# Computer Security for Data Collection Technologies

Camille Cobb, Samuel Sudar, Nicholas Reiter, Richard Anderson, Franziska Roesner, and Tadayoshi Kohno

University of Washington

paper: https://goo.gl/oYccxF

We investigate security in the developing world, specifically focusing on digital data collection.

# Motivation

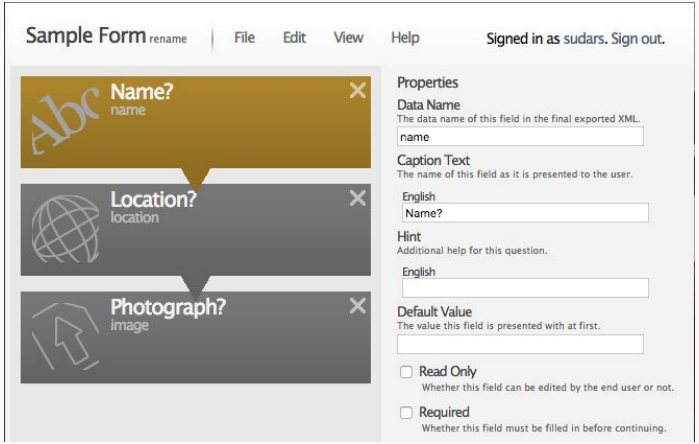

- Data collection very common in the developing world
  - NGOs, governments, researchers, charitable organizations
- Often intensely private information
  - ID numbers, sexual activity, HIV status, mental health indicators
- Is computer security a concern?
- Surveyed and interviewed groups performing digital data collection

1. Data Collection & ODK
2. Threat Modeling
3. Interviews
4. Takeaways

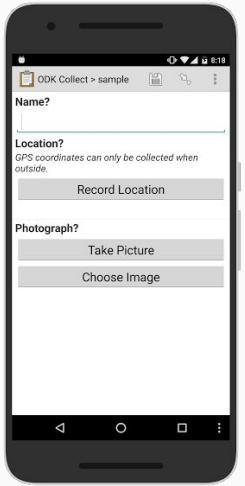
1. Data Collection & ODK
2. Threat Modeling
3. Interviews
4. Takeaways

# Digital Data Collection

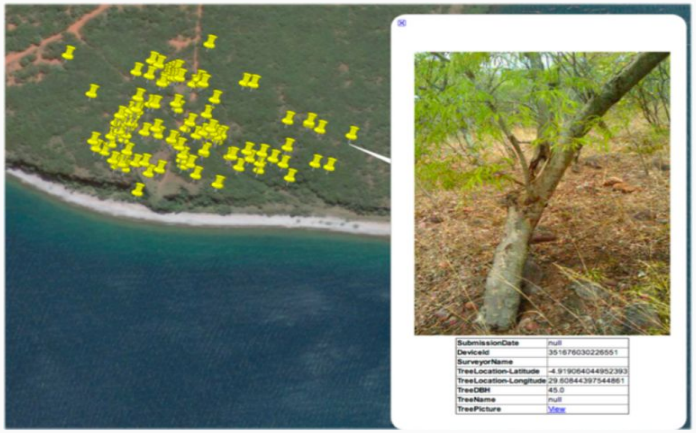
1. Build Form




2. Collect Data




3. Use Data



# Open Data Kit (ODK)

- Representative digital data collection platform
- Used in 125 countries
- Installed on 100,000s of devices
- PATH, AMPATH, Gates Foundation, and many others
- Millions of forms have been collected

