



International Center for Tropical Agriculture
Since 1967 / *Science to cultivate change*

Mainstreaming e- data collection in CIAT programs in Africa

14th October, 2015
Kigali, Rwanda

**Mercy Mutua,
Eliud Birachi**

E-mail mmutua@cigar.org



Introduction and motivation

- Data collection in Africa has been conducted using paper based approaches.
- Little attention has been paid to the possible costs and efficiencies of paper based approaches
- When both the data collection tool and samples are large (thousands) it can be time consuming to process data
- This affects timeliness of reporting and increases costs of data management

Introduction and motivation

- When projects cuts across regions or countries, difficulties arise in:
 - tracking the same variables/ indicators
 - generating a single database of all variables from a survey

Motivation

How can we generate timely reports while.....

- Maintaining data quality and integrity
- Saving costs
- Minding the environment

Is there a way not duplicate efforts across countries?

Shifting from PAPI to CAPI

- CAPI offers the option of using the exact same form to collect data in different countries
- Aggregate data through an automated process
- Run data quality/cleaning checks all at once
- No duplicated efforts in generating results

Repeated surveys – Baseline & midterm evaluations, value chain approach

- The challenge: omitted entries, errors introduced in manual transmission, time taken to transfer info from one wave for use in subsequent wave of the survey
- CAPI solution:
- Pre-loaded lists: data is automatically pulled from one wave to the other eg HH Member lists
- Eliminates errors of manual transmission/omissions,
- interviewer doesn't need to take time re-writing lists
- Automatically linking data hence: reduced effort in analysis, eliminate possibility of computation differences across time for tracked indicators

How can we reduce interviewer errors?

- Inconsistent information
 - built in checks,
 - skip logic/pattern,
 - not possible to type wrong answer code since enumerator selects from menu/radio buttons
- Missing data
 - required fields- can't proceed/can't send data if some data is missing
 - Skip logic
- Outliers- built in upper and lower limits

How can we reduce/eliminate data entry errors?

The problem:

- Entering data into wrong columns,
- keying in impossible figures,
- omitting to enter some data (intentional or otherwise),
- duplicating observations
- Manual DE is very time consuming and expensive

CAPI solution: manual DE is completely eliminated

Some important considerations

1) Software considerations:

- What platform to use:
 - Open Access, paid subscription, own customized or a mix- cost, time and data security implications
 - Size of questionnaire and types of questions- e.g can the platform accomodate long surveys with open ended questions??
 - Data hosting options – in a cloud or locally? Does the platform allow for either?
 - Data security

Some important considerations

2) Hardware considerations

- Screen size
- Cost- high end?
- Many mobile apps are specific to operating system- Android, windows

Examples of Cost and time savings

2) NMG survey in Rwanda (covering 9 districts), total interviews 8100; components of gender, nutrition, production, blood survey, Food consumption (using CsPro)

- Paper cost was estimated at over USD 40,000
- Cost of tablets: USD 30,000
- Cost of data entry and cleaning : USD 40,000
- Saved cost: All
- Time saved: PAPI 2 questionnaires per day
- CAPI: 3-4 per day
- Verification of data real time while interviewers are still in the field

CultiAF Pre-cooked bean project (Survey-CTO)

- Coverage and scope: Kenya and Uganda
- Approach: Value chain approach- interviewed bean farmers, traders, processors, consumer demand survey
- Other aspects: Nutrition and safety, gender

Implementation

- Generated tools for farm level production and marketing, gender analysis, consumer demand for pre-cooked products, processor, traders, nutrition and food safety
- Chose a platform with paid subscription- ODK based (SurveyCTO)
- Designed forms for the various tool components
- Collected farm level and gender data in Kenya using interviewers who were centrally trained- pull household list from HH level tool and only pre-load head and spouse into the gender tool

Implementation

- One week (6 days) to collect production and marketing data from 440 households using 22 enumerators
- One week (6 days) to collect data from 565 individuals (head and spouse where applicable) in 440 households– gender components (of which 64% female, 36% male)
- Average number of interviews per enumerator per day is 4– great improvement from PAPI where an enumerator would only administer 3 questionnaires of a similar size per day

Implementation

- One week (5 days) to clean and merge data files by 1 person
- One week (4 days) to generate descriptive statistics on both farm level and gender components (2 persons)
- One week to generate draft report write up (3 researchers)

Challenges

- Missing technical capacity for designing macro-
 - Choose platform that doesnt need complex programming, has community of users, offers technical support if paid subscription
 - invest in personel with right technical capacity else outsource. Problem with outsourcing is that consultant may not really capture what you need they way you want it
- Technical capacity for data collection
 - hire qualified interviewers (at least first degree and train them well, pre-test)

Challenges and lessons learnt

- Need to change form to capture new developments in the field
- Choose platform that allows for updating form as opposed to using new versions- to save time merging datasets later
- Adequately test your forms- during interviewer training and real life pre-test with respondents
- Data transmission
- Choose platform that doesnt require filling forms online- allows interview in areas with limited internet connectivity, save on data costs
- Mobile device battery life- need to have good quality gadgets or get external portable power sources

Challenges and lessons learnt

- Run thro training and pre-test data to ensure the form captures data in the way you want it before starting actual data collection

Areas of cost savings

- Copying and shipping costs eliminated
- Manual data entry costs eliminated
- Interviewer doesn't need to take much time flipping through papers to relevant questions, reduced interviewer verification time- hence able to administer slightly more interviews
- Less manhours needed for data processing and analysis

Added functionality

- GPS Coordinates capture
- Multimedia files



Member of the
CGIAR Consortium

www.ciat.cgiar.org
www.cgiar.org



CGIAR

Science for a food secure future