Open Data Kit 2.0: A Services-Based Application Framework for Disconnected Data Management

Waylon Brunette
University of Washington

CO-AUTHORS: Samuel Sudar, Mitchell Sundt, Clarice Larson, Jeffrey Beorse, Richard Anderson
Mobile Device Expansion

Mobile Device availability **DOES NOT** equate to Internet Access

Limited Internet Access


Note: High-speed internet (broadband) includes the total number of fixed-line broadband subscriptions (such as DSL, cable modems, fiber optics), and the total number of 4G/LTE mobile subscriptions, minus a correcting factor to allow for those who have both types of access. 4G = fourth generation; DSL = digital subscriber line; ICT = information and communication technology; LTE = Long Term Evolution.
GOAL: Magnify human resources through technology

1. Build form
2. Collect data
3. Aggregate results

http://opendatakit.org

Open Data Kit (ODK) Usage

- Last two months, ODK has been used in 177 of 193 in UN member countries
  - Used **DAILY** by thousands in 130+ distinct countries
- Google Play installs
  - 380,000+ devices have installed ODK Collect
  - **NOTE:** Many orgs setup their devices w/o using Play
- 50+ companies “support” the ODK tool suite
- Many companies use ODK as **core technology** that their products are based on
ODK Website (2010 – Current)

Overview

Sessions

Users

Pageviews

Pages / Session

Avg. Session Duration

Bounce Rate

One Week 2017: Jun-11 to Jun-17

<table>
<thead>
<tr>
<th>Country</th>
<th>Sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>1,087</td>
</tr>
<tr>
<td>Nigeria</td>
<td>993</td>
</tr>
<tr>
<td>Kenya</td>
<td>828</td>
</tr>
<tr>
<td>India</td>
<td>745</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>255</td>
</tr>
<tr>
<td>Uganda</td>
<td>243</td>
</tr>
<tr>
<td>Ghana</td>
<td>224</td>
</tr>
<tr>
<td>Malawi</td>
<td>216</td>
</tr>
<tr>
<td>Tanzania</td>
<td>191</td>
</tr>
<tr>
<td>South Africa</td>
<td>168</td>
</tr>
</tbody>
</table>
Deployment Architect

- Examine “Application Layer” from 3 perspectives
  - Software Developer
  - Deployment Architect
  - End User
TENSION: Generic vs. Customized

• Domain-specific/customized tools can be inflexible
  – Encourages the proverbial “re-inventing wheel”
  – Tool often cannot be reused in another similar domain
  – Keeps data silo-ed

• Users & Developers often find custom solutions easier
  – Can be modified to do exactly what the user wants
  – Developers can optimize performance and workflows
Open Data Kit (ODK)

• First release in 2009 (started in 2008)
• Mobile data collection tools for Android devices
• Modular, open architecture
• Open source (Apache2 license)

• KEY FEATURES TO SUCCESS:
  – Domain Independent Tools
  – Disconnected Operation
  – Targets a Deployment Architect

GOAL: Magnify human resources through technology
CASE STUDY: EU Refugee Crisis
ODK 2.0 Case Studies

- ODK 2.0 had an iterative requirements gathering process
  - Surveys
  - Pilot deployments in 18+ countries by a variety of organizations,
  - The ODK 2.0 tool suite went through a significant redesign from the original ODK 2.0 vision

- To validate the derived requirements we examined 6 case studies

Table 3: Case Study ODK 2.0 Feature Requirement Summary

<table>
<thead>
<tr>
<th>Feature Requirement</th>
<th>Childhood Pneumonia</th>
<th>Chimpanzee Behavior Tracking</th>
<th>HIV Clinical Trial</th>
<th>Disaster Response</th>
<th>Mosquito Infection Tracking</th>
<th>Tuberculosis Patient Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complex / Non-Linear Workflows</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Link Longitudinal Data To Collected Data</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Data Security and User Permissions</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Reuse of Data Fields Across Forms</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bidirectional Synchronization</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Customizable Form Presentation</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Custom JavaScript Apps</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Sensor Integration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Paper Digitization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Custom Data Types Update Multiple Fields in a Single User Action</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
CASE STUDY: Eliminate Dengue

- **Eliminate Dengue** - uses naturally occurring bacteria (*Wolbachia*) to reduce the ability of mosquitoes to transmit viruses (e.g., dengue, chikungunya, Zika)

