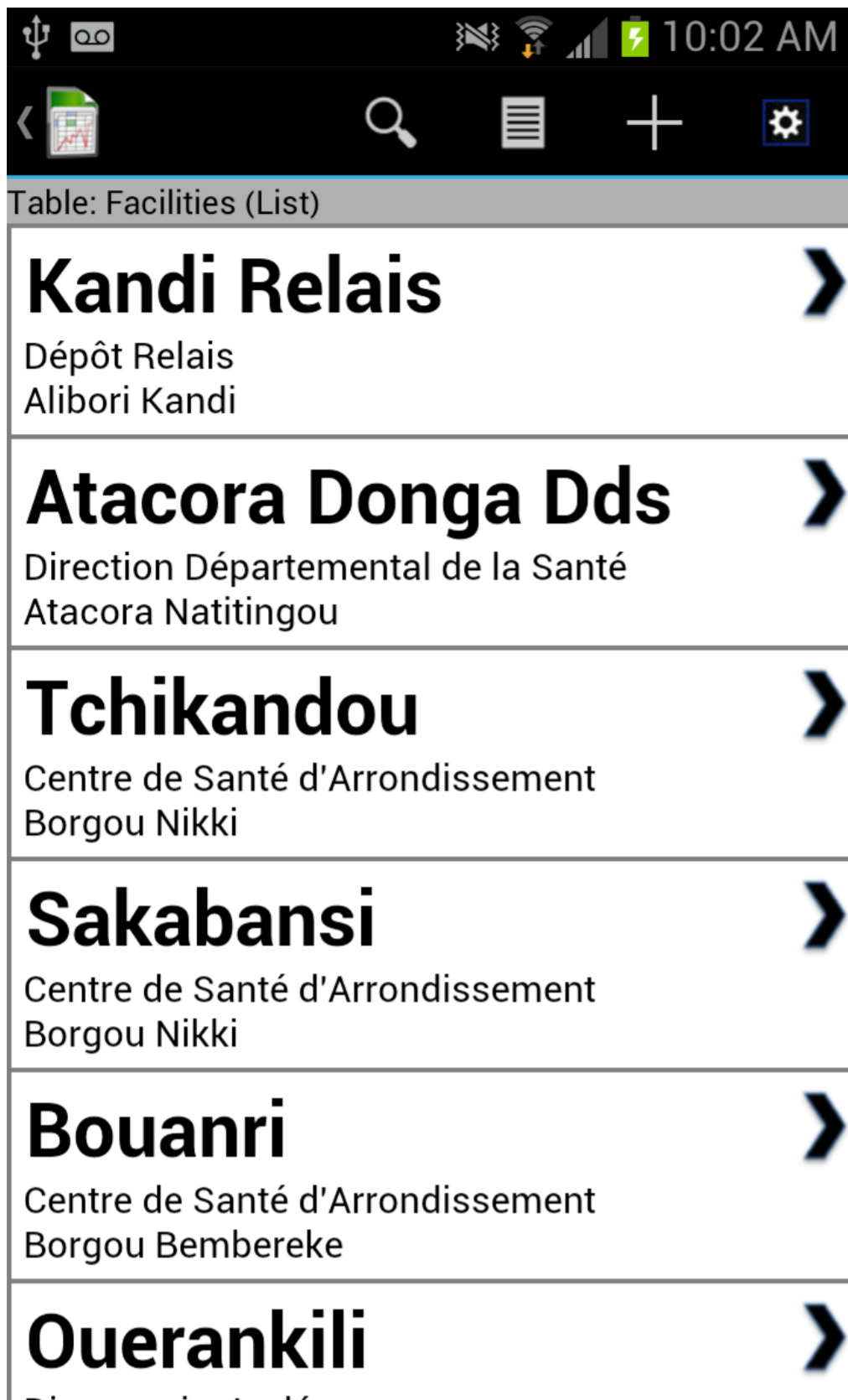
You can think of it like a big shared Access or Excel file that works everywhere. And it's a very powerful way for groups of people to collaborate data.

This screenshot is an example of a refrigerator inventory for a vaccine cold-chain for PATH.

So this data get synchronized to a server and to all the other devices. And it's designed to be mobile first, so you can do all your searching, curation, sharing right on the device.

And you don't just have table views. You can customize the views of the data.

ODK Tables: Generate easy to understand views on device.



