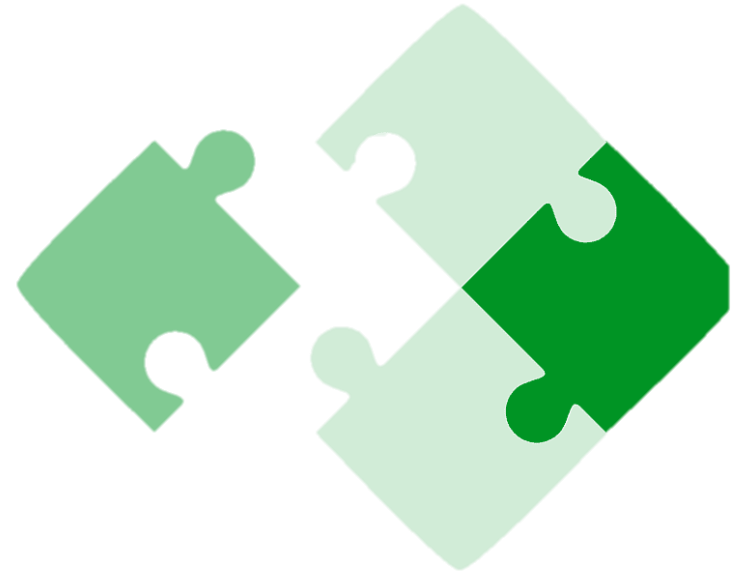


Integrating Data Sources to Complement and Improve Household Survey Data

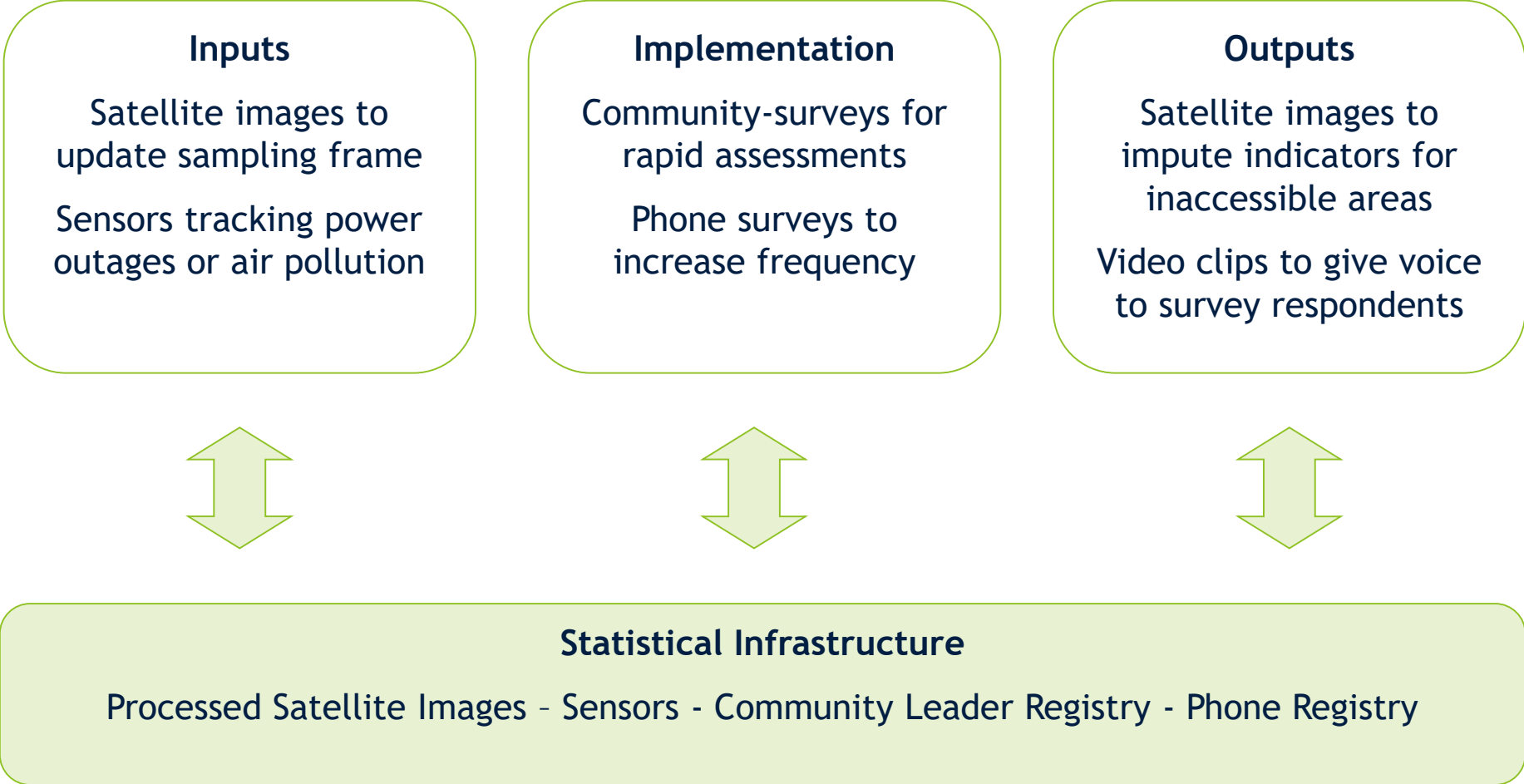


World Bank - Poverty & Equity Global Practice
February, 2020

Utz Pape

Modern statistical offices need to invest in new infrastructure and methods to respond to changing user demands.

