

# 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*

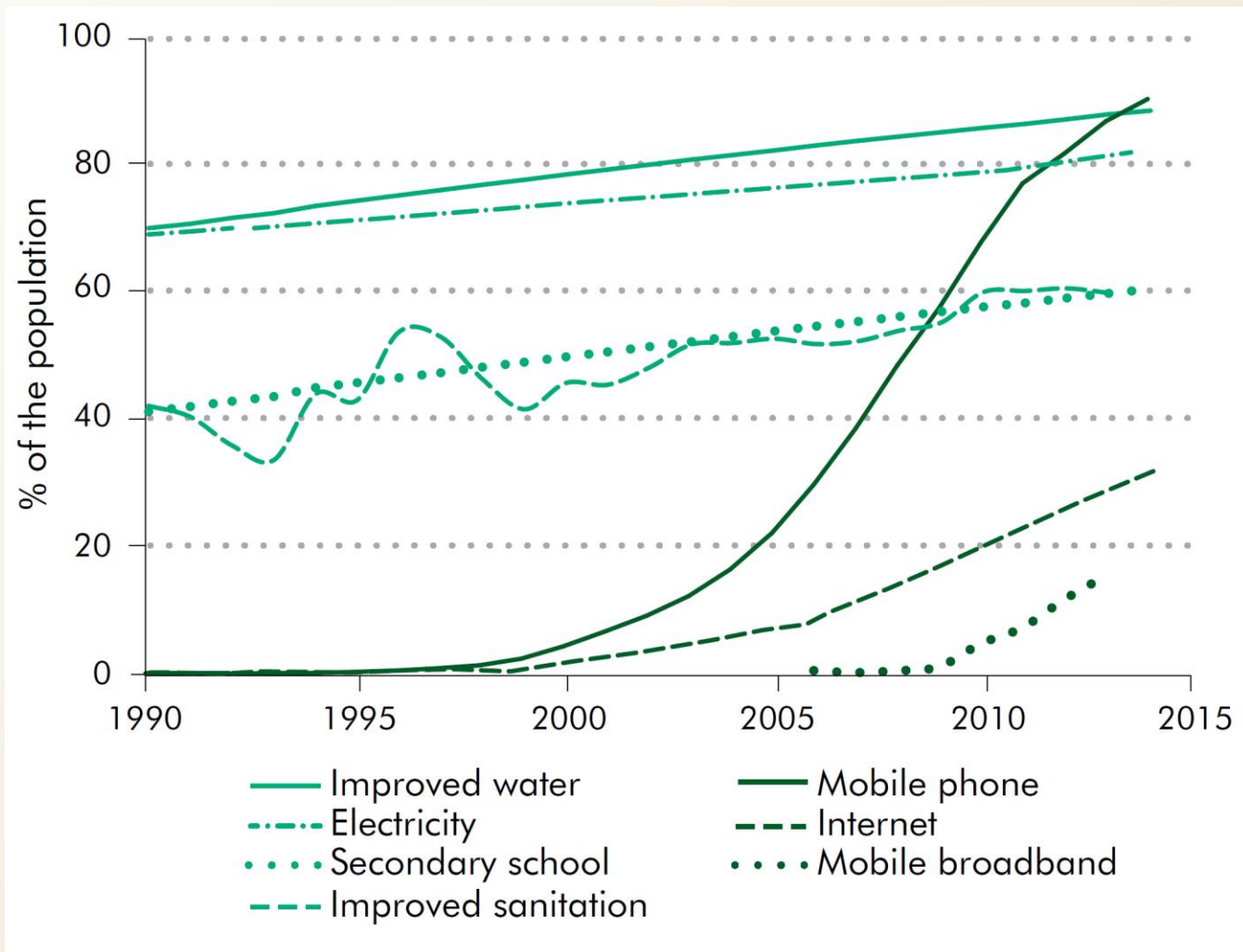






# Mobile Device Expansion

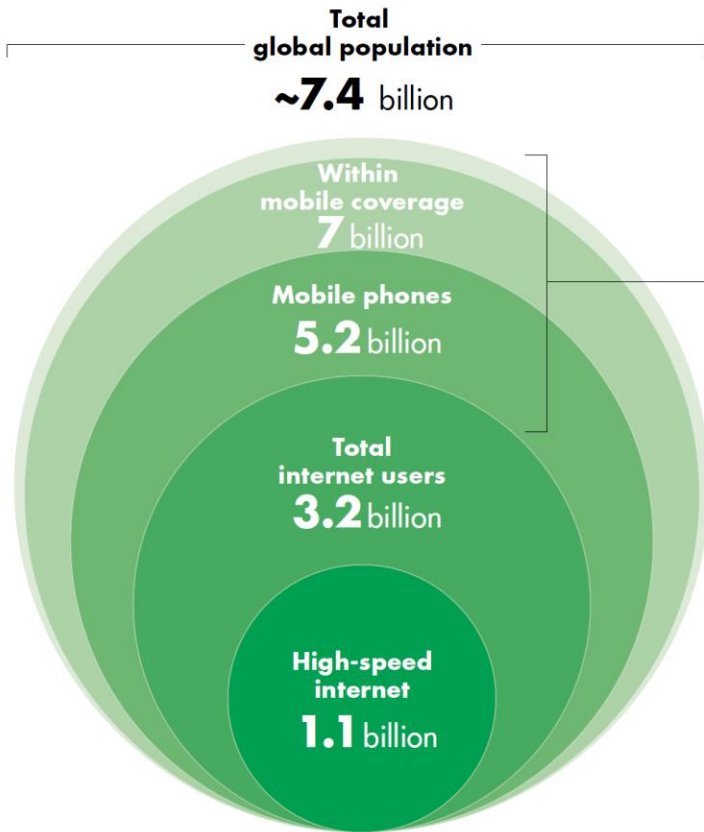
**Mobile Device availability *DOES NOT* equate to Internet Access**



