

# Development: Is There an App for That?

Luis Crouch

Presentation at International Conference on Big Data for Official Statistics

30 Aug – 1 Sept 2016

Dublin, Ireland



# Anecdote 1

Luis Crouch to son, 2 hours before a trip: “See you at 12<sup>th</sup> and F, around 3:00, we’ll then walk to the train station.”

Luis arrives at 3:20. Note the address. Seem like a familiar scene?

Son to Luis: “What did you use to do before cell phones?”

Luis: “We used to plan better.”



# Anecdote 2

- Crouch to Minister of Education:
  - “If we can put together your costs data, your exam results data, your teacher profile and training data, the country’s poverty data, and other data, we can do some nice modeling and predictive analytics, blah, blah...”
- Minister to Crouch:
  - “But Luis, we don’t even know how many schools we have...”
- Crouch:
  - “Ok, let’s then start with some basics...”

# Motivation

- RTI: dozens of apps, websites, software approaches, etc.
- Some, such as Tangerine, have become famous.
- But what is the logic?
- Internal framework for imposing logic on these efforts?
- Can we be more intentional and careful?
- Internal introspection: worth sharing?
- **Several development partners have asked us to share the results of our own thinking.**

# Assumption and Implications - 1

- Assumption
  - We are interested in how data handling applications can help produce progress towards the SDGs, not just track/measure them...
- Implications that challenge “Big Data:”
  - Understanding causality matters, correl not enough
  - Universal and purposeful measurement
  - Proper sampling characteristics
  - Demonstrated improvements in SDGs that follow from usage of data
  - Iterative improvement

# Assumption and Implications - 2

- **Solution implications**
  - Work on “little data:” build on legacy systems using tech, big data aspects
    - Including client feedback technologies
  - Better curation of what data exist already
    - Merge datasets: simple technique, but so far mostly untried in many Ministries
  - Demonstrate use in management
    - Have to create demand by clearly demonstrating SDG improvement through use of data, not just articulating the possibility – can use Big Data aspects (e.g. trackers), but key issue is managerial and accountability

# Presentation Outline

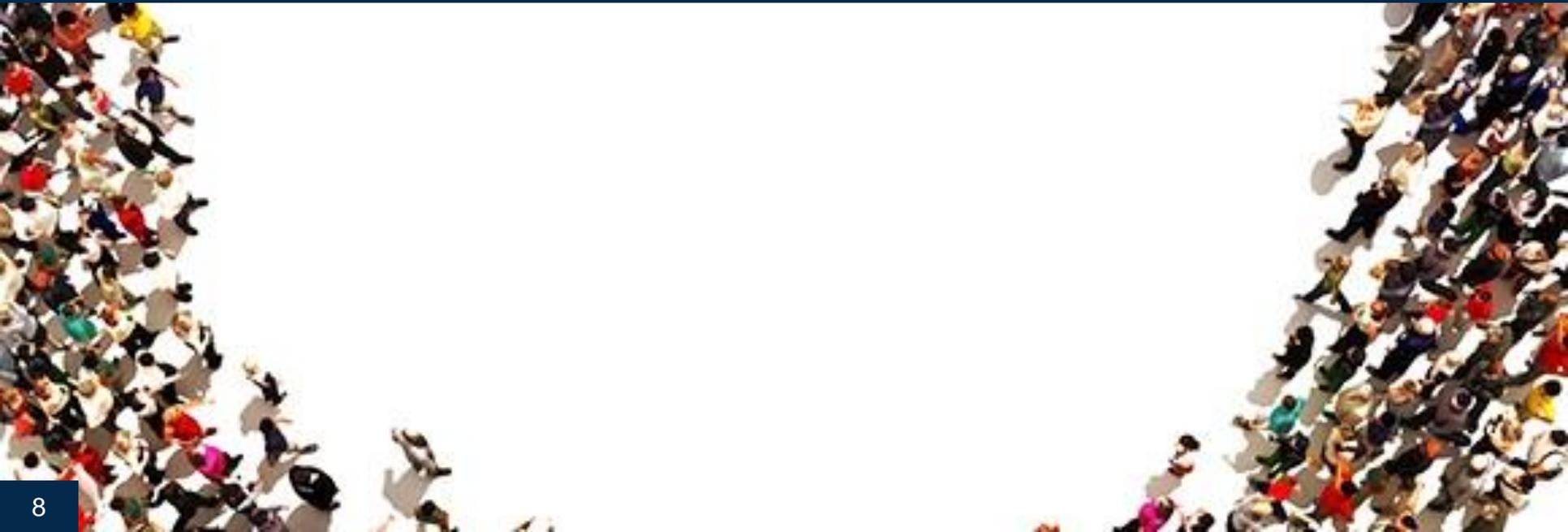
**1** Framework for selecting data tools

**2** Case studies to illustrate the framework

**3** Integrated illustrative example  
- DHIS2 health information systems approach  
(U of Oslo)

# 1

## Framework for Selecting Data and Information Tools



- There are many aspects to the Data Revolution
  1. Systematic reviews, other forms of research often using primary data
    - For new policy design
  2. Use of impact evaluations, also often using primary, purpose-gathered data
    - For new twists on policy
  3. Statistics, admin data, for planning, policy tracking
    - Mostly for policy implementation, accountability, tracking
    - Not just M&E; includes operational research

Today's chat focusing on this.



“Bleed” into each other, hard to separate.

For sake of focus: service provision, not overall business climate, market-shaping.

# Selecting Tools During the Data Revolution

## Context

- Technology
- Demand for data-driven insights

“We also call for a data revolution for sustainable development, with a new international initiative to improve the quality of statistics and information available to citizens. We should actively take advantage of new technology, crowd-sourcing, and improved connectivity to empower people with information on the progress towards the targets.”<sup>1</sup>

<sup>1</sup>United Nations (2013). A New Global Partnership: Eradicate Poverty and Transform Economies through Sustainable Development: The Report of the High-Level Panel of Eminent Persons on the Post-2015 Development Agenda. Executive Summary.

# Selecting Tools During the Data Revolution

## Context

- Technology
- Demand for data-driven insights

## New tools

- SMS, phone, IVR
- Administrative data
- Crowdsourcing
- Unstructured data
- Enhancements to face-to-face surveys

“There have been innovative initiatives to use mobile technology and other advances to enable real-time monitoring of development results. But this movement remains largely disconnected from the traditional statistics community at both global and national levels.”<sup>2</sup>

<sup>2</sup>UN (2013), p. 23

# Selecting Tools During the Data Revolution

## Context

- Technology
- Demand for data-driven insights

## New tools

- SMS, phone, IVR
- Administrative data
- Crowdsourcing
- Unstructured data
- Enhancements to face-to-face surveys

**Short-message service (SMS)**, or text messaging surveys

**Telephone surveys** with live interviewers

**Interactive voice response (IVR)**, where respondents listen to recorded questions and select numbers to indicate response

**Integrated administrative data**: Linking vital statistics, health records, school data to (1) each other and (2) survey data

**Crowdsourcing**, where people volunteer to collect data (e.g., tracking food prices in stories, mapping electricity availability)

**Unstructured data** such as audio, satellite data, photographs

**Enhancements to face-to-face surveys** with GIS sampling, audio-recorded interviews, and electronic data collection (e.g., *Survey Solutions*)

# Selecting Tools During the Data Revolution

## Context

- Technology
- Demand for data-driven insights

## New tools

- SMS, phone, IVR
- Administrative data
- Crowdsourcing
- Unstructured data
- Enhancements to face-to-face surveys

## Challenge

**When** should we use different tools?

## Solution

Develop framework to help guide decisions about best tool to use for a given objective

# Framework

## Objective

### Planning

- Standard sectoral planning
- Needs assessment to inform program

### Implementation

- Complement existing paper data collection
- Create communication mechanism
- Share information
- Allocate resources

### Program monitoring

### Rapid evaluation

### Impact evaluation

### Tracking development indicators

## Population

- General population
- Special population (beneficiaries, employees)



## Population profile

- Geography
- Language
- Socio-demographics
- Technology access, skills

## Quality standards

- Accuracy
- Timeliness
- Relevance
- Comparability
- Coherence
- Accessibility
- Interpretability

## Resources

- Budget
- Staff (skills, size, partners)
- Technology

## Study Design

- Survey length
- Cross-section versus panel
- Topic: sensitivity, complexity, interest

Don't have time to use the whole framework. Will use selected bits for illustrating. In real life applications: recommend using all aspects.