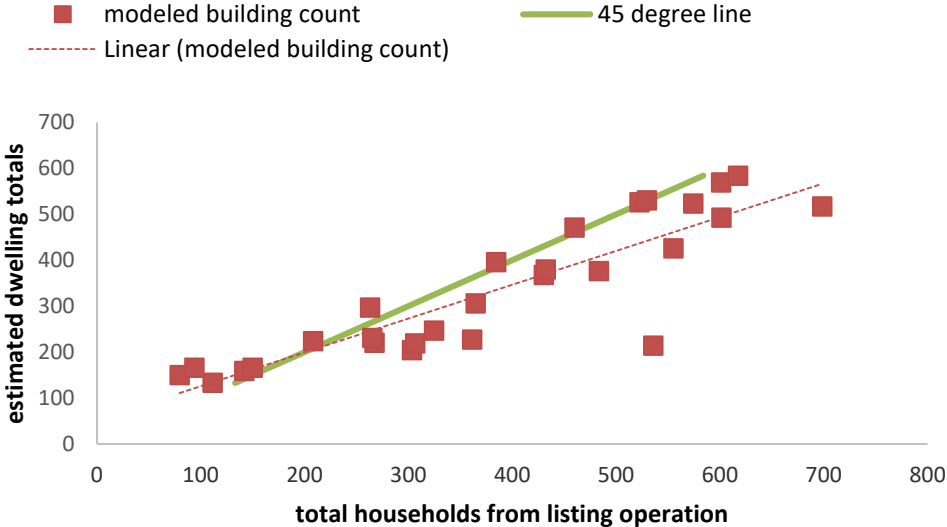
Changing User Demands for Low-Cost, High-Frequency, Timely and High-Quality Data



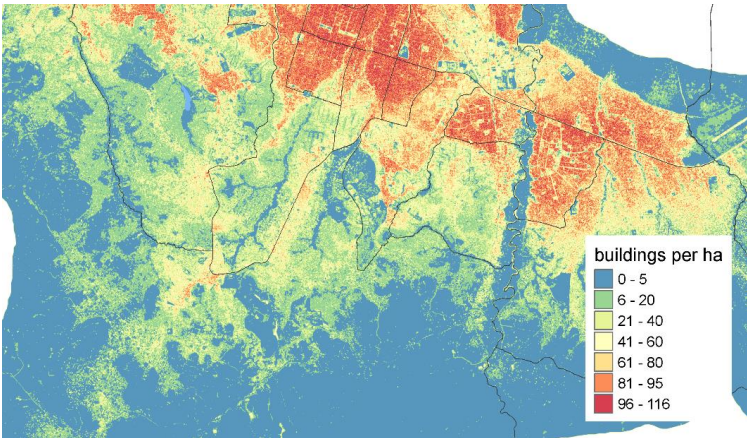
Satellite images can be used to update sampling frames rapidly and with low-cost especially after large population movements. Example: DRC.