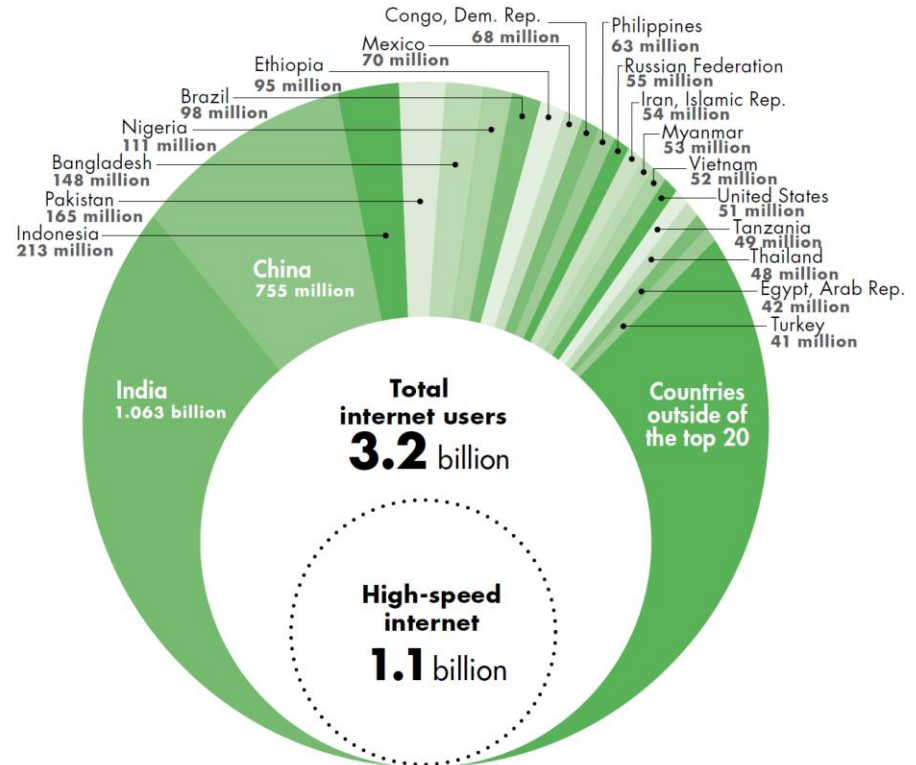


# Limited Internet Access

a. ICT access by population



b. A closer look at the world's offline population



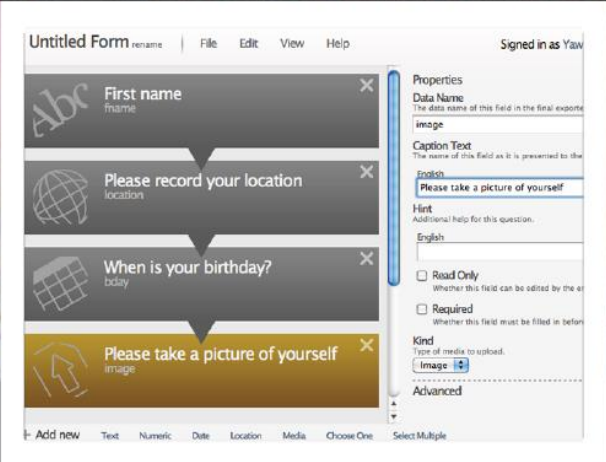
Sources: World Bank 2015; Meeker 2015; ITU 2015; GSMA, <https://gsmaintelligence.com/>; UN Population Division 2014. Data at [http://bit.do/WDR2016-FigO\\_5](http://bit.do/WDR2016-FigO_5).

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.

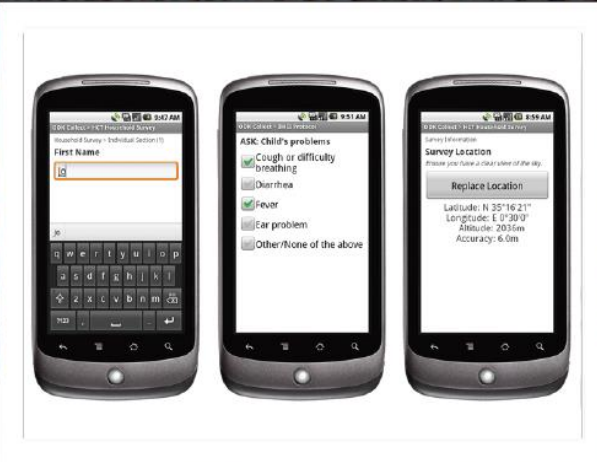
Graph from: *World Development Report 2016: Digital Dividends*. International Bank for Reconstruction and Development (World Bank), 2016.

541-700 ↓

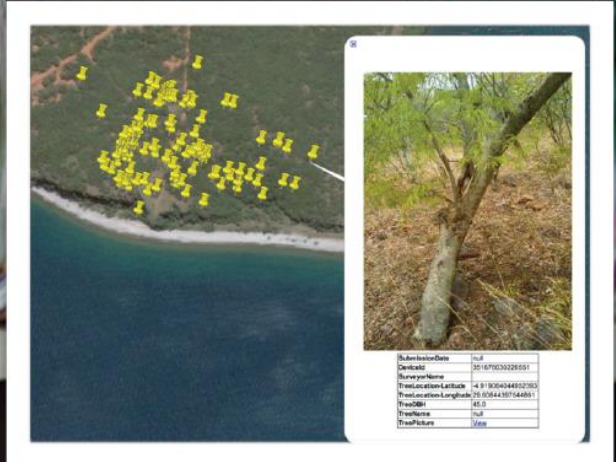
# 1. Build form



# 2. Collect data



# 3. Aggregate results



<http://opendatakit.org>

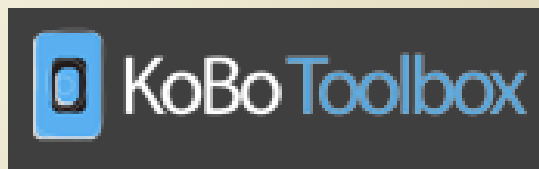
# GOAL: Magnify human resources through technology

- C. Hartung, Y. Anokwa, W. Brunette, A. Lerer, C. Tseng, and G. Borriello. Open Data Kit: Tools to Build Information Services for Developing Regions. In Proceedings of the 4<sup>th</sup> ACM/IEEE International Conference on Information and Communication Technologies and Development, ICTD '10, 2010.