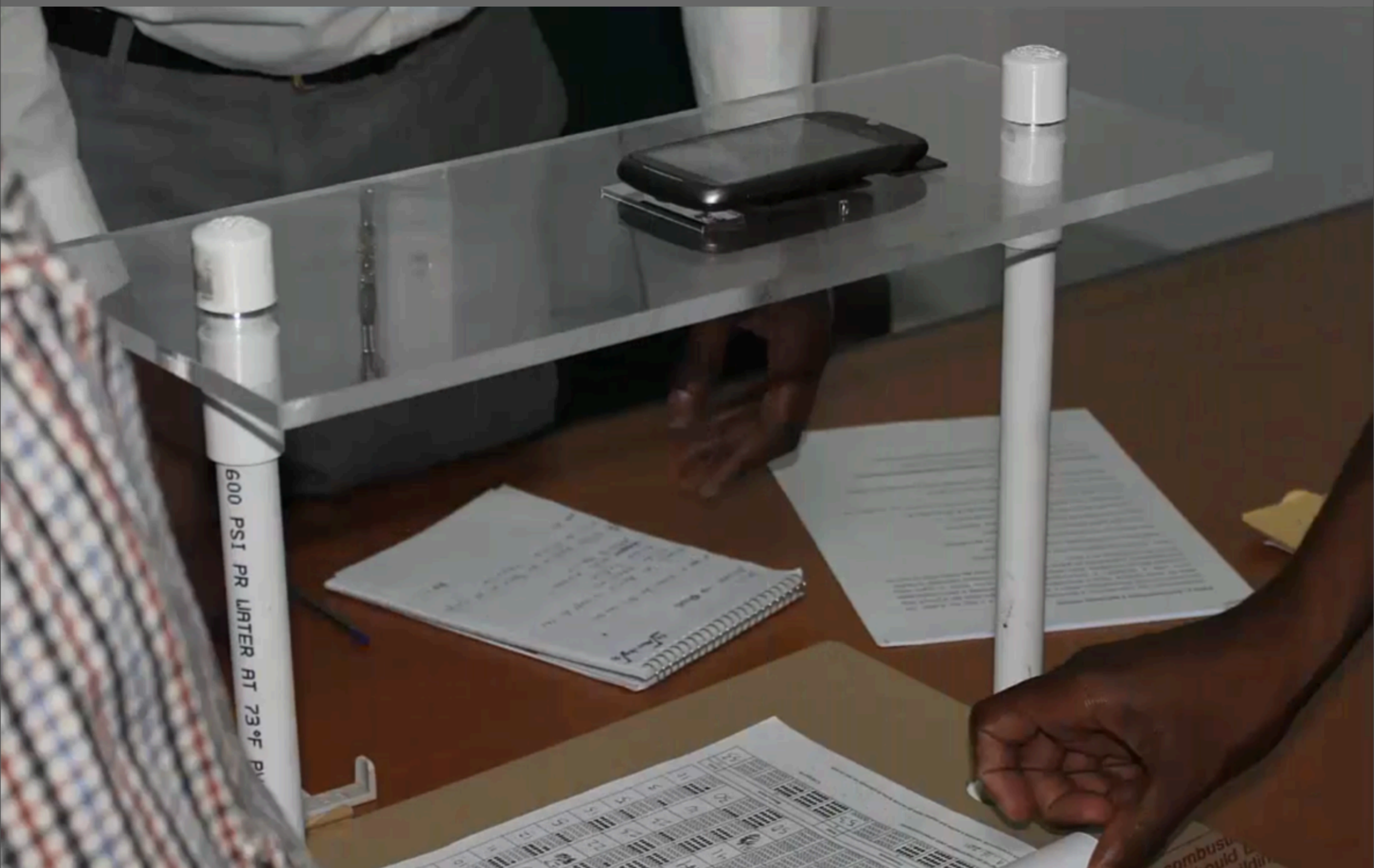
For example, here are a list of facilities. If you tap on one, you get all the information about that facility.

You can also do visualizations. So here are some recorded temperatures entered on the spreadsheet that then get shown on the device.

This makes it so much easier to do data collection where a bunch of people are updating data cross time and space. Everything from logistics to patient records.

So that's ODK Tables. This is work done by Waylon Brunette and Sam Sudar with PATH.

ODK Scan: Digitizing data from paper forms.



