

Data in the First Mile

Kuang Chen

kuangc@cs.berkeley.edu

With Joe Hellerstein and Tapan Parikh

CIDR 2011

NEW YORK TIMES BESTSELLER

TRACY KIDDER

WINNER OF THE PULITZER PRIZE

MOUNTAINS BEYOND MOUNTAINS

THE QUEST OF DR. PAUL FARMER,
A MAN WHO WOULD CURE
THE WORLD



"INSPIRING, DISTURBING, DARING AND COMPLETELY ABSORBING."
—ABRAHAM VERGHESE, THE NEW YORK TIMES BOOK REVIEW



NEW YORK TIMES BESTSELLER

TRACY KIDDER

WINNER OF THE PULITZER PRIZE

MOUNTAINS BEYOND MOUNTAINS

THE QUEST OF DR. PAUL FARMER,
A MAN WHO WOULD CURE
THE WORLD



"INSPIRING, DISTURBING, DARING AND COMPLETELY ABSORBING."
—ABRAHAM VERGHESE, THE NEW YORK TIMES BOOK REVIEW

“database research and the data management industry are at a turning point, with unusually rich opportunities for technical advances, intellectual achievement, entrepreneurship and impact on science and society”

- Claremont Report, 2008

Data management in developing regions

Talk outline

Challenges

Research directions
(and examples)