- Using an automated iterative process, the Primary Sampling Unit boundaries were first defined using administrative and physical boundaries such as rivers, highways, and roads.
- Population estimates were retrieved by a Random Forest Regression model based on contextual image information and a sub-sample of building locations

Primary Sampling Units by natural boundaries and administrative boundaries



Algorithmic Estimate of Population Density

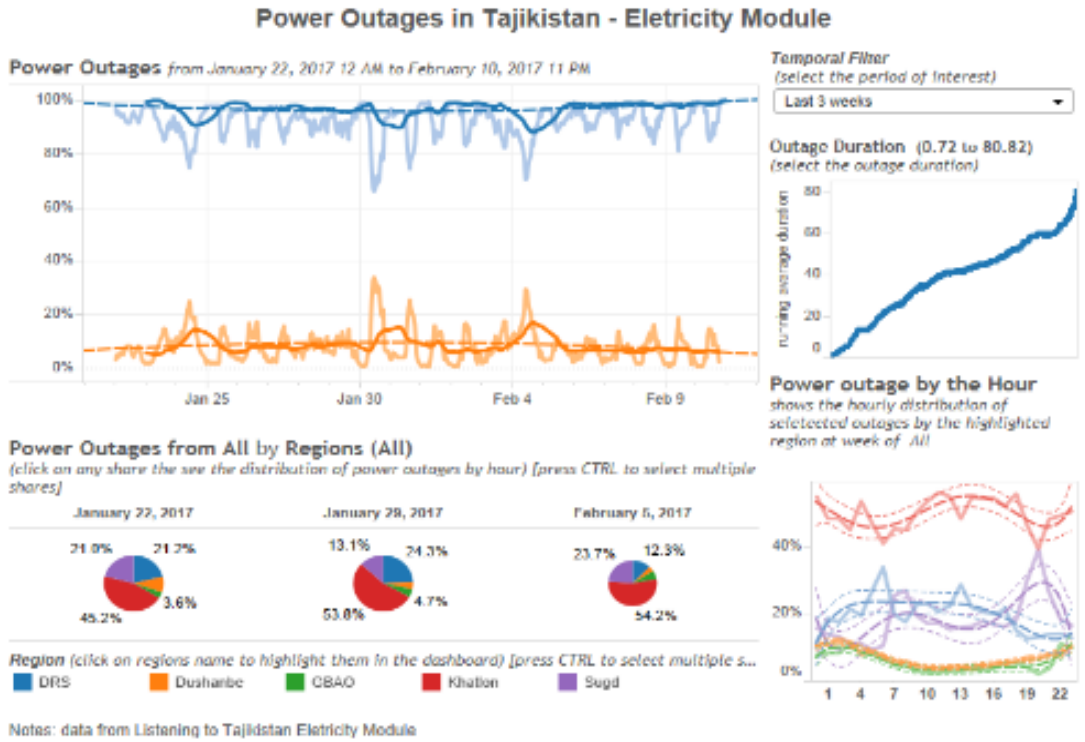


Sensors like smart survey boxes can track real-time indicators like air pollution and power outages. Example: Tajikistan.

Pilot project to monitor power outages automatically with a autonomous devices that send a signal if a power outage is detected.



- 300 boxes deployed.
- Less than 1 US\$ per day per box.
- Remote sensing enabling.



Community-surveys are rapid and low-cost data collection methods that can create deprivation maps within 3 months. Example: CAR.

To inform the Recovery and Peacebuilding Assessment (RPBA) in the Central African Republic (CAR) at a time when large parts of the country were still insecure, a district census was designed and fielded within a three-month period.

- The district census interviewed local officials in all 179 administrative centers of the country.
- The data were summarized in a composite indicator—the Local Development Index (LDI)—and were used to inform new development projects.
- The LDI offers a spatial assessment of the state of the nation (much like a poverty map) and has become an important tool for development planning and progress monitoring.

L'indice de développement local

Mesure de la capacité communale

Indice de développement local (IDL) : une mesure composite de la capacité de la commune

L'indice de développement local (IDL) est une mesure de la situation des services dans chaque commune. L'indice regroupe 12 indicateurs relatifs à trois piliers : administration locale, infrastructures, et services de base.