# 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



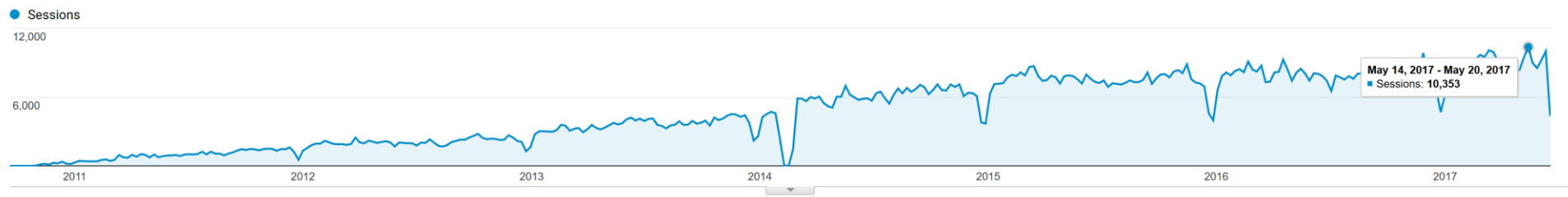


# ODK Website (2010 – Current)

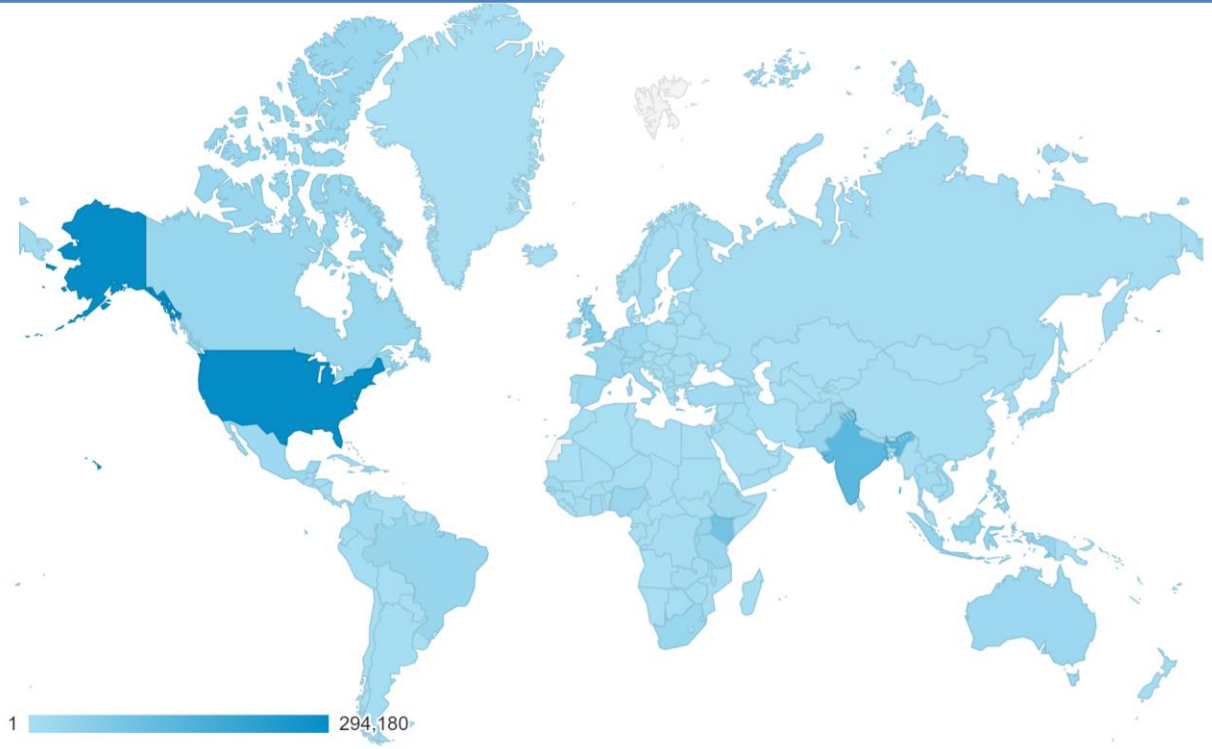
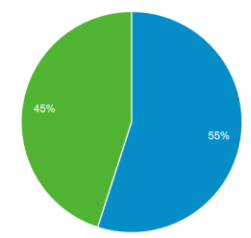
Overview

Sessions vs. Select a metric

Hourly Day Week Month



Returning Visitor New Visitor



## One Week 2017: Jun-11 to Jun-17

1.	United States	1,087 (10.87%)
2.	Nigeria	993 (9.93%)
3.	Kenya	828 (8.28%)
4.	India	745 (7.45%)
5.	United Kingdom	255 (2.55%)
6.	Uganda	243 (2.43%)
7.	Ghana	224 (2.24%)
8.	Malawi	216 (2.16%)
9.	Tanzania	191 (1.91%)
10.	South Africa	168 (1.68%)



# Deployment Architect

- Examine “Application Layer” from 3 perspectives
  - Software Developer
  - *Deployment Architect*
  - End User



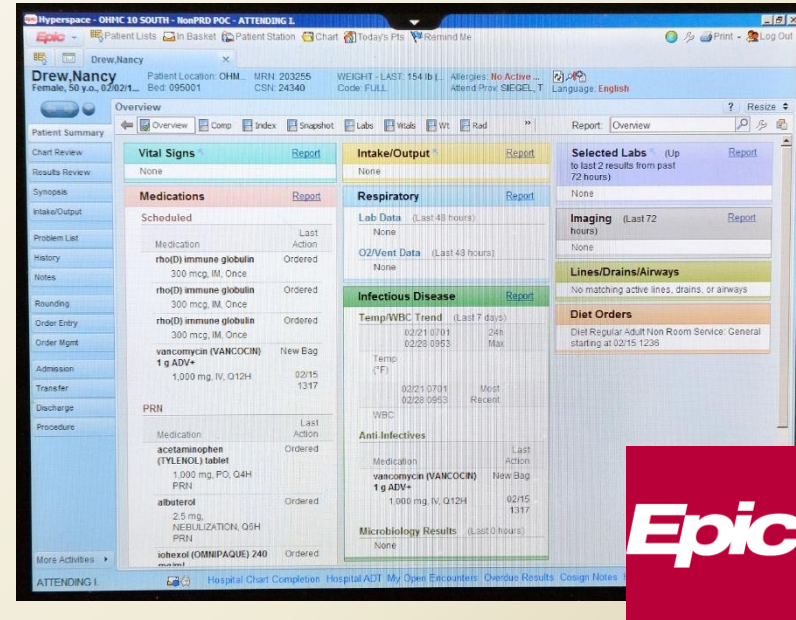




# TENSION: Generic vs. Customized



Versus



- 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





# CASE STUDY: EU Refugee Crisis





# 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**

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



# CASE STUDY: Eliminate Dengue

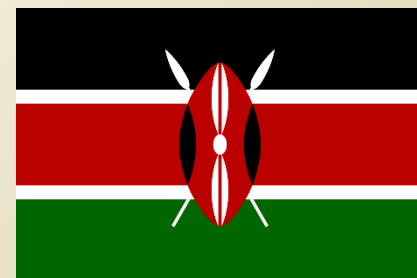
- **Eliminate Dengue** - uses naturally occurring bacteria (*Wolbachia*) to reduce the ability of mosquitoes to transmit viruses (e.g., dengue, chikungunya, Zika)
- **Using ODK 2.0** in Brazil, Columbia, Indonesia, Australia, and Vietnam
- Program Manager Feedback:
  - “quite easy to use and we haven’t 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