## Tools

- SMS: one way, two way
- IVR, phone
- Administrative data
- Crowdsourcing
- Unstructured data
- Enhancements to face-to-face surveys

## Contextual Factors

### Actors

- Government: national
- Government: sub-national
- Donors
- NGOs
- Businesses

Country

Sector

# Supply or Demand?

- Most countries already produce more data than they (creatively, properly) use
- Hence: importance of boosting the demand side
- Role of technology in boosting the demand side?
  - Or does technology have to be accompanied by other factors?

# What Data Do (Should) Different Actors Demand?

Actors	Type of data use
Politicians, high-level policy-makers	<ul style="list-style-type: none"><li>• Tracking of delivery promises, including SDGs</li><li>• Setting fundamental, innovative policy</li><li>• Budget approval and tracking</li></ul>
Technocrats, planners	<ul style="list-style-type: none"><li>• Policy proposals</li><li>• Policy nuances</li><li>• Design of policy implementation</li><li>• First-tier policy tracking of policy execution</li><li>• Budget proposals</li><li>• Detailed planning (facilities siting, funding formula design, citizen accountability design)</li></ul>
Citizens <ul style="list-style-type: none"><li>• Service receivers, “rights holders”</li><li>• Civil society watchdogs “for” the rights holders</li> <li>• Private sector</li></ul>	Monitoring policy execution, plans <ul style="list-style-type: none"><li>• Info for voice, choice</li><li>• Complaints “up”</li><li>• “Downward” flow of info on execution (info on both rights and delivery against rights)</li><li>• Policy proposals</li> <li>• Information on business climate, trade, macro issues, and overall government efficiency</li></ul>

# 2

## Case Studies to Illustrate the Framework

# Case Studies - Menu

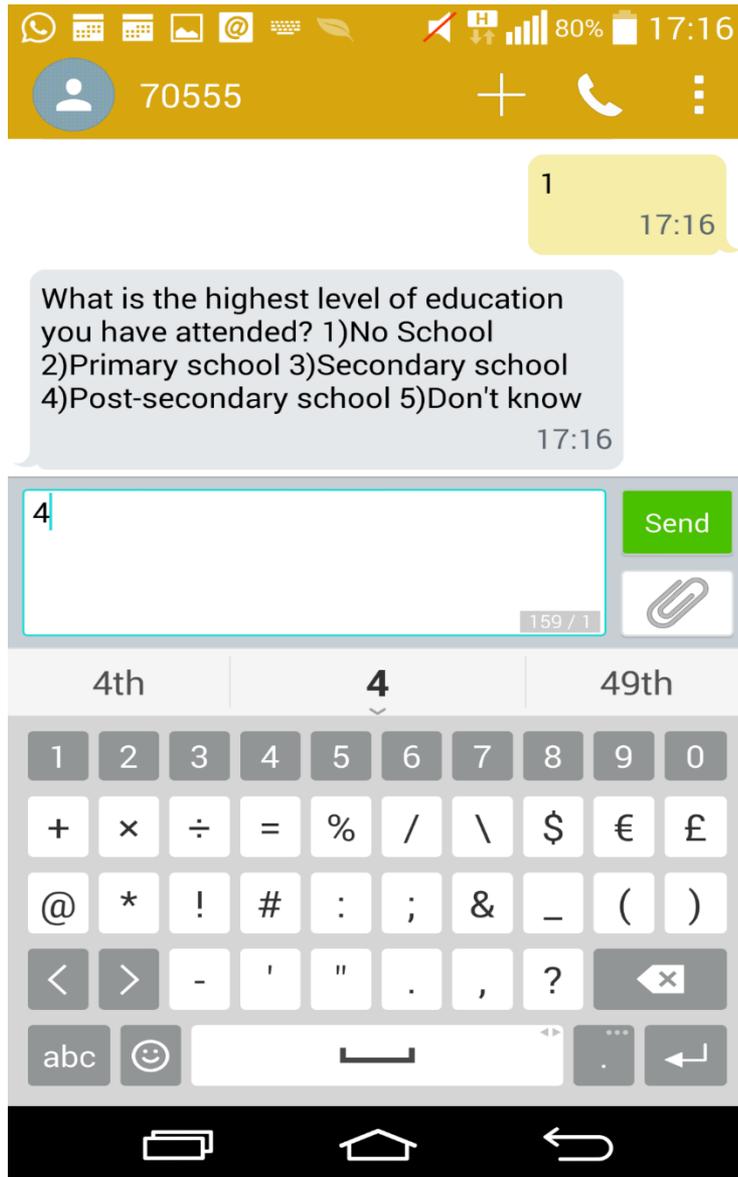
Objective	Case Study
1. Rapid assessment of policy execution via SMS	Fast, low cost SMS survey of general population in Ghana, Kenya, Nigeria, Uganda. <u>All SDGs.</u> 
2. Implementation assessment via “captive” populations	Collect data from special population (VDC) via smart-phone to monitor ante-natal and post-natal care in <u>Nepal.</u> <u>SDG 3.</u> 
3. Citizen’s complaints in local governance	In Uganda, experiment with citizens reporting problems with district service delivery via free SMS. Issues are routed to government officials to resolve cases. <u>All SDGs.</u> 
4. Target resource allocation via “professional” reporting	For efficient disease surveillance and monitoring, create two-way information flow where cases are reported <i>from and to</i> local level using SMS, and further case study is then executed. <u>SDG 3.</u> 
5. Electronic capture of learning assessment: Tangerine	Tracking student skills and teacher behavior via electronic applications for oral assessment of learner skills. <u>SDG 4.</u> 
6. Labor market information in near-real-time	“Transition to work” SMS survey in near-real time: experiment in Kenya. Intriguing info for \$10,000. <u>SDG 8.</u> 
7. Shadowing of “little data” paper systems with e-systems	Case study of Kenya EMIS: is e-collection feasible? <u>SDG 4.</u> 

# Case 1: Rapid Policy Execution Assessment Via SMS

Framework element	How it applies
Population	General population
Population profile	Not very literate
Speed	Is the whole point
Accuracy	Somewhat important to low importance
Cost	Important
1-way vs 2-way info flow	1-way

Consider SMS survey, but with caution

# Case 1: Rapid Policy Execution Assessment Via SMS



SMS surveys can rapidly\* collect data from a probability-based sample of the general population.

But how much accuracy are we sacrificing by using SMS for population-based surveys?

# Case 1: Rapid Policy Execution Assessment Via SMS

## **RTI and TNS' research on SMS**

- General population SMS surveys in Ghana, Kenya, Nigeria, and Uganda (n = 9,697).
- To evaluate SMS survey representativeness, compared with gold standard face-to-face data
- Experimental design sheds light on how to optimize SMS survey design.

# Case 1: Rapid Policy Execution Assessment Via SMS

## RTI and TNS' research on SMS

- General population SMS surveys in Ghana, Kenya, Nigeria, and Uganda (n = 9,697).
- To evaluate SMS survey representativeness, compared with gold standard face-to-face data
- Experimental design sheds light on how to optimize SMS survey design.

