

Data Collection II

Lecture 19: CSE 490c



OPEN DATA KIT

Announcements



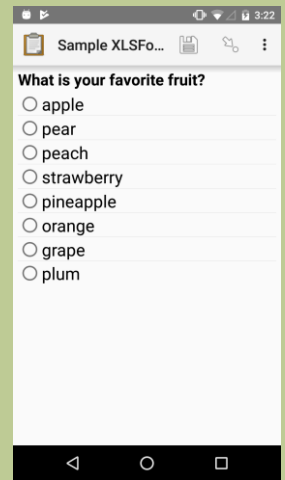
- New Homework Available
 - Programming Assignment 3, Low-literate UI for UW-Pesa, Due Wednesday, November 21
 - Homework 6, Icon Design for PA3, Due Wednesday, November 14

Topics

- Data Collection
- Open Data Kit
- Data Integrity
- Security
- ODK 2.0

Open Data Kit

- ODK 1.0: Forms based data collection
- ODK Collect: Android App, ODK Aggregate: Server



Sample XLSFo... 3:22

What is your favorite fruit?

apple

pear

peach

strawberry

pineapple

orange

grape

plum



- ODK 1.0 widely adopted by researchers and NGOs
 - Covers well defined use case of forms based data collection
 - First product to market
 - Established ecosystem and user community
 - Open source so now software costs or licensing

Subtitle:

Article type: Research Article

Authors: Koczela, Steve^{a,*} | Furlong, Cathy^b | McCarthy, Jaki^c | Mushtaq, Ali^d

Data Integrity

- Traditional problem with surveyors
 - What if enumerators cheat
 - This even has a name: curbstoning
- Data collectors make up data instead of doing surveys
- Methods for detection
 - Made up data often is not random enough
 - Consistent omission of data
 - E.g., Missing informant phone numbers
 - Made up data may not have appropriate means
 - E.g., Across sample, 40% of households might be away, while faker only identifies 20% of households being away

Tools for detecting bad data

- People are generally bad at making up random data
 - May contain patterns
 - May have non-random statistical properties
- Compare distribution of each collector's value with composite of other collectors
 - Multinomial means and variances
 - Very accurate if number of fakers is low
- Big brother
 - Record question times
 - Record GPS locations

Security

- Threat Model
 - CIA Goals: Confidentiality, Integrity, Availability
 - Adversaries
 - Governments, Thieves, Hackers, Partners, Enumerators
 - Potential Threats
 - Unauthorized access
 - Entering fake data
 - Coercing enumerators
 - Theft
 - Legal access to data
 - Instability of application
 - Information leakage on device
 - Fake ODK applications

Interview Study

- What do users care about
 - Data Loss
 - Encryption not used, because it risked data loss
 - Integrity
 - Enumerators answering “no” to shorten interview
 - Exploited data
 - Generally less of a concern, but there are some very sensitive ODK deployments
- Importance of device management
- Different levels of technical expertise
- Ethics board considerations
- Context: Comparison with Paper

Digression: About Names

- What does the Open in ODK mean
 - Open (Data Kit) vs. (Open Data) Kit
- ODK 1.0 and ODK 2.0
 - ODK 1.0 and ODK 2.0 are different projects that address different use cases
 - Naming suggests that the latter is replacement for the former
 - Not the case

ODK Research Projects

- ODK Clinic
 - Integrate with Medical Record System
- ODK Sensors
 - Provide drivers to allow sensors as input
- ODK Submit
 - Extend communication model to support a broader range of transport mechanisms to reduce cost
- ODK Scan
 - Use the camera for paper to digital input
 - Restricted forms of input: bubble forms and numbers

Limitations of ODK 1.0

- IMHO the success of ODK 1.0 is due to its focus on a core data collection use case
- But what else did people want?
 - Greater flexibility with user interface
 - Ability to collect data based on previous data
- ODK 2.0 developed as a data management platform requiring greater sophistication in application development and deployment

Open Data Kit learnings and philosophy

- Data collection projects require technical people in charge of managing forms, data, and devices
 - Deployment Architect
- The purpose of the Data Collection Platform is to give the Deployment Architect the ability create and manage survey process
- The Deployment Architect is not an Android developer
 - ODK 1.0 Deployment Architect: Forms designer
 - ODK 2.0 Deployment Architect: Javascript