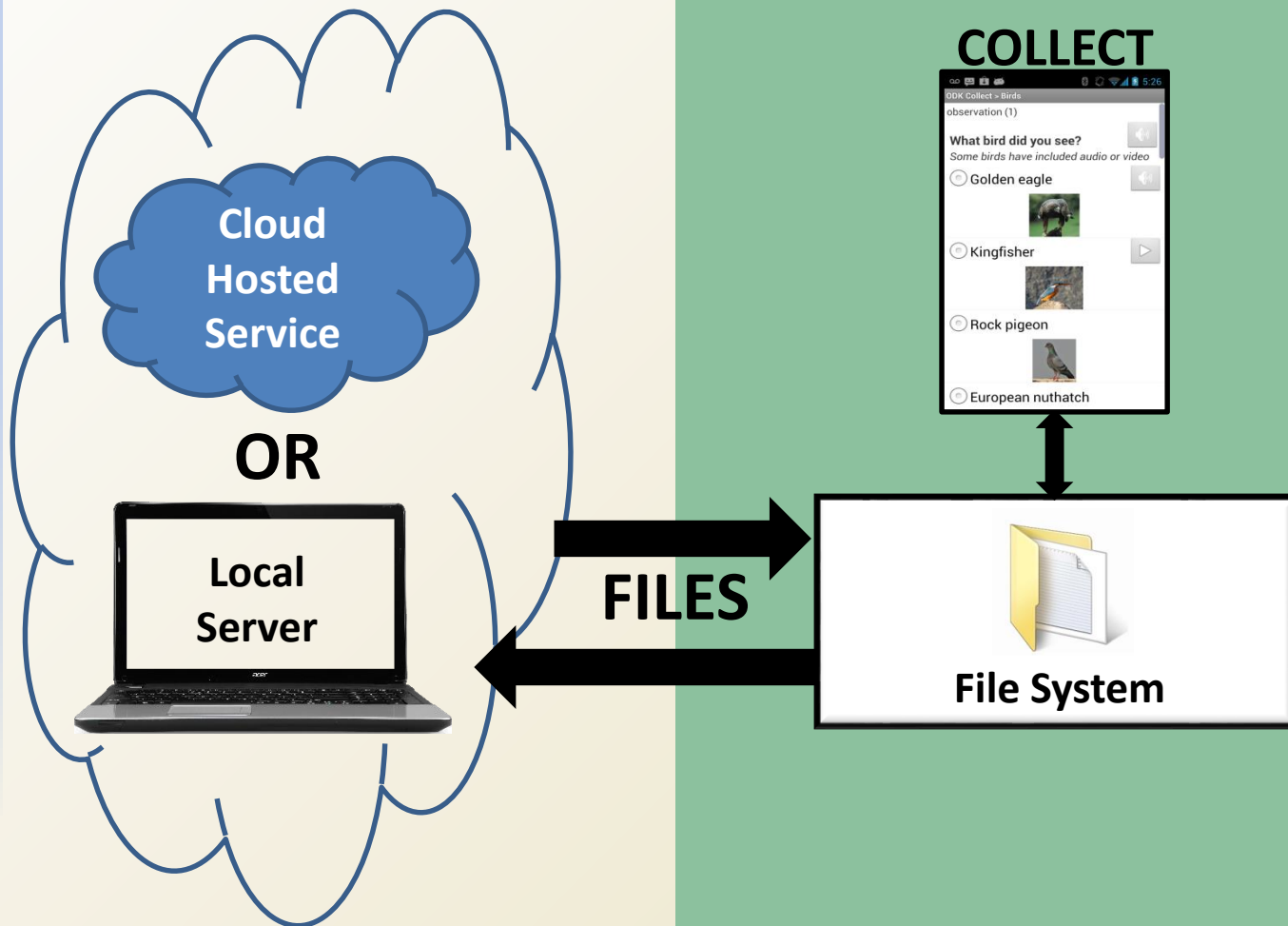
*"We needed a solution for capturing data from multiple forms and that would allow longitudinal follow-up of individual patients. We had experience with earlier versions of ODK, so the new features of 2.0 made it the only option for us if we wanted phone-based longitudinal form completion. Would definitely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