### Population

- Nationwide probability sample of mobile numbers, stratified by geography
- Offered survey in major languages
- Covers people with mobile phones

### Study Design

- Cross-sectional survey
- Non-sensitive topics
- Short survey
- Experiment: 8 and 16 question survey have same response rate

### Quality Standards

- Timeliness: Sent 1.1 million SMS invitations in Nigeria in 4 days
- Accuracy: Poor (see following slides)

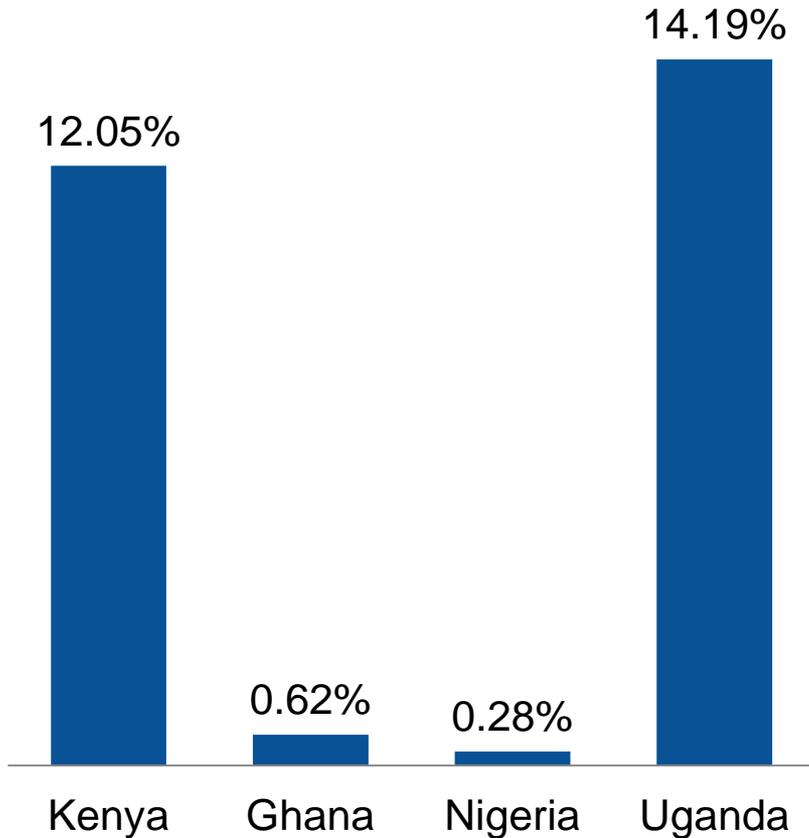
### Resources

- Low cost
- Quality vendor (GeoPoll)

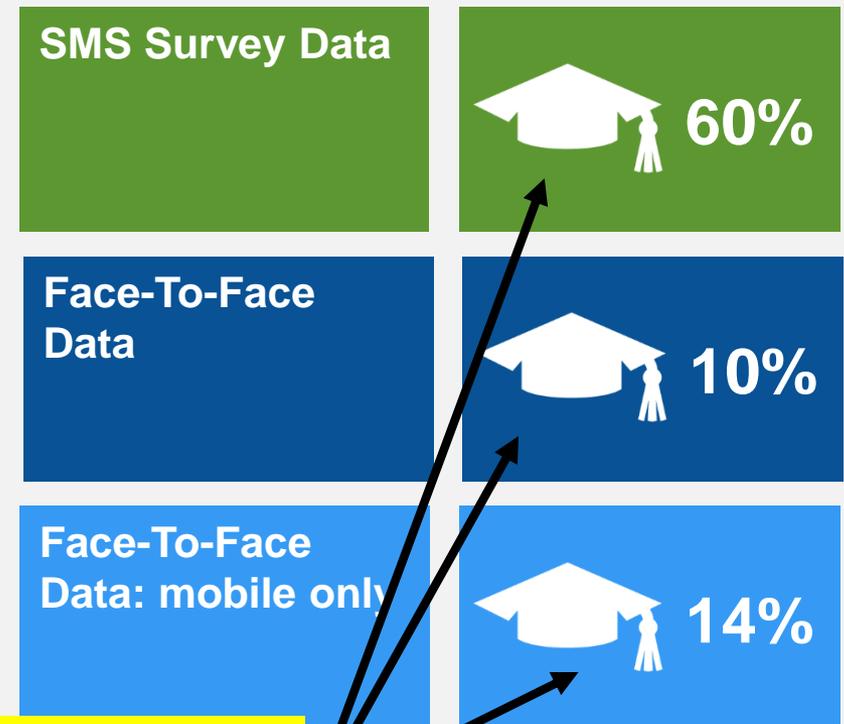
# Case 1: Rapid Policy Execution Assessment Via SMS

## Issue: Representativeness of SMS General Population Surveys

### Response Rates



### Post-Secondary Education (Nigeria)

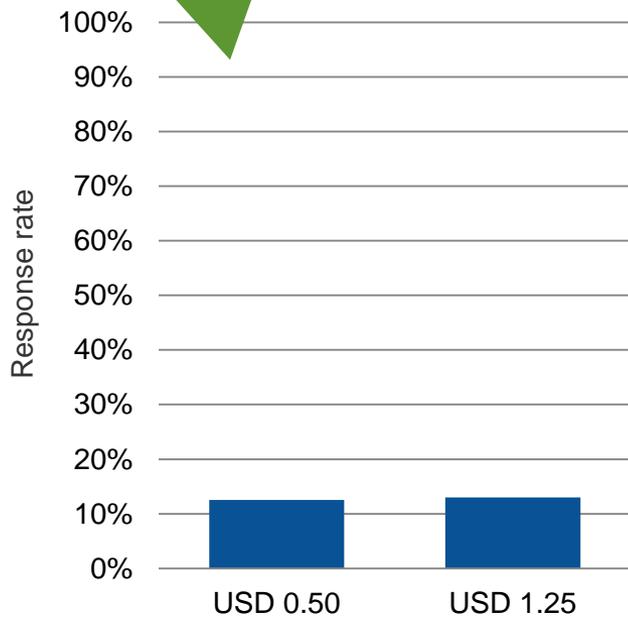


This suggests that the bias is introduced by non-response, not by cell phone ownership. Can mitigate by stratifying, weighting, but not perfect.

# Case 1: Rapid Policy Execution Assessment Via SMS

## How to counter non-response?

Increasing incentives doesn't raise the response rate (Kenya)



Do use reminders to reduce non-response bias



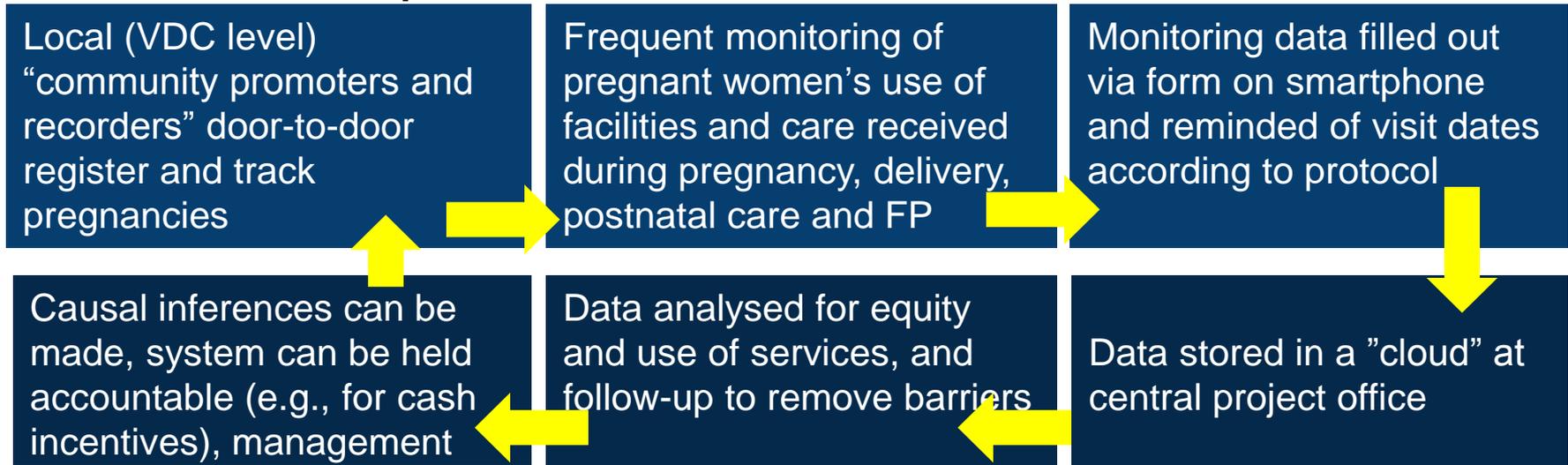
Study the bias, then stratify, weigh

# Case 2: Implementation Assessment Via “Captive” Populations

## Tracking pregnant women to increase equitable use of ANC, institutional delivery, PNC and postpartum FP in Nepal

Frequent outreach and monitoring needed; traditional facility-based monitoring passive, slow, error ridden, aggregated, and complex, and sometimes ignores important local issues, such as inequitable use.