L'IDL s'étend sur une échelle de 0 à 100, dans laquelle une valeur plus élevée indique un meilleur scénario. L'IDL est la somme pondérée des 12 indicateurs. Chaque indicateur n'est pas nécessairement attribué au même poids et certains indicateurs sont pondérés plus fortement dans l'IDL que d'autres. Les définitions de l'ensemble des indicateurs sont expliquées ci-dessous. Pour plus d'information sur l'IDL, veuillez-vous référer au site internet : <http://bit.ly/2k7wFqj>.

Classement des communes selon le quintile de l'IDL

- Les communes avec l'IDL le plus élevé
- Les communes avec l'IDL le deuxième plus élevé
- Les communes avec l'IDL le troisième plus élevé
- Les communes avec l'IDL le deuxième plus faible
- Les communes avec l'IDL le plus faible

Composants de l'indice

Indicateur	Score de l'indicateur	Poids	Score IDL
Indice de développement local (IDL)	100	1	100
Indice de développement local (IDL)	100	1	100
Indice de développement local (IDL)	100	1	100
Indice de développement local (IDL)	100	1	100
Indice de développement local (IDL)	100	1	100
Indice de développement local (IDL)	100	1	100
Indice de développement local (IDL)	100	1	100
Indice de développement local (IDL)	100	1	100
Indice de développement local (IDL)	100	1	100
Indice de développement local (IDL)	100	1	100
Indice de développement local (IDL)	100	1	100
Indice de développement local (IDL)	100	1	100

Classement des communes selon l'IDL

Commune	Ranking	Score IDL	Commune	Ranking	Score IDL
Bossembé	1	84	Bouar	7	65
Bossango	2	77	Yobé-Sangha	7	65
Babou	3	73	1er Arrondissement	9	64
Berberati	4	71	Carnot	10	63
Haute-Kadai	5	69	Pissa	10	63
Bimbo	6	67	Yaloké	10	63

Pour un classement de toutes les communes du pays, voir au verso de cette affiche

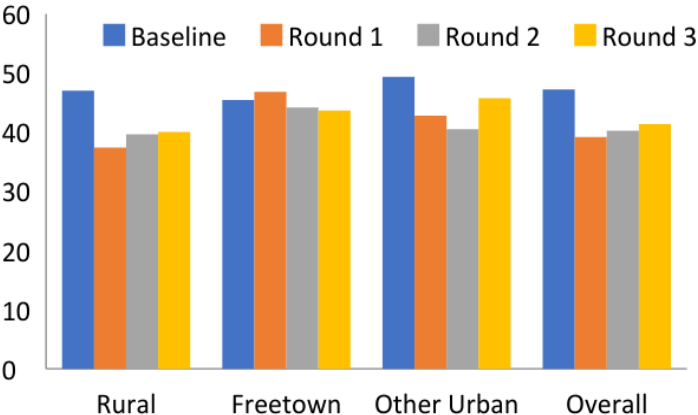
Phone surveys collect high-frequency data rapidly at low cost and can be deployed in the case of crisis. Example: Sierra Leone.

Using a pre-crisis face-to-face household survey together with follow-up phone interviews can provide powerful real-time crisis monitoring.

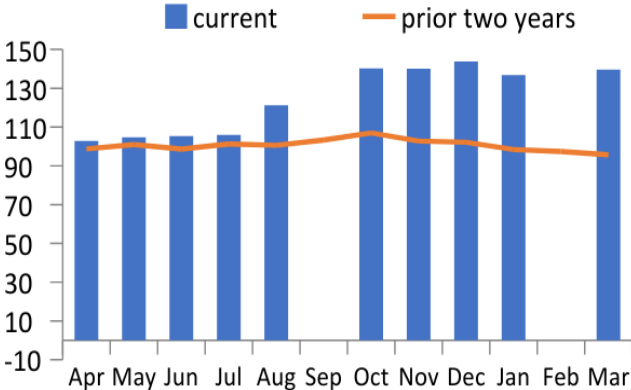
Suggests to pro-actively create sampling frames for phone surveys to allow rapid monitoring in the case of crises.



