

# Digitizing Paper Forms with Mobile Technologies

Paper forms are widely-used for data collection.



But accessing and analyzing data collected on paper is difficult.

Manual data entry is time-consuming  
and error-prone.

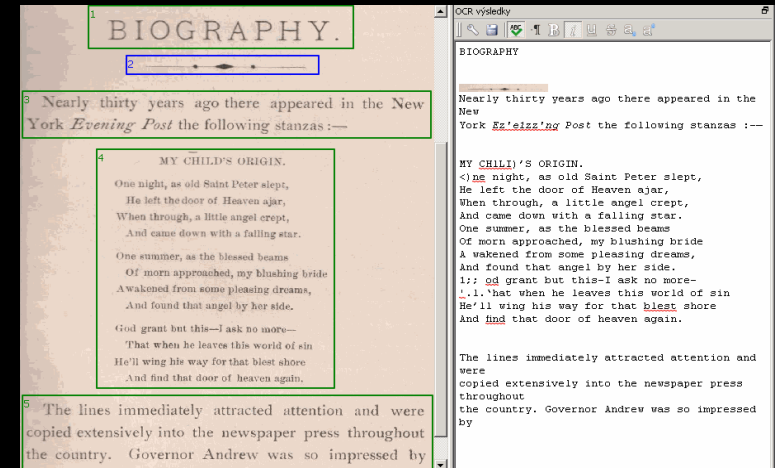


We need better ways to collect digital  
data from paper forms.

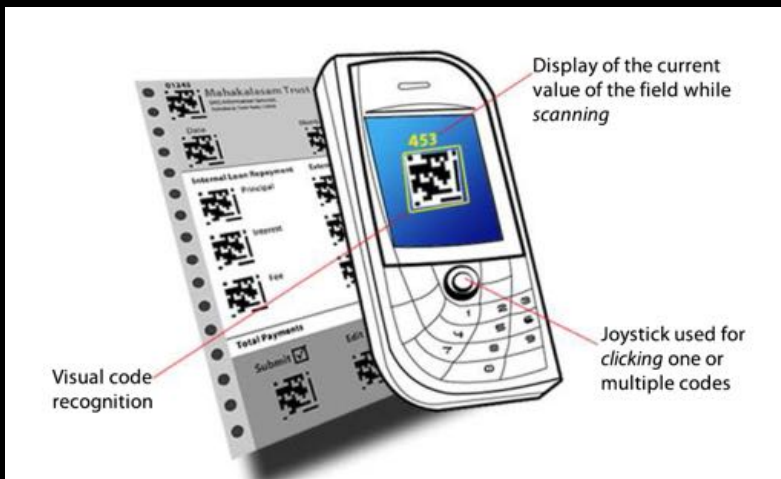
# Prior Work



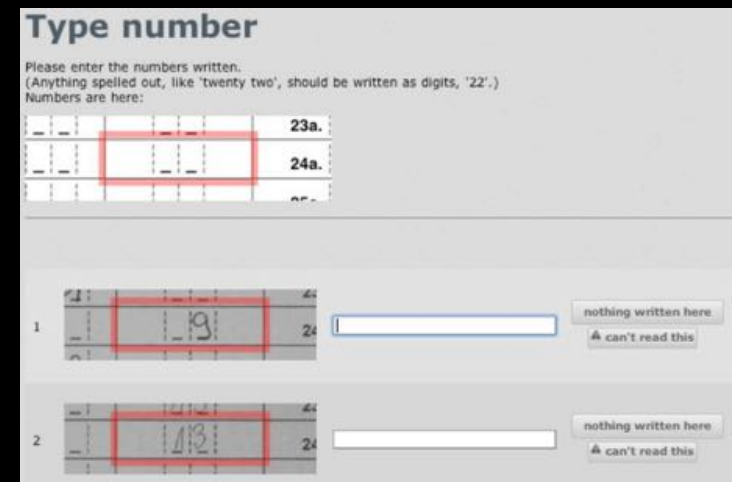
Optical Mark Recognition (OMR)



Optical Character Recognition (OCR)



CAM [Parikh et al.]



Shreddr/Captricity [Chen et al.]



## Goal:

Design and build a mobile system to efficiently digitize data from paper forms.

## Key idea:

Capture images of paper forms using the built-in camera on a mobile device.

Automatically interpret data using computer vision and machine-learning techniques.