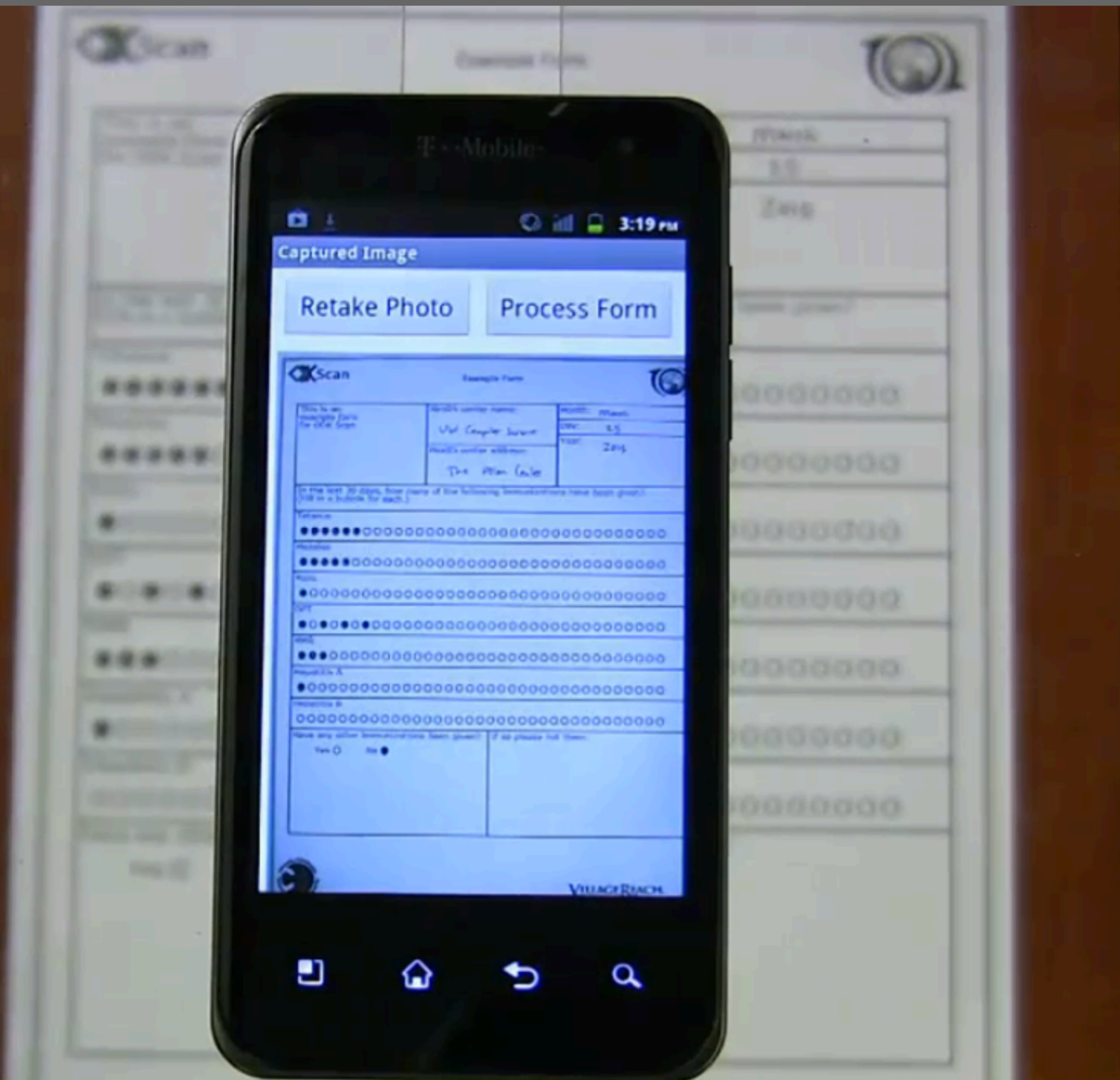
We know that many people already have paper forms and aren't ready to stop using them. Can we at least help to digitize these forms in the field?

That's what ODK Scan is about.

Turns out with the camera on the phone you can. We take a picture of the form and run a vision algorithm on it.