Hours worked last week among those employed in Sierra Leone



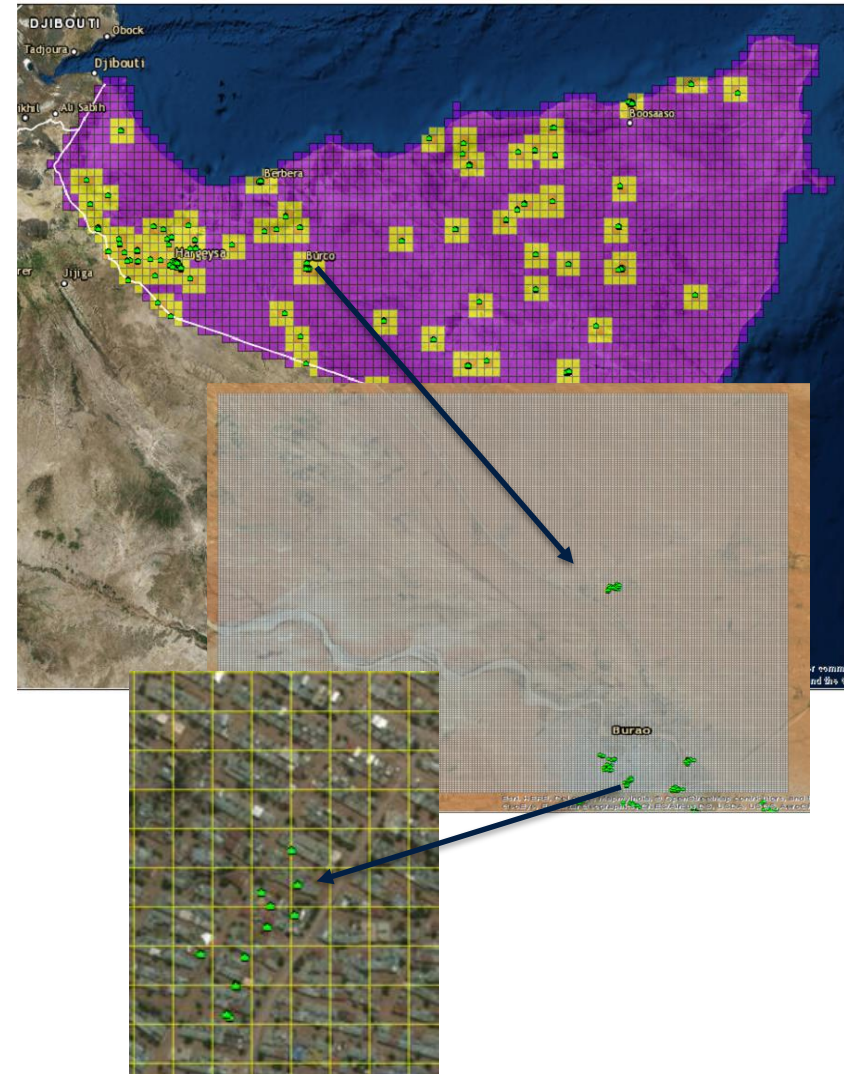
Price index for rice



Imputations based on satellite images can make estimates representative for inaccessible areas. Example: Somalia.

Field surveys sometimes lack access to certain (e.g. insecure) areas leading to under-representation of specific parts of the population. Imputations using satellite images can mitigate the under-representation to some extent.

- The imputation estimates a structural model on geographic features extracted from satellite images to estimate indicators for known locations of household respondents.
- The structural model can then be applied to areas without household respondents to estimate the indicators.
- The estimators might be biased if inaccessible areas are structurally different.
- Satellite images can be expensive; calculations are computationally intense.