### Solution Piloted in Nepal



# Case 2: Implementation Assessment Via “Captive” Populations

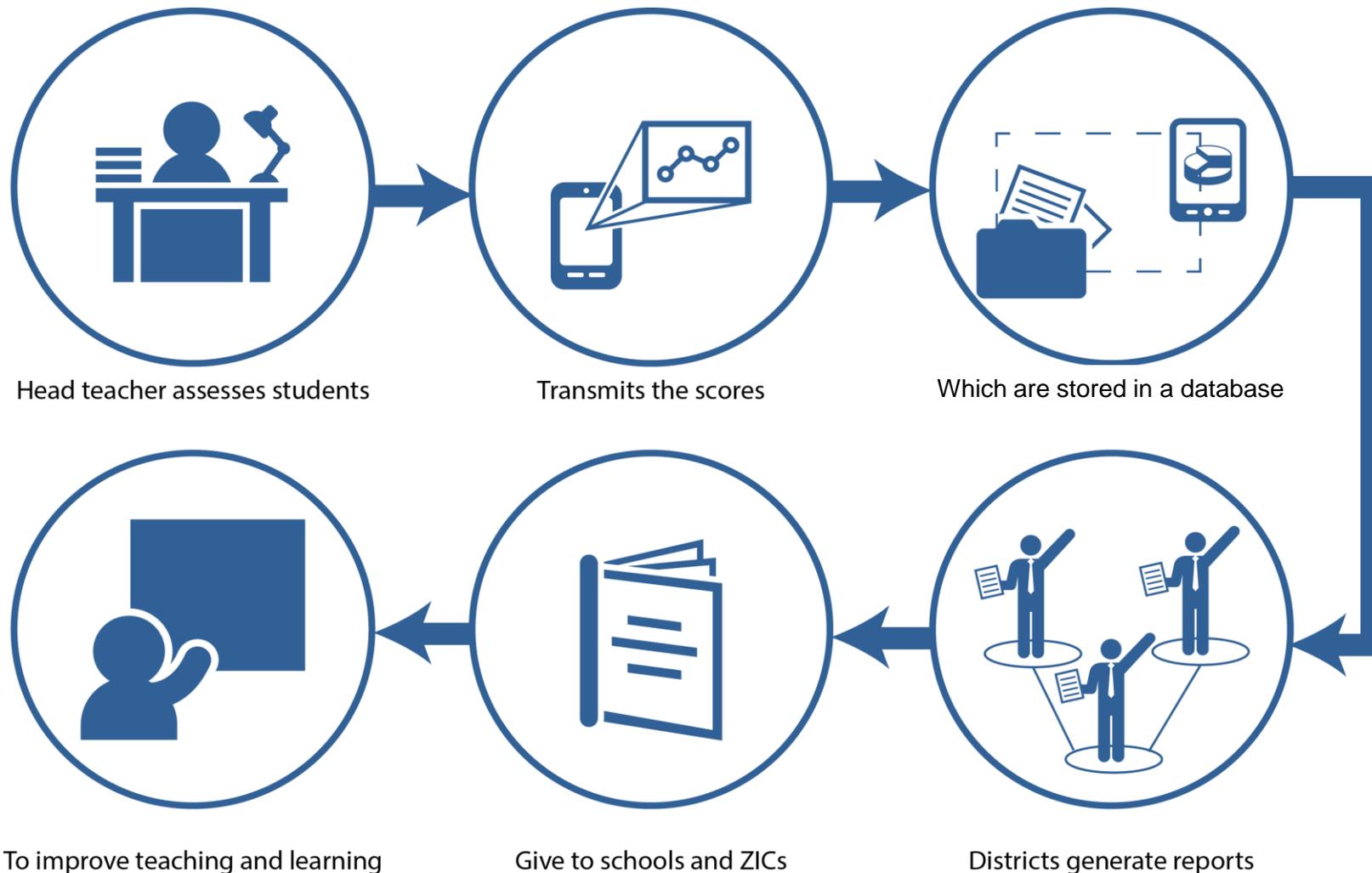
Framework element	How it applies
Population	<u>Two special</u> populations <ul style="list-style-type: none"><li>• Community Action Promoters and Recorders, “captive”</li><li>• Pregnant women (special but “non-captive” population)</li></ul> But ideal: <u>universal</u> coverage of the entire population, not sample
Population (reporter) profile	More literate, more likely to own phones
Speed	Relatively important
Accuracy	Very important
Cost	Important
1-way vs 2-way info flow	1-way, 2-way

Too complex for SMS. Smart phone allows “skip patterns.” Since reporter is not general pop., smart phones may be affordable. Reporting can lead to automatic tabulation, ranking of districts.

Can use for analysis, keeping health system accountable, tracking cases at individual or VDC level, communicating back down (in principle).

# Case 2: Implementation Assessment Via “Captive” Populations

Exactly parallel system in Zambia, for early grade reading



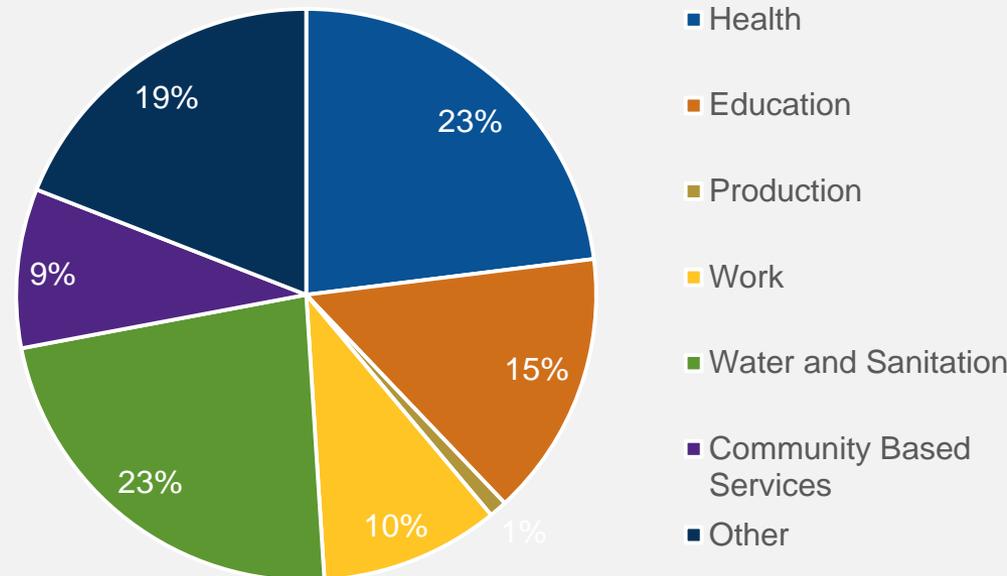
Source: M Rakusin, “Zambia School Gateway LQAS-modelled techniques for school feedback and support.”  
Presentation given at CIES 2015

# Case 3: Citizen's Complaints in Local Governance

## Citizen feedback on service delivery in Uganda: U-Bridge

How can you increase accountability at low cost?

Composition of Received U-Bridge Messages



### Solution: U-Bridge in collaboration with UNICEF's U-Report:

Free-to-citizen SMS →  
central location focal  
person (not automated)

Redirects to district head  
of department

Head of department acts  
and informs of case  
resolution

Sources:

"GAPP Project. U-Bridge Update." RTI International.

<http://www.ureport.ug/story/191/>

# Case 3: Citizen's Complaints in Local Governance

Framework element	How it applies
Population	General population
Population profile	More literate, more likely to own phones
Speed	Not key
Accuracy	Important? Maybe not!
Cost	Important
1-way vs 2-way info flow	1-way, <b><u>2-way</u></b>

SMSs originate with citizens (either any citizen or “reps of rights-holders”).