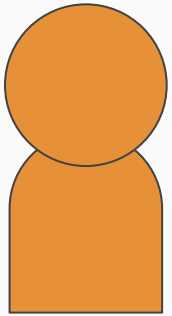
magpi

KoBo Toolbox

SurveyCTO

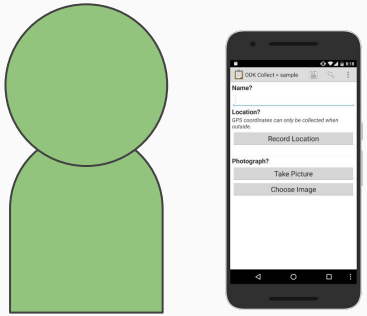
pendragon™
Software Corporation

# Deploying ODK



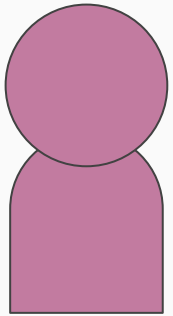
## Deployment Architect

- Designs forms
- Administers phones
- Employs Enumerators




## Enumerator

- Uses phone
- Speaks to Beneficiary

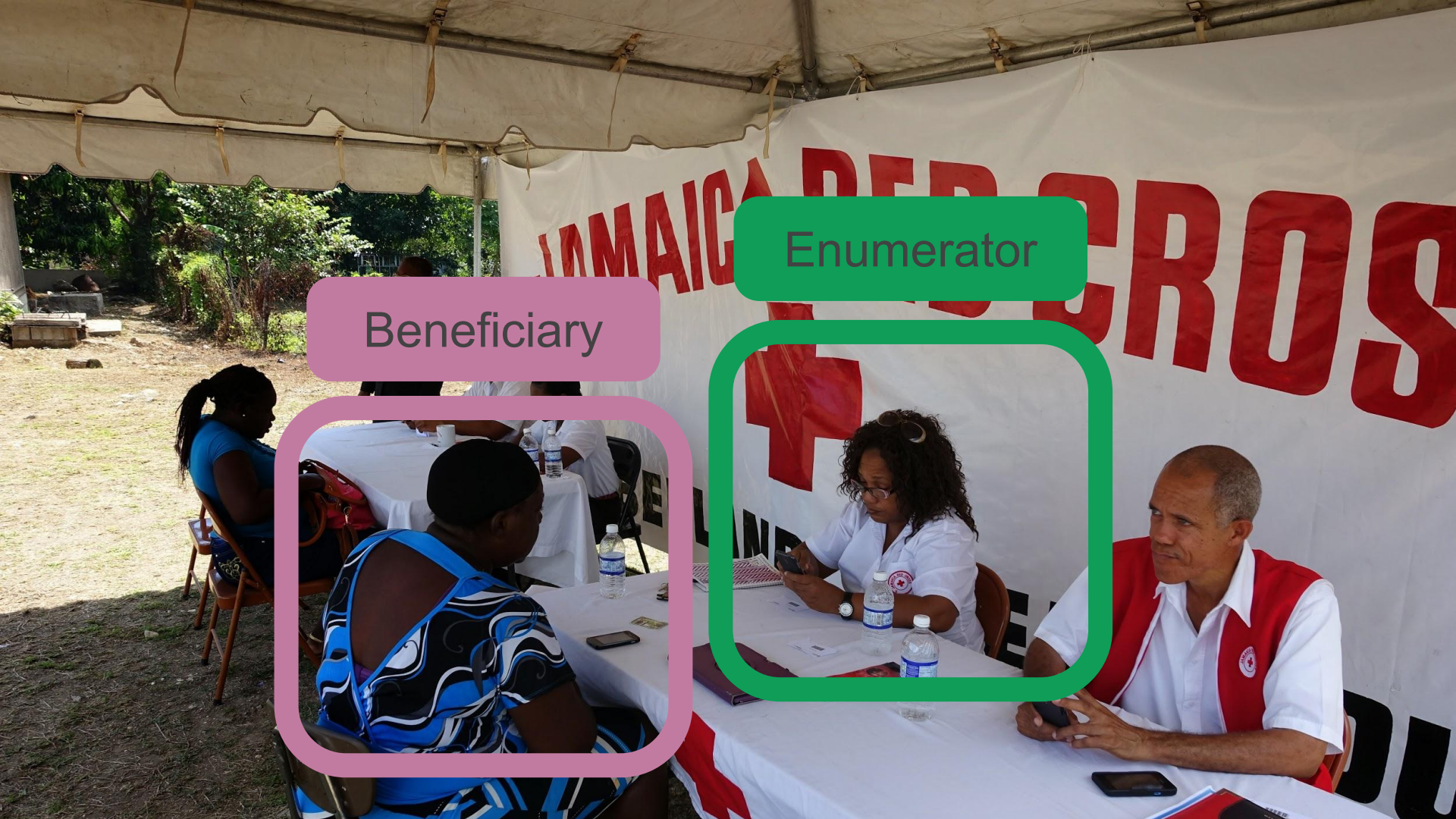


## Beneficiary

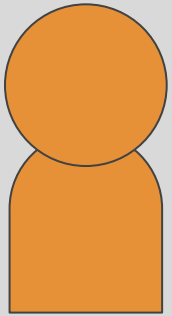
- Not associated with organization
- Gives data to Enumerator