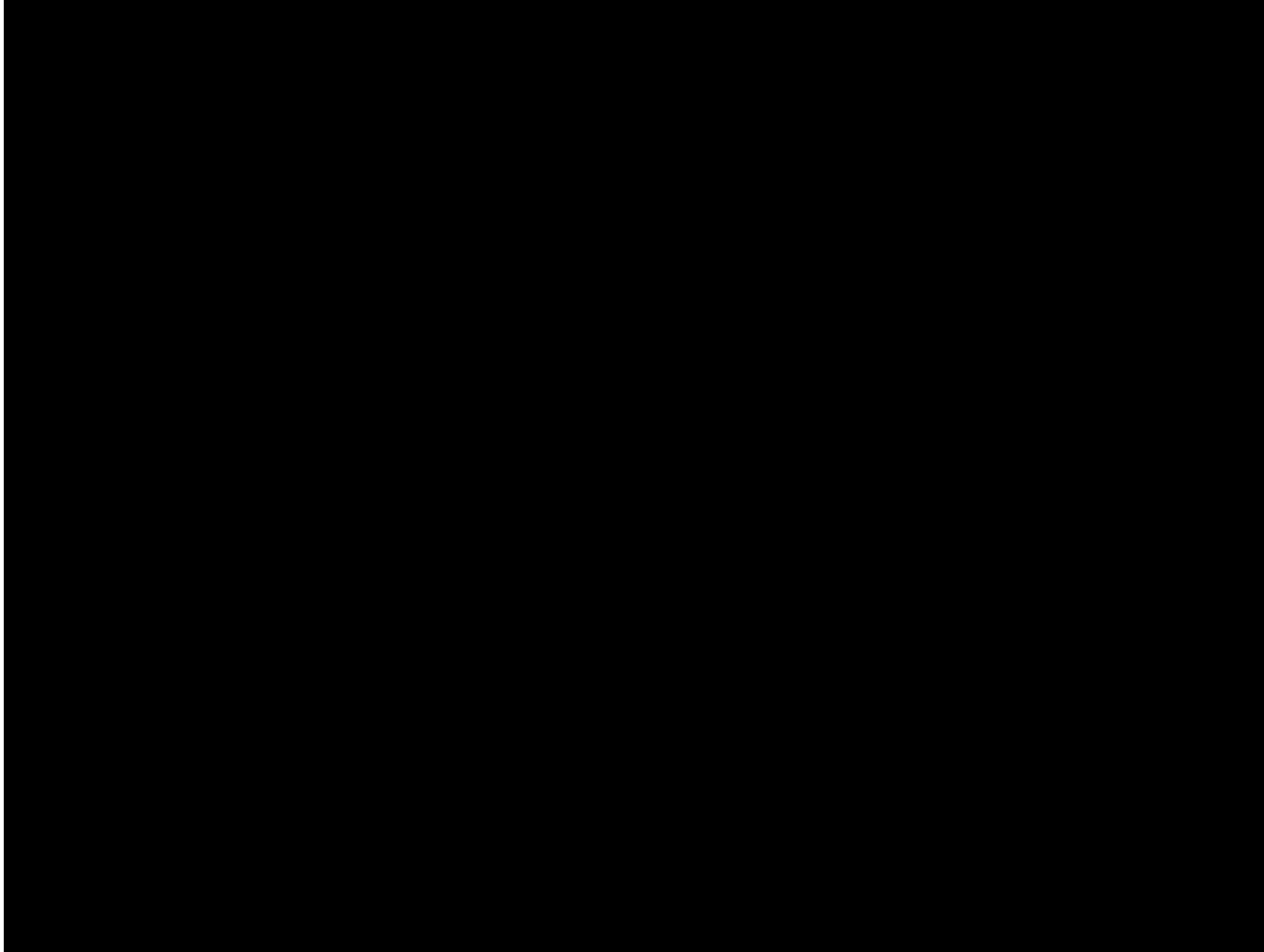
Video testimonials can be used to give a voice to survey respondents to share their subjective impressions. Example: South Sudan.



Video testimonials are a powerful low-cost tool to give voice to the most vulnerable populations.

www.thepulseofsouthsudan.com and www.thesomalipulse.com

Video testimonials can be used to give a voice to survey respondents to share their subjective impressions. Example: South Sudan.

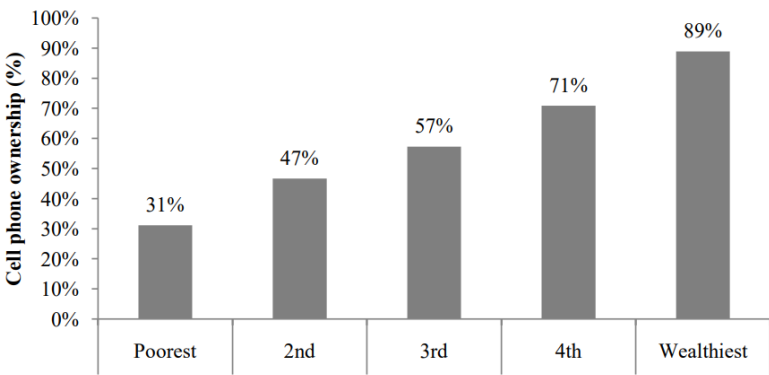


Video testimonials are a powerful low-cost tool to give voice to the most vulnerable populations.

