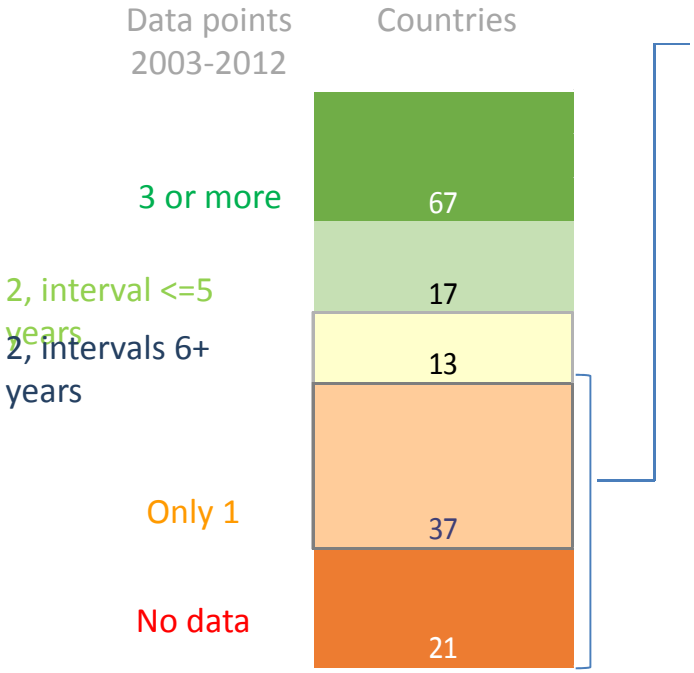
Household Surveys Network at the World Bank

Michael Lokshin

Data Gaps, Quality and Access



- 71 countries lack adequate data to monitor twin goals
  - No data: mainly in EAP and LAC small countries
  - Only 1 point: mainly in AFR
  - Irregular survey programs
- Also, issues of data quality (reliability, comparability) and accessibility

DDI Metadata Editor (Nesstar Publisher 4.0.9)

The IHSN Metadata Editor, also known as the Nesstar Publisher, is a specialized XML editor compliant with the Data Documentation Initiative (DDI) 2.n and the Dublin Core metadata standards

Statistical Disclosure Control (sdcMicro)

SDCMicro is free, R-based open-source package for the generation of protected microdata for researchers and public use. Data from statistical agencies and other institutions are mostly confidential. This package can be used for the generation of anonymized (micro)data,

i.e.

Microdata Cataloging Tool (NADA)

NADA is a web-based cataloging system that serves as a portal for researchers to browse, search, compare, apply for access, and download relevant census or survey information. It was originally developed to support the establishment of national survey data