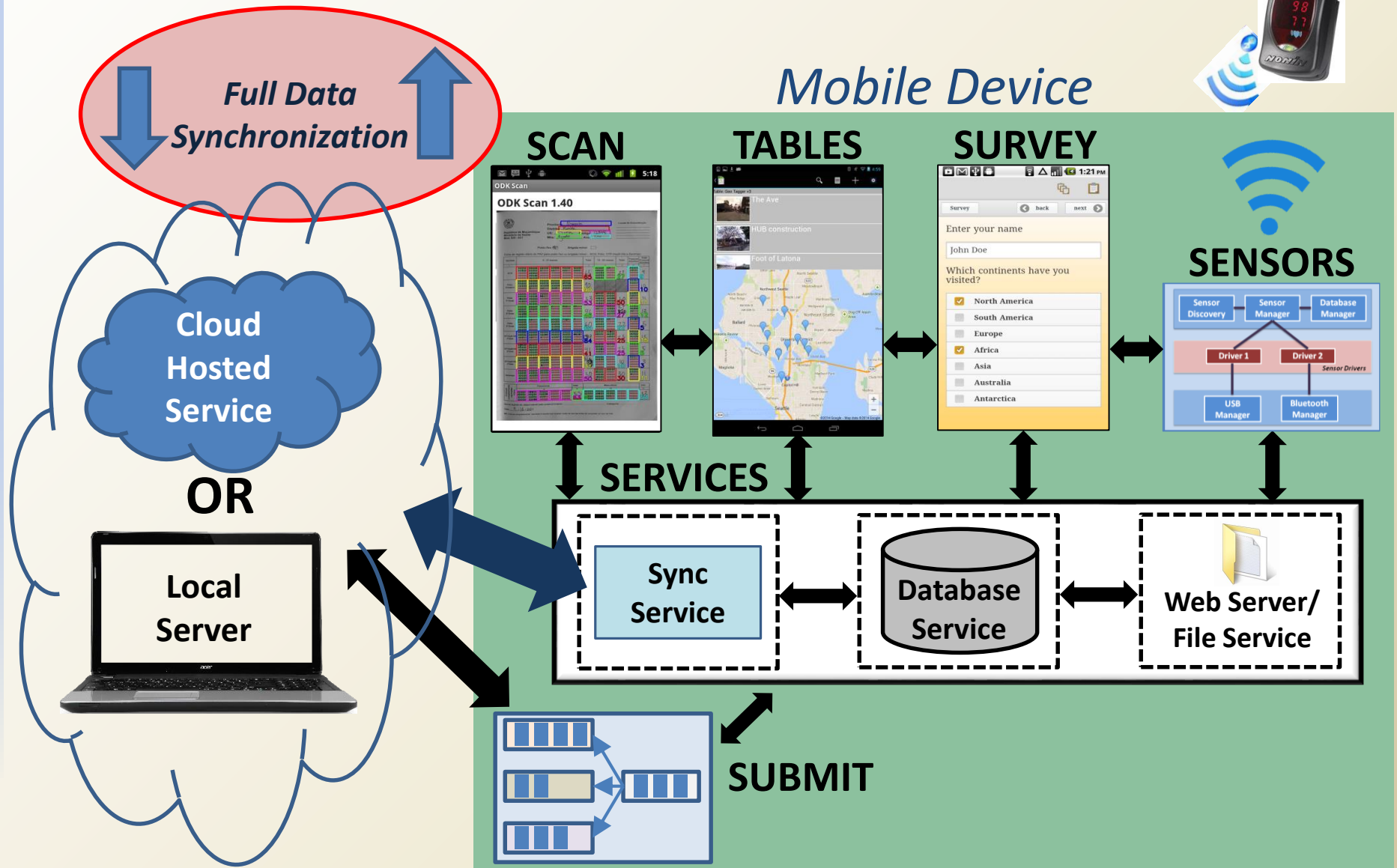
# ODK 1.x Architecture

## *Mobile Device*





# ODK 2.0 Architecture





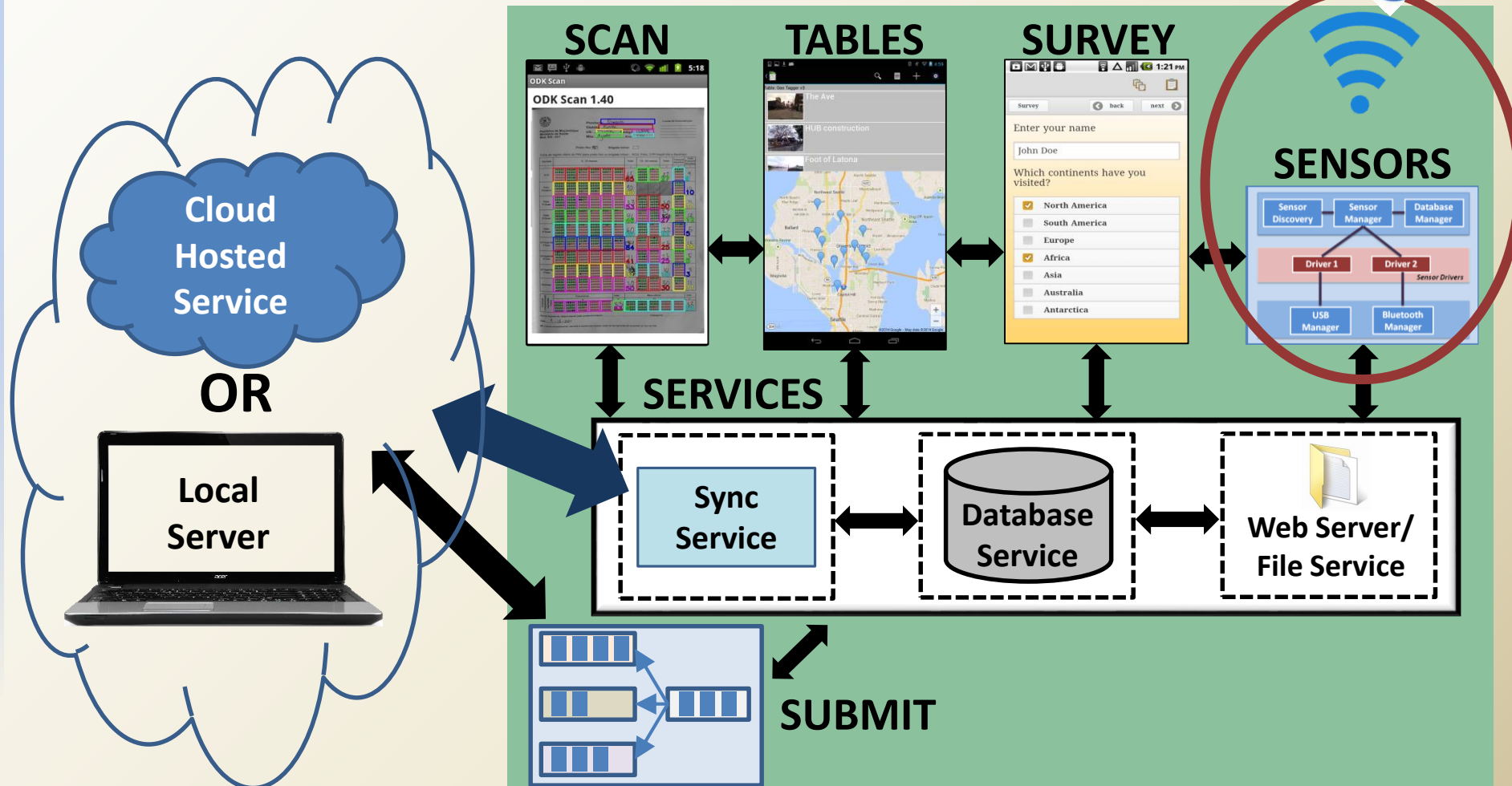
# ODK 2.0 Architecture

## Sensors Framework:

Framework to enable organizations to connect external sensors/hardware



Mobile Device



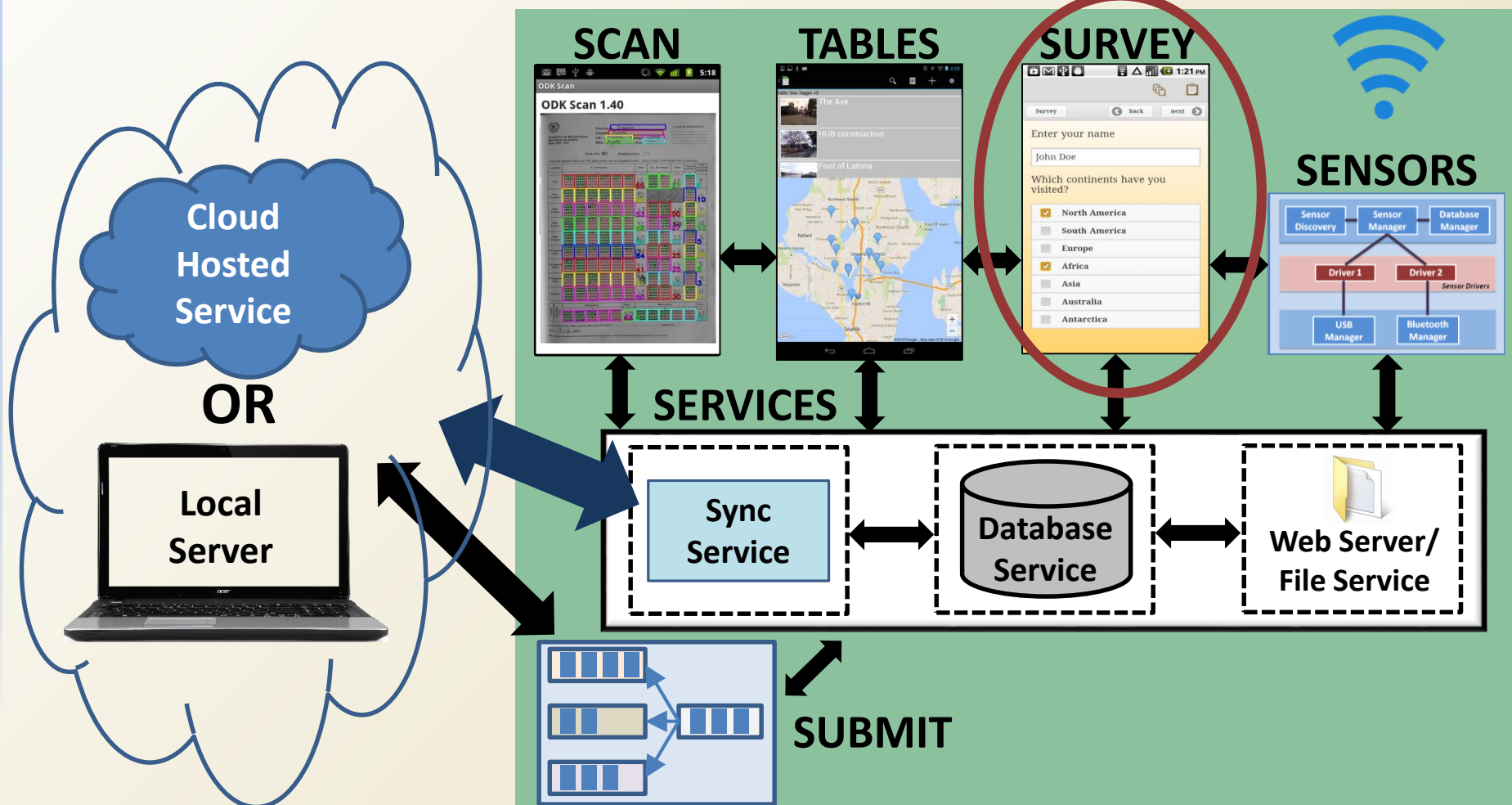