JAMAICA RED CROS


Enumerator


Beneficiary
Enumerator

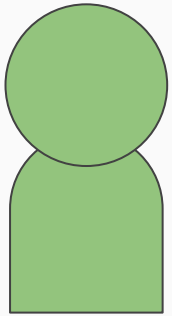
# Deploying ODK

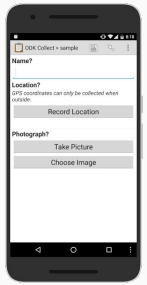


## Deployment Architect

- Designs forms
- Administers phones
- Employs Enumerators

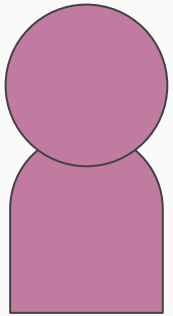
Interview






## Enumerator

- Uses phone
- Speaks to Beneficiary



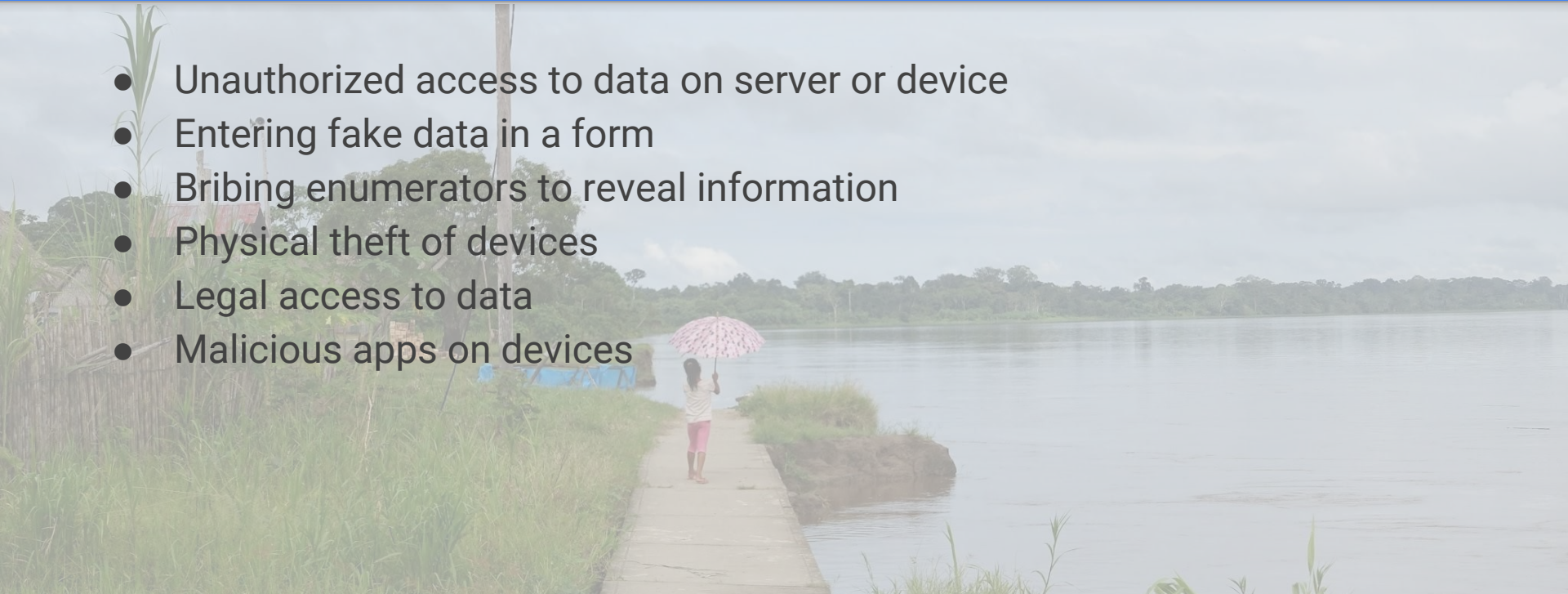
## Beneficiary

- Not associated with organization