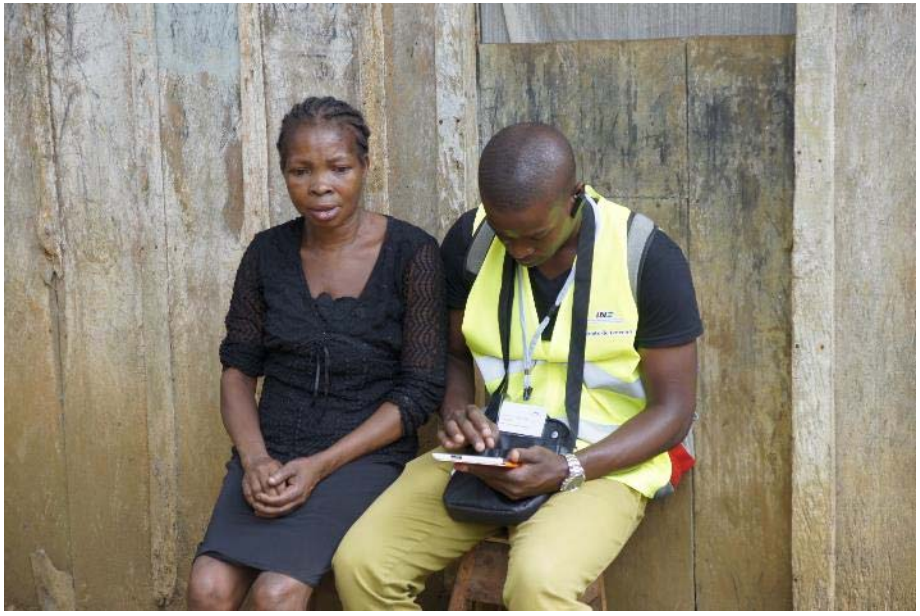
World Bank

Survey Solutions

Survey Solutions

COMPUTER-ASSISTED PERSONAL and WEB INTERVIEWING

Michael Lokshin



Benefits of CAPI

Improve timeliness of data collection

Ensure data quality and comparability

Allow collection of new types of information/data

Cost-effective, sustainable solution for NSOs

CAPI System Requirements

- Simple yet flexible system for the non-expert users. Typical clients – National Stat Offices

- Functionality for
  - **data capturing:** entering data on a tablet
  - **survey management**: managing teams of enumerators
  - **data management:** data aggregation, versioning, reporting

- Tablet-based with ability to display and navigate through multi-level large questionnaires.
- Support of panel surveys and complex validation algorithms.
- Cost effective system that can be used and supported by NSOs without external TA.

Main differences from other systems

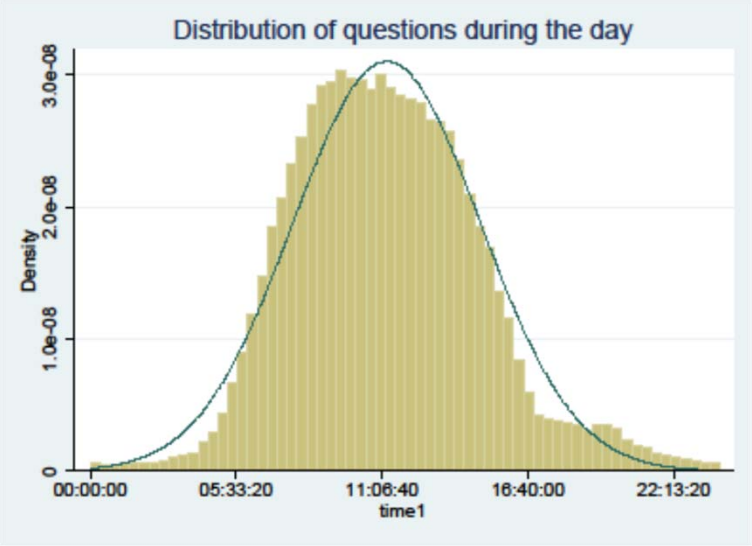
- **Out-of-the-box solution for survey data collection**: data capturing, data management, and survey management. No software on the market provides such a package. All other system focus mostly on data capturing.

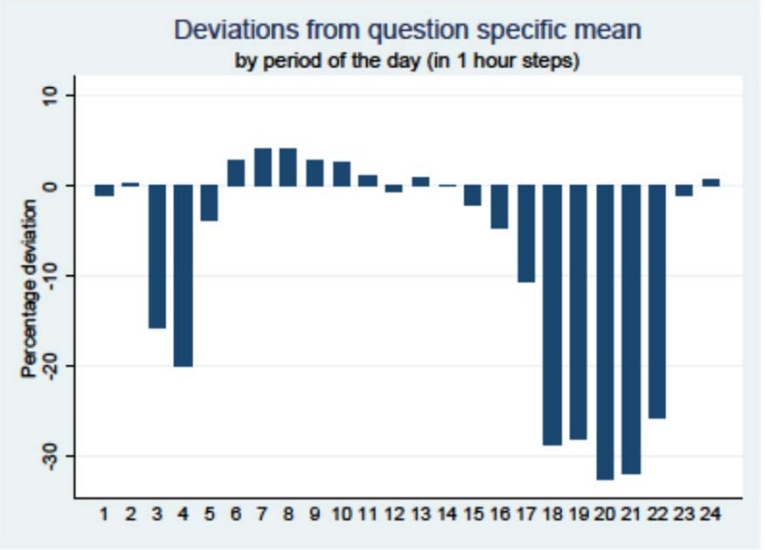
- **Minimum TA; Focus on Capacity Building**: Survey Solutions is designed to minimize the TA. Lowest leargning curve. Other systems require constant and significant TA. Expert versus User centered approach.