ODK 2.0 Features

- Fully customizable layout of prompts on the Android device
- More flexible, user-directed, navigation of a survey
- Bi-directional synchronization of data across devices
- Data curation and visualization on the device
- Row-level access filters

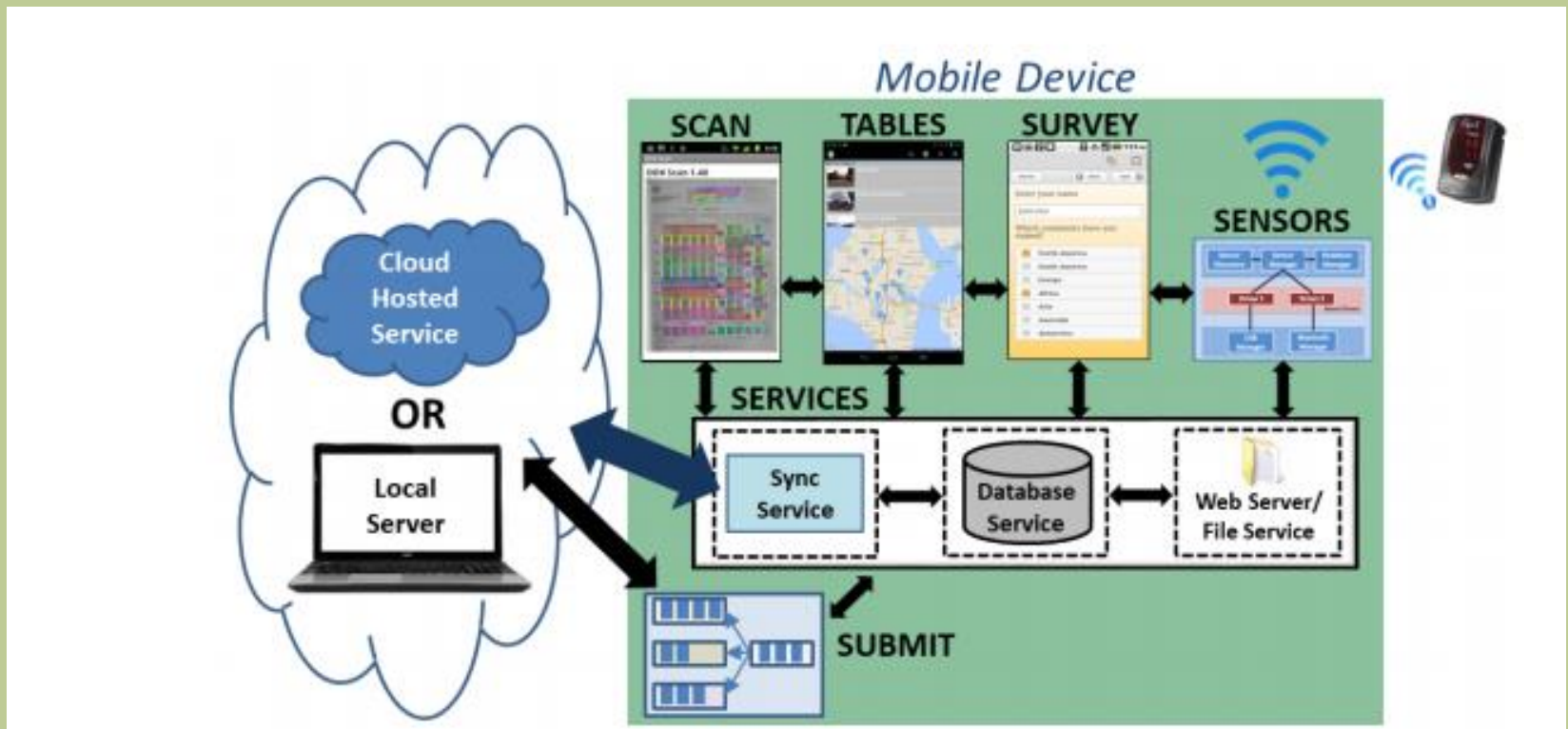
Note

The ODK 2 tool suite is targeted at advanced users who are unable to complete their workflows with the ODK 1 tools. If you find that the ODK 1 tools meet your needs then there is no reason to switch.