- Gives data to Enumerator

1. Data Collection & ODK
2. Threat Modeling
3. Interviews
4. Takeaways

# What threats could a deployment face?

# Threat Modeling in ICTD

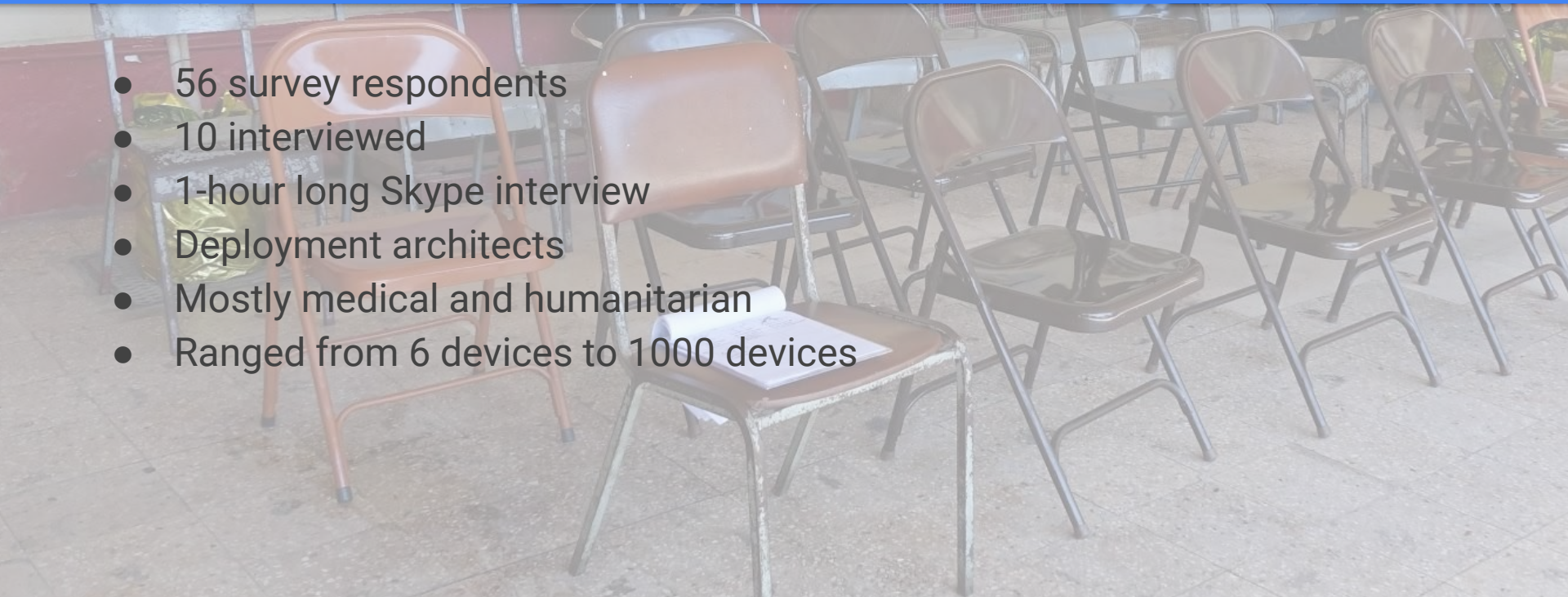


- Unauthorized access to data on server or device
- Entering fake data in a form
- Bribing enumerators to reveal information
- Physical theft of devices
- Legal access to data
- Malicious apps on devices

Threat models are highly context-dependent.

