



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.





The last mile

A photograph showing a woman with dark skin and short hair, wearing a white dress with yellow and black patterns and a red shawl. She is sitting on a stone path, looking down at a smartphone held by another person. The person holding the phone has dark hair and is wearing glasses. The background is a stone wall and a paved path. The text "The first mile" is overlaid in the center of the image.

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.

USAID | UGANDA

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

QAdo Quality Assurance Project

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 man in a plaid shirt standing and reading from a paper to a group of people seated in a room with barred windows. The text "Lack Data Expertise" is overlaid on the image.

Lack Data Expertise

Storytellers vs. Structured Data





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

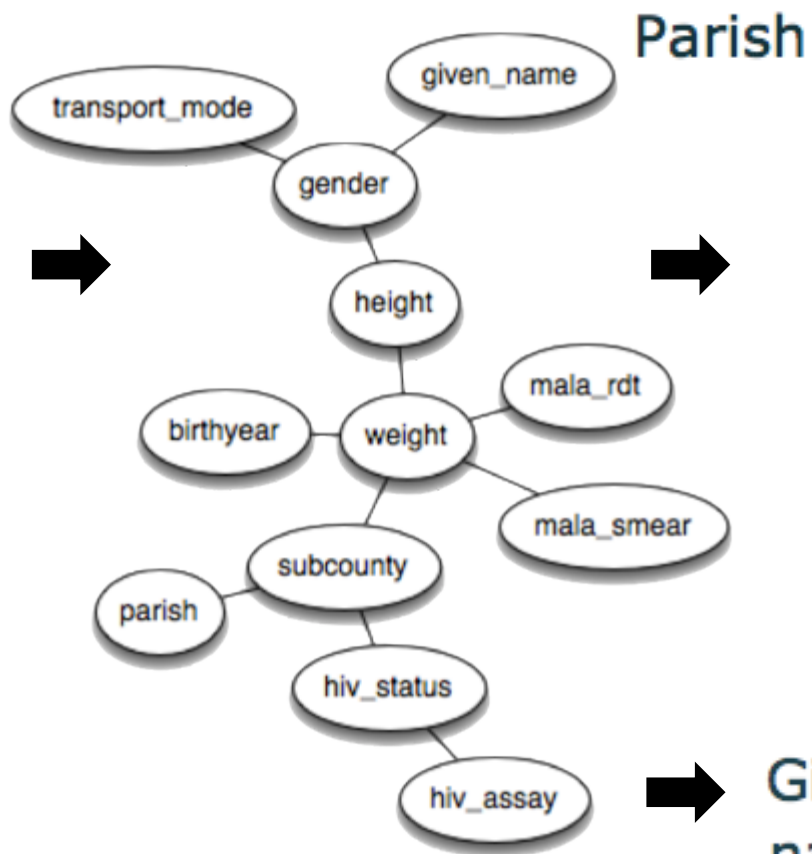
**Research directions
(and examples)**

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

Kisyo

Given
name

1234

WARNING! CHECK YOUR ANSWER!

Direction:

Intelligent Automation

Use data-driven
techniques to:

Reduce expertise and
training required

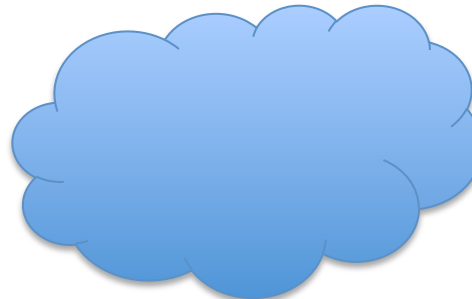
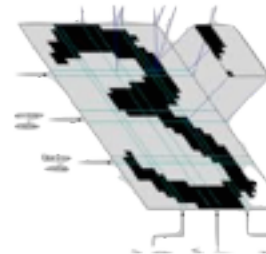
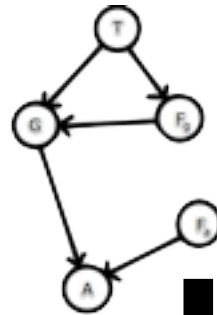
Improve data entry
efficiency & accuracy