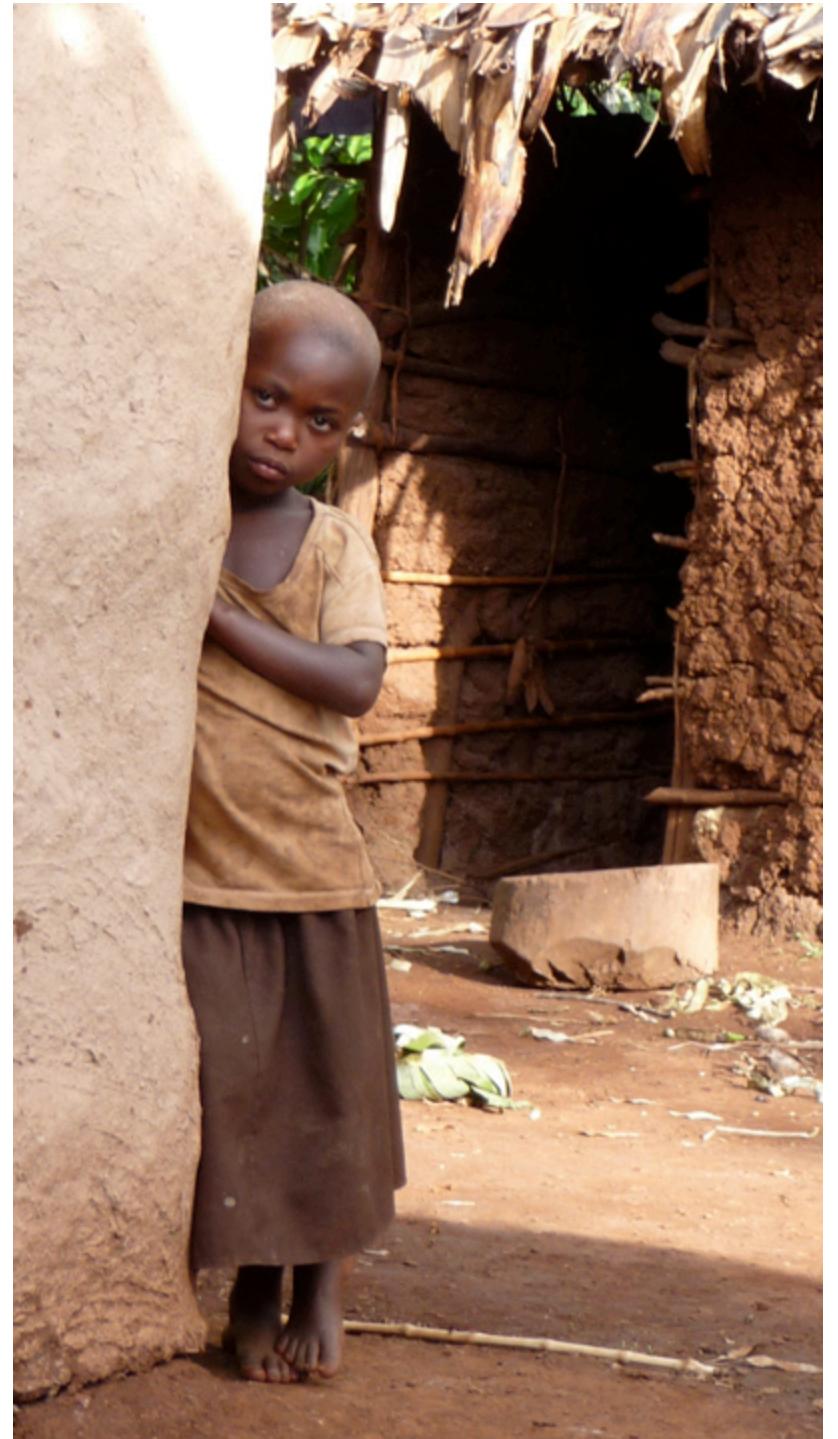
**RUHIIRA
HEALTH CENTRE**
ISINGIRO DISTRICT

SUPPORTED BY MILLENNIUM VILLAGES PROJECT


**Family Planning Services
Available Here**

76% of children born in sub-Saharan Africa are unregistered

- UNICEF, 2008.



A photograph of a person in dark clothing and a safety harness working on a tall, thin metal pole. The person is positioned near the top of the pole, which is surrounded by several large, white, ribbed satellite dish antennas. The background is a clear blue sky with a few wispy white clouds. The perspective is from below, looking up at the pole.

The last mile

A photograph showing a woman from the side, wearing a red and white patterned dress, holding a black smartphone. She is interacting with another person whose back is to the camera. The background is a stone wall.

The first mile

(< 200 cells/mm³)

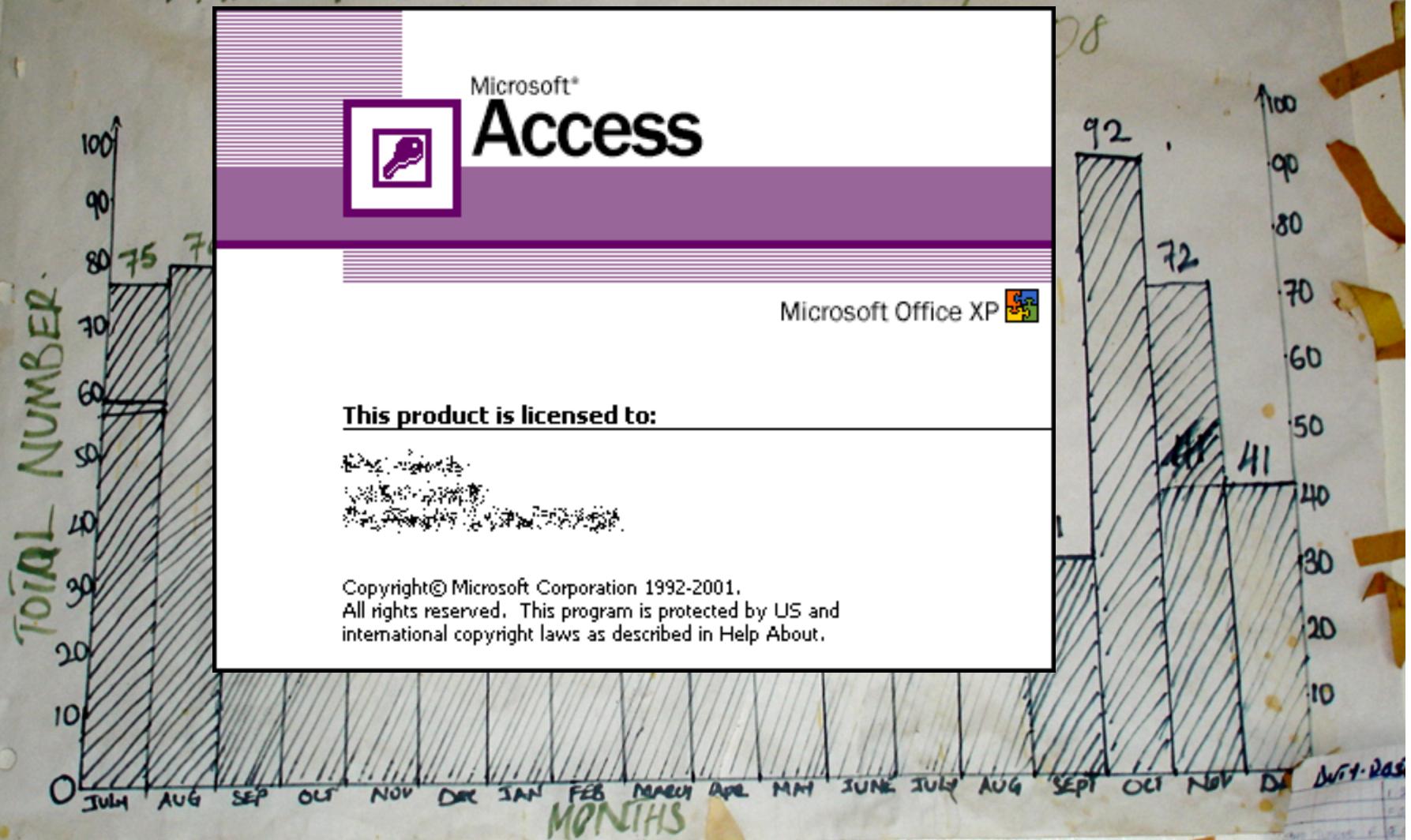
Note that these are interim recommendations and may be subject to change.
Total lymphocyte count (TLC) has been proposed as surrogate marker or
alternative to CD4 cell counts or CD4% in resource-constrained settings.