# Goal:

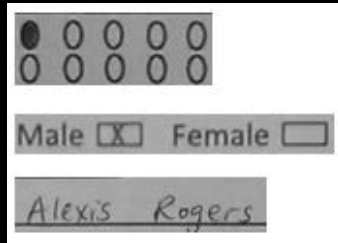
Design and build a mobile system to efficiently digitize data from paper forms.

# Challenges:

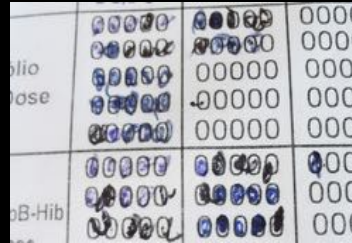
Wide range of domains and applications



Variety of data types



Messy forms, varied lighting



Limited connectivity, processing power



Human challenges

# Understanding paper-digital workflows in global development organizations.

Collected data from 23 development organizations in 16 countries.

## Design

- Decide survey content
- Decide layout and data types
- Approval processes (e.g., IRB)



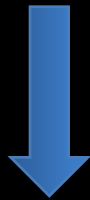
## Analysis

- Transmit data to analysts
- Data cleaning and aggregation
- Analysis and decision-making



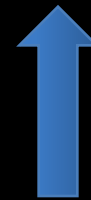
## Field Work

- Print and distribute surveys
- Field workers fill out surveys



## Data Entry

- Collect and transport filled surveys
- Manually transcribe data
- Validation and data checking



# Make it easy for people to create machine-readable paper forms

User designs form in browser

Automatically generate files

The screenshot shows the ODK Scan web interface. On the left, there's a 'Form Properties' sidebar with fields for Name, Label, Bubble size, Bubble type, Number of columns, Number of rows, Grid values, Include Border?, Verify Field?, Order Of Fields, and Margins. The main area displays a form titled 'VACCINES GIVEN AT BIRTH' with fields for Vaccine name, Vaccine given, Date given, and Nurse initials. The interface includes a menu bar with options like ODK Scan, File, Edit, Add, Copy, Delete Field, Undo Delete, Change Position, and Align Field.



Printable  
form image

The printable form image shows a form titled 'VACCINES GIVEN AT BIRTH' with fields for Vaccine name, Vaccine given, Date given, and Nurse initials. It includes a header with the Nigerian coat of arms and a footer with a logo.

Form  
description

```
{
  "type": "select1",
  "name": "given_bcg",
  "label": "given_bcg",
  "verify": "yes",
  "classifier": {
    "classifier_height": 20,
    "classifier_width": 20,
    "training_data_uri": "bubbles",
    "classification_map": {
      "empty": false
    }
  },
  "default_classification": true,
  "advanced": {
    "flip_training_data": false
  }
},
{
  "param": "yes_no",
  "grid_values": [
    "yes"
  ],
  "segments": [
    {
      "segment_x": 150,
      "segment_y": 200
    }
  ]
}
```

Database  
table definition

The diagram shows a green arrow pointing to a grid structure, representing the database table definition.

Editable  
digital form

begin screen			
begin screen	read_only	nextvaccination_at6weeks_image0	TRUE
begin screen	string	nextvaccination_at6weeks	
begin screen	read_only	bcg_datagiven_image0	TRUE
begin screen	string	bcg_datagiven	
begin screen	read_only	opv0_datagiven_image0	TRUE
begin screen	string	opv0_datagiven	
begin screen	read_only	hepb0_datagiven_image0	TRUE
begin screen	string	hepb0_datagiven	
begin screen	read_only	given_addtvacc_2_image0	TRUE
begin screen	select_one	given_addtvacc_2	TRUE
begin screen	read_only	nurse_notes_image0	TRUE
begin screen	string	nurse_note	
begin screen			



# Algorithm for digitizing forms

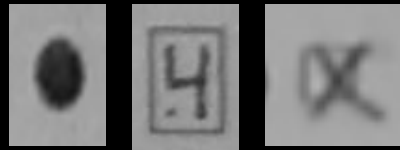
1. Capture form image



5. Save and visualize data.



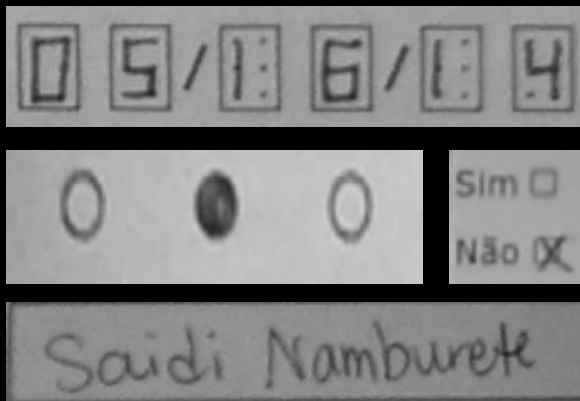
3. Classify machine-readable data types.