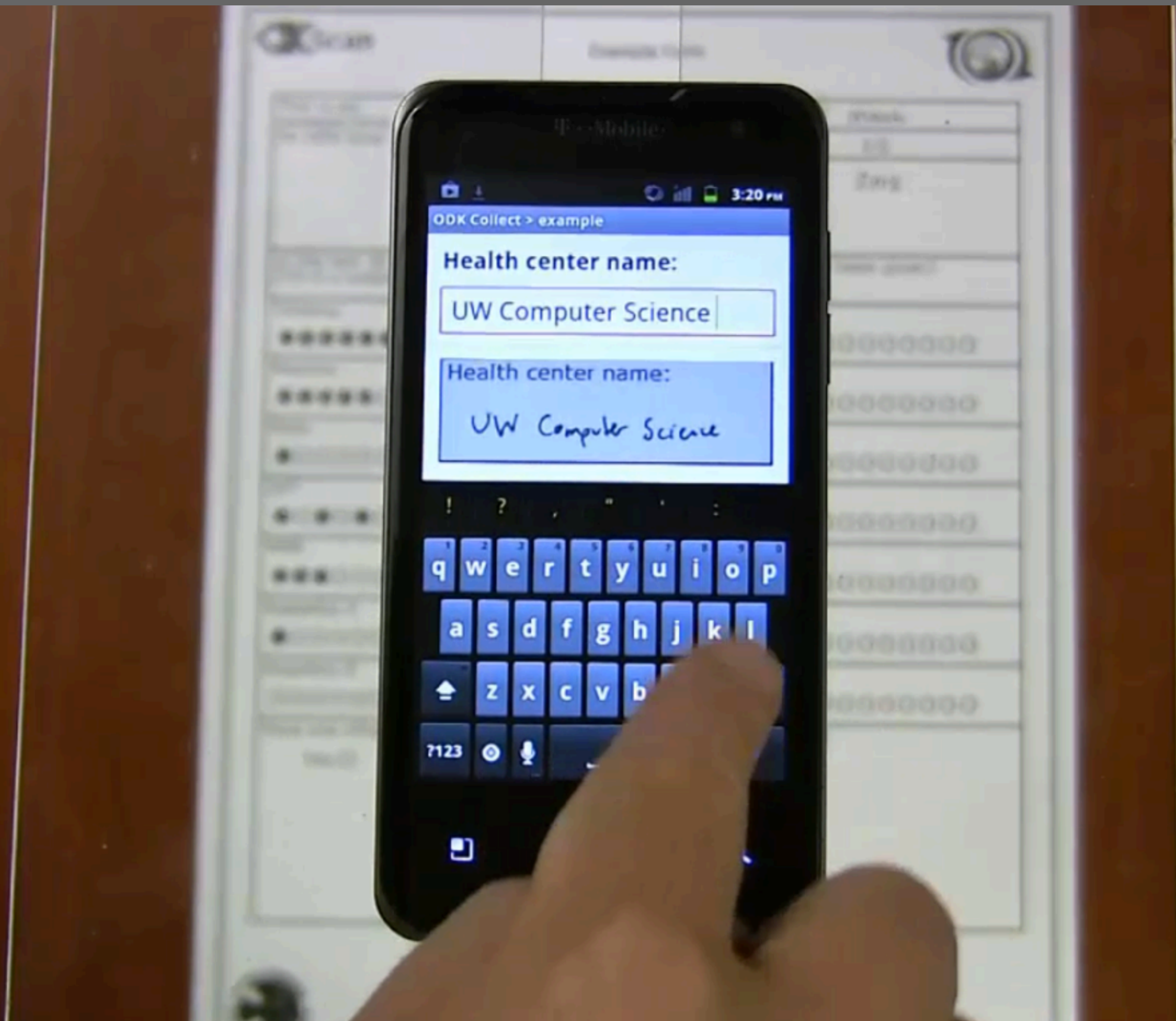
It looks like this.

ODK Scan: Over 99% accuracy on checkboxes, radio buttons



To use scan, you take a picture of the form. Then the phone processes the form and automatically detects things like radio buttons and checkboxes. It does this with 99% accuracy.

ODK Scan: UI for transcription of handwritten data.



For handwritten text that the phone can't process, ODK Scan takes the pictures of those sections, and the user can transcribe it right there on the phone. Or do it later, but either way you have a copy of the image.

You could also upload the images to the internet and get a crowd of volunteers to transcribe it. Because the form is cut up into many pieces, you can also preserve form privacy this way.

So that's ODK Scan. This is work done by Nicki Dell and Nathan Breit with Village Reach. It's being used in Mozambique.

Growing number of companies around the platform.