# ODK 2.0 Architecture

## Survey Framework:

*Framework for collecting data with verification using arbitrary workflows*

*Mobile Device*

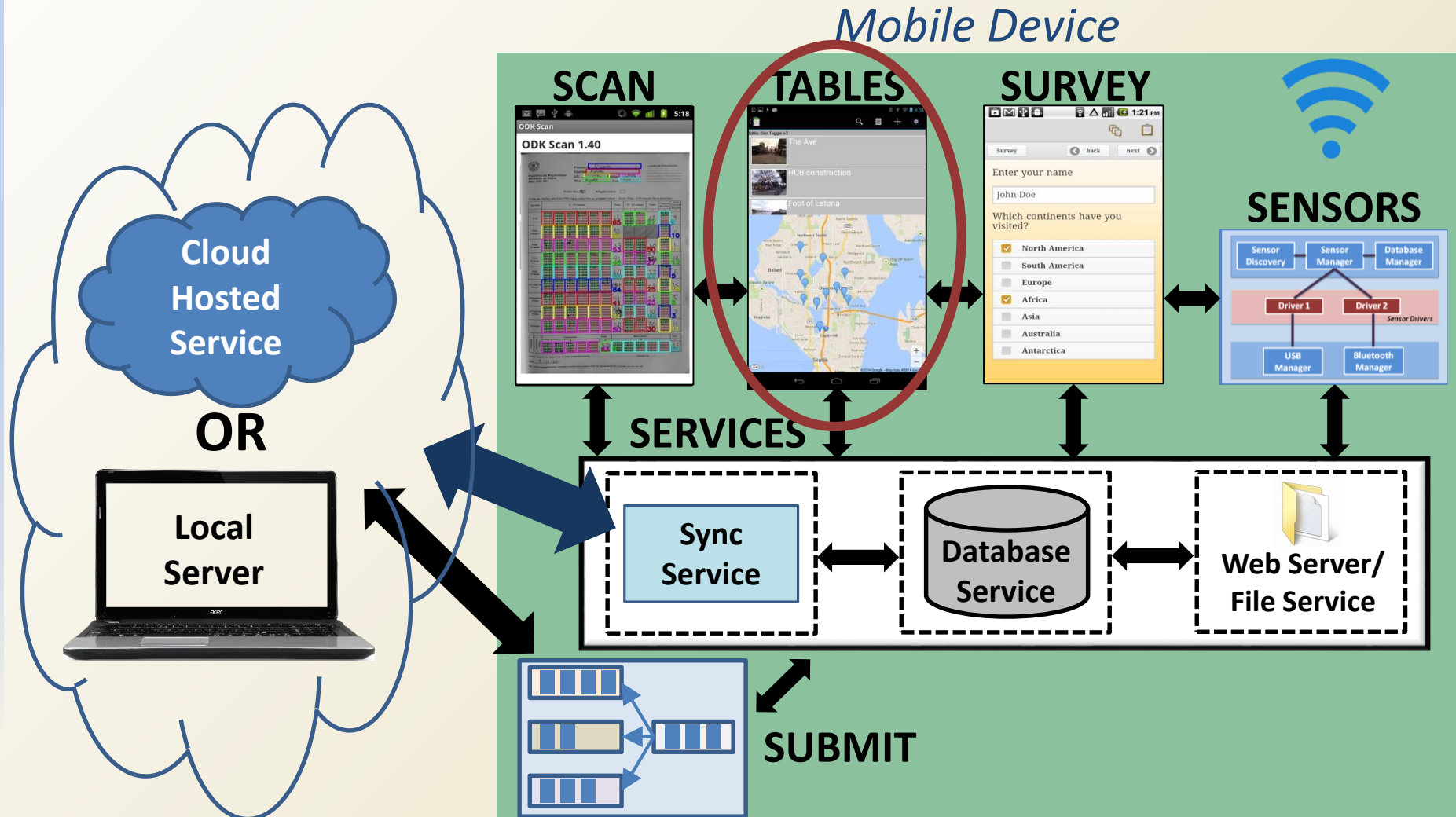




# ODK 2.0 Architecture

## Tables Framework:

*Framework to enable viewing and curating data on a disconnected device*



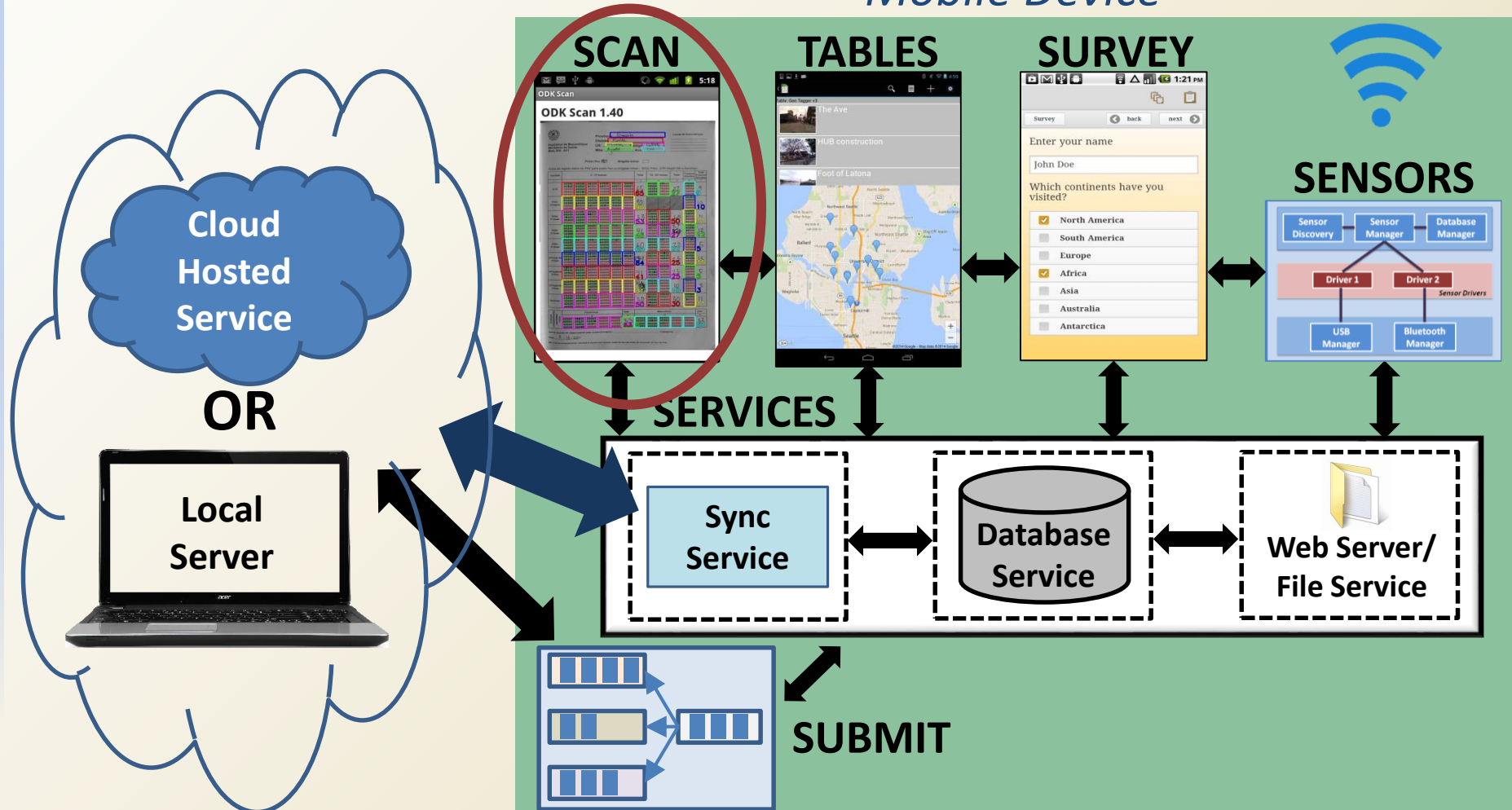


# ODK 2.0 Architecture

## Scan Framework:

*Framework to enable organizations to bridge paper to digital (Nicola Dell)*