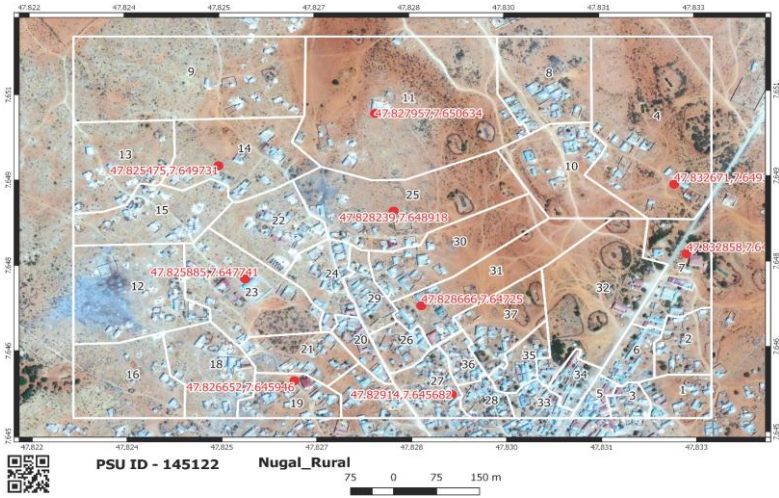
www.thepulseofsouthsudan.com and www.thesomalipulse.com

However, technology, big data and machine learning are not a panacea for data collection but have their own challenges.

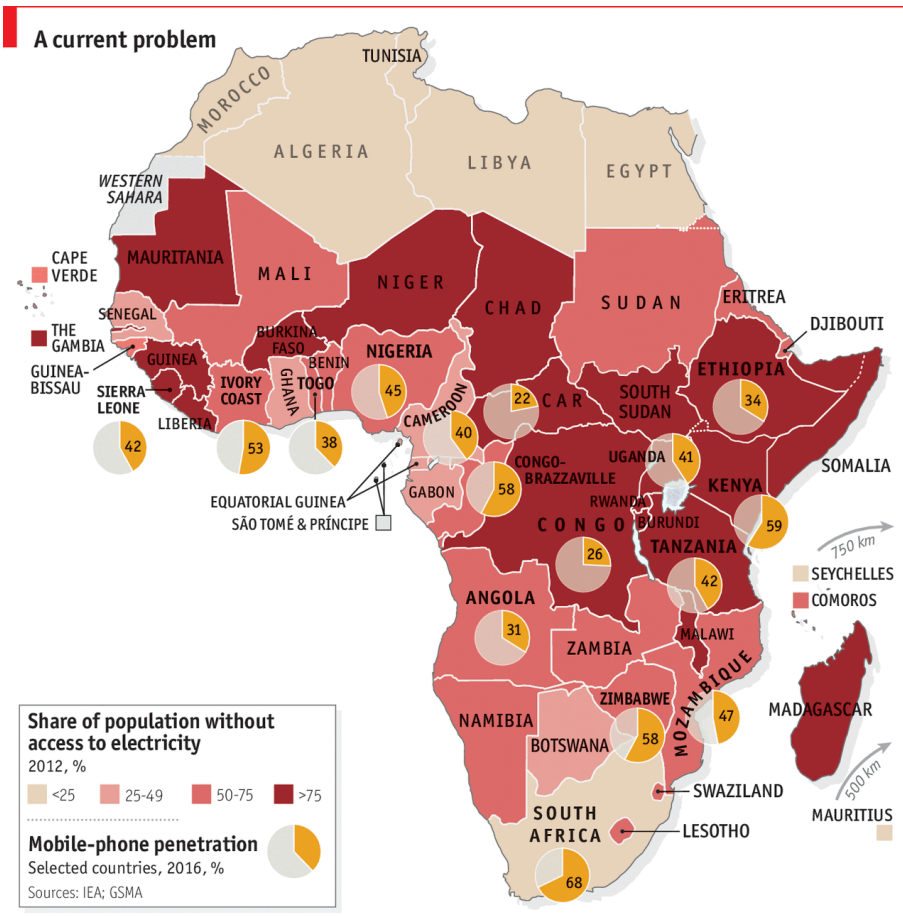
Cell Phone Ownership in Tanzania (2010/11)