1. Data Collection & ODK
2. Threat Modeling
3. Interviews
4. Takeaways

# Methodology



- 56 survey respondents
- 10 interviewed
- 1-hour long Skype interview
- Deployment architects
- Mostly medical and humanitarian
- Ranged from 6 devices to 1000 devices

# Data Security

# Data loss

**Main concern for several participants**

When asked about the greatest threats that could derail a deployment: **“far and away it is data loss.”**

**Data loss can trump other concerns, including confidentiality.**

**“[Encryption] did cause us some problems and that’s why we didn’t continue it**….You try to submit some data, some of them get lost along the way somehow.”

# Exploited data

**Collected data could lead to harm**

"We record violence that those children might have went through walking on the street. And actually it turned out that the highest perpetrator is, well I cannot mention the name now, but . . . **it is a very dangerous group. ...We don't actually email even this information, because if it gets intercepted, again, everyone would be in trouble.**"

**Device theft is not equated with data exploitation**

"We're hoping it's just about the hardware, that's fine. But I don't think it could be an issue about the data inside the tablet...**It's kind of fine, like 'take it, reset it, don't look at the data, and enjoy the tablet."**

# Perceived security of paper vs digital

**Security through obscurity**

“If someone’s really interested in the content, it’s easier to steal a stack of papers... So **relative to someone on the ground being able to steal the data content, tablets are much more secure than paper.”**

**Digital collects more sensitive information**

**“The only thing that is specifically unique to doing digital data collection is that you have more identifying information**...you could have photographs and audio recordings and all kinds of things.”

# Enumerators as adversaries

# Enumerators can pose threats

- Sell or leak data
- Fabricating data to avoid undesirable parts of their jobs
- Accidental installation of malware
- Honest mistakes with data entry

**Not all threats are technical**

“**It would be much easier to bribe or go see an enumerator and offer him a beer** [than hack the system]**.**”

# Defenses against enumerators

- GPS readings
- Photographs of relevant locations
- Timestamps to measure completion time
- Restrict phones to only ODK

**Explicit tradeoff of utility**

**“I couldn’t see myself limiting them from the benefits they could get from the tablet** in case they were in this kind of [dangerous] situation. Meaning having access to phone, having access to their emails.”

# Diversity of stakeholders

# Diffused responsibility

**Rely on external guarantees**

“**They say that everything is secure** and that the servers are…[are] underground providing maximum security.”

**Ethics boards do not have security expertise**

“[Encryption is] something [the IRB] should be requiring and...just **lack the technical sophistication to ask for it.**”