### Mobile Device



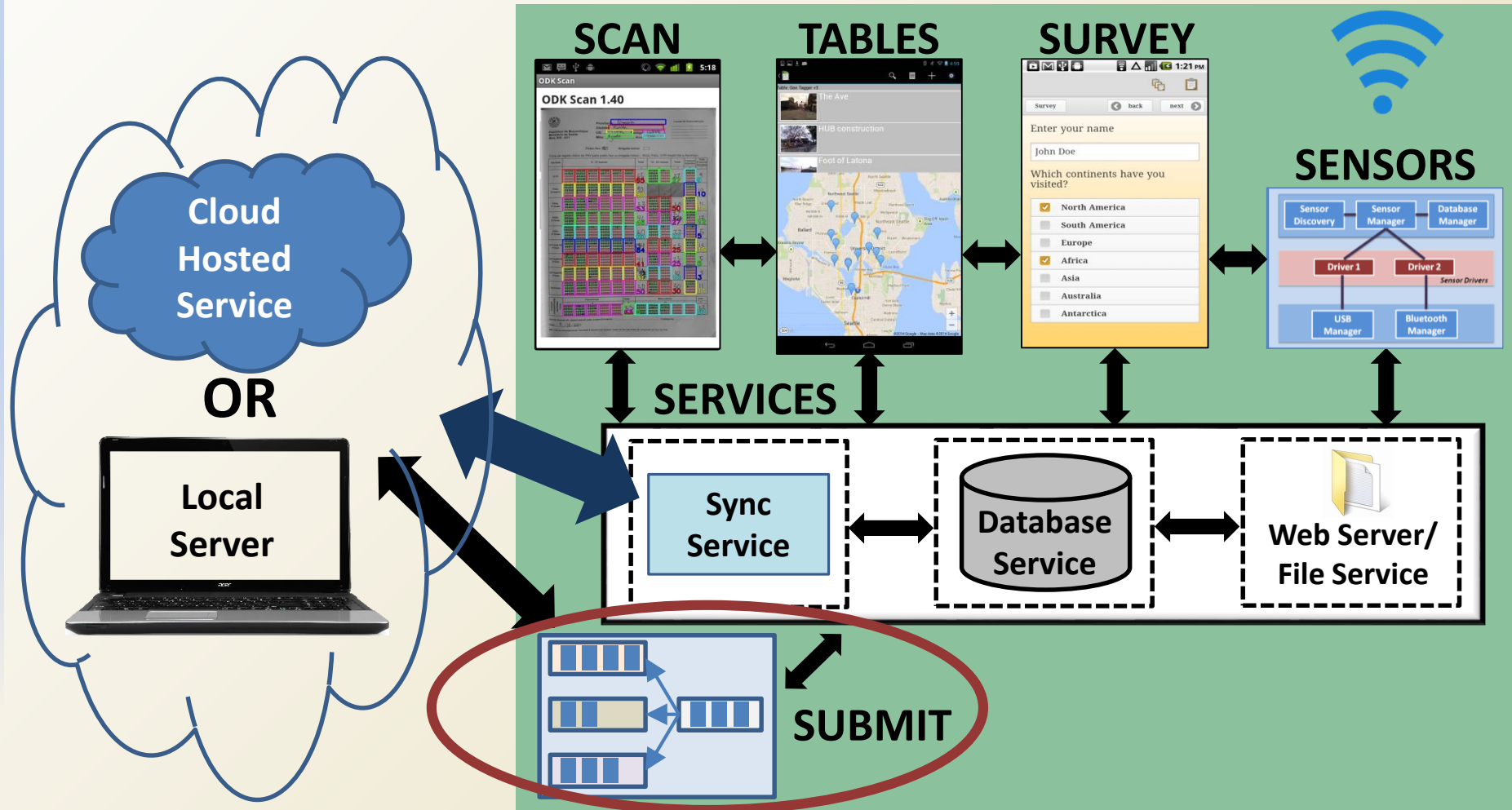


# ODK 2.0 Architecture

## Submit Framework:

*Framework to enable organizations to optimize data transmission*

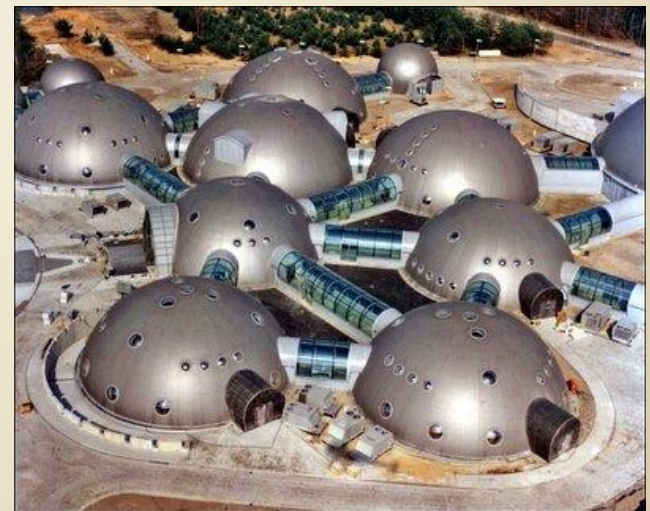
## Mobile Device





# 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*

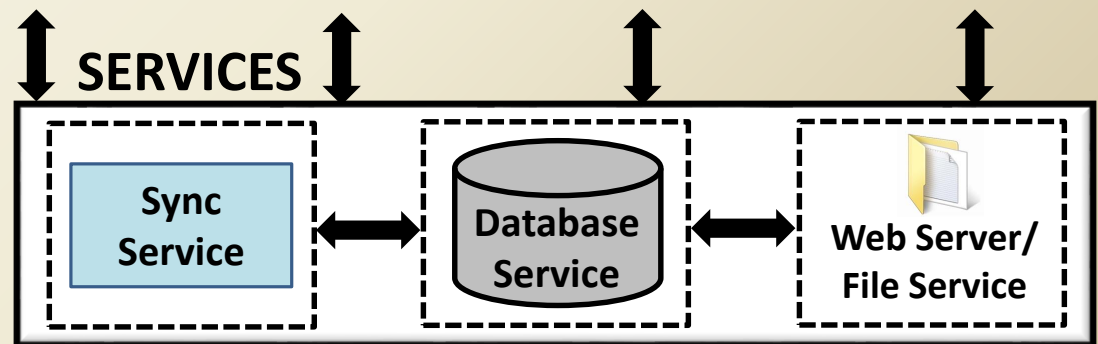