Looking for a company to build a form or customize an ODK tool?

- **Nafundi** - A consulting company started by the creators of the ODK project. Nafundi provides custom development, professional support, form design, implementation advice, and user training for ODK.
- **Dimagi** - Provides support for core and custom development on the Open Data Kit platform. They work closely with team that founded ODK.
- **Mindflow Associates** - Have experience customizing OpenMRS, ODK data collection solutions, building Android phone based applications, utilizing JavaRosa-based technologies amongst many others.
- **Afris** - Has developed and deployed end to end (mobile, touch-screen kiosks, SMS, desktop, web) software solutions across all provinces of Mozambique. They are quite familiar with JavaRosa-based technology.
- **Mega Six Solutions** - Based in Kenya, and specializing in Research and IT domains, we have expertise in data management processes. Using our Android development team, Mega Six will help you customize ODK for desired data collection and aggregation.
- **ULevel** - A company based in Brazil doing customizations on Collect for Android.
- **WebFirst** - Well versed in the development of data collection systems like JavaRosa and ODK in low-resource and standard environments.
- **EarlySail** - Have customized ODK Collect for a few pilot projects as well as assisted in form development. They will soon be porting ODK to BlackBerry.
- **Seeing Swans** - A company focusing on data collection and complex analytics. They have customized

ODK is more than some code, a website, and a mailing list.

The project is a community and we've spent a lot of time investing and growing in that community and it's paid off.

There are now companies that build entire products around ODK and that's great for spreading the word about ODK.

There are also a lot of consulting companies around the platform -- many of them in developing countries.



NAFUNDI

And now that I have graduated, I've jumped into the fray with Nafundi, a company I started about a year ago with Carl Hartung, one of my co-founders on ODK.

Nafundi is a software company with expertise in building solutions for challenging environments -- places with unreliable connectivity, no power, novice users, etc.

We have software running everywhere from the mountains of Afghanistan to the backcountry of Montana.

Half of our business is in providing services around ODK. Everything from form design, to implementation support, and user training. We are usually brought on as projects start or as they reach scale.

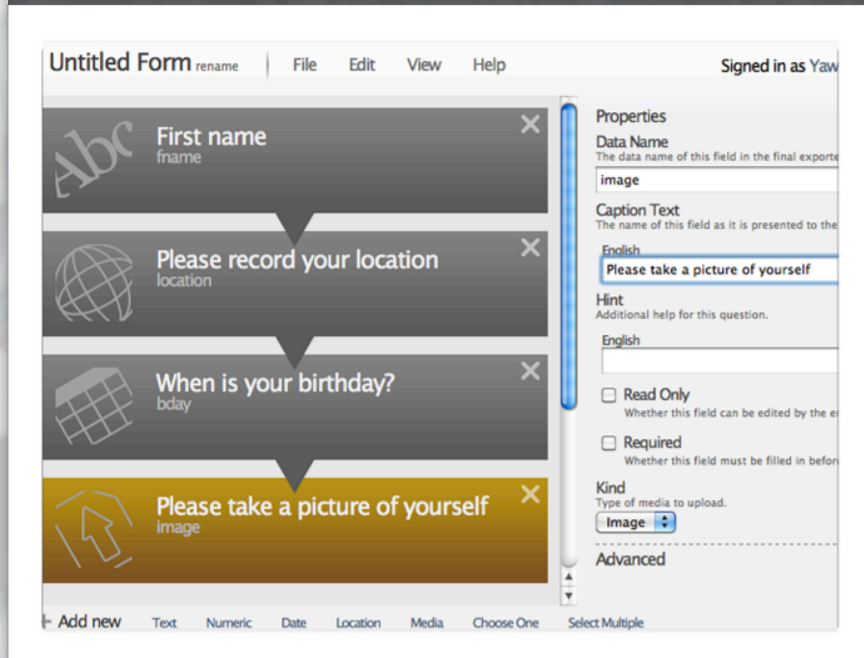
The other half is building mobile and server apps that will work well in these environments. We work with everyone from large tech companies that everyone knows to very small NGOs that you've never heard of.

Our experience on ODK helps us build software that just works.