Big biases so can't necessarily use for policy conclusions.

But discuss: should not governance respond to those most concerned? Or is that a bit like “local elite capture?”

Case disposition communication (2-way) or results back down is key.  
Communication may not need to be with the individual complainer.

# Case 4: Target resource allocation via “professional” reporting

## Monitoring epidemics and endemic diseases.

Surveillance is expensive. As malaria rates have decreased, need to identify individual cases has increased.

### Solution:

“Coconut” software developed by RTI

Passive: clinic reports individual case via simple SMS to central location;

Active: automatic SMS goes out to lowest level (village) for further case detection and action; key: detect unsymptomatic cases

Village workers have tablets so as to capture more info

Data then uploaded to central server when 3G connection available; data  
▶ analysis, centralized statistics

Geo-location ▶ automated “hot spots”

Over 3 years, malaria prevalence decreased from 33% to < 1%

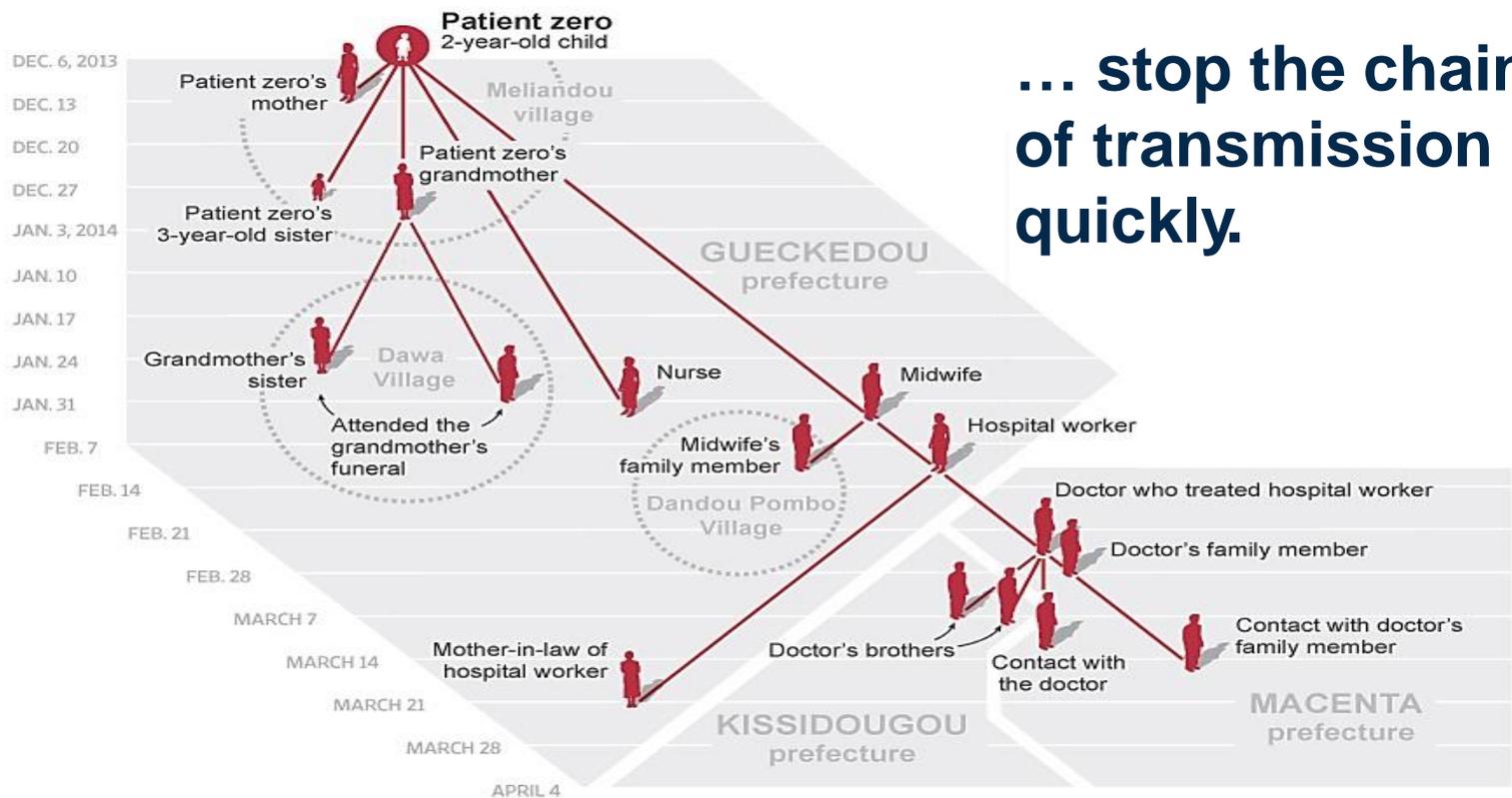
# Case 4: Target resource allocation via “professional” reporting

- Ideal:
  - Require minimum local infrastructure
  - Work offline or online
  - Synchronize case records across mobile devices, even when they are only occasionally connected
  - Effect both passive and active surveillance

**... stop the chain of transmission quickly.**

Reported, passive

Active



Sources: New England Journal of Medicine

The Wall Street Journal

# Case 4: Target resource allocation via “professional” reporting

Framework element	How it applies
Population	General, but respondent is “professional” or “empowered citizen”
Population profile	Slightly more literate, can be given a phone – not a sustainability issue
Speed	Key
Accuracy	Extremely important
Cost	Important
1-way vs 2-way info flow	1-way, <b>2-way</b>

SMS reporting system by “professionals”