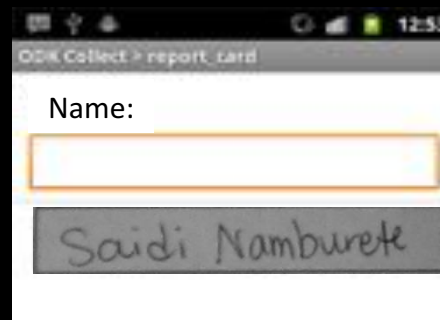
6. Sync with server.



2. Align and segment image.



4. (Optionally) transcribe text



# Technical Evaluation

## Classification accuracy:

Bubbles: >99%

Checkboxes: > 99%

Numbers: >99%



## Data Transcription:

>10% faster with image snippets

## Pilot Test:

Digitize clinics' vaccine statistics with >99% accuracy in 30 seconds.



# Field Deployment: Tracking health worker usage of medical supplies in Mozambique

Community health workers provide essential services to rural communities.

Health workers need a reliable supply of medicines.



## Challenges:

No standardized resupply process.

No system for health workers to track their supply usage.

No system for collecting and reporting usage data.

# Field Deployment: Tracking health worker usage of medical supplies in Mozambique

Health workers collect supplies and paper forms.



Health workers dispense supplies to patients and fill out the form.



The supervisor digitizes the forms and transmits the data.



Health workers bring completed forms to their supervisor.

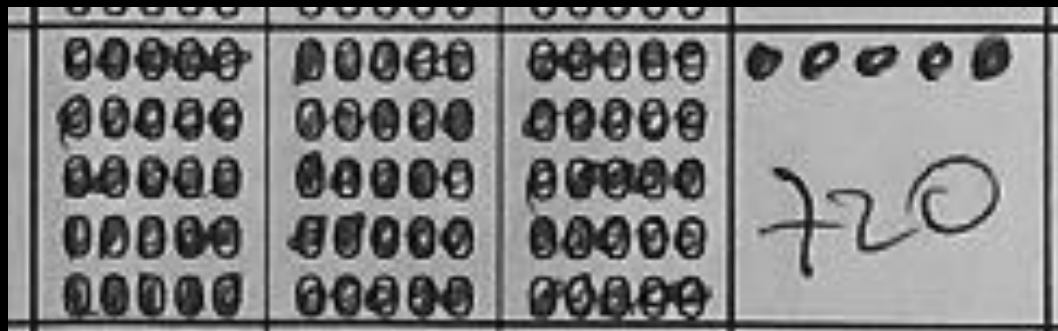
Medicina		Stock no Início do Mês (incl. MR Q Comprimidos)	Total Recebido no Mês Comprimidos	Número de Tratamentos Usados	Total Tratamentos Usado	Inventário no Fim do Mês Comprimidos
1.)	6 meses à 35 meses (6x1 Tratamentos AL)			0000000000000000		
				0000000000000000		
				0000000000000000		
				0000000000000000		
				0000000000000000		

# Field Deployment: Tracking health worker usage of medical supplies in Mozambique

- Four-month deployment
- Two districts, 45 health workers
- 16 different medical supplies

## Results

- Data entry accuracy > 98% for all data types.





# The system made it easier to track supplies

*“This form provides the information all in one place. It is very useful for me. I can just see the form and know what happened. Before, I would have to go to different data sources to confirm everything.”*



## The system improved efficiency

*"Using [the system] it is quicker to send the data to anywhere. When I have written data, I need to write everything down, then take it for [an approval] signature, then find transport to send that information to [the provincial office]. It can take days to do that."*



# Broader Impact



Track student attendance at 300 schools in Kyrgyzstan.



Digitizing patient records in Zambia for the “Better Immunization Data” initiative.



Digitizing patient registers at health camps in Nigeria.

## Summary:

Paper forms are widely used for data collection.

We need better ways to digitize data from paper.

## This project:

1. Formative study to better understand paper-digital workflows in global development.
2. A new mobile system that efficiently digitizes data from paper forms.
3. Deployments to evaluate the system with health workers in Mozambique.
4. Artifact ODK Scan: Enables broader impact.