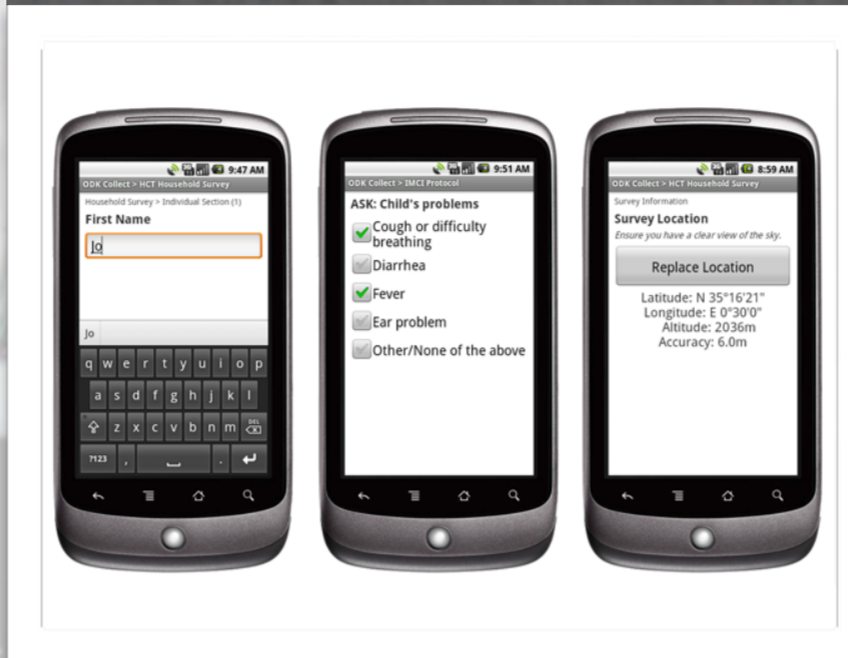
So with that, let me just summarize the key take away points about ODK and take some questions...

Collecting data accurately and quickly on paper is difficult.
ODK uses phones and servers to digitize data collection.

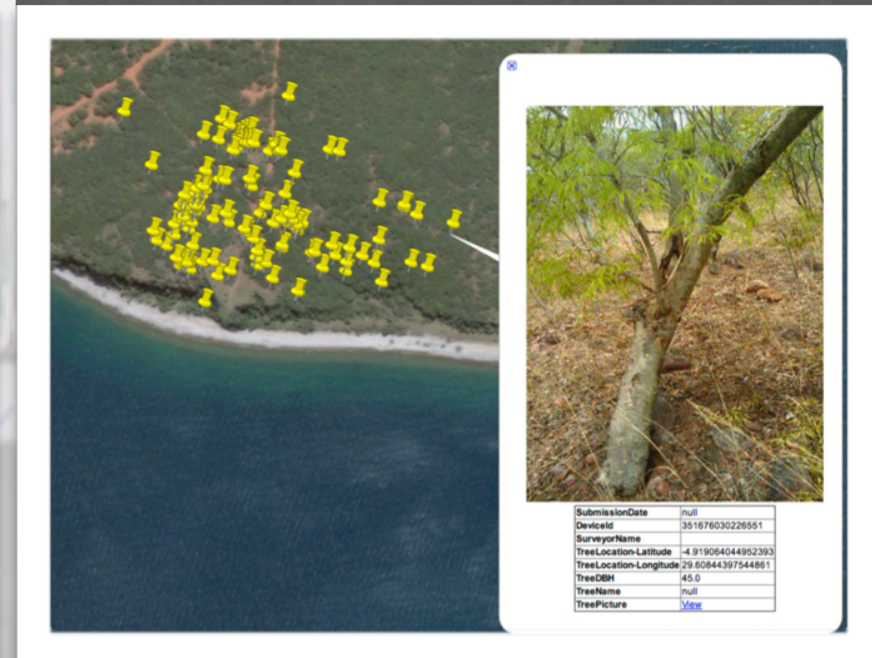
1. Build form



2. Collect data



3. Aggregate results



<http://opendatakit.org> (@opendatakit)

<http://nafundi.com> (@nafundi)

<http://anokwa.com> (@yanokwa)

Using paper to collect data is difficult and inefficient. ODK is a free and open-source set of tools which use smartphones and cloud servers to digitize data collection.

ODK provides an out-of-the-box solution for users to:

1. Build a data collection form or survey;
2. Collect the data on a mobile device and send it to a server; and
3. Aggregate the collected data on a server and extract it in useful formats.

So with that, are there are any questions I could answer? (e.g., how is this free, what about security, what other tools are there?)