Heavy two-way communication, leading to even more communication

Use tablets for complex final data report up

3G as opposed to SMS, update asynchronously

# 1 Notifying of Cases

Uses very simple, very cheap dumb-phone SMS or other means

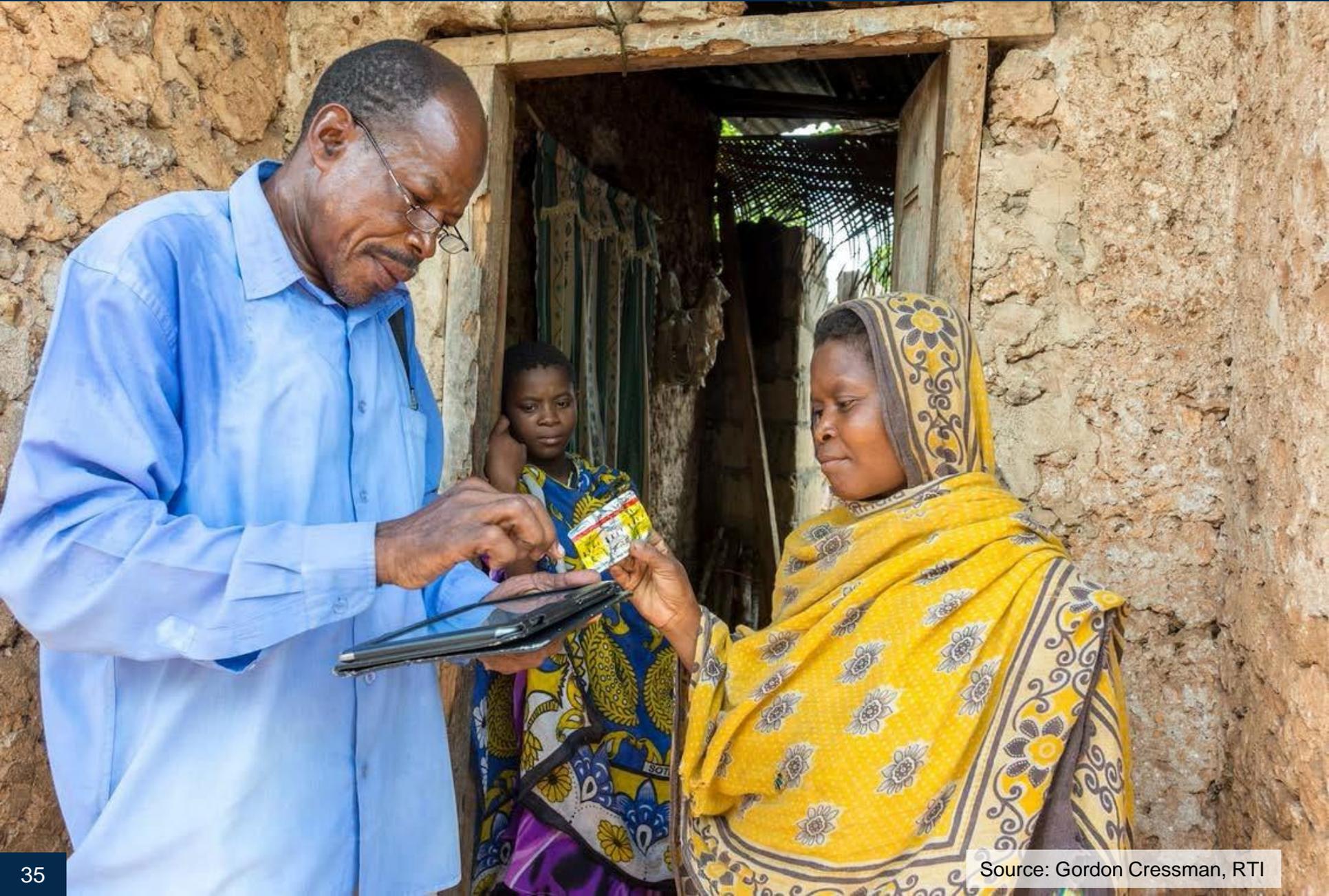


## 2 At a Facility

But the response requires more intensive data, recorded on tablet.



# 3 4 At a Household



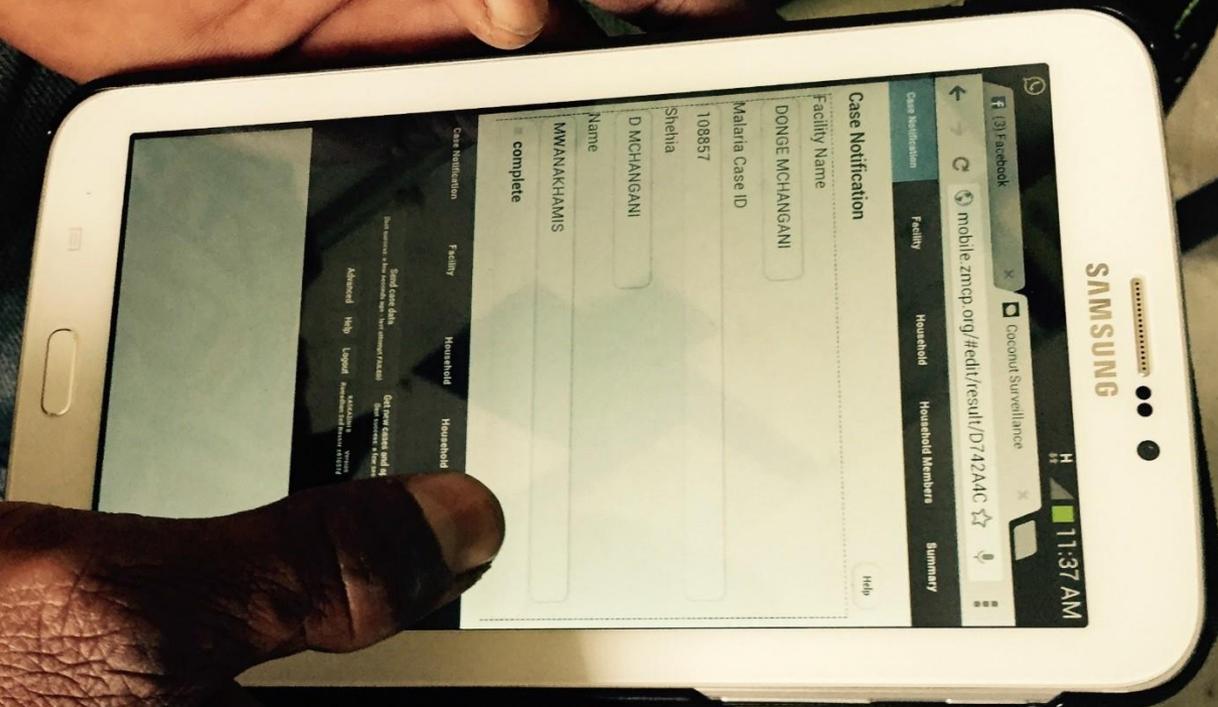
# 5 Tracing Contacts



Source: Gordon Cressman, RTI

# 6 Synchronizing the Data

Can be done using 3-G or 4-G, afterwards, asynchronously.



# Case 5: Electronic capture of learning assessment: Tangerine

## Tracking learner skills with paper tests is slow, error-prone

In addition, for targets pertaining to younger children, they may not be able to fill out forms.

### Solution: Tangerine

Tool developed online,  
downloaded to tablets.

Prototypes exist.

Students get paper  
stimulus

Tablet-based system for  
recording student response  
by assessor, not student

Data are uploaded via cell  
phone or WiFi.

Skip patterns and error  
traps are automatic

Results tabulation  
essentially instantaneous

Data can be hosted  
centrally if the user  
organization desires

Sampling uses standard  
random sampling  
approaches

Tool used in:

49 countries

64 languages

36 organizations

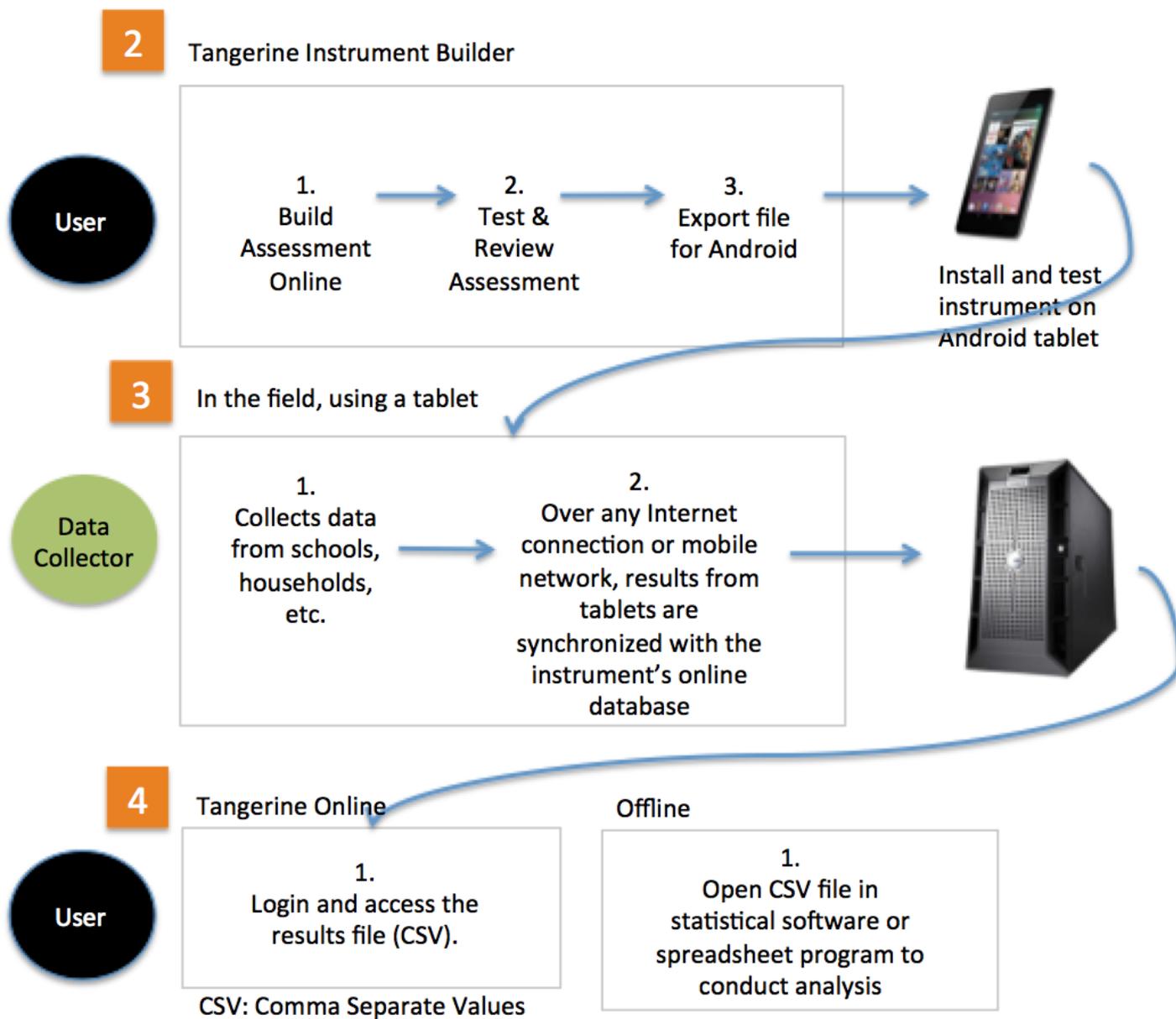
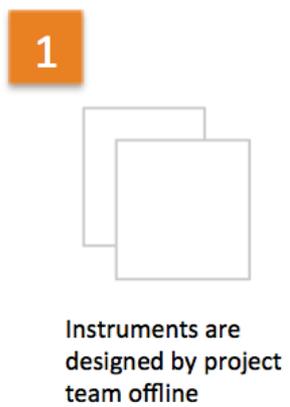
> 1000000 student  
records