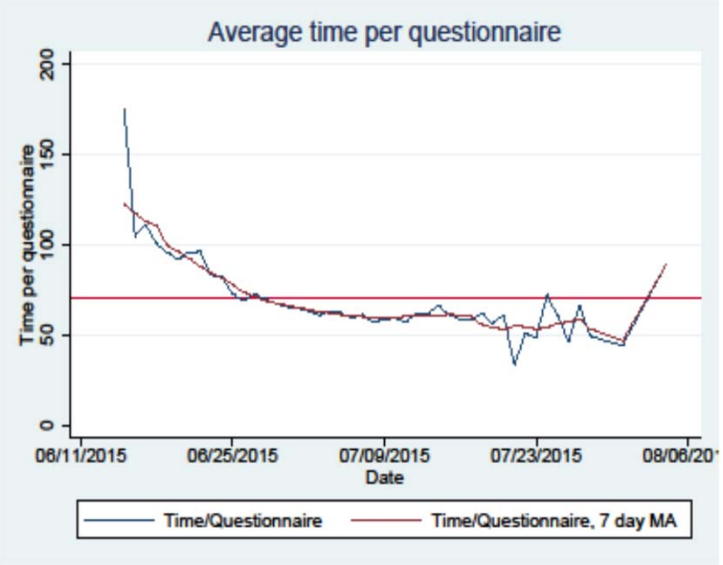
- **Designed for large surveys:** Survey Solutions is specifically designed for LSMS and HBS-types of surveys: Nested Rosters, Cascading and Linked questions, Roster-specific validations. Online collaboration.

- **Data security:** Survey Solutions allows storing data on the local servers of NSO thus complying with the local data privacy and anonymity laws. No other software used by the NSOs has such functionality (SurveyCTO can do that for a high fee, but we know of no NSO that is using it – different market (Kenya)).