# LESSON: Modularity Too Far

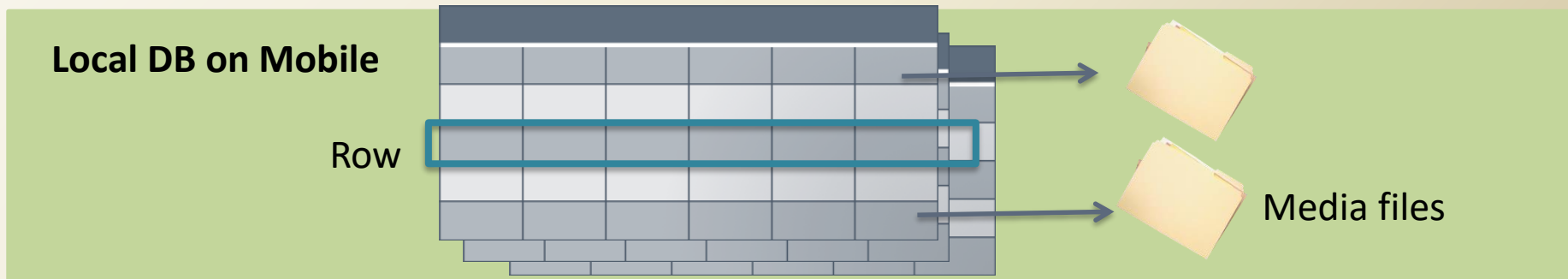
- Originally all frameworks were *completely separate* no shared infrastructure (PRO: Easy for users)
  - System encountered sporadic database and file lockout issues, timing issues, etc.
- Shifted to a Services Oriented Architecture for shared core services. (CON: Users confusion)
  - Performance tests 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>