*Conditions requiring diagnosis by doctor or medical officer—should be referred to a medical facility to confirm HIV infection. ART may be initiated.
In a child with Presumptive diagnosis of severe HIV disease, where it is not possible to confirm HIV infection, ART may be initiated.

QA Quality Assurance Project

USAID|UGANDA

BAR GRAPH SHOWING MALARIA CASES



DUTY ROSTER. MAT. OCTOBER. 2009

Data management in developing regions

Talk outline

Challenges

Research directions
(and examples)

A photograph of a classroom or meeting room. A man in an orange and white striped shirt stands on the right side, holding a piece of paper and reading from it. He is facing a group of people who are seated on the floor, also holding papers. The room has yellow walls and several windows with bars. A sign on the wall above the windows reads "GOD IS WITH US" and "JESUS CHRIST".

Lack Data Expertise

Storytellers vs. Structured Data





Local Data Usability

Access DBs
everywhere...

Periodic report vs.
Actionable datum

Report success vs.
Analyze failure

Database Research Assumptions



| Database research | Developing world |
|-----------------------------|-----------------------------------|
| Too much data – privacy | Not enough data – data collection |
| Competent DBAs | Lacking technical skills |
| Many research opportunities | No novel research in applications |

Data management in developing regions

Talk outline

Challenges

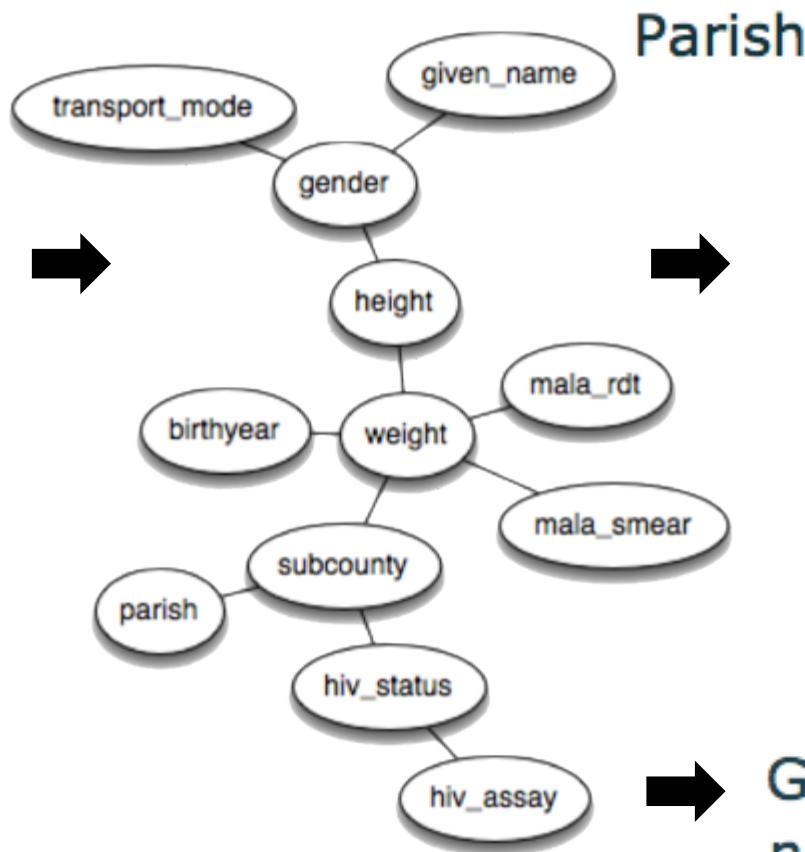
**Research directions
(and examples)**

Goal:

**Automatically increase quality and
reduce expertise required in local
data management**

USHER

- CHEN ET AL. ICDE 2010, UIST 2010



Parish

Select value

- Nyakitunda
- Ruhiira
- NA--
- Bugongi
- Kabugu
- Kisyoro

Given
name

1234

WARNING! CHECK YOUR ANSWER!

Direction: **Intelligent Automation**

Use data-driven
techniques to:

Reduce expertise and
training required

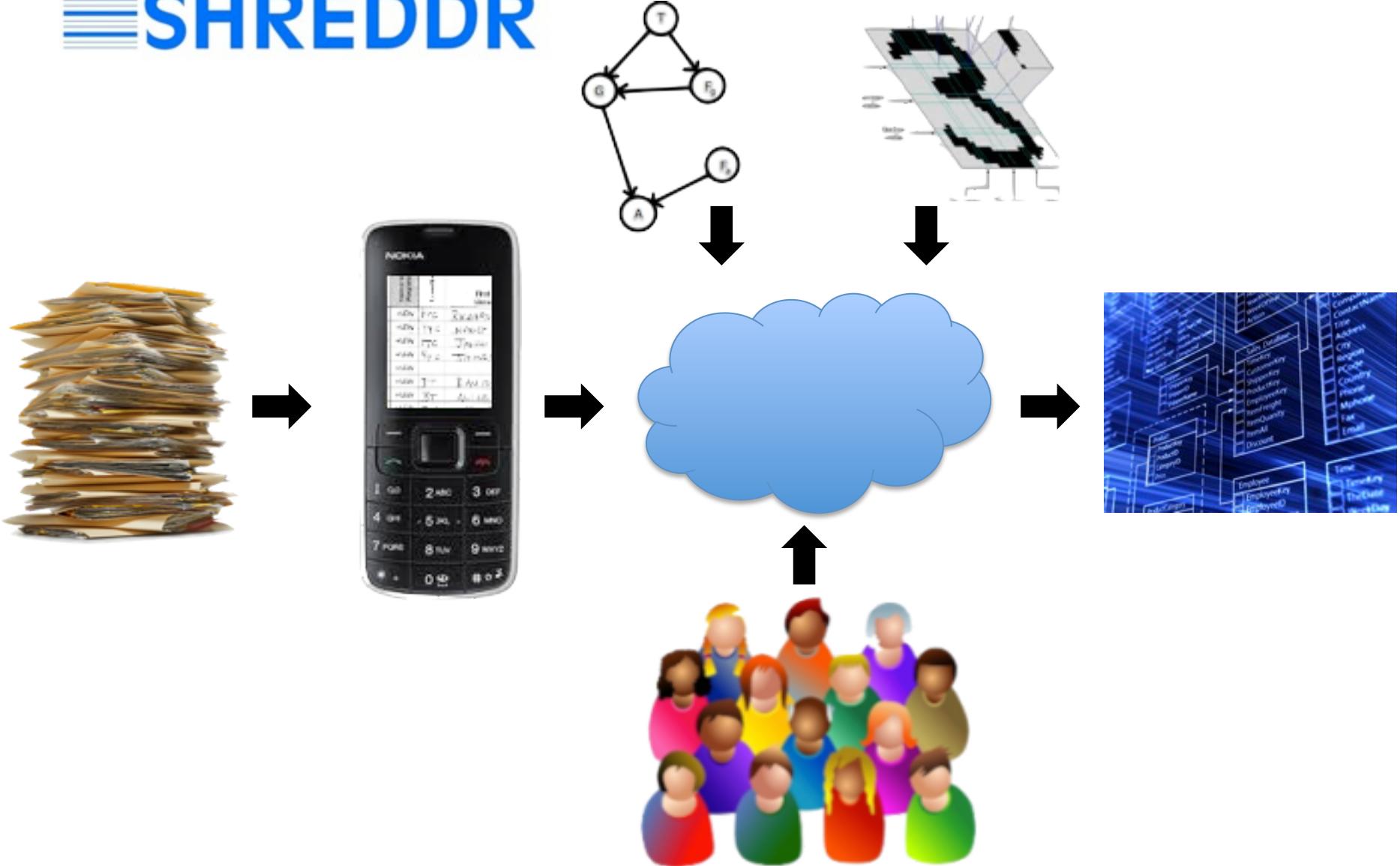
Improve data entry
efficiency & accuracy