# Differing privacy norms

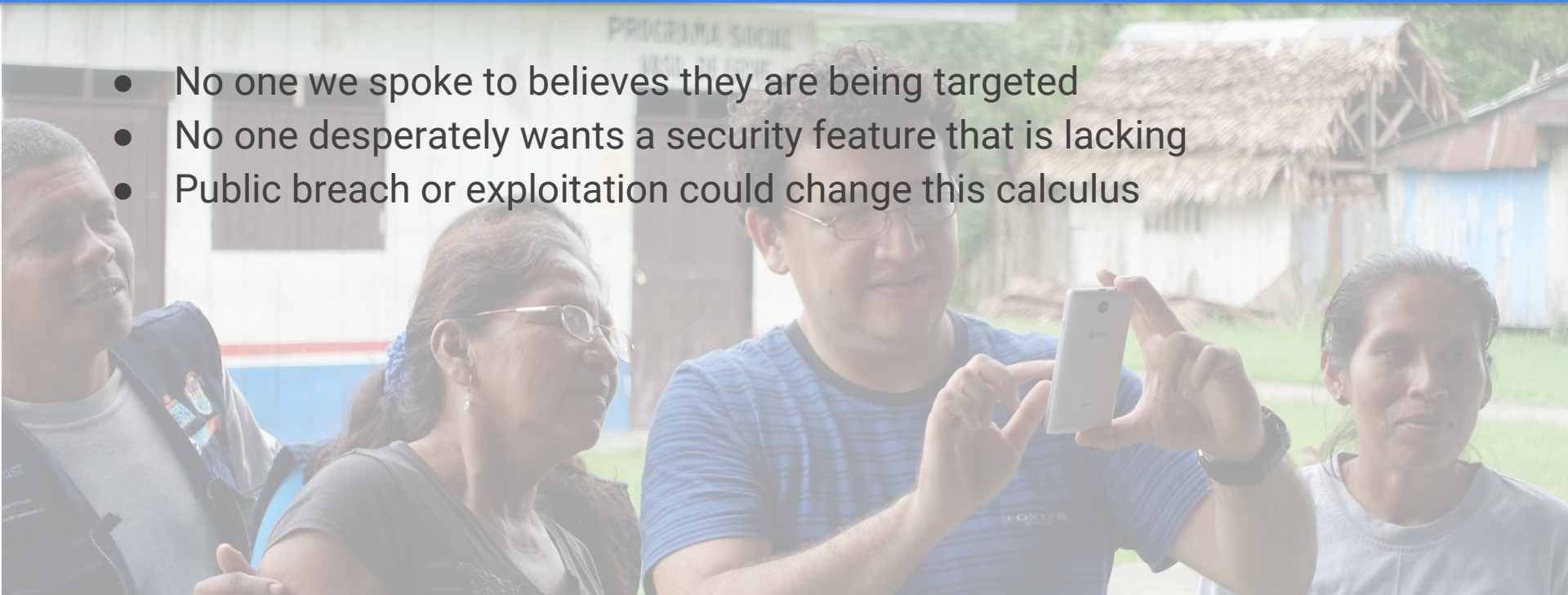
**Privacy norms can differ between organizations and local communities**

"When we've requested summary statistics from them, we've often received a lot more details than I would expect them to be comfortable sharing. . . . In general I think, in the rural areas where we mostly work, a lot of these things are kind of treated as common knowledge. **Within the village everybody knows who's poor, everybody knows if you have some sort of special needs**, so I don't think it's really on the forefront of their minds."

1. Data Collection & ODK
2. Threat Modeling
3. Interviews
4. Takeaways

Given the surfaced threat models, security measures are broadly appropriate.