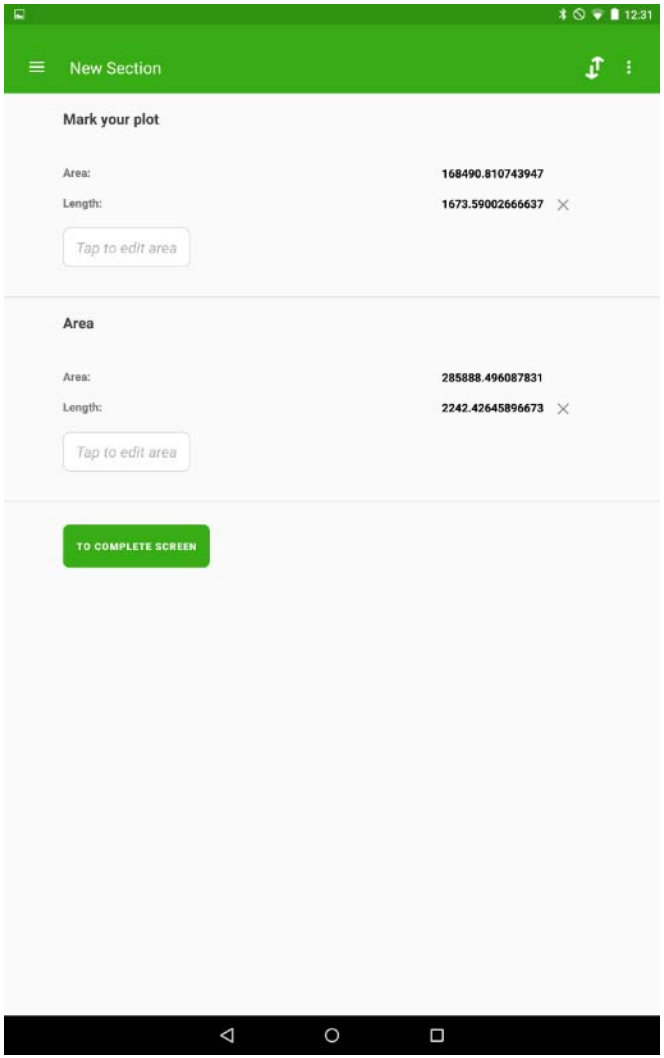
Paradata: improving data quality

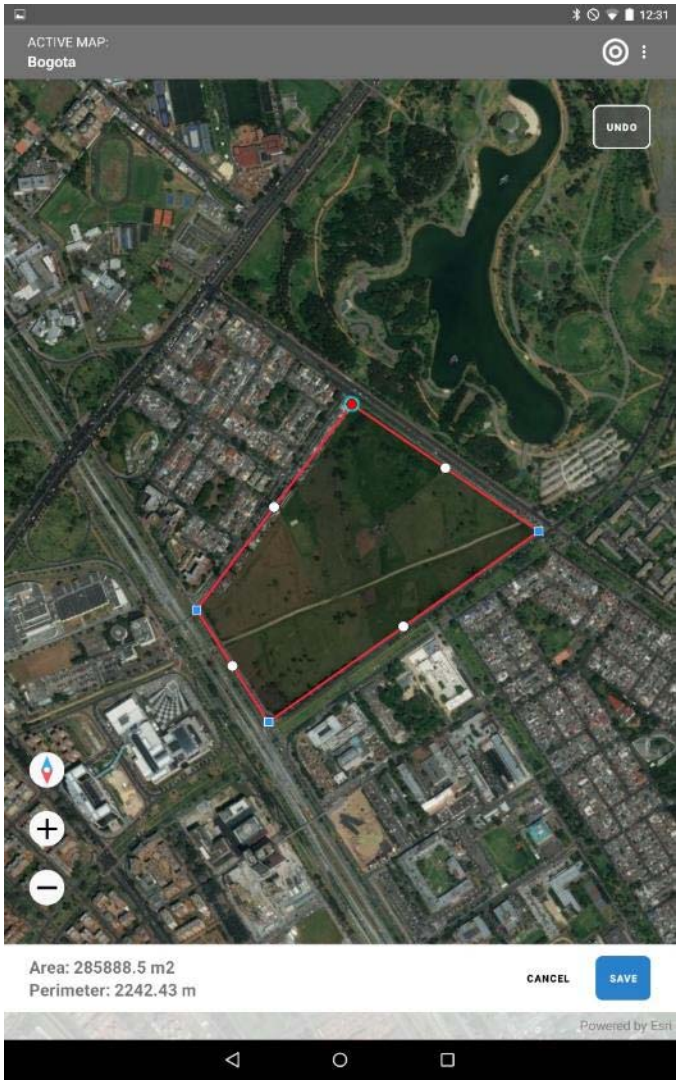


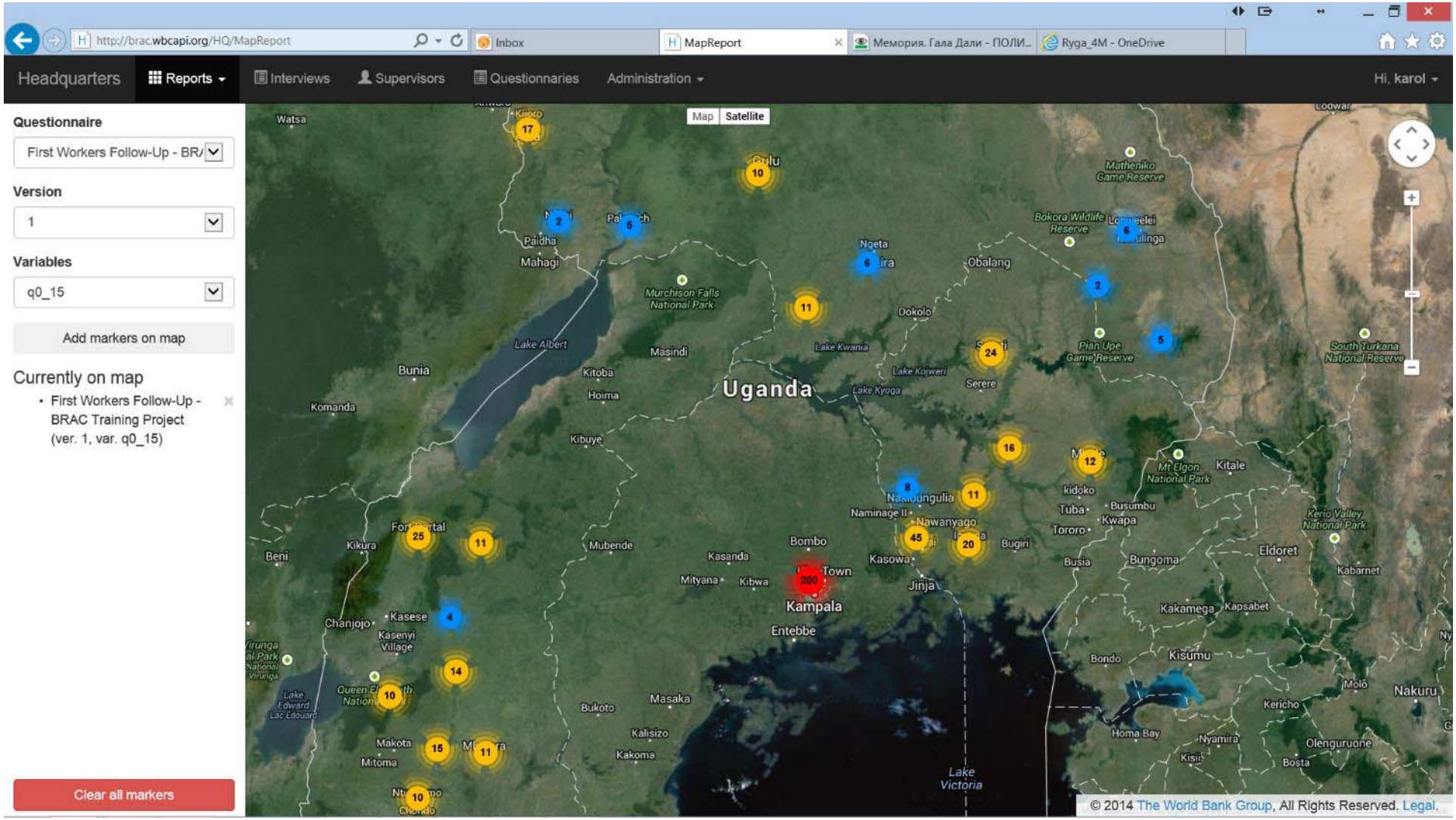




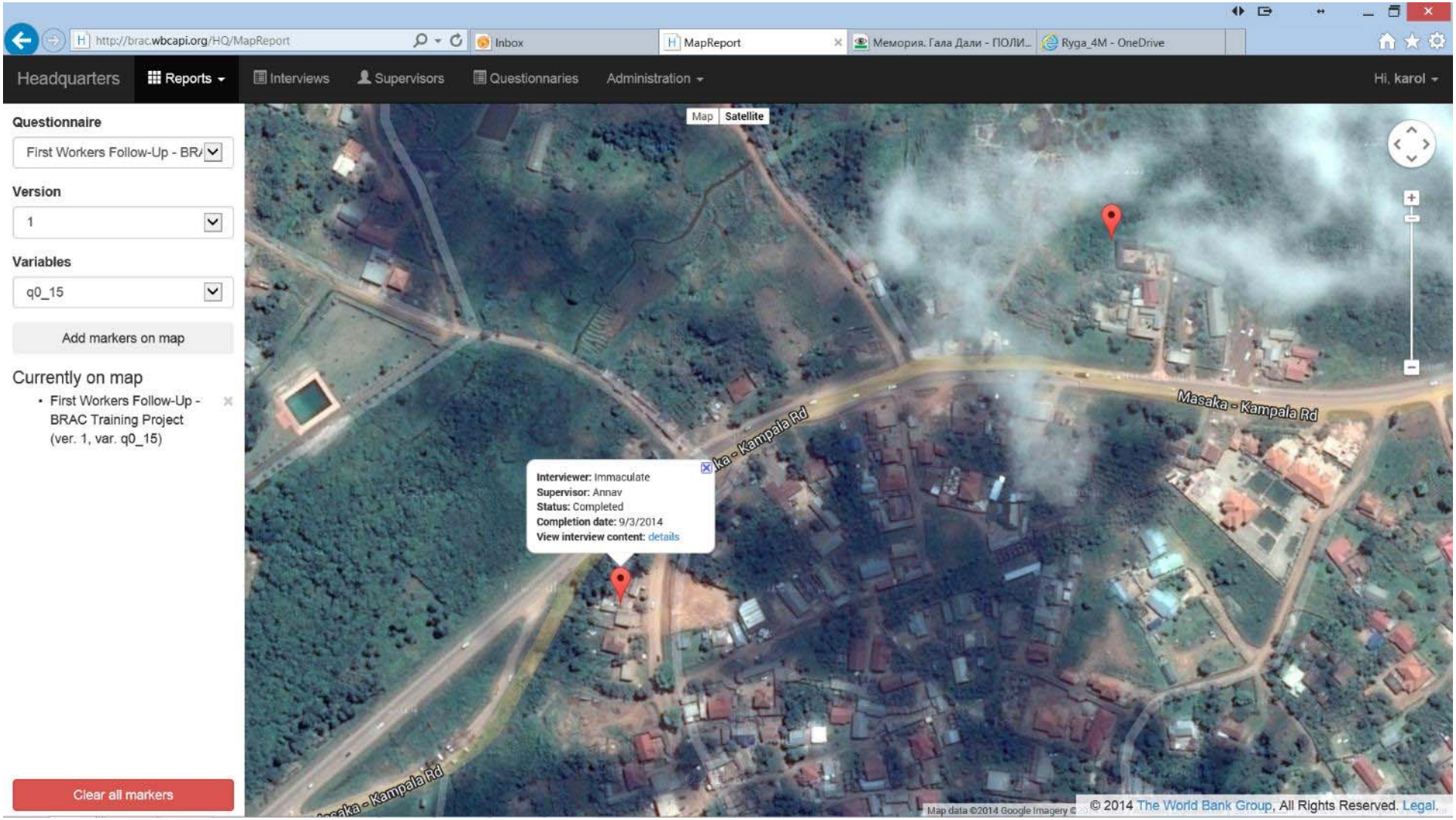
Collaboration with ESRI on GIS







Map of the survey



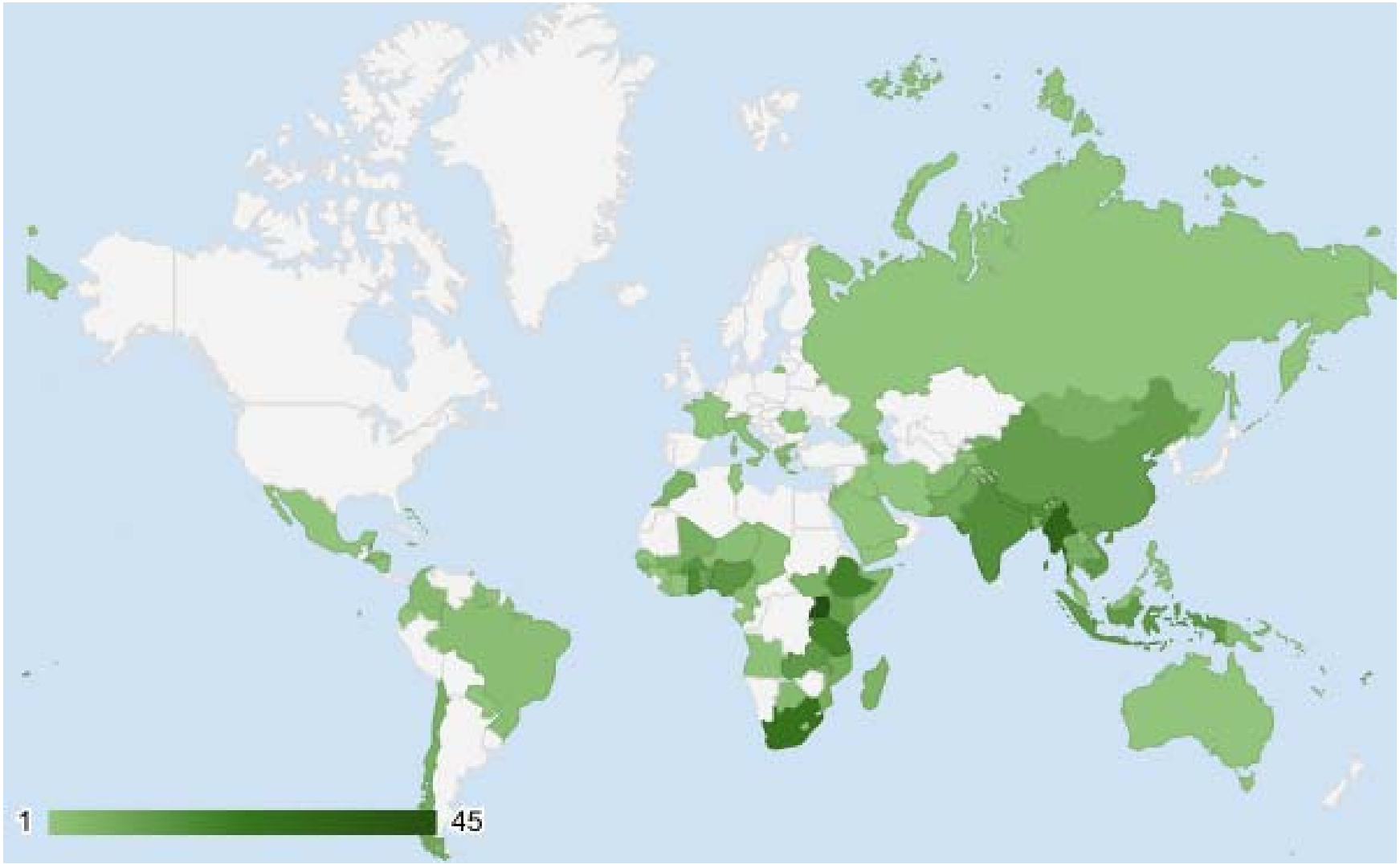
Monitor the survey by checking the GPS location of where and when the interview took place.

Collaboration with AWS and Facebook to Reduce Costs of Data

- Sampling based on population estimates using high resolution satellite images and ML
- Anthropometric measurements using image recognition
- Verification of individuals in panel surveys using face recognition and face aging algorithms.
- Experimentation in determining the reliability of answers based on facial expressions.
- ML for non standard units
- Sign language recognition

SURVEY PIPELINE: 600+ *NATIONAL SURVEYS IN 108 COUNTRIES*

6,500,000 FACE TO FACE INTERVIEWS



SURVEY SCOPE

- **Survey types**: LSMS, HBS, LFS, Enterprise Survey, EDU, Health

- **Large Clients:** India NSSO; Stats SA; Thailand NSO; FAO; IFAD; OECS; AfDB; IADB; OECS; IFS, LSTD, …, BRAC; OPM; Mathematica

- **Largest survey:** SA – 2,600,000 households, 12,000 enumerators, 500 supervisors, 60K questionnaires per day
- **Questionnaire:** Malawi – 3,000+ questions, 94 rosters

- **Other large surveys**:
  - Indonesia Sakarnas 200,000
  - Ethiopia, Uganda, Malawi LSMS
  - Nigeria LSMS
  - WAEMU Household surveys in 12 countries of West Africa
  - Many others