- **Using ODK 2.0** in Brazil, Columbia, Indonesia, Australia, and Vietnam

- Program Manager Feedback:
  - “quite easy to use and we haven’t had any acceptance issues.”
  - “the app is scaling quite well”
CASE STUDY: HIV Patient Tracking

• Adaptive Strategies for Preventing and Treating Lapses of Retention in Care (AdaPT-R)
  – UCSF Randomized Control Trial in Kenya

• ODK 2.0 deployed in 5 clinics for multiple years
  – Clinics serve ~65,000 patients
  – ~17,000 HIV Patients
  – 18 clinical employees using ODK 2.0

“We needed a solution for capturing data from multiple forms and that would allow longitudinal follow-up of individual patients. We had experience with earlier versions of ODK, so the new features of 2.0 made it the only option for us if we wanted phone-based longitudinal form completion. Would definitely recommend ODK 2.0!“ - Primary Investigator
Cloud Hosted Service OR Local Server

Mobile Device

Collect

File System

Files
ODK 2.0 Architecture

Full Data Synchronization

Cloud Hosted Service

OR

Local Server

Mobile Device

SCAN

TABLES

SURVEY

SENSORS

Cloud

Hosted

Service

Sync Service

Database Service

Web Server/

File Service

SUBMIT
ODK 2.0 Architecture

Sensors Framework:
Framework to enable organizations to connect external sensors/hardware

Mobile Device

Cloud Hosted Service

OR

Local Server

Cloud Hosted Service

OR

Local Server

SENSORS

Sync Service

Database Service

Web Server/File Service

SUBMIT
ODK 2.0 Architecture

Survey Framework:
Framework for collecting data with verification using arbitrary workflows

Cloud Hosted Service
OR
Local Server

Mobile Device

Scan
Tables
Survey

Services
Sync Service
Database Service
Web Server/File Service

Sensors

Submit
ODK 2.0 Architecture

Tables Framework:
*Framework to enable viewing and curating data on a disconnected device*
Scan Framework:
Framework to enable organizations to bridge paper to digital (Nicola Dell)
Submit Framework:

*Framework to enable organizations to optimize data transmission*

**ODK 2.0 Architecture**

**Mobile Device**

**Cloud Hosted Service**

**Local Server**

**Scan**

**Tables**

**Survey**

**Services**

**Sync Service**

**Database Service**

**Web Server/File Service**

**Sensors**

**Submit**

**Submit Framework:**

*Framework to enable organizations to optimize data transmission*
Challenges/Lessons Learned

• Challenges involved in designing 5 mobile frameworks to work together seamlessly on the mobile device
  – Part of modularity and open-source ecosystem goal
  – Goal frameworks can work independently or together to make a more complex system
LESSON: Modularity Too Far

• Originally all frameworks were *completely separate* no shared infrastructure *(PRO: Easy for users)*
  – System encountered sporadic database and file lockout issues, timing issues, etc.

• Shifted to a Services Oriented Architecture for shared core services. *(CON: Users confusion)*
  – Performance tests revealed SOA added ~150 ms overhead to response time
LESSON: Schema helped

• ODK 2.0 is **database-centric** instead of file-centric
  – DB rows are the basic unit of storage of ODK 2.0
  – View definitions and settings stored in files (not data)

• **Helps Deployment Architects understand how to use the 5 frameworks together because they are in control of the common DB schema for all tools/frameworks**
  – Also helps with deployment issues
    • Pushes Deployment Architects to avoid collecting data and not understanding how to process the data
    • Deployment Architects can avoid possible conflicting disconnected updates between users in a specific row by partitioning data into separate tables
    • Can easily separate out important data to transmit more quickly rather than data that is less important
Data Synchronization

- Designed to be adaptable in networking environments with high latencies, low bandwidths, and long periods of disconnected operation.
  - DB Rows enable smaller data transmission size enabling adjustments to network conditions
  - DB Rows provide small granularity for change-tracking to simplify conflict resolution
  - Only need to sync row changes
  - Designed to be Idempotent
Summary

• ODK 2.0 provides multiple modular frameworks to build data collection and management services for resource-constrained environments.
  – Focused on a Deployment Architect

• Building a real system used by real organizations in differing domains shows the general applicability of ODK’s mobile frameworks in challenged or disconnected network environments.
Questions?

Thank you for your attention

http://www.opendatakit.org