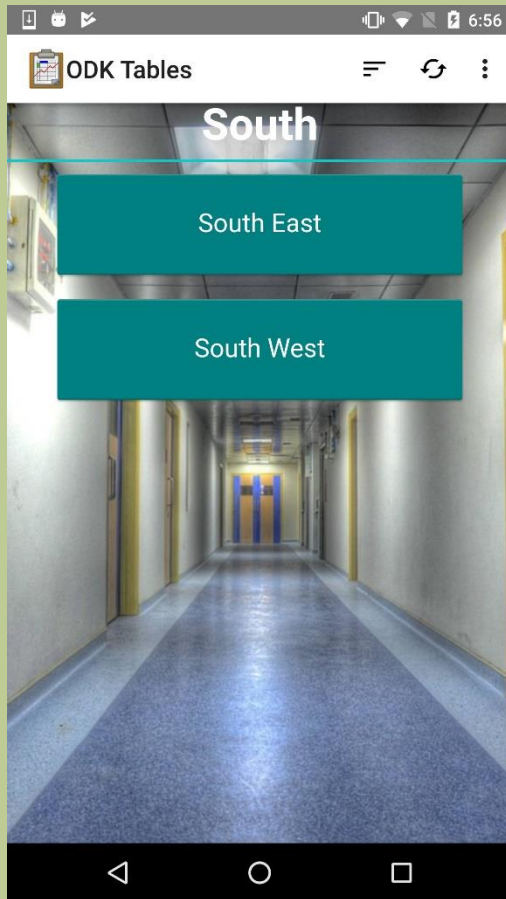
ODK 2.0

- ODK Application Designer
- ODK Survey
 - Data Collection app based on HTML, CSS, JavaScript
- ODK Tables
 - Data Curation and Visualization app
- ODK Services
 - Application for syncing data across multiple devices and with server
- ODK Cloud Endpoints

ODK 2.0 Architecture



ODK 2.0 Cold Chain Demo I



ODK 2.0 Cold Chain Demo II

ODK Tables

Chisitu Dispensary

Basic Facility Information

Health Facility ID: 1019-0025	Population: 2498
Facility Type: Health Center	Coverage: 100%
Ownership: Public	Admin Region: Mulanje

Power Information

Electricity Source: Grid	Kerosene Availability: Not Available
Grid Availability: More Than 16	Solar Suitable Climate?: Yes

ODK Tables

Stock Information

Distance To Supply Point: 10 km	Vaccine Reserve Stock Req: 1
Vaccine Supply Interval: 4	Vaccine Supply Mode: Delivered

Refrigerator Inventory (2)

Add Refrigerator

Edit Facility

ODK Tables

Refrigerators

Search

Prev 10 Showing 1-2 Of 2 Next

Refrigerator: 32951031
Catalog ID: E3108m
Health Facility: Chisitu Dispensary [Edit](#)

Refrigerator: 941077
Catalog ID: E391m
Health Facility: Chisitu Dispensary [Edit](#)

ODK 2.0 Cold Chain Demo III

ODK Tables

Refrigerator 941077

Basic Refrigerator Information

Facility: Chisitu Dispensary	Model ID: E391M
Year Installed: 2016	Refrigerator ID: R898
Status: Functioning	Voltage Regulator? Unknown
Reason Not Working: Not Applicable	Date Serviced: 2017-05-05
Service Priority: Not Applicable	

View Model Information

ODK Tables

View Facility Information

Add Maintenance Record

View All Maintenance Records

Edit Refrigerator Status

Edit Refrigerator

Refrigerator Status

Back Next

OPEN DATA KIT

ODK Survey

Form name: Refrigerator Status

Form version: 20171011

You are at the start of a new instance.

Last saved:

Wed Oct 10 2018 19:12:00 GMT-0700 (Pacific Daylight Time)

Go to next prompt

ODK 2.0 Cold Chain Demo IV

Refrigerator Status ← Back Next →

Choose the current use status:

In Use

Not In Use

In Store For Allocation

Working status:

Functioning

Awaiting Repair

Unservicable

Reason not working:

Needs Spare Parts

No Finance

No Fuel

Unservicable

Reason not working:

Needs Spare Parts

No Finance

No Fuel

Dead

Not Applicable

Choose priority for maintenance if applicable:


Not Applicable

Low

Medium

High

Refrigerator Status ← Back Next →



ODK Survey

Form name: Refrigerator Status

Form version: 20171011

You are at the end of instance:

""

Finalize Incomplete

ODK 2.0 Cold Chain Demo V