EGRA and EGMA are orally administered.



The assessor records the responses of the pupil using a tablet and Tangerine software.



Note: Tangerine data files are very lightweight so as to facilitate a fast transfer.

# Case 5: Impact Evaluation Via Detailed Reporting and Feedback

Framework element	How it applies
Population	Captive population
Population profile	Not literate at all, oral assessment, recorded electronically
Speed	Vital
Accuracy	Extremely important
Cost	Less important
1-way vs 2-way info flow	Partial 2-way flow

Use classical sampling and learner assessment and psychometric techniques

Use technology (tablets, cell phone networks, WiFi) for speed, accuracy

# Case 6: Labor Market Information in Near-Real-Time

## Traditional Labor Market surveys can be slow, expensive

Typical case: maybe 30 pages of questions, \$100,000, 8 months to complete

### In Kenya

#### Approach

- Self-targeting survey
- Using SMS for speed and self-targeting
- 1266 responses
- 10 weeks from design to tabulation
- \$10,000 in cash costs
- In partnership with Geopoll platform

#### Results

- 46% earn < 2000 shillings/week: slightly above the poverty level
- 50% work in field other than their education and training
- 55% “dissatisfied” or “very dissatisfied” with job
- 62% temporary
- 45% part time.
- Took on average of 11 months to find work

#### Results

- Yet, 65% would recommend their training institution
- Why?
- We don't know, but a \$10,000 survey raised intriguing questions
- Can be followed up with more traditional studies

# Case 6: Labor Market Information in Near-Real-Time

Framework element	How it applies
Population	Special but not “ <i>captive</i> ” population (youth 16-35) with some secondary education, <b><i>employed</i></b>
Population profile	Quite literate, likely to be cell-phone literate
Speed	Quite important
Accuracy	Rather important
Cost	Important
1-way vs 2-way info flow	1-way electronically, 2-way in person

Try SMS: very efficient at targeting: population can opt out if not the right special population

But correct for biases using more stratification and weighting

# Case 7: Shadowing of “Little Data” Paper Systems with e-Systems

**Traditional EMIS or HMIS systems: slow, inaccurate, no 2-way comms**  
Ministers, others: move fast to e-Systems: is it wise? Feasible?

## In Kenya

### Approach

- Reporting via paper improved in recent years (UNICEF, MOEST)
- Move suddenly?
- Perhaps pilot
- With cell phones
- With tablets
- ***Not an RCT!***

### Results

- Electronic submission mostly but ***not*** perfectly feasible
- Usual expected problems: connectivity, loss of equipment, malfunctions
- Key: e-reporting more accurate when based on paper forms.
- Recommend dual approach for now
- Accuracy seems to improve with electronic submission, feedback can correct
- Many of the issues encountered apply to paper systems as well, not specific to e-systems
- Key: e-reporting more accurate when based on paper forms.

# Case 7: Shadowing of “Little Data” Paper Systems with e-Systems

Framework element	How it applies
Population	Special, “captive,” population: head teachers, but have to have <i>universal</i> coverage
Population profile	Quite literate, likely to be cell-phone literate
Speed	Quite important
Accuracy	Very important
Cost	Important
1-way vs 2-way info flow	1-way electronically, 2-way with paper

Try SMS, tablets, using 3G or WiFi.

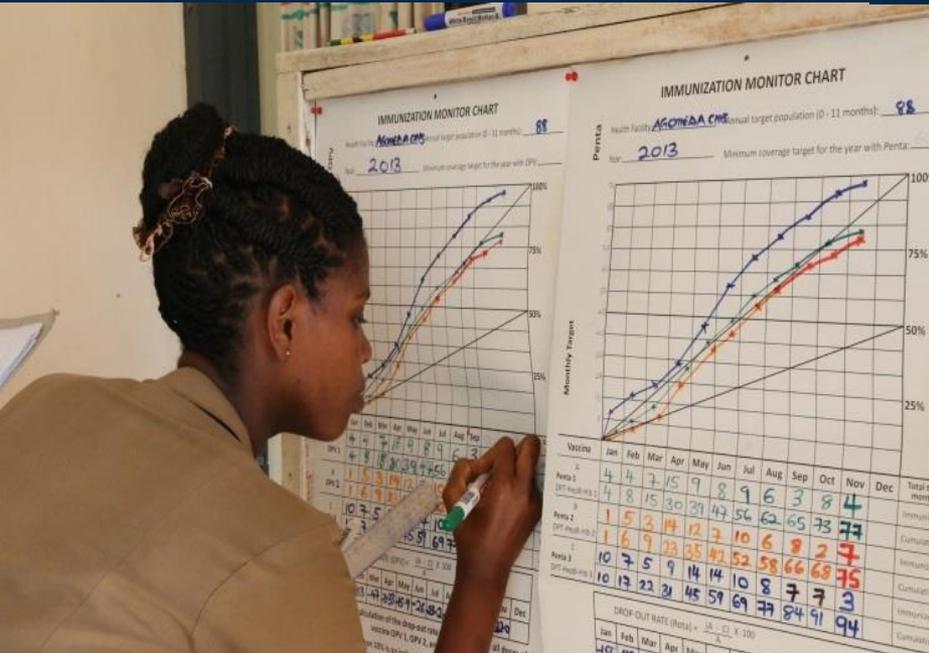
No need to correct for response biases: “captive” population