Identify actionable
anomalies

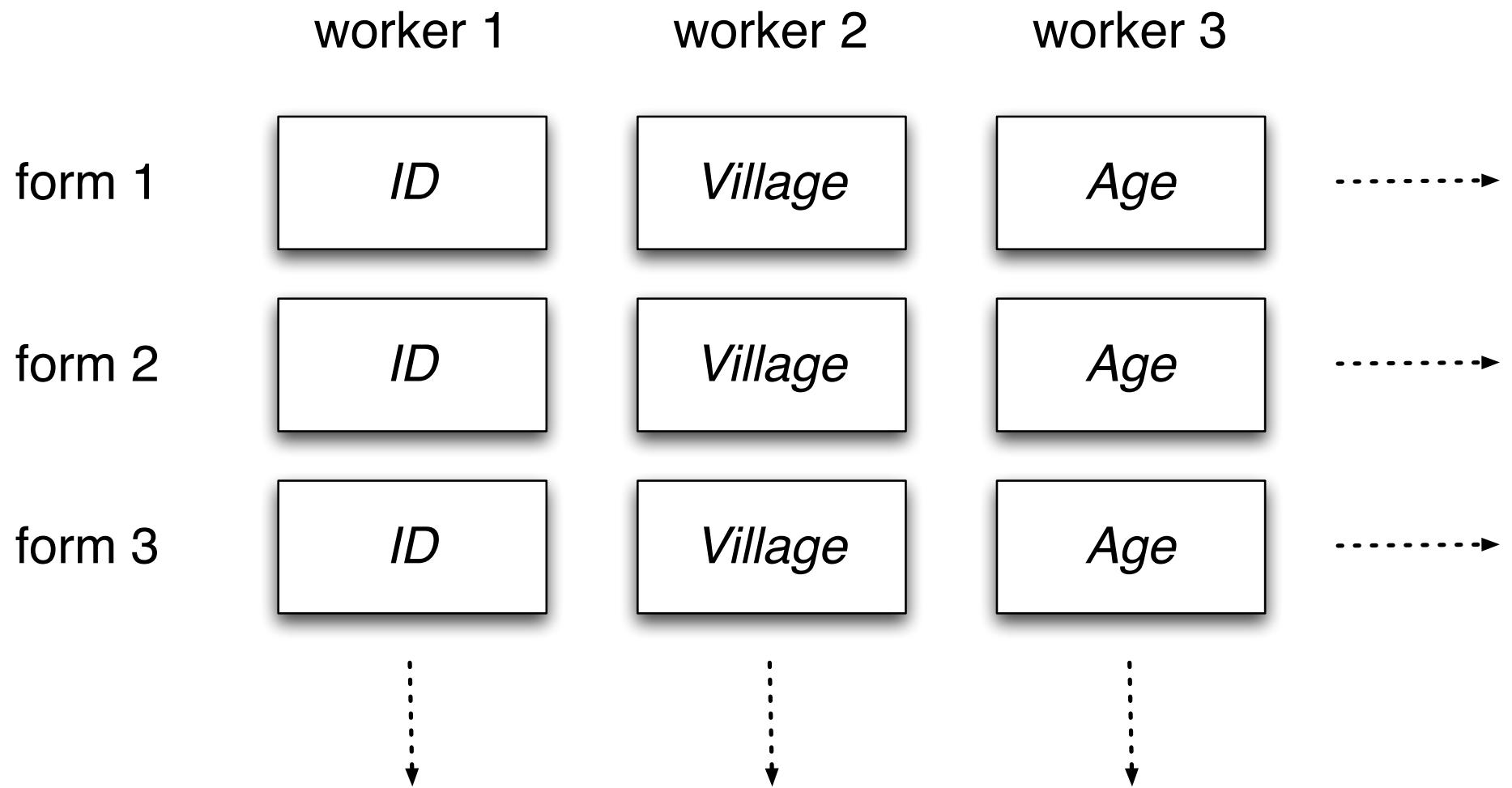
Goal:

**Improve digitization efficiency and
local data access**

SHREDDR



Column-oriented Data Entry



Column-oriented Data Entry

Select the values are not: Michael

| | | | | | | |
|---------|---------|---------|---------|---------|---------|---------|
| Michael | MICHAEL | Michael | Michael | PHILLIP | Michael | Michael |
| Michael | Mike | Michael | Michael | Michele | Michael | Steve |
| Michael | Michael | Wesley | Michael | Michael | Michael | Michael |
| Michael | Michael | Michael | Michael | Michal | Michael | Michael |
| Michael | Michele | Michael | Mike | Michael | Michael | MICHAEL |

Direction: **Break apart the workflow**

More appropriate **capture**
interfaces for field workers:
paper, audio, sensors

Move **structuring** to
appropriate workers

Pool little data to make big
data

Create new opportunities to
add value

Summary

First Mile Challenges

Lack of data expertise

Oral cultures

Local data usability

Data research assumptions

Directions

Intelligent automation

Breaking up workflow to leveraging the cloud and crowd

A photograph of a sunset over a city skyline. The sky is filled with warm orange and yellow hues, with some darker clouds at the top. In the foreground, there are silhouettes of trees and the dark roofs of buildings. The sun is low on the horizon, casting a bright glow.

How to Get
Started

Find a local
partner

Collaborate

Solve a real
problem



A photograph of a sunset over a city skyline. The sky is filled with warm orange and yellow clouds. In the foreground, there are silhouettes of trees and the dark roofs of buildings. Overlaid on the center of the image is the word "Questions" in a large, white, sans-serif font.

Questions

Backup



Innovation pile-up

SHREDDR

ChildCount+ Form A: REGISTRATION
Focus on: Household Heads, Children Under 5, and Pregnant Women

GENERAL REGISTRATION For any household member without a HEALTH ID that needs to be registered in the CHW data system.

| Date | Health ID | First Name | Family Name (Surname) | Date of Birth (DD/MM/YY) | Age (in yrs) | Household Head HEALTH ID |
|------|-----------|------------|-----------------------|--------------------------|--------------|--------------------------|
| | H-A001 | | | | | |
| | H-A002 | | | | | |
| | H-A003 | | | | | |
| | H-A004 | | | | | |
| | H-A005 | | | | | |
| | H-A006 | | | | | |
| | H-A007 | | | | | |
| | H-A008 | | | | | |
| | H-A009 | | | | | |
| | H-A010 | | | | | |
| | H-A011 | | | | | |
| | H-A012 | | | | | |
| | H-A013 | | | | | |
| | H-A014 | | | | | |
| | H-A015 | | | | | |
| | H-A016 | | | | | |
| | H-A017 | | | | | |
| | H-A018 | | | | | |
| | H-A019 | | | | | |
| | H-A020 | | | | | |

DEATH --> STILL BORN / MISCARRIAGE:
Death with HEALTH ID Death without HEALTH ID
 Death with HEALTH ID Death without HEALTH ID
 Health ID Reg. Date of Death (DD/MM/YY) First Name (Surname) Date Birth Date (DD/MM/YY) Age (in yrs) Date of Death (DD/MM/YY) Household Head HEALTH ID
 H-OB1 +OB1 H-OB2 +OB2 H-OB3 +OB3 H-OB4 +OB4 H-OB5 +OB5 H-OB6 +OB6 H-OB7 +OB7 H-OB8 +OB8 H-OB9 +OB9 H-OB10 +OB10 H-OB11 +OB11 H-OB12 +OB12 H-OB13 +OB13 H-OB14 +OB14 H-OB15 +OB15 H-OB16 +OB16 H-OB17 +OB17 H-OB18 +OB18 H-OB19 +OB19 H-OB20 +OB20

CHW Name: _____

CHW Number: _____

ded in the CHW data system.

| Mother's HEALTH ID [U = Unknown] | Delivered in Health Facility? (Y-N-U) | Weight at Birth [in KG] | Mobile Phone | Mobile Phone Number |
|--|---|----------------------------|--------------|------------------------|
| U | X | X | +MOB | X |
| U | X | X | +MOB | |
| U | X | X | +MOB | |