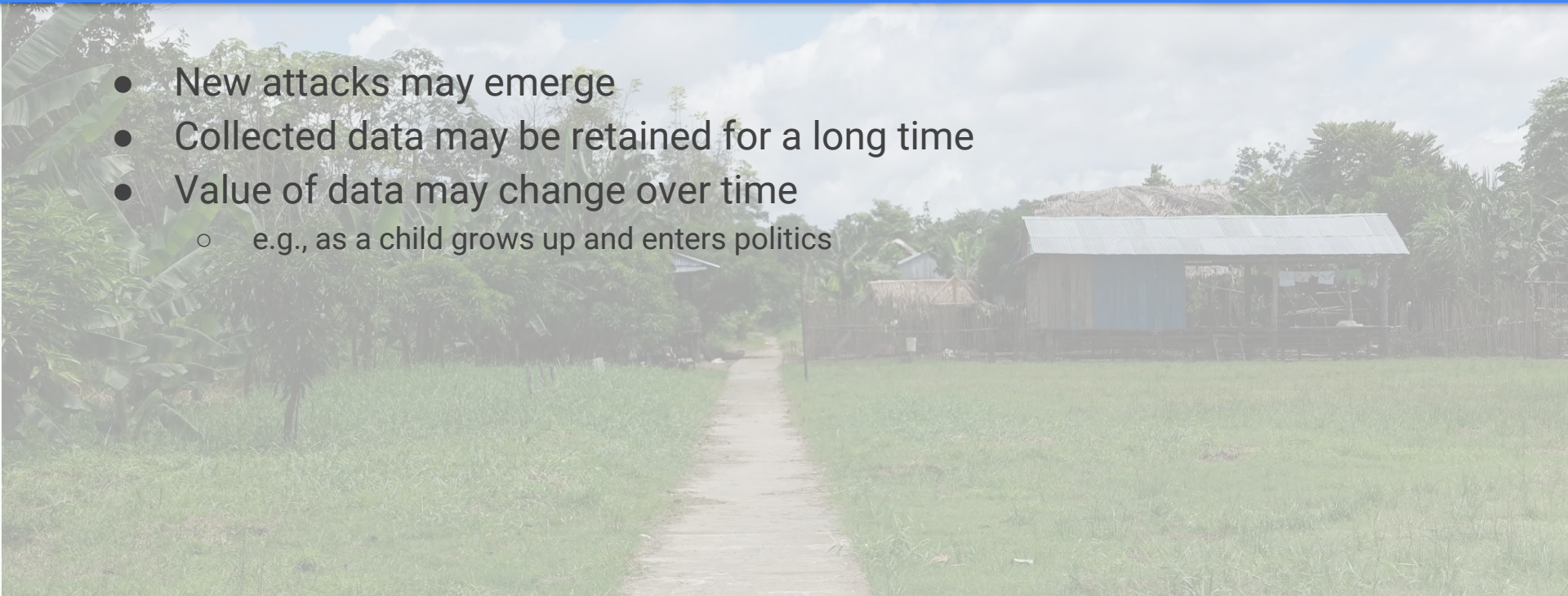
# Reasonably secure



- No one we spoke to believes they are being targeted
- No one desperately wants a security feature that is lacking
- Public breach or exploitation could change this calculus

Threat models must be periodically re-evaluated.