Mix with feedback back down with paper or e-systems

# 3

## Integrated Illustrative Example

### DHIS2 health information system (U of Oslo)

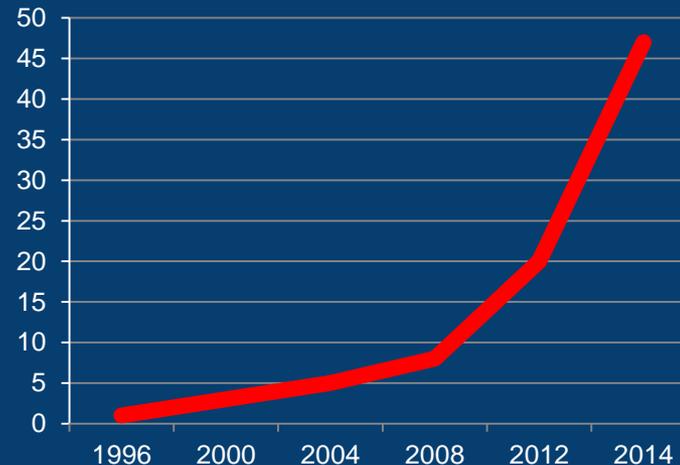
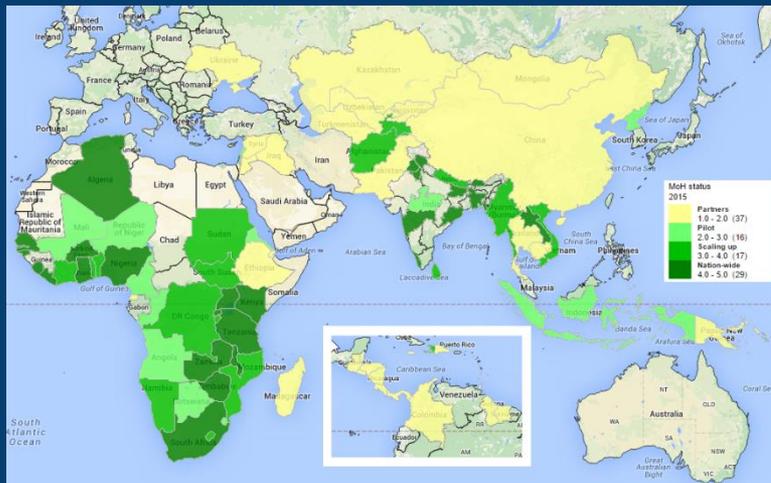


# Illustrative Integrated Example

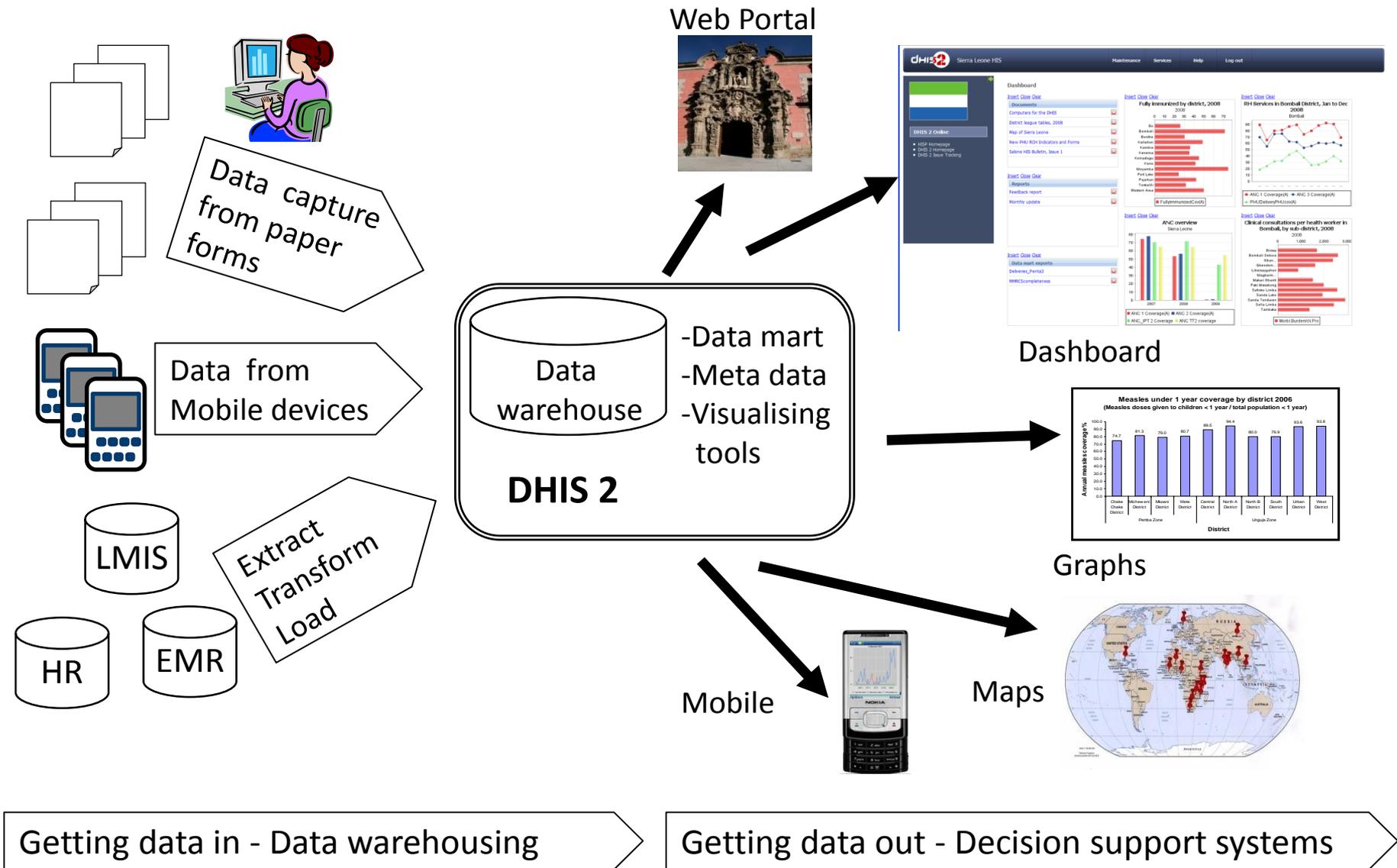
**Problem:** Traditional, paper-based HMIS systems very slow. But efforts by international agencies to computerize → fragmentation, non-ownership by governments, non-sustainability.

**Solution:** DHIS 2 is a HMIS developed at U of Oslo that:

- Grew on demand-led basis (first: post-apartheid decentralization in South Africa)
- Started small, solved real problems
- Relied on demand-side for spread, not top down imposition or selling
- Grew through “network externalities”
  - The more users and producers, the more existing users benefit
  - From one district to 50 countries



# How the system works



# Other important features or aspects

- Not a software “product” but more of an ecosystem
- Built on action research model
- Open source, many developers
- Builds on HMIS legacy that exists
- Ecosystem includes “academies” and training processes
  - Including PhDs and Master’s at U of Oslo
- Use mHealth to extend and strengthen the existing national health systems
  - Support the widest range of available technologies, but led from the need and demand
- Implement a scalable approach from day one
- Use the local Ministry of Health as a coordination body
- Pursue partnerships to gain momentum for integration
- Avoid exotic features and nice-to-have
- Release often: fixed time - flexible scope

# Conclusions - 1

- Demand for data: bigger constraint than supply
- Yet, technology: supply
  - Can be used to stimulate demand
- Need to control for biases, when accuracy is important
  - But it is not always important, or as important as speed and low cost
  - Can control for biases using various techniques such as stratification and weighing, but only once you know the biases
  - Bias in some cases (dumb phones) comes from non-response rather than access to technology
  - Bias with social media likely to be due to technology, for now
    - Pulse Lab: interesting things like digitizing community FM radio for natural language sentiment analysis
- Note: SMS or phone surveys may be cheap per completed questionnaire, but not per completed high-quality question\*
- “Things are changing” – but technologists always ahead of people: permanent “supply push” problem and need to be cautious and skeptical, yet optimistic
- Cases of appropriate, demand-led, sustainable use exist

\*This is not often noted. A 20 question SMS survey can cost \$10,000, but a 300-question survey that costs \$100,000 can have 15 times more info for 10 times the cost. Question is: do you really need 300 questions?

# Conclusions - 2

- A good use of technology is to bolster “little data” and legacy systems (EMIS, HMIS)
- Integration of current legacy (and new) systems to do predictive analytics is an useful frontier—not difficult
- Data enhancement: needs tie-in to demonstrations of precisely how improved management and accountability use the data, and how it improves results
  - Requires tying to existing accountability systems
  - Cannot typically “invent” (out of whole cloth) the accountability approach to pressure, if none exists, so as to also run a technology effort
  - Rely on existing technology: don’t try to create the technology platform and then use it for improving management (i.e., giving average citizens cell phones)
- Combinations, even within same project, are probably the way to go for now:
  - Paper and cell phones, depending on circumstances and use
  - Apps and radio, especially community radio
  - Face to face traditional surveys and SMS surveys, etc.

# More Information

## Luis Crouch

Vice President and Chief Technical Officer,  
International Development Group

202.728.2058

lcrouch@rti.org