Satellite-based Creation of Enumeration Areas

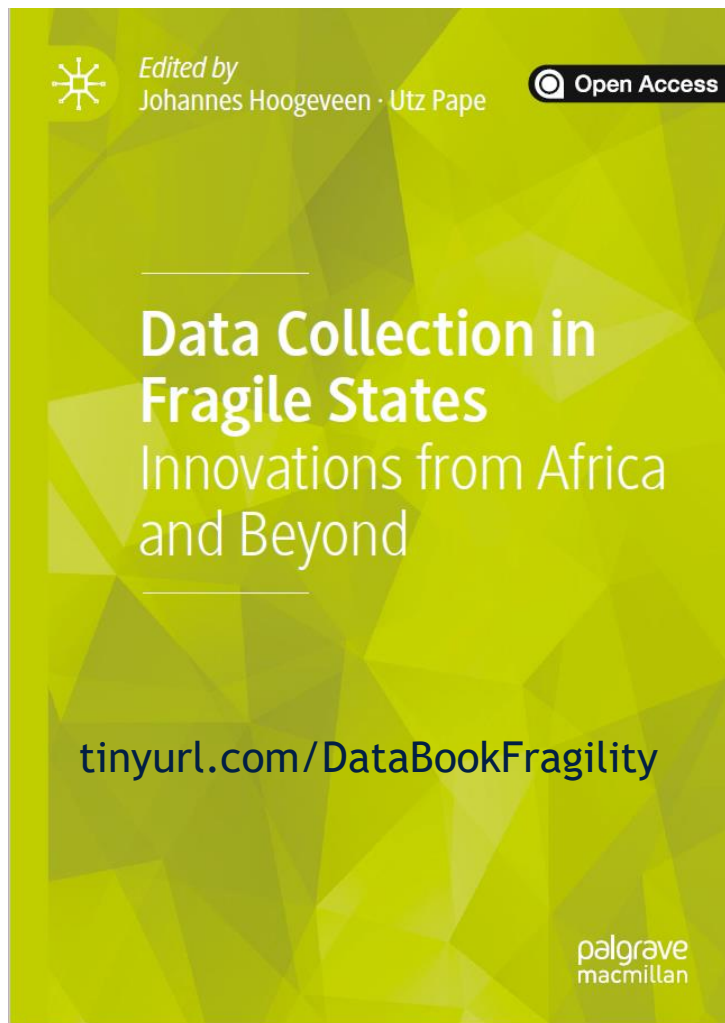


Electricity Access and Mobile Phone Penetration



Economist.com

More innovative ideas - scalable and applicable in non-fragile settings - are described in 'Data Collection in Fragile States'.



- Fragility and Innovations in Data Collection
- **Part I: Innovations in Data Collection**
 - Monitoring the Ebola Crisis Using Mobile Phone Surveys
 - Rapid Emergency Response Survey
 - Tracking Displaced People in Mali
 - Resident Enumerators for Continuous Monitoring
 - A Local Development Index for the CAR and Mali
- **Part II Methodological Innovations**
 - Methods of Geo-Spatial Sampling
 - Sampling for Representative Surveys of Displaced Populations
 - Rapid Consumption Surveys
 - Studying Sensitive Topics in Fragile Contexts
 - Eliciting Accurate Consumption Responses from Vulnerable Populations
- **Part III Other Innovations**
 - Using Video Testimonials to Give a Voice to the Poor
 - Iterative Beneficiary Monitoring of Donor Projects
- Concluding Remarks: Data Collection in FCV Environments

Contributors: Ana Aguilera, Mohamed Coulibaly, Stephanie Eckman, Alvin Etang, Saad Gulzar, Kristen Himelein, Johannes Hoogeveen, Mohammad Isaqzadeh, Lennart Kaplan, Roy Katayama, Nandini Krishnan, Johan Mistiaen, Juan Munoz, Utz Pape, Falvio Russo Riva, Jacob Shapiro, Dhiraj Sharma, Andre-Marie Taptue, Tara Vishwanath, James Walsh, Gervais Chamberlin Yama