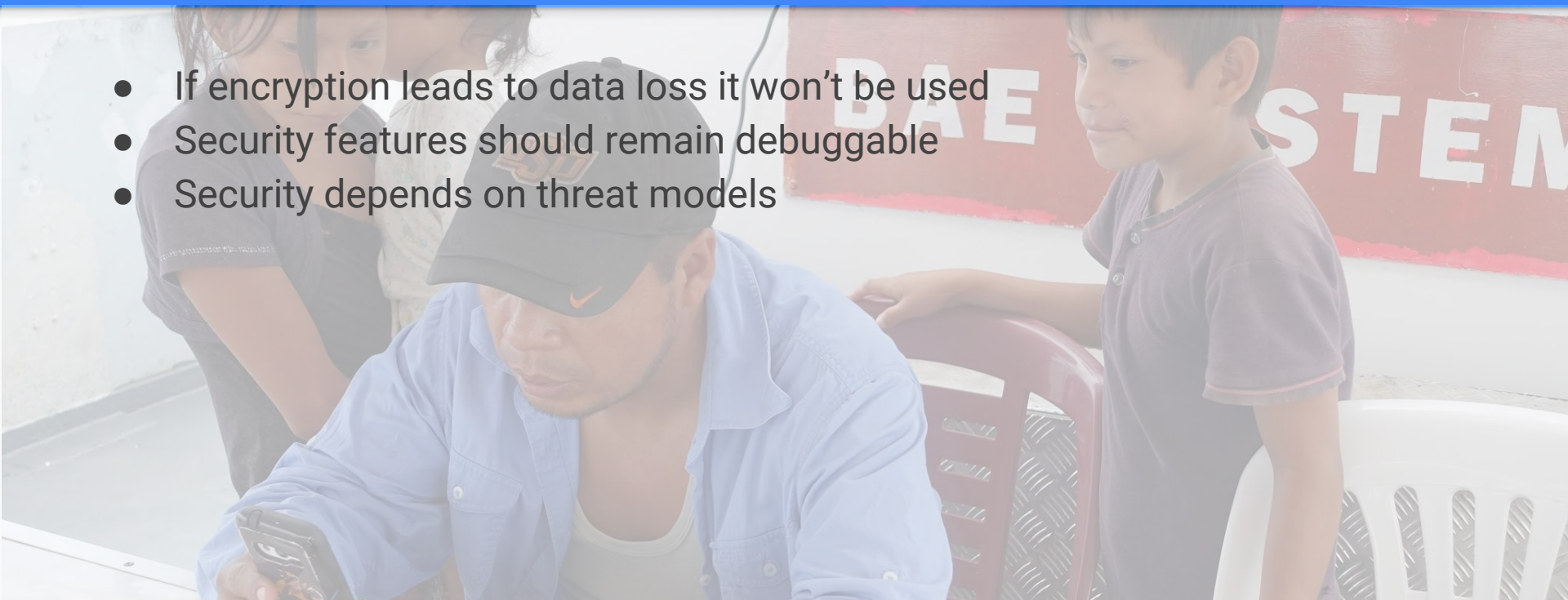
# Threats evolve



- New attacks may emerge
- Collected data may be retained for a long time
- Value of data may change over time
  - e.g., as a child grows up and enters politics

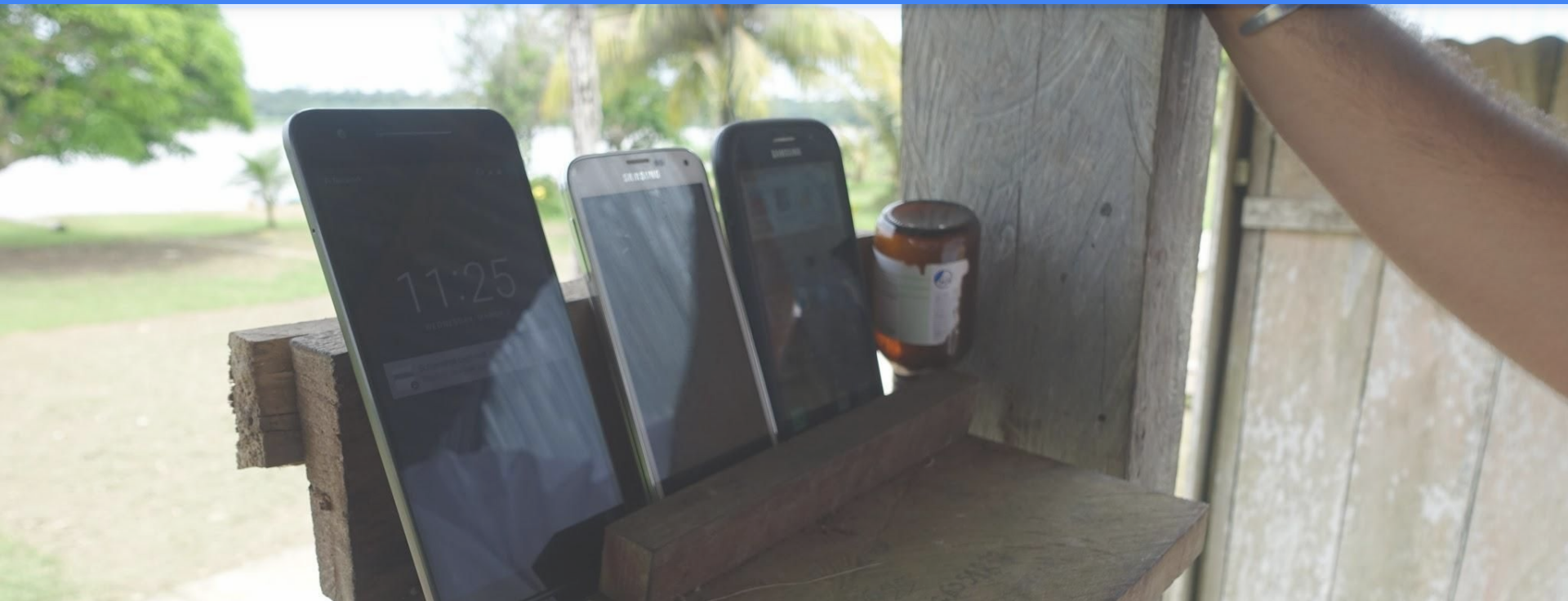
Usability can trump security.

# Adding “security” won’t work



- If encryption leads to data loss it won’t be used
- Security features should remain debuggable
- Security depends on threat models

# Conclusion



# Questions?

Thank you.

Sam Sudar

sudar.sam@gmail.com

https://samsudar.com