Identify actionable
anomalies

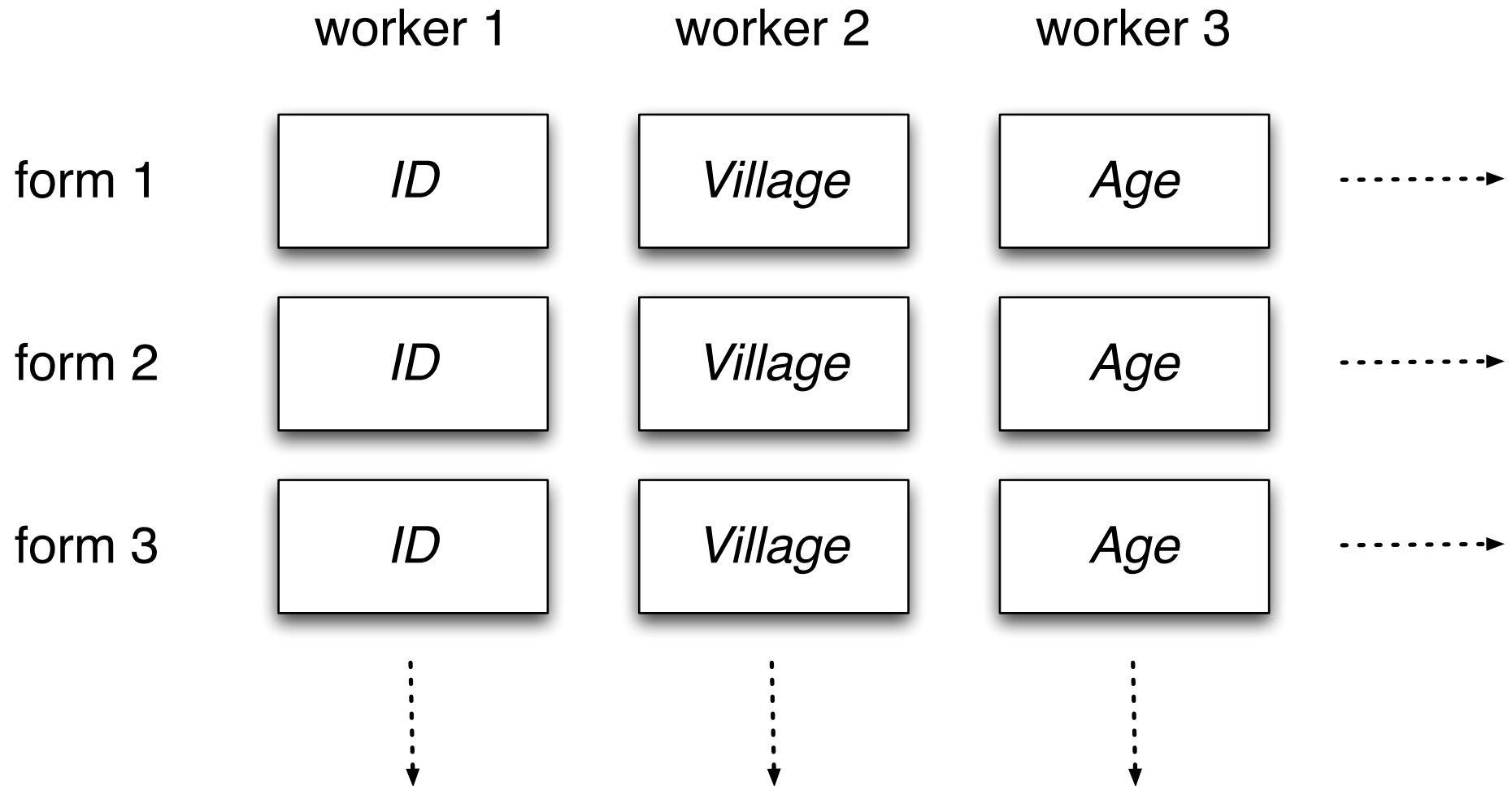
Goal:

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	Thirford	Michele	Michael	Steve
Michael	Michael	Wesley	Michael	Michael	Michael	Michael
Thirford	Michael	Thirford	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. The sun is a bright orange circle on the horizon, surrounded by a ring of orange and yellow clouds. The sky is a mix of orange, yellow, and grey. In the foreground, there is a dark, corrugated metal roof. In the background, there are silhouettes of trees and buildings.

**Find a local
partner**

**How to Get
Started**

Collaborate

**Solve a real
problem**

Questions

A photograph of a sunset or sunrise over a city. The sun is a bright orange-yellow orb partially obscured by a thin layer of clouds. The sky is filled with soft, wispy clouds in shades of orange, pink, and light blue. In the foreground, the dark silhouettes of trees and a corrugated metal roof are visible. The word "Questions" is written in a large, white, sans-serif font with a subtle drop shadow, centered over the sun.

Backup



SHREDDR

CHW Name:

CHW Number:

ded in the CHW data system.

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 recorded in the CHW data system.

Date	HEALTH ID	Registration Code	Location Code	First Name	Family Name (Surname)	Sex (M/F)	Birth Date (DDMMYY)	Age (in Years)	Household Head HEALTH ID (If = Person in HH Head)	Mother's HEALTH ID (If = Unknown)	Delivered in Health Facility (Y/N/U)	Weight at Birth (in KG)	Mobile Phone Number
		+M00											
		+M00											
		+M00											
		+M00											
		+M00											
		+M00											
		+M00											
		+M00											
		+M00											
		+M00											
		+M00											
		+M00											
		+M00											
		+M00											
		+M00											
		+M00											
		+M00											
		+M00											
		+M00											
		+M00											

DEATH --or-- STILLBORN / MISCARRIAGE

Death with HEALTH ID

HEALTH ID	Sex	Date of Death (DDMMYY)
+00A		
+00A		
+00A		

Death without HEALTH ID

Sex	First Name	Family Name (Surname)	Birth Date (DDMMYY) or Age (in Years)	Date of Death (DDMMYY)	Household Head HEALTH ID

Stillbirth / Miscarriage

Mother's HEALTH ID	Sex	Date of Death (DDMMYY)	Type
+00B			
+00B			
+00B			

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